

# Recognizing, Modeling, and Responding to Users' Affective States

Helmut Prendinger<sup>1</sup>, Junichiro Mori<sup>2</sup>, and Mitsuru Ishizuka<sup>2</sup>

<sup>1</sup> National Institute of Informatics  
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan  
`helmut@nii.ac.jp`

<sup>2</sup> Dept. of Information and Communication Engineering, University of Tokyo  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan  
`ishizuka@miv.t.u-tokyo.ac.jp`

**Abstract.** We describe a system that recognizes physiological data of users in real-time, interprets this information as affective states, and responds to affect by employing an animated agent. The agent assumes the role of an Empathic Companion in a virtual job interview scenario where it accompanies a human interviewee. While previously obtained results with the companion with were not significant, the analysis reported here demonstrates that empathic feedback of an agent may reduce user arousal while hearing interviewer questions. This outcome may prove useful for educational systems or applications that induce user stress.

## 1 Introduction

Computers sensing users' physiological activity are becoming increasingly popular in the human-computer interface and user modeling communities, partly because of the availability of affordable high-specification sensing technologies, and also due to the recent progress in interpreting physiological states as affective states or emotions [10]. The general vision is that if a user's emotion could be recognized by the 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 particular interest concerns interfaces that employ animated or embodied agents as interaction partners of the user. By emulating multi-modal human-human communication and displaying social cues including (synthetic) speech, communicative gestures, and the expression of emotion, those agents may trigger social reactions in users, and thus implement the "computers as social actors" metaphor [14]. This type of social and affect-aware interface has been demonstrated to enrich human-computer interaction in a wide variety of applications, including interactive presentations, training, and sales [2, 12].

In this paper, we propose an interface that obtains information about a user's physiological activity in real-time and provides affective feedback by means of an

embodied agent. The interface is intended to respond to the user’s emotion by showing concern about user affect, sometimes called *empathic* (or *sympathetic*) behavior. Empathic interfaces may leave users less frustrated in the case of a stressful event related to the interaction [5]. Potential application fields include software (assuming unavoidable software-related failures), computer-based customer support, and educational systems. The web-based (virtual) job interview scenario described here 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. In this paper, we will extend and complement our previous investigations on empathic agents.

- *Virtual Quizmaster*. An agent providing empathic feedback to a deliberately frustrated user can significantly reduce user arousal or stress when compared to an agent that ignores the user’s frustration [13].
- *Empathic Companion*. An empathic agent has no overall positive effect on the user’s interaction experience in terms of lower levels of arousal [11].

The rest of this paper is organized as follows. In Sect. 2, we describe related work. Section 3 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. 3 discusses the decision-theoretic agent that is responsible for selecting the Empathic Companion’s actions. In Sect. 4, we illustrate the structure an interaction with the Empathic Companion in the setting of a virtual job interview, and provide new results of an experiment that recorded users’ physiological activity during the interaction. Section 5 concludes the paper.

## 2 Related Work

There are various research strands that share the methodology and motivation of our approach to affective and empathic interfaces. The tutoring system developed by Conati [3] demonstrates that the user’s physiological state can play a key role in selecting strategies to adapt an educational interface. When the user’s frustration is detected, an interface agent can try to undo the user’s negative feeling. Bickmore [1] investigates empathic agents in the role of health behavior change assistants that are designed to develop and maintain long-term, social-emotional relationships with users, so-called ‘relational agents’.

The investigation of Klein et al. [5] 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 into the course of the game. The Empathic Companion interface differs from the one used in [5] in two aspects. First, the user in our system is given feedback in a more timely fashion, i.e. shortly after the emotion



**Fig. 1.** Job Interview Scenario.

actually occurs, and not after the interaction session, in response to the subject's questionnaire entries. While providing immediate response to user affect 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 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 an embodied agent, rather than a text message. Although the study of Klein and co-workers [5] 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 [12].

