

A STUDY IN USERS' PHYSIOLOGICAL RESPONSE TO AN EMPATHIC INTERFACE AGENT

HELMUT PRENDINGER

*National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
helmut@nii.ac.jp*

CHRISTIAN BECKER*

*Faculty of Technology, University of Bielefeld,
33594 Bielefeld, Germany*

MITSURU ISHIZUKA

*Graduate School of Information Science and Technology,
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-8656, Japan
ishizuka@i.u-tokyo.ac.jp*

Received 4 November 2005

This paper presents a novel method for evaluating the impact of animated interface agents with affective and empathic behavior. While previous studies relied on questionnaires in order to assess the user's overall experience with the interface agent, we will analyze users' physiological response (skin conductance and electromyography), which allows us to estimate affect-related user experiences on a moment-by-moment basis without interfering with the primary interaction task. As an interaction scenario, a card game has been implemented where the user plays against a virtual opponent. The findings of our study indicate that within a competitive gaming scenario, (i) the absence of the agent's display of negative emotions is conceived as arousing or stress-inducing, and (ii) the valence of users' emotional response is congruent with the valence of the emotion expressed by the agent. Our results for skin conductance could also be reproduced by assuming a local rather than a global baseline.

Keywords: Physiological user information; life-like characters; affective behavior; empathy; evaluation.

1. Introduction and Motivation

While life-like characters (or agents) with affective behavior are becoming ever more popular as virtual interaction partners,^{1–3} little is known about their effect

*The collaboration was performed while the author was a Pre-Doctoral Fellow of the Japan Society for the Promotion of Science (JSPS) at the National Institute of Informatics, Tokyo.

on human interaction partners. This situation is surprising given the large body of literature asserting the key importance and positive effect of emotion-based virtual agents on users' interaction experience.^{4–13} Studies that systematically investigated and compared the impact of various types of life-like agents report on often inconsistent results.^{14–17}

Only recently have some researchers started to question whether the classical method of questionnaires is adequate to evaluate affective human–computer interactions.^{18,19} The main argument is that post-experiment “self-report” methods are deficient in estimating the moment-by-moment experience of a subject,²⁰ and hence unreliable in assessing how the user felt at a particular moment during the interaction.

As a promising alternative, we will advocate a physiology-based approach to evaluating affective interactions with life-like agents since human physiology provides rich information regarding a person's emotional experience.²¹ An early study has been conducted by Ekman *et al.*,²² who investigated the effects of six emotions (surprise, disgust, sadness, anger, fear, and happiness) on four types of physiological signals: heart rate, skin temperature, skin resistance, and muscle tension. Their findings include the larger increase of heart rate with anger and fear than with happiness, and a higher decrease of skin resistance (leading to higher skin conductance) for fear and disgust as opposed to happiness, among other results. More recently, research in “affective computing”^{23–26} is offering sound results on interpreting human physiological information as emotions.

The key advantages of using human physiological response as an evaluation for human–computer interaction are: (i) the dynamic moment-to-moment nature of users' interaction experience can be estimated; (ii) physiological response is usually not within the conscious control of users, preventing fake attitudes or body expressions (e.g. simulated facial expressions); (iii) physiological information provides insight into the user's affective state without relying on cognitive judgements¹⁸ or the ability to memorize past emotions; (iv) the recording of physiological signals does not interfere with the primary interaction task. A potential drawback of using sensors is that they can be seen as intrusive.

We will apply our new method to evaluate the effect of an interface agent that displays affective behavior and also responds to the (hypothesized) emotion of the user. The latter type of behavior is known as *empathic* response.^{27–29}

The rest of the paper is organized as follows. Section 2 reports on related work. Section 3 describes the life-like agent and game scenario used in the study. In Sec. 4, we formulate our hypotheses. Section 5 explains the method and setup of the experiment. Section 6 presents our results based on the analysis of the physiological data recorded during game play. Section 7 discusses and concludes the paper. Appendix A contains the explanation of the game that was given to the subjects.

2. Related Work

Investigations on the impact of affective agent behavior on users did not exist until very recently. Brave *et al.*²⁸ claim that they conducted the first study of this kind. In their study, a user plays the Blackjack game against a disembodied dealer. The casino-style interaction scenario also features another player that is represented by a photograph depicting a human face. The photo-graphical agent displays four types of emotional responses: (i) self-oriented emotional (happy for having won, sad for having lost); (ii) empathic (happy for the winning user, sorry for the losing user); (iii) both self-oriented emotional and empathic; (iv) neither. Each response is accompanied by a text box adjacent to the agent that contains an appropriate message, e.g. an empathic comment like “That’s great! I’m really happy that you won.” The authors found that subjects rated the empathic photographic agent as more likeable and trustworthy, and perceived it as more caring. Those opinions were not seen by subjects in the self-oriented emotional condition.

Although the results of Brave *et al.*²⁸ offer valuable support for the utility of empathic agents, the study has some limitations, which we want to address in our paper. Most importantly, situations where users interact with an agent seem to be more typical (and interesting) than those where a user and an agent assume the same view as co-players (against the dealer). Secondly, animated agents provide a richer set of communicative modalities than photographic agents, and are more likely used as part of intelligent interfaces. Thirdly, as argued above, questionnaires may be useful for estimating users’ opinions on dimensions such as likeability, trustworthiness, or intelligence, but they fall short in assessing users’ emotional moment-to-moment experience.

Berry *et al.*¹⁷ investigated the impact of a 3D facial agent on various dimensions (likeability, helpfulness, recall, and so on) in the setting of an advice-giving task. It was found that only if the emotional expression of the agent is consistent with the presented message (some advice), then memory performance was comparable to the other conditions (a human actor, voice only, text only). One key result of this work is to highlight the importance of an agent’s consistency in its behavior.

In our own previous work, we demonstrated that empathic (“apologizing”) agent behavior in a quiz game with deliberately induced frustration may significantly reduce user arousal or stress measured by skin conductance.¹⁹ The current study significantly extends this work by a wider variety of induced and displayed emotions in the context of interactive gaming rather than a simple quiz.

