x30 x4

ishizuka@miv.t.u-tokyo.ac.jp

Abstract

In this paper, we report on our efforts in developing affective character-based interfaces, i.e. interfaces that recognize and measure affective information of the user and address user affect by employing embodied characters. In particular, we describe the Empathic Companion, an animated interface agent that accompanies the user in the setting of a virtual job interview. This interface application takes physiological data (skin conductance and electromyography) of a user in real-time, interprets them as emotions, and addresses the user's affective states in the form of empathic feedback. The Empathic Companion is conceived as an educational agent that supports job seekers in preparing for a job interview.

We also present results from an exploratory study that aims to evaluate the impact of the Empathic Companion by measuring users' skin conductance and heart rate.

CONTEXT

CONTEXT can also be used for Keith Lam's emotional room

#1 p. 1

1 Introduction

The idea of a computer 'sensing' the user's autonomic nervous system (ANS) activity is becoming increasingly popular in the human–computer interface community, partly because of the availability of affordable high-specification sensing technologies, and also due to the recent progress in interpreting users' physiological states as affective states or emotions (Picard, 1997). The general vision is that if a user's emotion could be recognized by a computer, human–computer interaction would become more natural, enjoyable, and productive. The computer could offer help and assistance to a confused user or try to cheer up a frustrated user, and hence react in ways that are more appropriate than simply ignoring the user's affective state as is the case with most current interfaces.

Our interest in this paper concerns interfaces that employ embodied agents, or life-like characters, as interaction partners of the user and serve an educational goal. By emulating multi-modal human–human communication and displaying social cues including (synthetic) speech, communicative gestures, and the expression of emotion, those characters may also trigger social reactions in users, and thus implement the "computers as social actors" metaphor (Reeves and Nass, 1998; Bickmore, 2003). This type of 'social interface' has been demonstrated to enrich human–computer interaction in a wide variety of applications, including interactive presentations, tutoring, and sales (see Prendinger and Ishizuka (2004) for a recent overview).

This part addresses the question I raised in Sharing Emotions and Space — Empathy as a basis for cooperative spatial interaction. Machine can trigger social reactions of people if they emulate multi-model human-human communication.

While virtual tutor applications constitute a well-established application area of animated agents (e.g. (Johnson et al., 2000; Conati, 2002)), the broader class of educational agents is increasingly gaining importance (Hayes-Roth, 2001; Marsella et al., 2004; Paiva et al., 2004). The role of educational agents is not confined to that of a 'virtual teacher' instructing a student, but rather to support the user in achieving an educational goal in a wider sense, such as receiving support in coping with a problem. In the case of the Empathic Companion, the educational aspect consists in preparing job seekers for an interview situation. The underlying assumption is that better prepared interviewees are less stressed and can thus perform better when being interviewed.

The Empathic Companion application is an interface that obtains information about a user's physiological activity in real-time and provides affective feedback by means of a life-like character. The interface is intended to address the user's emotion by showing concern about user affect, sometimes called *empathic* (or *sympathetic*) behavior (Paiva et al., 2004). Empathic interfaces have been shown to improve human–computer interaction by leaving users less frustrated in the case of a stressful event related to the interaction (Klein et al., 2002). The web-based (virtual) job interview scenario described in this paper serves as a simple demonstrator application that allows us to discuss the technical issues involved in real-time emotion recognition as well as the implementation of an empathic agent.

This paper also aims to illustrate two approaches to using human physiology to *evaluate* empathic embodied interfaces. Following Ward and Marsden (2003), we distinguish

#2 p. 2

real-time. Hence, in order to assess a user's emotional state in online, we implemented a system that takes physiological signals of the user during the interaction with the computer. *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 supports the argument that embodiment is not necessary to achieve social response, 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' (van Mulken et al., 1998).

3 Empathizing with a Frustrated User

#3 p. 4

Netica (2003). Norsys Software Corp.. URL: http://www.norsys.com.

Paiva, A., Dias, J., Sobral, D., and R. Aylett (2004). Caring for agents and agents that

#4 p. 18

The Empathic Companion: A Character-based Interface that Addresses Users' Affective States

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

Mitsuru Ishizuka
Dept. 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
ishizuka@miv.t.u-tokyo.ac.jp

Abstract

CONTEXT

In this paper, we report on our efforts in developing affective character-based interfaces, i.e. interfaces that recognize and measure affective information of the user and address user affect by employing embodied characters. In particular, we describe the Empathic Companion, an animated interface agent that accompanies the user in the setting of a virtual job interview. This interface application takes physiological data (skin conductance and electromyography) of a user in real-time, interprets them as emotions, and addresses the user's affective states in the form of empathic feedback. The Empathic Companion is conceived as an educational agent that supports job seekers in preparing for a job interview.

CONTEXT can also be used for Keith Lam's emotional room

We also present results from an exploratory study that aims to evaluate the impact of the Empathic Companion by measuring users' skin conductance and heart rate. While an overall positive effect of the Empathic Companion could not be shown, the outcome of the experiment suggests that empathic feedback has a positive effect on the interviewee's stress level while hearing the interviewer question.

Submitted to: Applied Artificial Intelligence Journal

Abbreviated title: "The Empathic Companion"

1 Introduction

The idea of a computer 'sensing' the user's autonomic nervous system (ANS) activity is becoming increasingly popular in the human–computer interface community, partly because of the availability of affordable high-specification sensing technologies, and also due to the recent progress in interpreting users' physiological states as affective states or emotions (Picard, 1997). The general vision is that if a user's emotion could be recognized by a computer, human–computer interaction would become more natural, enjoyable, and productive. The computer could offer help and assistance to a confused user or try to cheer up a frustrated user, and hence react in ways that are more appropriate than simply ignoring the user's affective state as is the case with most current interfaces.