### 3 The Empathic Companion

The Empathic Companion is an embodied agent 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. 1). Being interviewed is likely to elicit emotions in the user, especially when the interviewer (Fig. 1, 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. 1, 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

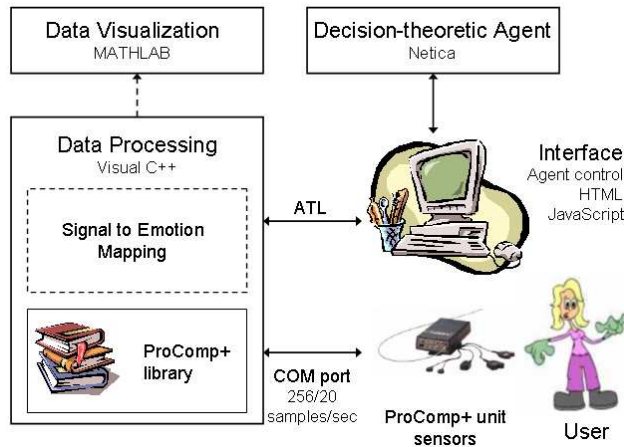


Fig. 2. System architecture.

stress level of a face-to-face or phone interview, it provides a convenient training platform for job seekers.

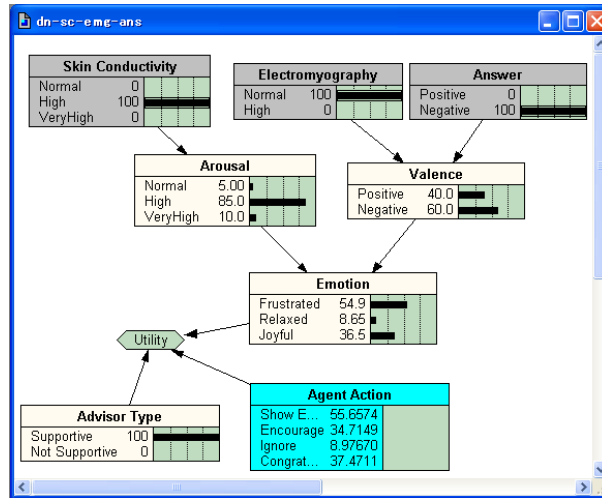
### 3.1 System Architecture

Since the Empathic Companion application assumes real-time emotion recognition, the system architecture depicted in Fig. 2 has been implemented on the Windows XP platform.

**Data Capturing and Processing.** The user is attached to sensors of the ProComp+ unit [15]. 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. Data capturing is achieved by a module written in Visual C++ that employs the ProComp+ data capture library.

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 physiological activity are computed by comparing the current mean signal values to the baseline value. The baseline is obtained during a relaxation period preceding the interaction. The current mean value is derived from a segment of five seconds, the average duration of an emotion [7]. 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.

**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 [8] to control the verbal and non-verbal behavior (such as gestures or facial displays) of characters.



**Fig. 3.** Simple decision network.

This package includes an animation engine to trigger about 50 pre-defined 2D animation sequences and a text-to-speech engine.

**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 (see Fig. 3). The decision-theoretic agent will be discussed in Sect. 3.3. Before that, we will explain the modeling and interpretation of the user's physiological activity as emotions.

### 3.2 Relating Physiological Signals to Emotions

Lang [6] claims that all emotions can be characterized in terms of judged valence (pleasant or unpleasant) and arousal (calm or aroused), and therefore named emotions can be identified as coordinates in the arousal–valence space. For instance, the “Angry” emotion is located in the ‘arousal=high’ & ‘valence=negative’ segment. The relation between physiological signals and arousal/valence is established in psychophysiology that argues that the activation of the autonomic nervous system changes while emotions are elicited [7]. The following two signals have been chosen for their high reliability (other signals are discussed, e.g. in [10]). Galvanic skin response (GSR) is an indicator of skin conductance (SC), and increases linearly with a person’s level of overall arousal. Electromyography (EMG) measures muscle activity and has been shown to correlate with negatively valenced emotions [6].

### 3.3 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 [9], a software package that allows solving decision problems and provides convenient tools, including an API in Java.