3. The Game Scenario

As an affective gaming scenario, the classical cards game “Skip-Bo” has been implemented as a face-to-face interaction scenario between a human player and a virtual

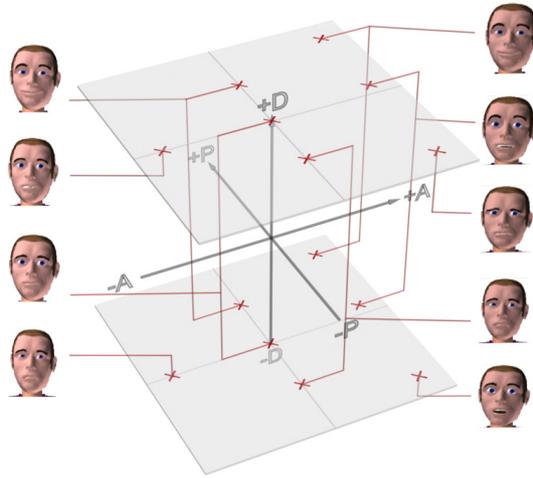


Fig. 1. The ten emotion categories in PAD-space and the seven attributed facial expressions.³² Instead of the term “friendly,” we will often use the more common term “joyful.”

agent (see Appendix A).^a In the game, players have the conflictive goals of getting rid of the eight cards on the pay-off piles to the right side of the table by playing them to the shared white center stacks. As on these center stacks the order of cards (from one to twelve) is relevant, the hand and stock cards must be used strategically by the players to achieve this overall goal, and win the game.

The “physical” objects necessary to play the game were modeled as 3D objects and enriched by semantic information, so that intuitive point-and-click interaction by the human player as well as natural gestural interaction by the agent (e.g. moving cards on the table) were easily realized.³¹

3.1. The virtual opponent

As a virtual opponent, we used a 3D agent called “Max” that is developed by the Artificial Intelligence Group at the University of Bielefeld.³³ (A screenshot can be found in Appendix A.) The Max agent has basic abilities for multi-modal interaction such as synchronized auditory speech, facial and bodily gestures, e.g. the agent may display different types of facial emotions (see Fig. 1) within the pleasure–arousal–dominance (PAD) space that reflect its current emotional state.³⁴

Speech was not seen as necessary in the cards game setting and has therefore not been implemented. However, Max utters various types of “affective sounds” such

^aThe game scenario was briefly sketched before,³⁰ together with questionnaire results. The current paper is complementary in that it provides the results of an extended analysis of users’ physiological signals and puts the research in the context of evaluations of virtual humans. A demo of the gaming interaction can be found at: http://research.nii.ac.jp/~prendinger/demo/JapanMax_02.wmv

as grunts and moans. Moreover, the agent continuously simulated breathing and eye-blinking, giving the user the impression of interacting with a life-like agent.

Visual and auditory feedback was also given whenever the human player was selecting or moving cards. Moreover, the agent gave visual feedback to the user by dynamically looking at the objects (cards) selected by itself or the user for a short period of time, and then looking straight ahead again in the direction of the user. Max also performed a simple type of turn-taking by nodding whenever completing its move. These behaviors were intended to increase the user's perception of interacting with an agent that is aware of its environment and the actual state of the game.

3.2. Emotion recognition

If the agent is supposed to respond in an empathic way, it is of paramount importance that emotions of the user are interpreted in real-time, and input to the agent's emotion simulation system. We employed our previously developed system,²⁹ that derives user emotions from skin conductance, electromyography, and situational context parameters (e.g. the game state). In short, the emotion recognition component builds on the two-dimensional (arousal, valence) model of Lang,³⁵ who claims that all emotions can be characterized in terms of judged valence (positive or negative) and arousal (high or low). As skin conductance increases with a person's level of overall arousal or stress, and electromyography correlates with negatively valenced emotions,^{21,23} named emotions can be identified in the arousal–valence space (see Ref. 29 for details). Besides exploiting contextual information, our work currently relies on a single bio-signal for each dimension, which seems justified given the rich evidence for their reliability found in the literature.^b However, as demonstrated in the work of Kim *et al.*²⁶ multi-modal analysis may lead to significant improvements in emotion recognition.

In the Skip-Bo game, the behavior of the agent is modulated by both its own and the human interlocutor's emotional state. However, in situations where user emotions are interpreted in order to determine adequate agent response, the agent's behavior will be solely determined by the user's affective state, thereby overriding the influence of its own emotion simulation model.

4. Theory

Empathy has recently been found as an important aspect in human–computer interaction.^{27,28,36} Batson characterizes empathy as “an other-oriented emotional response congruent with another's perceived welfare” [cited from Brave *et al.*²⁸ (p. 162)]. Other definitions see empathy as the cognitive act of taking another person's perspective, without making assumptions on felt sympathy.^{28,36} In our study,

^bNote that both arousal and stress lead to an increased level of skin conductance.^{21,23,24} Hence our results should always be understood to cover both interpretations.

empathy refers to the agent's response to the user's assumed emotion and covers both positive (emotional) response (e.g. sorry for the user's distress) and negative response (e.g. happy about the user's distress).

While the expression of emotion and empathy has well-known positive effects in social life, little is known about the importance of affect when expressed by a virtual agent. Reflecting the experience of Berry *et al.*¹⁷ and recasting the suggestion of Dehn and van Mulken¹⁶ (p. 19), we want to provide a partial answer to the question "What kind of animated agent used in what kind of domain influences what kind of user's physiological state?" rather than simply "Does an animated agent improve human-computer interaction?"

Since Skip-Bo is a competitive game, we expect that users will perceive the agent as an opponent. Hence, our primary hypothesis is: *If the virtual game opponent behaves "naturally" in that it follows its own goals and expresses associated positively or negatively valenced affective behaviors, users will be less aroused or stressed than when the agent does not do so.* Our secondary hypothesis is: *If the game opponent is oriented only towards its own goals and displays associated behaviors, users will be less aroused or stressed than when the agent does not express any emotion at all.*

The study was also motivated by the question whether the expression of negative emotions would induce negatively valenced responses in the user, or analogously, the expression of positive emotions would induce positively valenced user emotions. Following the argument in Ref. 37 (p. 23), positively valenced physiological response (a state of "relaxed happiness") is characterized by the absence of negative response. Although we are not aware of dedicated literature, we believe that an agent's ability of reciprocity-inducing behavior would be an important aspect for educational and gaming applications.

5. Method

5.1. Subjects

The study included 14 male and 18 female subjects. All but one subject were Japanese. Their age ranged from 22 to 55 years and the average age was 30 years. Subjects were given a monetary reward of 500 Yen for participation. They were told in advance that they would receive an extra reward of 500 Yen if they won against the agent. Subjects were randomly assigned to four experimental conditions (eight in each condition).

5.2. Design

In order to assess the effect of simulated emotions and empathic feedback in the context of human-computer interaction, we designed the following four conditions within the proposed gaming scenario:

- (i) *Non-Emotional* condition: The agent does not display any emotional behavior.

- (ii) *Self-Centered Emotional* condition: The agent only appraises its own game play, e.g. by displaying (facial) joy when it is able to move cards.
- (iii) *Negative Empathic* condition: The agent shows (a) self-centered emotional behavior, and (b) responds to the user in a “negative” way. The agent will display distress or fear when the user performs a good move, and will express joy when the user is recognizably distressed about a bad move.
- (iv) *Positive Empathic* condition: Here, the agent is (a) self-centered emotional, and (b) user actions are appraised “positively” so that the agent will be happy for the user’s game progress. If the user is detected to be distressed, the agent will be sorry for the user and will display sadness.