Our interest in this paper concerns interfaces that employ embodied agents, or life-like characters, as interaction partners of the user and serve an educational goal. By emulating multi-modal human–human communication and displaying social cues including (synthetic) speech, communicative gestures, and the expression of emotion, those characters may also trigger social reactions in users, and thus implement the "computers as social actors" metaphor (Reeves and Nass, 1998; Bickmore, 2003). This type of 'social interface' has been demonstrated to enrich human–computer interaction in a wide variety of applications, including interactive presentations, tutoring, and sales (see Prendinger and Ishizuka (2004) for a recent overview).

This part addresses the question I raised in Sharing Emotions and Space — Empathy as a basis for cooperative spatial interaction. Machine can trigger social reactions of people if they emulate multi-model human-human communication.

While virtual tutor applications constitute a well-established application area of animated agents (e.g. (Johnson et al., 2000; Conati, 2002)), the broader class of educational agents is increasingly gaining importance (Hayes-Roth, 2001; Marsella et al., 2004; Paiva et al., 2004). The role of educational agents is not confined to that of a 'virtual teacher' instructing a student, but rather to support the user in achieving an educational goal in a wider sense, such as receiving support in coping with a problem. In the case of the Empathic Companion, the educational aspect consists in preparing job seekers for an interview situation. The underlying assumption is that better prepared interviewees are less stressed and can thus perform better when being interviewed.

The Empathic Companion application is an interface that obtains information about a user's physiological activity in real-time and provides affective feedback by means of a life-like character. The interface is intended to address the user's emotion by showing concern about user affect, sometimes called *empathic* (or *sympathetic*) behavior (Paiva et al., 2004). Empathic interfaces have been shown to improve human–computer interaction by leaving users less frustrated in the case of a stressful event related to the interaction (Klein et al., 2002). The web-based (virtual) job interview scenario described in this paper serves as a simple demonstrator application that allows us to discuss the technical issues involved in real-time emotion recognition as well as the implementation of an empathic agent.

This paper also aims to illustrate two approaches to using human physiology to *evaluate* empathic embodied interfaces. Following Ward and Marsden (2003), we distinguish between

- the paradigm that measures short-time (<5 seconds) physiological changes in response to specific events, and
- the paradigm that performs comparisons of ANS readings across longer (>5 minutes) periods of time under different control conditions.

While the first mentioned approach is applicable to situations where experimental conditions can be tightly controlled (see Sect. 3), interactions for which tight control conditions are not possible can be evaluated by following the latter approach (see Sect. 5).

The rest of this paper is organized as follows. In Sect. 2, we will describe work related to our own. Section 3 reports on the result of our previous study showing that empathic embodied feedback may reduce (deliberately frustrated) users' level of arousal. Section 4 is dedicated to introducing the Empathic Companion. There, we first describe our system for real-time emotion recognition, and then explain how physiological signals are mapped to named emotions. The final part of Sect. 4 discusses the decision-theoretic agent that is responsible for selecting the Empathic Companion's actions. In Sect. 5, we illustrate the structure an interaction with the Empathic Companion in the setting of a virtual job interview, and discuss the results of an experiment that recorded users' physiological activity during the interaction. Furthermore, issues in real-time assessment of physiological signals will be discussed. Section 6 concludes the paper.

2 Related Work

Here we will report on research strands that share the methodology or motivation of our approach to affective interfaces. A review of the literature suggests that it is possible to distinguish at least six modes of usage of a user's physiology for (embodied) interfaces:

- A user's physiological data can be used to *track* the impact of the interface on the affective state of the user. As shown in Ward and Marsden (2003), recording users' physiological data and associating them with interface events is an important methodology for testing software, e.g. to measure the effect of different types of web page design.

- A user's ANS activity can be used in order to *reflect* (or 'mirror') the user's affective state by means of an embodied agent. In this way, the user may gain insight into his or her physiological responses (bio-feedback). This type of application has been shown to bear considerable relevance for tele-home care systems (Lisetti et al., 2003).

- The user's physiological state can play a key role in selecting strategies to *adapt* the interface. When the user's frustration is detected, an interface agent can try to undo the user's negative feeling. A main application field of adaptive interfaces are tutoring systems that aim at tailoring their behavior in accord with the student's affective state and learning goal (Conati, 2002).

- User bio-signal data can be used to *address* the user's affective state. Major work has been done by Bickmore (2003) who proposes the term 'relational agents' to investigate animated agents that are designed to develop and maintain long-term, social-emotional relationships with users. Specifically, he describes an agent that addresses human affect in the role of an health behavior change assistant. The Empathic Companion application illustrated in this paper also falls under this category.

- A user's physiological responses may become increasingly important to *learn* the user's situation-dependent affective responses and hence allow for the acquisition of predictive user models (André and Müller, 2003). Learning of emotional behavior is also crucial for the previously described relational agents that are intended to enable fertile interactions with human users over extended periods of time (Bickmore, 2003).

- Physiological user information can also be used to *disambiguate* dialogue acts (Bosma and André, 2004). Here, emotions derived from physiological data help to identify the user's intention where the verbal utterance cannot be interpreted unambiguously.

The investigation of Klein et al. (2002) is most closely related to our work on empathic interfaces. They 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.

The Emphatic Companion interface differs from the one used by Klein in two 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 subject's questionnaire entries. While providing immediate response to the user's affective reaction is certainly preferable in terms of natural interaction, it assumes that affect is processed in real-time. Hence, in order to assess a user's emotional state in online, we implemented a system that takes physiological signals of the user during the interaction with the computer. *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 supports the argument that embodiment is not necessary to achieve social response, 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' (van Mulken et al., 1998).