The decision network depicted in Fig. 3 represents a simple decision problem. A decision-theoretic agent selects actions that maximize the outcome in terms of some utility function [4]. 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 [6]. The probabilities have been set in accord with the literature (whereby the concrete numbers are made up). “Relaxed (happiness)” is defined by the absence of autonomic signals, i.e. no arousal (relative to the baseline), and positive valence. “Joyful” is defined by increased arousal and positive valence, whereas “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 utility function is set to the effect that negatively aroused users receive empathic feedback, by assuming that negative states are a hindrance to performing successfully in stressful situations (including interviews).

Table 1 lists some responses of the Empathic Companion associated to the action types (see also Fig. 3). 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 respond to the user’s affective state. “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.

## 4 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.

**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.).

#### 4.1 The Structure of the Interview

An interview session is composed of (interview) episodes, whereby each episode consists of four segments (see below). 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.

- *Segment 1:* The interviewer agent asks a question, e.g. “Tell me about your previous work experience”.
- *Segment 2:* The user chooses an answer from the set of given options (see Fig. 1, 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.
- *Segment 3:* The interviewer responds to the user’s answer, e.g. “Then you are not the kind of person we are looking for” or “I am happy to hear that you have extensive experience in the field”.
- *Segment 4:* The companion agent responds to the emotion derived from the data gathered during the third segment and the user’s answer given in the second segment.

#### 4.2 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. A signal processor has been developed in-house that reads users’ skin conductance (SC) and heart rate (HR). Like EMG, HR also correlates with negative emotions and Lang’s [6] model can be applied.

Observe that unlike the experiment reported in [13], tight experimental controls are not practicable in the job interview application as the interaction 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 [16] thus propose to compare signal values for whole interaction periods rather than for specific interface events.

Following this paradigm, we *initially* hypothesized that, averaged over the entire interview period (see Fig. 4), 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 was used, where the “Ignore” action is always selected. However, no significant results could be obtained.

The main reason might be that the user interacts with the interviewer rather than the empathic agent most of the time. Other possible reasons include: (i) The responses intended to have a calming effect on the user might actually not do so; (ii) heart rate might not be a reliable indicator of negative valence for all users; (iii) a measurement spanning the whole interaction period is too coarse.

Extending the analysis of [11], a more fine-grained data analysis has been carried out, 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 compared the normalized SC and HR data from the period when the interviewer asks the question (Segment 1) for the “Supportive” and “Not Supportive” versions of the Empathic Companion application, abbreviated as *Em* and *NEm*, respectively.

In the study subjects are connected both to the GSR sensors of the Pro-Comp+ unit with the first two fingers of their non-dominant hand,<sup>3</sup> and to our in-house encoder that provides a wristband for SC and an ear-clip to measure HR. Participants were 10 staff and students from the University of Tokyo, aged 23–40, who were randomly assigned to the two versions (5 subjects in each).

The result for SC is depicted in Fig. 5. Here the *t*-test (two-tailed, assuming unequal variances) demonstrates a significant effect of the companion in the “Supportive” (*Em*) version ( $t(8) = -5.49$ ;  $p = 0.0002$ ). The companion display-

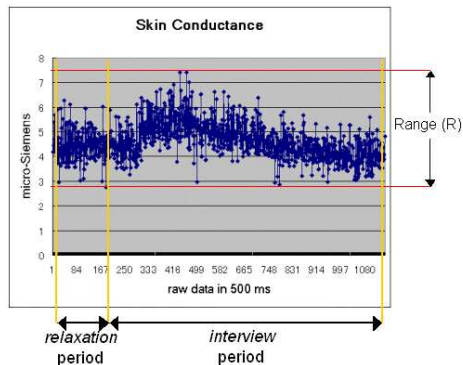
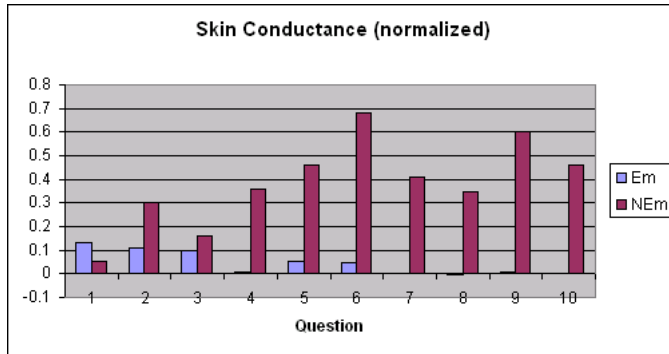


Fig. 4. SC data of one subject.