Those conditions should be seen as two pairs of conditions: (i) self-centered emotional (only) versus absence of self-centered emotional behavior (non-emotional behavior), and (ii) negative empathic versus positive empathic behavior. The first set will also be called *non-empathic* conditions, and the latter set *empathic* conditions. It is important to recall that our conditions are subtly different from the conditions in Ref. 28 and extend their conditions by also considering negative empathy.

5.3. Procedure

Subjects received written instructions of the card game (in Japanese) with a screenshot of the starting condition before they entered the room with the experimental setup. Subjects entered the room individually and were seated in front of a 50 inch plasma display with attached loud speakers on both sides (see Fig. 2). They were briefed about the experiment, in particular that they would play a competitive game. Then, subjects could play a short introductory game against a non-emotional

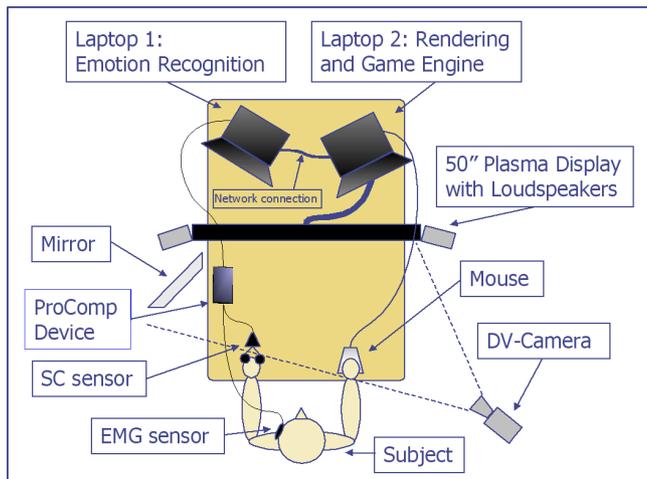


Fig. 2. Experimental setup.³⁰

Max, which allowed them to get used to the mouse-based point-and-click interface, and also provided subjects the possibility to ask clarifying questions about the game. Each subject won this first game easily.

Next, the biometrical sensors of the ProComp Infinity encoder³⁸ were attached to the subject and the subject was assured that these sensors were not harmful.^c Upon consent, a skin conductance (SC) sensor was attached to the index finger and the small finger of the non-dominant hand. The electromyography (EMG) sensor was attached to the subject's left (mirror-oriented) cheek to measure the activity of the masseter muscle. Then a relaxation phase of three minutes started, with Max leaving the display and the subject being advised not to speak. This phase was necessary to obtain a baseline for the normalization of the bio-signals, since values may greatly vary depending on subject.

From now on, the experimenter remained visually separated from the subject (behind the screen) only to supervise the experiment. After the baseline was set, the agent re-entered to the screen and the subject was asked to start the game. After the game was completed, the subjects were asked to fill in a questionnaire in English presented on the screen, together with a Japanese translation on hard-copy. The questionnaire contained 25 questions that were related to the participant's subjective experience while playing the game.³⁰

The whole interaction was recorded with a digital video camera positioned to the right behind the subject. In order to capture both the interaction on the screen as well as the human player's facial expression, a mirror was set up to acquire in indirect image of the human players face. Facial expressions were not analyzed in the current study. The rationale for the mirror was to be able to identify artifacts in the EMG values due to "laughing" behaviors of subjects. Each game lasted for about ten minutes. A protocol of the progression of the game, the acquired physiological data, and the video data were recorded for later analysis.

6. Results

This section presents our findings obtained from the analysis of biometrical data (SC and EMG) under the assumption of both global and local baselines.

6.1. *Analysis of winning situations*

We first focused on game situations where emotional reactions in the human player were likely to occur. Specifically, we hypothesized emotional reactions whenever *either* of the players (user or agent) was able to play at least two pay-off pile cards in a row — which are moves towards winning the game — and found eighty-seven such situations.

^cA detailed description of how measures were taken using the ProComp Infinity device can be found in the manual of the encoder that is available at the website.³⁸

Determining the exact duration of emotions is a notoriously hard problem.³⁷ We chose to analyze periods of ten seconds, consisting of five seconds before the last pay-off card was played, and the following five seconds. For those segments the arithmetic means (averages) were calculated for both normalized SC and normalized EMG values. For each data set (each subject and each signal type), normalization was performed by applying the equation

$$x_{\text{norm}} = \frac{x_{\text{current}} - \bar{x}_{\text{base}}}{x_{\text{max}} - x_{\text{min}}},$$

i.e. by first subtracting the average baseline value \bar{x}_{base} from the current signal value x_{current} (in the relevant segment) and dividing the resulting value by the entire range of values applicable to each subject. Note that the analysis here assumes a global baseline, as described in Sec. 5.3.

Although named emotions could have been computed from SC and EMG data by applying the model of Lang,³⁵ we will treat signal types separately, in order to retain detailed physiological information about the user.

6.1.1. Skin conductance

The results for skin conductance are shown in Fig. 3.

Agent winning move. Regarding the user's response to the agent's behavior when the agent performed a winning move, we found a significant difference between the

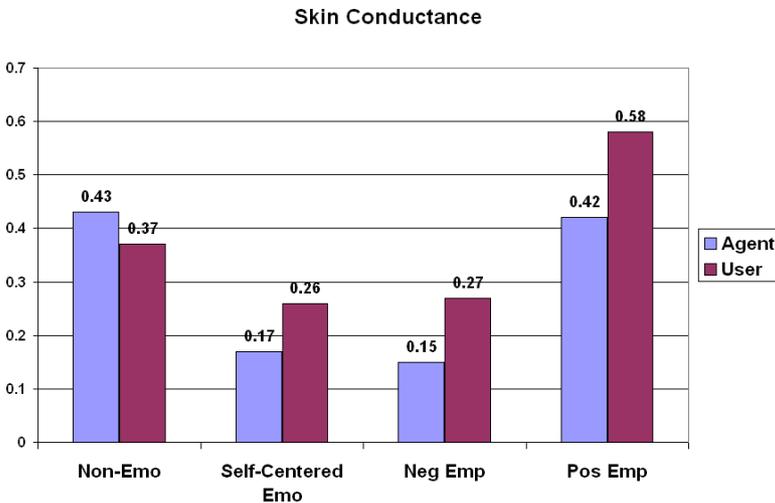


Fig. 3. The average values of normalized skin conductance data within dedicated segments of the interaction in the four conditions: Non-Emotional (“Non-Emo”), Self-Centered Emotional (“Self-Centered Emo”), Negative Empathic (“Neg Emp”), and Positive Empathic (“Pos Emp”). “Agent” refers to situations where the agent performs a winning move; “User” refers to winning move situations of the user.