3 Empathizing with a Frustrated User

In this section, we want to mention the main result of a previously conducted study that had the purpose of evaluating the effect of empathic embodied feedback on deliberately frustrated users (Prendinger et al., 2003). The impact of a life-like character's empathic response was measured by comparing the skin conductance readings of subjects that received empathic feedback with the skin conductance level of subjects that did not.

A simple mathematical 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 ($i \leq 4$). The instruction is given by the "Shima" character, an animated cartoon-style 2D agent, using synthetic speech and appropriate gestures (see Fig. 1). 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 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 induced 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 other (non-affective) version the occurrence of the delay was ignored by the animated instructor.

The main result of this experiment can be summarized as follows:

> If an embodied character shows empathy to a deliberately frustrated user, then the user's skin conductance level is significantly lower than when the character does not display empathy, as compared to the period of induced frustration (the delay period).

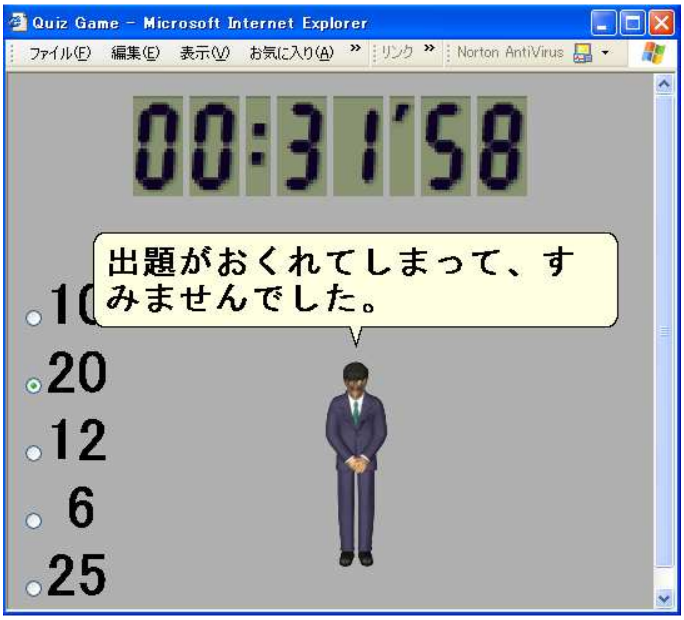

Figure 1: Shima character: "I apologize that there was a delay in posing the question."

If the level of skin conductance is interpreted as the user's level of stress or frustration, then this result indicates that empathic feedback may undo some of the user's negative emotions. Since the parameters in the study were tightly controlled, it was possible to apply the first type of evaluation paradigm described in Ward and Marsden (2003), namely, the identification of short-term ANS changes in response to specific interface events.

4 Addressing Users' Affective State

The Empathic Companion is a life-like character that was developed in the context of a web-based job interview scenario, where it addresses the user's emotion resulting from an interview situation (see Fig. 2). Being interviewed is likely to elicit emotions in the user, especially when the interviewer (Fig. 2, left) asks potentially unpleasant or probing questions, such as "What was your final grade at university?" or "Are you willing to work unpaid overtime?", and comments pejoratively upon the interviewee's (i.e. the user's) unsatisfactory answer. In order to emphasize the training aspect of the interview situation, the user is led by a companion agent (Fig. 2, right) that addresses the user's (negative) emotions by giving empathic feedback, e.g. "It seems you did not like this question so much" or "Maybe you felt a bit bad to be asked this kind of question". The user is told that the companion is invisible to the interviewer and present for his or her comfort only. Although a web-based (virtual) interview cannot induce the stress level of a face-to-face or phone interview, it provides a convenient training platform for job seekers.

4.1 System Architecture for Real-time Emotion Recognition

Since the Empathic Companion application assumes real-time emotion recognition, the system architecture depicted in Fig. 3 has been implemented on the Windows XP platform. Below, we will explain each of its components.

Figure 2: Job Interview Scenario.

4.1.1 Data Capturing

The user is attached to sensors of the ProComp+ unit from ThoughtTechnology (2003). The ProComp+ encoder allows to use input from up to eight sensors simultaneously. Currently, we only use galvanic skin response (GSR) and electromyography (EMG) sensors. Sensor input is digitally sampled by the ProComp+ unit and transmitted to the computer via a fiber-optic cable using the RS232 COM port. Although the ProComp+ unit enables data sampling up to 256 samples/second, GSR and EMG signals allow for a much lower rate, at 20 samples/second. Data capturing is achieved by a module written in Visual C++ that employs the ProComp+ data capture library.

4.1.2 Data Processing

When prompted by the application (i.e. interface events), the Data Processing component retrieves new data every 50 milliseconds, stores and evaluates them. Given the baseline information for skin conductance (GSR signal) and muscle activity (EMG signal), changes in ANS activity are computed by comparing the current mean signal values to the baseline value. The baseline is obtained during an relaxation period preceding the interaction. The current mean value is derived from a segment of five seconds, the average duration of an emotion (Levenson, 1988). If skin conductance is 15–30% above the baseline, is assumed as "high", for more than 30% as "very high". If muscle activity is more than three times higher than the baseline average, it is assumed as "high", else "normal". Emotions are hypothesized from signals using a Bayesian network (as part of the decision network discussed below), but optionally, a "Signal to Emotion Mapping" module is available, if no decisions are required on the interface side.[1]

The connection between the Data Processing component and the User Interface is es-