<sup>3</sup> For simplicity, the EMG sensors have not been used.





**Fig. 5.** Normalized SC data for anticipatory emotion.

ing 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 even higher in the *Em* version than in the *NEm* version (contrary to our expectation). Hence the core finding of the experiment can be stated as: *Users (as interviewees) receiving empathic feedback are significantly less aroused when hearing the interviewer’s questions, independently of whether they are able to give a satisfying answer or not.*

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

## 5 Conclusions

This paper describes the Empathic Companion, an animated agent based interface that takes physiological signals of the user in real-time, models and interprets them as affective states, and addresses user emotions derived from those signals. A virtual job interview serves as an exploratory application that can be seen as an instance of stress-inducing interaction scenarios such as educational or training interfaces.

While results of statistical relevance of the Empathic Companion could not be obtained for the whole interview period [11], interestingly, empathic feedback is shown to have a significant impact on users’ arousal level while being queried. This result relates to the importance of the timing of data assessment in emotion recognition [7] – when an emotion occurs. The Empathic Companion only responds to a user’s emotional reaction that happens when the interviewer responds to the user’s answer to the interviewer question. A more advanced system, however, must allow to react user emotion in a more flexible manner.

The study described in this paper was not designed to test a model of emotion recognition but employed a widely used two-dimensional emotion theory [6].

Future work will be directed toward a richer emotion model that takes into account situational and task-specific parameters of emotion elicitation during human-computer interaction.

**Acknowledgments.** This research is partly supported by the JSPS Research Grant (1999-2003) for the Future Program and a Memorandum of Understanding between the National Institute of Informatics and the Univ. of Augsburg.

## References

1. T. Bickmore. *Relational Agents: Effecting Change through Human-Computer Relationships*. PhD thesis, Massachusetts Institute of Technology, 2003.
2. J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors. *Embodied Conversational Agents*. The MIT Press, Cambridge, MA, 2000.
3. C. Conati. Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16:555–575, 2002.
4. F. Jensen. *Bayesian Networks and Decision Graphs*. Springer, Berlin New York, 2001.
5. J. Klein, Y. Moon, and R. Picard. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14:119–140, 2002.
6. P. J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.
7. R. W. Levenson. Autonomic specificity and emotion. In R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, editors, *Handbook of Affective Sciences*, pages 212–224. Oxford University Press, Oxford, 2003.
8. Microsoft. *Developing for Microsoft Agent*. Microsoft Press, Redmond, WA, 1998.
9. Norsys Software Corp. Netica, 2003. URL: <http://www.norsys.com>.
10. R. W. Picard. *Affective Computing*. The MIT Press, 1997.
11. H. Prendinger, H. Dohi, H. Wang, S. Mayer, and M. Ishizuka. Empathic embodied interfaces: Addressing users' affective state. In *Tutorial and Research Workshop on Affective Dialogue Systems*, LNAI 3068, pages 53–64, Berlin Heidelberg, 2004. Springer.
12. H. Prendinger and M. Ishizuka, editors. *Life-Like Characters. Tools, Affective Functions, and Applications*. Cognitive Technologies. Springer Verlag, Berlin Heidelberg, 2004.
13. H. Prendinger, J. Mori, and M. Ishizuka. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies*, 62(2), 2005.
14. B. Reeves and C. Nass. *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, 1998.
15. Thought Technology Ltd., 2002. URL: <http://www.thoughttechnology.com>.
16. R. Ward and P. Marsden. Physiological responses to different WEB page designs. *International Journal of Human-Computer Studies*, 59:199–212, 2003.