Negative Empathic condition and the Positive Empathic condition [$t(20) = 2.1$; $p < 0.03$].^d The non-empathic conditions were not statistically different [$t(11) = 2.36$; $p = 0.13$].

Given that high skin conductance is an indicator of high arousal or stress,^{21,24} the human player was seemingly most aroused or stressed in the Non-Emotional condition and in the Positive Empathic condition. Although counter-intuitive at first sight, it is important to notice that in the setting of a competitive game, the lack of emotional expression or positive empathy are quite unnatural behaviors and may thus have induced user stress. The result supports the argumentation that inappropriate behavior (relative to an interaction task) may lead to higher stress levels.

User winning move. Users' physiological response to the agent when users are in a winning situation showed a somewhat similar pattern. Observe that the agent's behavior is not independent of the user's (favorable) game moves since the physiological reaction of the user triggers emotional behavior in the agent in accord with the respective condition.

The Positive Empathic condition was experienced as significantly more arousing or stressful than the Negative Empathic condition [$t(26) = 2.07$; $p < 0.01$]. However, there was no significant difference between the Non-Emotional and Self-Centered conditions [$t(21) = 2.09$; $p = 0.46$]. The result and its explanation are related to the previous ones; e.g. in the Positive Empathic condition the agent was happy for the human player's success and gave positive feedback by displaying sorriness for the user, which constitutes an unusual behavior in a competitive game.

Our findings are also consistent with the corresponding questionnaire item asking whether the agent's behavior is seen as irritating.³⁰ The agent was perceived as most irritating in the Non-Emotional condition, followed by the Positive Empathic condition.

6.1.2. Electromyography

Electromyography results are shown in Fig. 4. Observe that most values are below zero, meaning that the baseline period was experienced as negatively valenced rather than as "relaxing" in terms of muscle tension.

Agent winning move. The Negative Empathic condition differs significantly from the Positive Empathic condition [$t(20) = 2.2$; $p < 0.04$], indicating that users seemingly "reflect" the valence of the agent's emotion expression on a physiological level. There was no statistical difference between the non-empathic conditions [$t(11) = 2.23$; $p = 0.85$].

User winning move. Comparable to the result for the agent, the Negative Empathic condition is significantly different from the Positive Empathic condition [$t(26) = 2.2$;

^dAll p values were obtained with two-tailed t -tests assuming unequal variances. The confidence level α was set to 0.05.

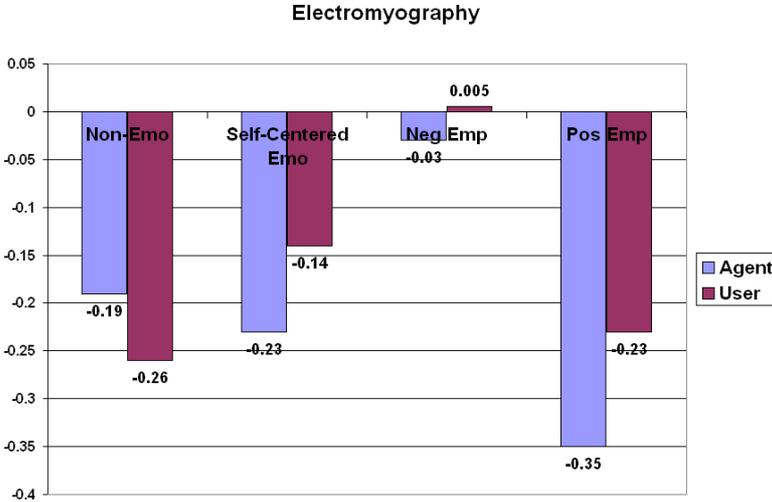


Fig. 4. The average values of normalized electromyography data within dedicated segments of the interaction in the four conditions: Non-Emotional (“Non-Emo”), Self-Centered Emotional (“Self-Centered Emo”), Negative Empathic (“Neg Emp”), and Positive Empathic (“Pos Emp”).

$p < 0.04$]. Again, the non-empathic conditions were not statistically different [$t(21) = 2.07$; $p = 0.35$].

High values of electromyography are primarily an indicator of negative valence.²¹ The highest values are achieved in the Negative Empathic condition, where the agent was designed to evoke negative emotions in the human player by showing negative emotions, e.g. a mocking smile (a “joyful” facial expression with an appropriate affective sound) to the user’s (recognized) frustration. Notably, the lowest EMG values can be observed in the Positive Empathic condition where the agent performed a “calm down” gesture (slow up and down movement of hands) if the human player was detected to be frustrated or angry.

We may tentatively conclude that the agent could trigger user responses of the same “sign” (with respect to valence) as its expressed emotions, which indicates a certain reciprocity in user responses.

An interesting observation is that users seemingly do not respond significantly different (for both skin conductance and electromyography signals) when empathic agent behavior is absent. This result demonstrates the discriminative effect of the type of empathic behavior displayed to the user, and underlines the importance of an agent caring about user feelings in an appropriate fashion.

6.2. Analysis of situations where particular agent emotions are expressed

Besides situations where either the agent or the user is in a winning (game) situation, we also investigated situations where the agent expressed some particular emotion.

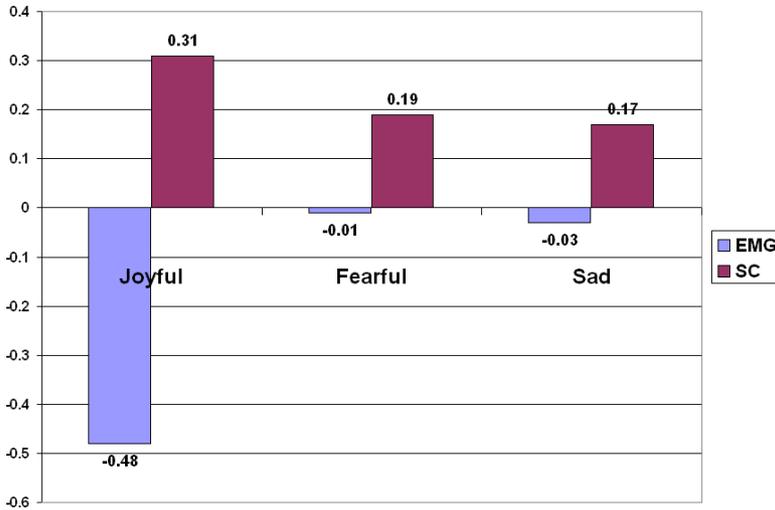


Fig. 5. The average values of normalized skin conductance and electromyography data for the three emotions “joy”, “fear”, and “sadness”.