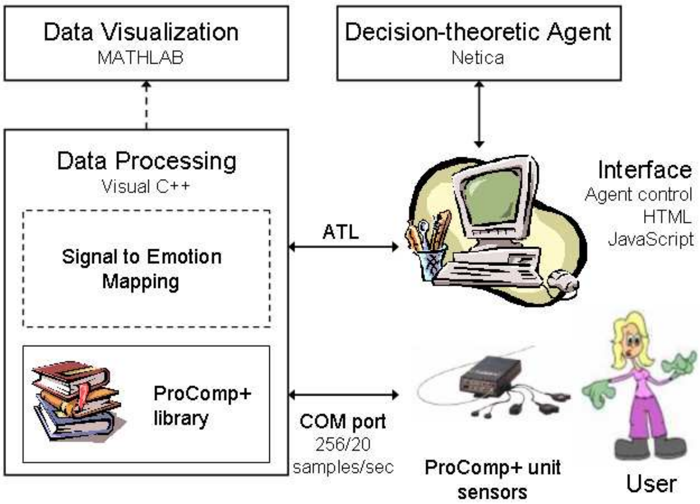

Figure 3: System architecture.

tablished by the Active Template Library (ATL) which requires functions including *Init* (initializes the ProComp+ encoder), *Start* (initializes data retrieval), *Finish* (de-allocates memory), *GetBatteryLevel* (retrieves current battery level), and *DataTransmission* (assigns data to variables).

4.1.3 User Interface

The User Interface component contains the job interview scenario and runs under Internet Explorer 5.5 (or higher). It is written in HTML and JAVASCRIPT, and utilizes the Microsoft Agent package (Microsoft, 1998) to control the verbal and non-verbal behavior of characters. This package includes an animation engine to trigger about 50 pre-defined 2D animation sequences and a text-to-speech engine. Prendinger et al. (2004) provide a detailed explanation of available character behaviors and character control features.

4.1.4 Data Visualization

In order to visualize the physiological signals obtained during an interaction session, the data can alternatively be exported to MATHLAB.

4.1.5 Decision-theoretic Agent

A decision network is used to combine bio-signals and other facts about the interaction, and relate them to emotions as well as agent decisions. The decision-theoretic agent will be discussed in Sect. 4.3. Before that, we will explain the interpretation of the user's physiological activity as emotions.

4.2 Relating Physiological Signals to Emotions

Lang (1995) claims that all emotions can be characterized in terms of judged valence (pleasant or unpleasant) and arousal (calm or aroused). Figure 4 shows some named emotions as coordinates in the arousal–valence space. The relation between physiolog-

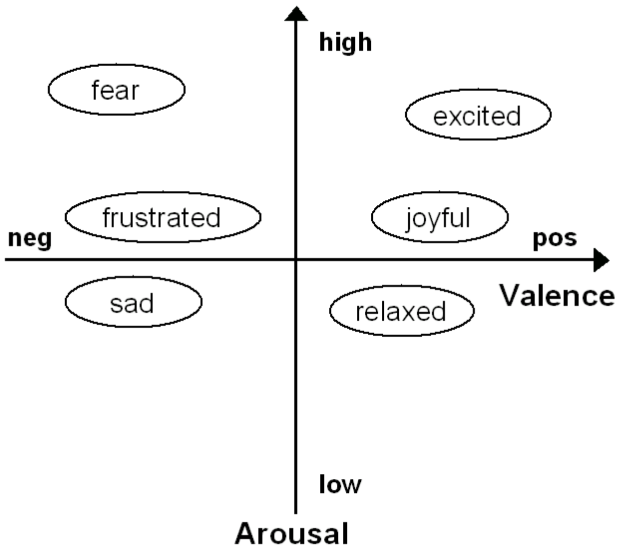

Figure 4: Some named emotions in the arousal–valence space.

ical signals and arousal/valence is established in psychophysiology that argues that the activation of the ANS changes while emotions are elicited (Levenson, 1988).

The following two signals have been chosen for their high reliability (other signals are discussed, e.g. in Picard (1997)). Galvanic skin response (GSR) is an indicator of skin conductance (SC). Under certain circumstances, the glands in the skin produce ionic sweat which changes the electrical resistance. By passing small voltage across two electrodes, the conductance between them can be measured. The electrodes can be attached to two fingers. SC increases linearly with a person's level of overall arousal (Lang, 1995). Electromyography (EMG) measures muscle activity by detecting surface voltage that occurs when the tiny muscle fibers are contracted by means of electrical impulses (lower arm or masseter muscle). Mean muscle activity has been shown to correlate with negatively valenced emotions (Lang, 1995).

4.3 Decision-theoretic Agent

The decision-theoretic agent is responsible for deriving the user's emotion given physiological data and the valence of the user's answer (to the question of the interviewer), and to suggest an appropriate action. The agent is implemented with Netica (2003), a software package that allows solving decision problems and provides convenient tools, including an API in JAVA, which has been used to implement the agent.

The decision network depicted in Fig. 5 represents a simple decision problem. A decision-theoretic agent selects actions that maximize the outcome in terms of some utility function (Jensen, 2001). The subnet consisting only of chance nodes is the Bayesian network used to derive the user's emotional state. It relates physiological signals (GSR, EMG) and the user's answer to arousal and valence which are employed to infer the user's emotional state by applying the model of Lang (1995).

The probabilities have been set in accord with the literature (whereby the concrete numbers are made up). The emotions are characterized as follows:

- "Relaxed (happiness)" is defined by the absence of autonomic signals, i.e. no arousal (relative to the baseline), and positive valence.