This allows us to directly associate particular agent behaviors to user responses. Note that this type of analysis is different from the previous one in that we do not take into account the experimental condition in which the emotion occurred.

The effect of the expression of three emotions (joyful, fearful, sad) could be analyzed (see Fig. 5). Occurrences of the expression of other emotions (angry, bored, surprised) were too little for statistical analysis (fewer than six) and were hence discarded.

With regard to SC, a between-subjects analysis of variance (ANOVA) showed that subjects were significantly more aroused or stressed when the agent expressed “joy” than when the agent expressed “fear” or “sadness” [$F(2, 120) = 3.9; p < 0.03$]. Again, it can be argued that users seemingly consider joyful agent reactions as unnatural in a competitive gaming scenario and hence as arousing or stressful.

The main effect of negative emotions on EMG was even more clear cut. Users showed a significantly less negatively valenced response to joy than to fear or sadness [$F(2, 120) = 33.78; p < 0.0001$]. The high statistical significance of the outcome might have to be partly attributed to the nature of the EMG signal, where values typically rise beyond 300% over the baseline when the masseter muscle contracts. The result indicates that the expression of a positive emotion (joy) induces a significantly less negatively valenced response than the expression of negative emotions (fear, sadness).

6.3. Pearson’s correlation

We also computed the Pearson Product Moment Correlation r between normalized SC and EMG values but no such effect was found for any of the discussed

segments. Pearson's r was always smaller than 0.6. Although intuitively arousal or stress responses and negatively valenced responses are positively correlated, this was not true for the given data. This result could be interpreted as evidence for the independence of SC and EMG within our interaction scenario.

6.4. Skin conductance startle response

The bio-signal analysis presented so far assumed the recording of a (global) baseline preceding the game interaction, where the subject is assumed to be in a state of (relative) relaxation or has a moderate level of autonomic nervous system activity. Although taking a global baseline is the preferred method to compensate for individual differences in signal levels and situational parameters (e.g. room temperature), Levenson³⁷ (pp. 23–26) pointed out the possibility of methodological problems with global baselines and motivated the recording of local baseline as an alternative approach. The main rationale for assuming a local baseline is that although biometric signals are “center-seeking” (homeostatic), there might be slight shifts in the center point over time.

We follow the method of Healey²⁴ (p. 25, pp. 111–113), who developed an automatic startle detection algorithm that establishes a local baseline at the onset level of the (second) response when the first derivative exceeds a certain threshold (to distinguish high from low arousal), and then finds the local maximum following that point (a peak). The algorithm directly measures startle magnitude $S_M = x_{\text{peak}} - x_{\text{onset}}$ and startle duration $S_D = t_{\text{peak}} - t_{\text{onset}}$. A derived feature of the response is the frequency of occurrence S_F defined as the sum of startle durations per minute.

In our data, however, we could only find one (segment) instance of a startle response,²⁴ where values for S_M , S_D , and S_F were output. Here, the agent expresses “joy” to Subject 29 (see Fig. 6).

A possible reason is that Ref. 24 requires that two responses have to be more than one second apart, which could not be seen in any of the other segments. On the other hand, we were able to detect instances of (local) arousal startle crossovers by using the threshold of 0.093 micro-Siemens per second which was empirically determined.²⁴ Startle crossovers indicate situations where the user gets aroused or stressed.

Figure 7 depicts a segment where the agent displays “sadness” to Subject 20. No startle crossover is detected here.

In order to systematically investigate whether there are significant differences in the number of startle crossovers for the most frequently expressed emotions (joyful, fearful, sad), we performed another between-subjects analysis of variance. Figure 8 depicts the mean values for segments in which the considered emotions were displayed. It could be shown that there are significantly more startle crossovers when the agent expresses “joy” than when it expresses “fear” or “sadness” [$F(2, 105)$; $p < 0.005$]. This result partly resembles our previous result as described in Sec. 6.2,

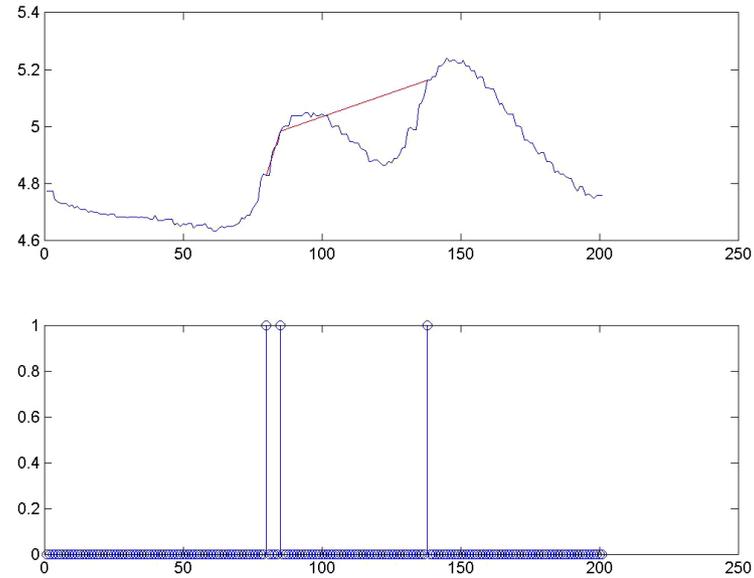


Fig. 6. Agent expresses “joy” to user. Upper graph: skin conductance signal at 20 samples/second (samples at x -axis); values in micro-Siemens (y -axis). Lower graph: starting point and the ending point of crossing startle threshold.

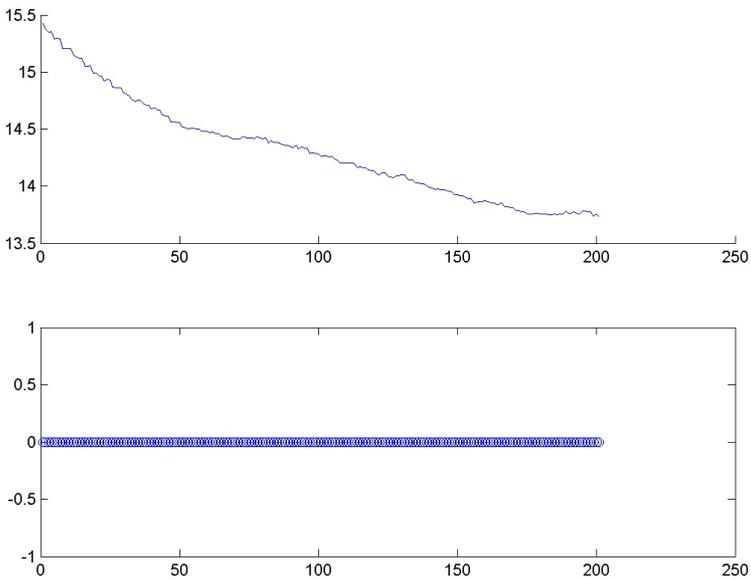


Fig. 7. Agent expresses “sadness” to user (skin conductance signal in micro-Siemens at 20 samples/second).

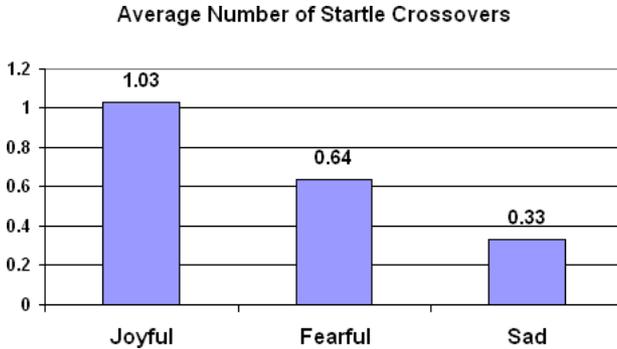


Fig. 8. Comparison of average number of startle crossovers for segments where the agent expresses “joy”, “fear”, or “sadness.”

where, assuming a global baseline, the agent expressing “joy” was experienced as most arousing.

We also calculated startle crossovers for the agent and user winning situations (discussed in Sec. 6.1), but did not find any significant differences between conditions. It should be noticed that only 18% of the considered segments showed a startle crossover. (For the “Expressed Emotions” segments, 43% had a startle crossover.)

In sum, the results in this section indicate that (given an adequate mass of data) both local and global baselines lead to a similar outcome regarding the detection of a user’s arousal. This is practically important since in most real-world applications a global baseline is hard to obtain.

7. Discussion and Conclusions

This paper advocates a novel approach to evaluate the interaction between a user and a virtual humanoid agent. Specifically, we systematically investigated the impact of different types of empathic behavior displayed by an animated interface agent on human physiology.

The results support the supposition that agent behavior has to be adequate with respect to the given interaction task.^{16,17} While previous similar studies only considered positive empathic response,^{19,27–29} our experiment also analyzed the utility of displaying negative emotions. Regarding our primary hypothesis the chief finding is that — within a competitive game scenario — displaying positive affect is conceived as significantly more arousing or stressful than displaying negative effect (derived from skin conductance), as it might also be experienced when playing against a human player. Our secondary hypothesis was not supported in the study. If the agent does not care about the users’ emotions (the non-empathic conditions), users do not care either, i.e. their physiological response is not significantly different.

A complementary result is that negative empathic agent behavior induces negatively valenced emotions (derived from electromyography) in users, and analogously, positive empathic behavior is characterized by the absence of negatively valenced

emotions. This finding indicates a certain reciprocity between the agent’s display of affect and user physiological response.

Another interesting observation of our study is that the agent’s expression of a positive emotion (joy) is experienced as more arousing or stressful than the expression of a negative emotion, such as “fearful” or “sad.” On the other hand, the expression of negative emotions seemingly induces negatively valenced response, unlike the investigated positive emotion.

A crucial aspect of the game scenario was to estimate and respond to user emotions in real-time. Although our emotion recognition system²⁹ cannot be easily validated directly, the systematic analysis of physiological information seems to be a promising approach to evaluate its effect within the interaction loop with humans.

Systems implementing real-time affective response in real-world applications will generally not provide the time to take a global baseline that precedes the interaction. Consequently, we started to experiment with a different method that establishes a local baseline. Although we obtained promising results for skin conductance, where the agent expressed some emotions, we will have to investigate more robust and task-specific methods. The parameters of the startle algorithm, originally developed for car driving,²⁴ might have to be fine-tuned for the interface domain.

The chief advantage of recording physiological signals for interface evaluation is that they provide information about user affect on a moment-to-moment basis without interfering with the interaction task. While this is also true for video recordings of, for example, the user’s facial expression, our method does not rely on skillful annotation. However, interpreting physiological signals remains a challenging problem despite of recent progress.^{23,25} It also has to be noted that questionnaires are more flexible in the information that can be assessed from users, whereas human physiology is typically restricted to estimating user states related to affect and task-load. Thus, we recently started to conduct complementary research on detecting users’ focus of attention while interacting with an embodied interface by using an eye tracker.³⁹ Similar to the evaluation method described in this paper, eye tracking may capitalize on (mostly) involuntary responses on a moment-to-moment basis.

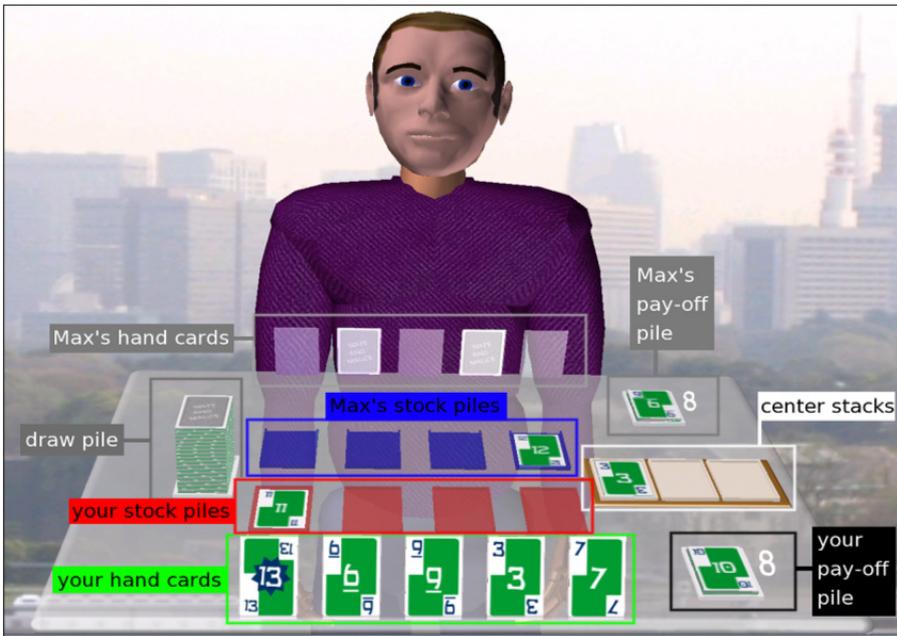
Given our results for evaluating the impact of empathic behavior displayed by an embodied interface agent, we can only speculate about whether a similar effect can be obtained for humans interacting with robotic agents expressing affect and empathy. One of the co-authors of this paper has started to experiment with the Honda humanoid robot ASIMO,⁴⁰ and we are eager to conduct further studies to possibly repeat our results for humanoid robots.