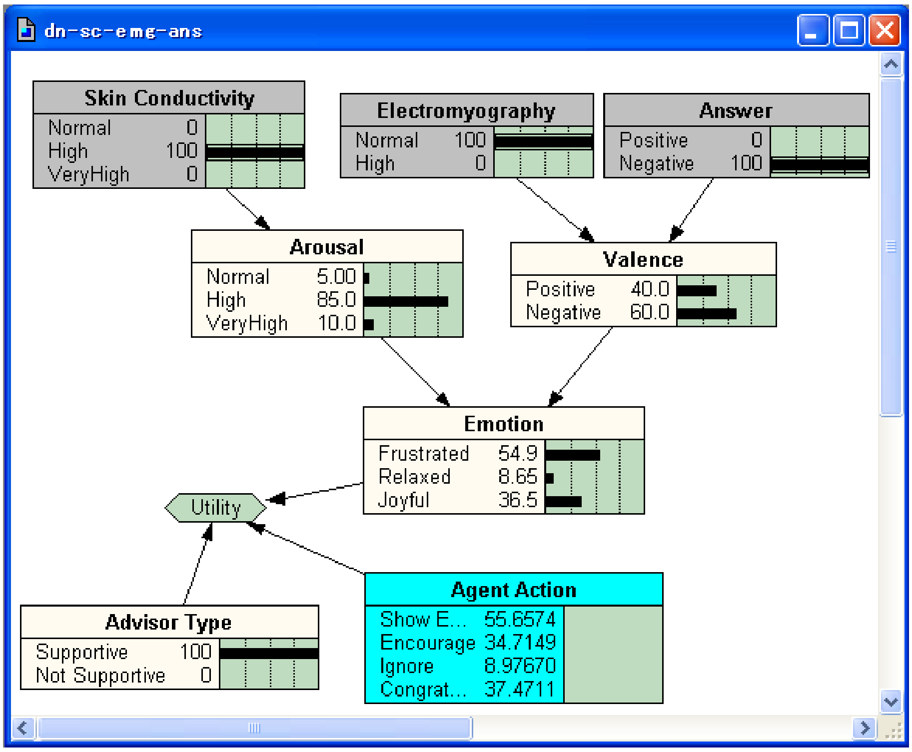

Figure 5: Simple decision network.

- "Joyful" is defined by increased arousal and positive valence.
- "Frustrated" is defined by increased arousal and negative valence.

The node "Answer" in the network represents situations where the user gives a 'positive answer' (that satisfies the interviewer's question) or a 'negative answer' (that does not satisfy the interviewer's question). This ('non-physiological') node was included to the network in order to more easily hypothesize the user's positive or negative appraisal of the question, as the user's EMG value changes (in this application) are often too small to evaluate valence.

Besides nodes representing probabilistic events in the world (chance nodes), decision networks contain nodes representing agent choices (decision nodes), and the agent's utility function (utility or value node). The decision node in Fig. 5 shows four types of actions, and Table 1 lists some responses of the Empathic Companion associated to the action types. The actual implementation of the job interview scenario provides linguistic variations for each response category. If the advisor type is supportive, the utility function is set to give priority to empathic responses.

"Advisor Type" is a deterministic (rather than chance) node that allows us to characterize the agent as supportive or non-supportive. If set to "Not Supportive", the "Ignore" action is selected for all inputs. This node is needed to compare empathic vs. non-empathic versions of the companion.

Table 1: Example responses of the Empathic Companion.

Actions	**Example Response**
Show Empathy	The agent displays concern for a user who is aroused and has a negatively valenced emotion, e.g. by saying "I am sorry that you seem to feel a bit bad about that question".
Encourage	If the user is not aroused, the agent gives some friendly comment, e.g. by saying "You appear calm and don't have to worry. Keep going!".
Ignore	The agent does not address the user's emotion, and simply refers to the interview progress, by saying, e.g. "Let us go on to the next question".
Congratulate	If the agent detects that the user is aroused in a positive way, it applauds the user ("Well done!", "Good job! You said the right thing", etc.).

5 Interacting with the Empathic Companion

In an interaction session with the Empathic Companion, the user is seated in front of a computer running the job interview, with the GSR sensors attached to two fingers of the non-dominant hand, and the EMG sensors attached to the forearm of the same body side. The baseline for subsequent bio-signal changes is obtained during an initial relaxation period of one minute, where the user listens to music from Caf del Mar (Vol.9), as the mean of GSR and EMG values.

5.1 The Structure of the Interview

An interview session is composed of (interview) episodes, whereby each episode consists of four segments (see Table 2). The entire interview session contains ten episodes, and concludes with the interviewer agent's acceptance or rejection of the user as a new employee of the company, depending on how many 'credits' the user could collect. A certain number of credits is assigned to each answer of the interviewee, reflecting the degree to which the reply satisfies the interviewer's expectation.

5.2 Exploratory Study

While a questionnaire method is certainly possible to evaluate the impact of the Empathic Companion agent, we are using physiological data to assess the user's perception of the interface. As a technical limitation of the ProComp+ unit is that it cannot be simultaneously employed for real-time data assessment and monitoring (tracking), we used the signal processor designed by our colleague Hiroshi Dohi that reads users' skin conductance (SC) and heart rate (HR). Like EMG, heart rate also correlates with negatively valenced emotions. Since SC and HR are slowly changing signals, it was sufficient to set the signal processor to 2 samples/second.

Observe that unlike the experiment reported in Sect. 3 (and Prendinger et al. (2003)), tight experimental controls are not practicable in the job interview application as the in-

Table 2: The four segments of an interview episode.