Acknowledgments

We would like to express our cordial thanks to Arturo Nakasone for fruitful discussion and for implementing the emotion recognition module, and Avinash Lavania for helping with the bio-signal analysis. Christian Becker is supported by a Pre-Doctoral Fellowship of the Japan Society for the Promotion of Science (JSPS).

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

Appendix A. Instructions for the card game Skip-Bo



1. Purpose of the game



Both players try to be the first to get rid of a pile of “pay-off cards” by playing them to center stacks which are begun with a one and continue in upward sequence to a twelve. The players take alternate turns and they play with nine series of cards each ranging from 1 to 13, which makes a total of 117 cards. **Thirteens are wild cards (jokers) and may be played in place of any card you desire.**

At the beginning you will be dealt five cards to your hand (so called hand cards) which are placed at the bottom of the screen facing to you (see above in the screenshot). These cards are only visible to you. Then eight cards are dealt to make up the pay-off piles which are placed to the right side of the table. Only the top cards of the two piles are face up and therefore visible to you as well as your opponent Max. You will go first.

2. The Play

The object of the game is to be the first to **get rid of all the cards in your pay-off pile** by playing them to the three white center stacks. The first card in each center

stack must be a 1 (or a 13), then 2, 3, and so on in sequence up to 12, each card played being one higher than the card it covers.

It is also possible to play a card from your hand to a center stack or to one of your four red stock piles in front of you, or to move a card from one of your stock piles to a center stack. There is no restriction on the ranks of cards which can be played on the stock pile.

You may play as many cards to center stacks as you want in any order, but as soon as you move a card from your hand to one of your stock piles your turn ends, and Max takes a turn. You must play a card to a stock pile at the end of each turn.

If during your turn you manage to play all five cards from your hand, without playing to a stock pile, you immediately draw five more cards from the draw pile and continue playing.

If in the beginning of your turn you have fewer than five cards in your hand, the required number of cards will be drawn from the draw pile to bring your hand up to five cards again.

If you complete a center stack by playing a twelve (or a thirteen representing a twelve) to the center, Max shuffles the completed stack into the draw pile, creating a space for a new center stack, and you can continue playing.

3. Summary

- Who is first to get rid of all cards in his pay-off pile wins the game.
- In the beginning of your turn the required number of cards will be drawn from the draw pile to bring your hand up to five cards automatically.
- You may play as many cards as you want from either your pay-off pile, your hand or one of your four red stock piles to any of the three white center piles as long as you follow the order of cards.
- Whenever you run out of hand cards without having played a card to one of your red stock piles you are immediately dealt five new hand cards.
- You finish your turn by playing one of your hand cards to one of your red stock piles.

4. Useful Strategies and Hints

- (i) Always keep in mind the number of your current pay-off card!
- (ii) You may try to keep an eye on the current pay-off card of Max. Sometimes it might be better not to play a card if this lets Max play his pay-off card afterwards.
- (iii) You may try to get rid of your hand cards first.
- (iv) When playing your last card to one of your red stock piles you may try to keep the following strategy in mind:
 - Always play high cards on empty stock piles.
 - If there are cards on some stock piles already, you may try to play cards on top of them in descending order, e.g. an 11 on top of a 12 or a 7 on top of an 8.

Good luck!

References

1. J. Cassell, J. Sullivan, S. Prevost and E. Churchill (eds.), *Embodied Conversational Agents* (The MIT Press, Cambridge, MA, 2000).
2. J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan and N. Badler, Creating interactive virtual humans: Some assembly required, *IEEE Intell. Syst.* **17**(4) (2002) 54–63.
3. H. Prendinger and M. Ishizuka (eds.), *Life-Like Characters. Tools, Affective Functions, and Applications*, Cognitive Technologies Series (Springer, Berlin, Heidelberg, 2004).
4. J. Bates, The role of emotion in believable agents, *Commun. ACM* **37**(7) (1994) 122–125.
5. J. D. Velásquez, Modeling emotions and other motivations in synthetic agents, in *Proc. 14th Nat. Conf. Artificial Intelligence (AAAI-97)*, Providence, Rhode Island (AAAI Press/The MIT Press, 1997), pp. 10–15.
6. E. André, M. Klesen, P. Gebhard, S. Allen and T. Rist, Integrating models of personality and emotions into lifelike characters, in *Proc. Int. Workshop Affect in Interactions — Towards a New Generation of Interfaces*, Siena, Italy (1999), pp. 136–149.
7. G. Ball and J. Breese, Emotion and personality in a conversational agent, in *Embodied Conversational Agents*, eds. J. Cassell, J. Sullivan, S. Prevost and E. Churchill (The MIT Press, Cambridge, MA, 2000), pp. 189–219.
8. A. Paiva (ed.), *Affective Interactions: Towards a New Generation of Computer Interfaces*, LNAI 1814 (Springer, Berlin, 2000).
9. H. Prendinger, S. Descamps and M. Ishizuka, Scripting affective communication with life-like characters in web-based interaction systems, *Int. J. Appl. Artif. Intell.* **16**(7–8) (2002) 519–553.
10. A. Ortony, On making believable emotional agents believable, in *Emotions in Humans and Artifacts* eds. R. Trappl, P. Petta, and S. Payr (The MIT Press, Cambridge, MA, 2003), pp. 189–211.
11. C. Pelachaud and M. Bilvi, Computational model of believable conversational agents, in *Communication in Multiagent Systems: Background, Current Trends and Future* (Springer, 2003), pp. 300–317.
12. R. Trappl, P. Petta and S. Payr (eds.), *Emotions in Humans and Artifacts* (The MIT Press, Cambridge, MA, 2003).
13. L. Canamero, Emotion understanding from the perspective of autonomous robots research, *Neural Networks* **18**(4) (2005) 445–455.
14. 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 *Proc. CHI-97*, Atlanta, Georgia (ACM Press, 1997), pp. 359–366.
15. S. van Mulken, E. André and J. Müller, The persona effect: How substantial is it?, in *Proc. Human-Computer Interaction (HCI-98)*, Sheffield, England (Springer, Berlin, 1998), pp. 53–66.
16. D. M. Dehn and S. van Mulken, The impact of animated interface agents: A review of empirical research, *Int. J. Human-Comput. Studies* **52** (2000) 1–22.
17. D. C. Berry, L. T. Butler and F. de Rosis, Evaluating a realistic agent in an advice-giving task, *Int. J. Human-Computer Studies* **63**(3) (2005) 304–327.
18. R. W. Picard and S. B. Daily, Evaluating affective interactions: Alternatives to asking what users feel, in *CHI-05 Workshop on Evaluating Affective Interfaces: Innovative Approaches*, Portland, Oregon (2005).
19. H. Prendinger, J. Mori and M. Ishizuka, Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game, *Int. J. Human-Comput. Studies* **62**(2) (2005) 231–245.