Segment	Description
1	The interviewer agent asks a question, e.g. "Tell me about your previous work experience".
2	The user chooses an answer from the set of given options (see Fig. 2, lower part), by clicking on the button next to the selected answer, e.g. the user admits the lack of experience by clicking the lower button.
3	The interviewer responds to the user's answer, e.g. "Then you are not the kind of person we are looking for" or otherwise, "I am happy to hear that you have extensive experience in the field".
4	The companion agent responds to the interviewee's emotion derived from the data gathered during the third segment and the user's answer given in the second segment.

teraction is not designed to invoke specific emotions at specific moments. In particular, depending on their answers to the interviewer's questions, users may receive positive or negative feedback. Facing a comparable situation – users' physiological responses to different web page designs – Ward and Marsden (2003) thus propose to compare signal values for whole interaction periods rather than for specific interface events.

Following this paradigm, we hypothesize the following:

> *Averaged over the entire interaction period, the presence of a (supportive) Empathic Companion will have users with lower levels of arousal and less negatively valenced affective states.*

As the control condition, the "Not Supportive" advisor type is used (see Fig. 5), where the "Ignore" action is always selected.

We thus conducted an exploratory study on the overall effect of the presence of the Empathic Companion. The subjects are connected both to the GSR sensors of the ProComp+ unit with the first two fingers of their non-dominant hand,[2] and to our in-house signal encoder unit that provides a wristband for SC (using the dominant hand) and an ear-clip to measure HR. The SC and HR data for one subject performing the interview with the (supportive) Empathic Companion agent are shown in Fig. 6 and Fig. 7, respectively. In the study, participants were 10 staff and students from the University of Tokyo, aged 23–40, who were randomly assigned to the "Supportive" and "Not Supportive" version of the Empathic Companion application, abbreviated as *Em* and *NEm*, respectively (5 subjects in each version). In the following, "AM_{relax}" refers to the (arithmetic) mean of the signal values obtained during the initial relaxation period, whereas "AM_{int}" refers to the mean of the interview period. "Range" refers to the range of data points of the whole interaction period.

First we compared AM_{relax} and AM_{int} without distinguishing between the two versions. For HR, AM_{relax}=77.83 and AM_{int}=82.05, and for SC, AM_{relax}=6 and AM_{int}=7.81. However, none of those results were statistically significant. A possible reason might be that a significant signal change only occurs after the interview starts, and the difference between the relaxation and interview periods gradually diminishes during the rest of the

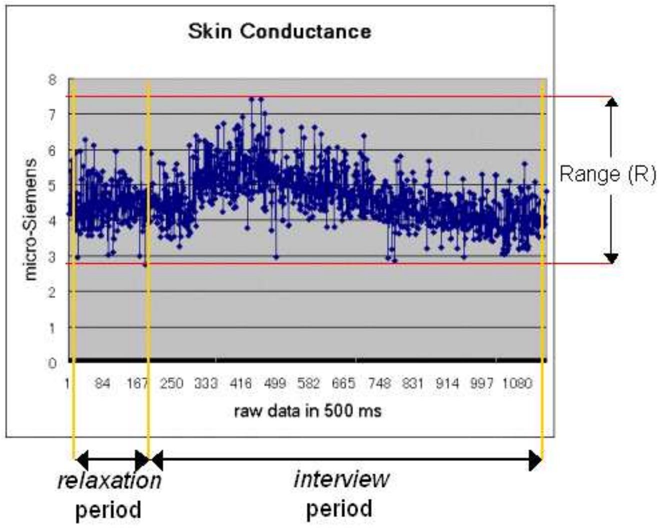

Figure 6: Sample SC data of subject interacting with the Empathic Companion.

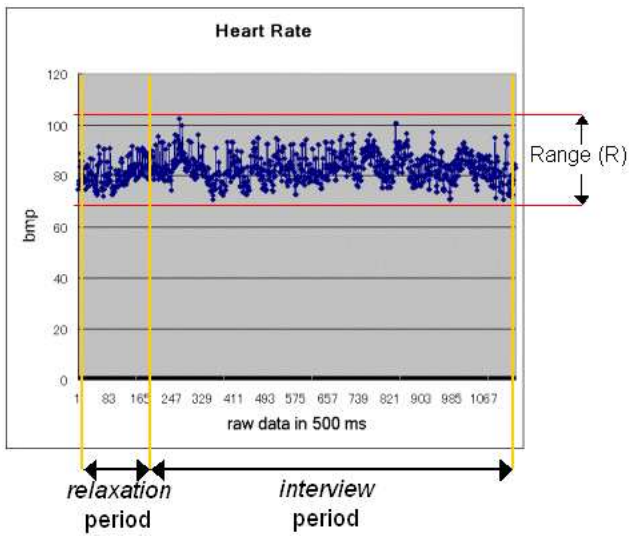

Figure 7: Sample HR data of subject interacting with the Empathic Companion.

interview. While it is not representative, the pattern of SC data depicted in Fig. 6 indicates this tendency.

Then we compared the *Em* and *NEm* versions by applying the equation $(\mathrm{AM}_{int} - \mathrm{AM}_{relax})/\mathrm{Range}$.[3] The intuition here is that a smaller value indicates that the character has a more calming effect on the user (SC) or decreases negative feelings (HR) to a higher extent. Our observations, however, do not show this. In the case of SC, the mean value of *Em* is 0.11 and 0.08 for *NEm* (contrary to our expectation). For HR, the mean value of *Em* is 0.04, and that of *NEm* is 0.06. Hence there is no significant overall positive effect of the supportive companion. However, it is worth noting that the current application (the job interview) has the empathic character interact with the user only in a very limited way. Most of the time, the user interacts with the interviewer, whose behavior is the same in both versions. Therefore, we consider to design another experiment that allows a more direct comparison between the *Em* and *NEm* versions. For instance, the character could perform as a virtual medical health assistant that asks the user about his or her health-related lifestyle, and then comments on the user's answer in a neutral or empathic fashion.

Certainly, other reasons could be responsible for the lack of significance of our results:

- The empathic responses intended to have a calming effect on the user might actually not do so.
- Heart rate might not be a reliable indicator of negative valence for all users.
- The analysis is too coarse.

In the following, we will report on a more fine-grained analysis of the data that inspects physiological response to each of the ten interview questions. The analysis is based on an affective concept that we call "anticipatory emotion". This type of emotional response occurs when a person expects a certain event to happen that will likely elicit a particular emotion. In the interview scenario a user might be assumed to experience stress when being asked a question for which he or she will not be able to give a satisfying answer. In order to investigate the effect of the Empathic Companion on subjects' anticipatory emotion, we computed the normalized SC and HR data from the period when the interviewer asks the question (Segment 1) for each version (*Em*, *NEm*).

The results are depicted in Fig. 8 and Fig. 9. For SC, a significant effect of the companion in the "Supportive" version (*Em*) could be shown ($t(10) = -5.49$; $p = 0.0002$). The companion displaying empathy effects a decrease in the subject's arousal level for the period of being questioned by the interviewer, which eventually stabilizes at the baseline level. On the other hand, when the companion ignores the subject's emotion, the arousal level increases over the interview session. In the case of HR, the average (normalized) value is higher in the *Em* version than in the *NEm* version (contrary to our expectation).

The core finding of the experiment can thus be stated as:

> *Users (as interviewees) receiving empathic feedback are significantly less aroused when hearing the interviewer's questions.*

In summary, the results of the experiment indicate that while an 'overall' positive effect of the Empathic Companion cannot be shown, the presence of a character that 'cares' can have a positive effect on the way users perceive questions in terms of lower levels of arousal (and stress).

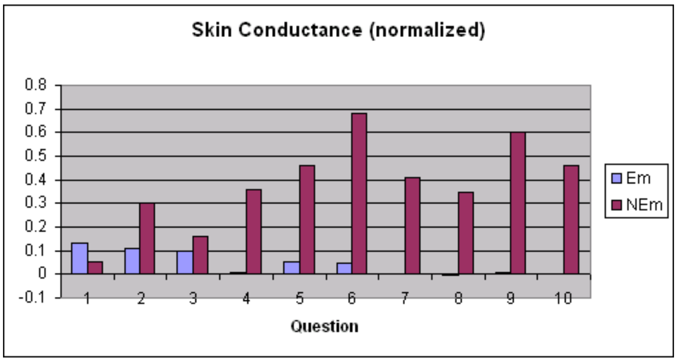

Figure 8: Normalized SC data for anticipatory emotion.

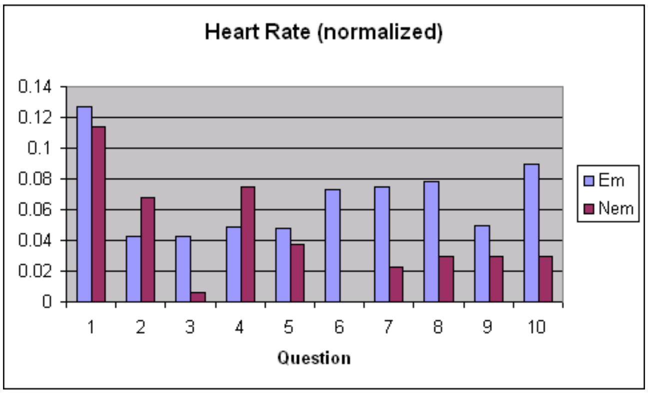

Figure 9: Normalized HR data for anticipatory emotion.

5.3 Issues in Real-time Assessment of Physiological Data

The psychophysiological literature discusses a range of problems related to the (real-time) assessment of a person's physiological information (Levenson, 1988). Most important among them are the 'Baseline Problem', the 'Timing of Data Assessment Problem', and the 'Intensity of Emotion Problem'.

5.3.1 Baseline Problem

The Baseline Problem refers to the problem of finding a condition against which physiological change can be compared – the baseline. An obvious choice is a 'rest' period where the subject can be assumed to have no particular emotion. However, as Levenson (1988, p. 24) notes, emotion "is rarely superimposed upon a prior state of 'rest'. Instead, emotion occurs most typically when the organism is in some prior activation." Consequently, he suggests to adopt a baseline procedure that generates a moderate level of ANS activity. Hence, in our experiment, we used an initial relaxation period of 40 seconds with a moderate level of ANS activity, where the subject listen to calm music. This procedure guarantees some independence of subjects' individual ANS activity levels as well as independence of situational factors, such as room temperature.

Note that in a more general setting, e.g., when using information about human physiology in a pervasive (ubiquitous, ambient) computing environment (Satyanarayanan, 2001), the assumption of a 'relaxation' period is impractical as users cannot be expected to provide a baseline measurement before entering the environment. Levenson (1988) argues that a no-baseline approach is only viable to the extent that previous data (relating ANS activity to emotions) for a given user exist. Therefore, the use of physiological data in pervasive environments will require a phase of initial tuning for each user.

5.3.2 Timing of Data Assessment Problem

The Timing of Data Assessment Problem refers to the temporal dimension of emotion elicitation, including onset[4] and duration of emotions. Levenson (1988, p. 30) suggests 0.5–4 seconds as an approximation for the duration of emotions, which locates them duration-wise between (orienting) reflexes (i.e. an organism's response to novelty) and moods. In our study, we take values every 50 milliseconds, for a period of five seconds.