20. R. E. Nisbett and T. D. Wilson, Telling more than we know: Verbal reports on mental processes, *Psychol. Rev.* **84** (1977) 231–259.
21. J. L. Andreassi, *Psychophysiology: Human Behavior & Physiological Response*, 4th edn. (Lawrence Erlbaum Associates, Mahwah, NJ, 2000).
22. P. Ekman, R. W. Levenson and W. V. Friesen, Autonomic nervous system activity distinguishes among emotions, *Science* **221** (1983) 1208–1210.
23. R. W. Picard, *Affective Computing* (The MIT Press, Cambridge, MA, 1997).
24. J. A. Healey, Wearable and automotive systems for affect recognition from physiology, Ph.D. thesis, Massachusetts Institute of Technology (2000).
25. C. Lisetti and F. Nasoz, Using noninvasive wearable computers to recognize human emotions from physiological signals, in *EURASIP J. Appl. Signal Process.* **11** (2004) 1672–1687.
26. J. Kim, E. André, M. Rehm, T. Vogt and J. Wagner, Integrating information from speech and physiological signals to achieve emotional sensitivity, in *Proc. 9th Eur. Conf. Speech Communication and Technology*, Lisbon, Portugal (2005).
27. A. Paiva, J. Dias, D. Sobral, R. Aylett, P. Sobrepelez, S. Woods, C. Zoll and L. Hall, Caring for agents and agents that care: Building empathic relations with synthetic agents, in *Proc. 3rd Int. Joint Conf. Autonomous Agents and Multi-Agent Systems (AAMAS-03)*, Melbourne, Victoria, Australia (ACM Press, New York, 2004), pp. 194–201.
28. S. Brave, C. Nass and K. Hutchinson, Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent, *Int. J. Human-Comput. Studies* **62**(2) (2005) 161–178.
29. H. Prendinger and M. Ishizuka, The empathic companion: A character-based interface that addresses users' affective states, *Int. J. Appl. Artif. Intell.* **19**(3) (2005) 267–285.
30. C. Becker, H. Prendinger, M. Ishizuka and I. Wachsmuth, Evaluating affective feedback of the 3D agent Max in a competitive cards game, in *Proc. 1st Int. Conf. Affective Computing and Intelligent Interaction (ACII-05)*, Beijing, China, LNCS 3784 (Springer, Berlin, 2005), pp. 466–473.
31. M. E. Latoschik, P. Biermann and I. Wachsmuth, Knowledge in the loop: Semantics representation for multimodal simulative environments, in *Proc. 5th Int. Symp. Smart Graphics* (Frauenwoerth Cloister, Germany, 2005).
32. C. Becker, N. Lessmann, S. Kopp and I. Wachsmuth, Connecting feelings and thoughts — Modeling the interaction of emotion and cognition in embodied agents, in *Proc. 7th Int. Conf. Cognitive Modelling (ICCM-06)*, Trieste, Italy (2006), pp. 32–37.
33. S. Kopp, B. Jung, N. Lessmann and I. Wachsmuth, Max — A multimodal assistant in virtual reality construction, *KI Zeitschrift (German Magazine of Artificial Intelligence)* **4** (2003) 11–17 [Special Issue on Embodied Conversational Agents].
34. C. Becker, S. Kopp and I. Wachsmuth, Simulating the emotion dynamics of a multimodal conversational agent, in *Proc. Tutorial and Research Workshop on Affective Dialogue Systems (ADS-04)*, Kloster Irsee, Germany, LNAI 3068 (Springer, Berlin, 2004), pp. 154–165.
35. P. J. Lang, The emotion probe: Studies of motivation and attention, *Am. Psychol.* **50**(5) (1995) 372–385.
36. J. Klein, Y. Moon and R. W. Picard, This computer responds to user frustration: Theory, design, and results, *Interacting Comput.* **14** (2002) 119–140.
37. R. W. Levenson, Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity, in *Social Psychophysiology and Emotion: Theory and Clinical Applications*, ed. H. L. Wagner (John Wiley & Sons, Hoboken, NJ, 1988), pp. 17–42.

38. Thought Technology Ltd., <http://www.thoughttechnology.com> (2005).
39. H. Prendinger, C. Ma, J. Yingzi, A. Nakasone and M. Ishizuka, Understanding the effect of life-like interface agents through eye users' eye movements, in *Proc. 7th Int. Conf. Multimodal Interfaces (ICMI-05)*, Trento, Italy (ACM Press, New York, 2005), pp. 108–115.
40. K. Kushida, Y. Nishimura, H. Dohi, M. Ishizuka, J. Takeuchi and H. Tsujino, Humanoid robot presentation through multimodal presentation markup language MPML-HR, in *AAMAS-05 Workshop on Creating Bonds with Humanoids*, Utrecht, The Netherlands (2005), pp. 23–29.



Helmut Prendinger is Associate Professor at the National Institute of Informatics (NII), Tokyo. He previously held positions as a JSPS research associate and JSPS post-doctoral fellow at the University of Tokyo, Ishizuka Laboratory. Earlier he worked as a junior specialist at the University of California, Irvine. He received his M.A. and Ph.D. degrees from the University of Salzburg, Department of Logic and Philosophy of Science and Department of Computer Science.

His research interests include artificial intelligence, intelligent multi-modal interfaces, and affective computing, areas in which he has published more than 75 papers in international journals and conferences. He is a co-editor (with M. Ishizuka) of the book “Life-Like Characters, Tools, Affective Functions, and Applications” that appeared in the Cognitive Technologies Series by Springer in 2004.



Christian Becker received his M.S. degree from the University of Bielefeld, for his work on emotional dynamics of the articulated communicator Max. He is currently a Ph.D. student and research assistant in the Artificial Intelligence Group at the University of Bielefeld, Faculty of Technology. In his research, he concentrates on the simulation of complex emotions for conversational artifacts.



Mitsuru Ishizuka is a Professor of the Graduate School of Information Science and Technology, University of Tokyo. Previously, he worked at NTT Yokosuka Laboratory and Institute of Industrial Science, University of Tokyo. During 1980–1981, he was a visiting Associate Professor at Purdue University. He received his B.S., M.S. and Ph.D. degrees in electronic engineering from the University of Tokyo.

His research interests are in the areas of artificial intelligence, multi-modal media with life-like agents, Web intelligence, and next-generation Web foundations. He is a member of IEEE, AAAI, the Japanese Society for Artificial Intelligence (currently, president), IPS Japan, among others.