The generalization to environments that process an ongoing stream of autonomic activity rather than a specified segment (as in our job interview application) may pose additional challenges for determining the occurrence of emotions correctly. As pointed out in Levenson (1988), when measuring at the wrong time the emotion might be missed or, different emotions might be covered when too long periods are measured. While the ANS is sometimes considered a slow reacting system, latency of onset for autonomic activity related to emotions can be very short, e.g. with surprise. On the other hand, an emotion like anger may build up over time and blur the actual 'start' of the anger emotion.

5.3.3 Intensity of Emotion Problem

The Intensity of Emotion Problem concerns the question how the intensity of an emotion is reflected in the physiological data. While at a low level of emotion intensity no informative ANS activity occurs, a very high intensity level may destroy the pattern of ANS activity associated with an emotion (Levenson, 1988). In our job interview application, negatively valenced emotions (indicated by EMG) would hardly be indicated, whereas different arousal levels (indicated by GSR) could be easily shown. In practice, emotions with little autonomic activity ('relaxed happiness') or moderate intensity levels seem to occur most frequently. To date, issues in emotion intensity remain largely unsolved (Levenson, 1988).

6 Conclusions

This paper describes the Empathic Companion, a character-based interface that takes physiological signals of the user in real-time and addresses user emotions derived from those signals. A virtual job interview serves as an exploratory application scenario. While results of statistical significance could not be obtained for the overall (positive) impact of the Empathic Companion, it could be shown that users receiving empathic feedback are less stressed when being asked an interview question. Although this finding cannot be directly mapped to improving job interview training, we expect that the Empathic Companion will be beneficial in applications where negative feelings are involved or likely to occur, such as online customer support or computer-based education. A particularly interesting application field is tele-home health care where the Empathic Companion may play a crucial role in addressing patients' affective and social needs (Lisetti et al., 2003). In the near future, we plan to implement a more straightforward use of the Empathic Companion, i.e. in the role of the main and only interaction partner of the user rather than a secondary interlocutor as was the case in the interview scenario. Regarding evaluation, we will also consider to use text-based (empathic) feedback in order to compare "persona" vs. "no persona" conditions.

Acknowledgments

This research was supported by the JSPS Research Grant (FY1999–FY2003) for the Future Program.

Notes

[1]This module is used in another application, the *Emotion Mirror*, where the user's emotions are reflected back to the user. The mirror metaphor is realized by an animated interface agent that displays the user's presumed emotion both verbally and non-verbally. The aim of this application is to facilitate the training of emotion management and regulation (Prendinger et al., 2003).

[2]For simplicity, the EMG sensors have not been used.

[3]This equation is a practical version of the range correction proposed by Schandry (1998, p. 199). In order to achieve values that are largely independent of a person's individual skin conductance level, he uses the equation

$$\Phi_{SC_x} = \frac{SC_x - SC_{min}}{SC_{max} - SC_{min}}$$

such that Φ_{SC_x} is the corrected skin conductance value, and SC_x, SC_{min}, and SC_{max} are the current, minimal, and maximal skin conductance value for the subject. As noted by Schandry (1998), the problem of applying this equation is to obtain the SC_{min} and SC_{max} values that require rather severe (obtrusive) tests.

[4]The onset of an emotion indicates how fast an emotion is elicited.

References

André, E., and M. E. Müller (2003). Learning affective behavior. In *Proceedings HCI International 2003 (Vol.4)*, pages 512–516.

Bickmore, T. (2003). *Relational Agents: Effecting Change through Human-Computer Relationships*. PhD thesis, Massachusetts Institute of Technology.

Bosma, W., and E. André (2004). Exploiting emotions to disambiguate dialogue acts. In *Proceedings of the 9th International Conference on Intelligent User Interfaces (IUI-04)*, ACM Press, pages 85–92.

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

Hayes-Roth, B. (2001). Adaptive learning guides. In *Proceedings of the IASTED Conference on Computers and Advanced Technology in Education*.

Jensen, F. (2001). *Bayesian Networks and Decision Graphs*. Springer, Berlin New York.

Johnson, W. L., Rickel, J. W., and J. C. Lester (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11:47–78.

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

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

Levenson, R. W. (1988). Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. In H. L. Wagner, editor, *Social Psychophysiology and Emotion: Theory and Clinical Applications*, pages 17–42. John Wiley & Sons, Hoboken, NJ.

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

Marsella, S., Gratch, J., and J. Rickel (2004). Expressive behaviors for virtual worlds. In H. Prendinger and M. Ishizuka, editors, *Life-like Characters. Tools, Affective Functions and Applications*, Cognitive Technologies, pages 317–360. Springer Verlag, Berlin Heidelberg.

Microsoft (1998). *Developing for Microsoft Agent*. Microsoft Press, Redmond, WA.

Netica (2003). Norsys Software Corp.. URL: `http://www.norsys.com`.

Paiva, A., Dias, J., Sobral, D., and R. Aylett (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-04)*, New York, ACM Press.

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

Prendinger, H., Descamps, S., and M. Ishizuka (2004). MPML: A markup language for controlling the behavior of life-like characters. *Journal of Visual Languages and Computing*, 15(2):183–203.

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

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

Reeves, B., and C. Nass (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.

Satyanarayanan, M. (2001). Pervasive computing: Vision and challenges. *IEEE Personal Communications*, pages 10–17. August.

Schandry, R. (1998). *Textbook Psychophysiology. Bodily Indicators of Psychological Events.* (3rd Edition). Beltz Psychologie Verlags Union, Weinheim, Germany. (In German).

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

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

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