PAPER *Special Issue on Software Agent and its Applications*

Character-based Interfaces Adapting to Users' Autonomic Nervous System Activity

Helmut PRENDINGER†, **Junichiro MORI**†, **Sonja MAYER**††, ***Nonmembers***, *and* **Mitsuru ISHIZUKA**†, ***Regular Member***

SUMMARY This paper explores some possibilities of using real-time processing of users' physiological signals in order to improve the interaction with character-based interfaces. Specifically, interfaces that adapt to or reflect a user's affective state as derived from the user's autonomic nervous system activity will be discussed. The 'Emotion Mirror' is an example of a system where the user's emotions are reflected back to the user, and hence allows to train emotion management and regulation. The 'Emphatic Companion' is an extension of the previous system employing a decision-theoretic embodied agent that adapts its behavior depending the recognized emotional state of the user, e.g. by giving support and encouragement. A simple virtual job interview scenario has been implemented to illustrate the potential of considering physiological data of users in real-time.

key words: *Affective Computing, adaptive user interfaces, real-time emotion recognition, autonomic nervous system (ANS), bio-signals, life-like characters, decision networks, web-based environments.*

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 (Picard [13]). 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 more appropriate ways.

Our particular interest concerns interfaces that feature an embodied agent, or life-like character, as an interaction partner of the user. This type of 'social interface' has been shown to enrich human-computer interaction in a wide variety of applications (see Prendinger and Ishizuka [15] for a recent overview). We may distinguish at least four modes of usage of a user's ANS activity for character-based interfaces.

First, a user's physiological data can be used to *track* the impact of a character-based interface on the affective state of the user.* We recently conducted an experimental study which shows that emphatic agent behavior after the occurrence of a frustrating event significantly decreases the arousal level of subjects (Prendinger et al. [16]). Recording users' physiological data and associating them with computer interaction events might become an important technology for testing software.

Second, 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. This type of application bears considerable relevance for the medical sector, including biofeedback, e-Healthcare, and tele-home care which includes distant monitoring of patients' well-being (Lisetti et al. [9]).

Third, the user's physiological state can play a key role to select strategies to *adapt* the interface to the user. 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 according to the student's affective state and the learning goal (Conati [3]).

Forth, a user's physiological responses may become increasingly important to *learn* the user's affective situated responses and hence allow for the acquisition of predictive user models (André and Müller [1]). Machine learning of emotional behavior is also crucial for 'relational agents' that are intended to enable fertile interactions with human users over extended periods of time, e.g. as health behavior change assistants (Bickmore [2]).

In this paper, we will focus on character-based interfaces that process physiological data in *real-time* and hence discuss applications where a life-like character reflects or adapts to the autonomic nervous system activity of the user. In particular, we will describe the Emotion Mirror, a character-based job interview scenario where the user's affective state is 'mirrored' to

†The authors are with the Department of Information and Communication Eng., Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.

††The author is an internship student visiting the first author, supported by a grant of INWENT GmbH, Germany (formerly called Carl Duisberg Society).

*Although we currently focus on character-based interfaces, our ideas can be generalized to any type of human-computer interaction, including human-robot interaction.

the user in the role of an interviewee. An extension of that scenario will serve as an example of an adaptive interface, where a character adjusts its behavior depending on the recognized emotions of the user in order to maximize the expected utility of its response, e.g. as a support-oriented interface agent.

The remaining paper is organized as follows. In the following section, we will explain the meaning of physiological signals and relate them to a person's emotional state. In Section 3, the Emotion Mirror application and an architecture for real-time emotion recognition will be described. Section 4 introduces the Emphatic Companion application, which extends the Emotion Mirror by a providing the Emotion Mirror agent with utility-based response selection. Finally, Section 5 summarizes and concludes the paper.

2. Background

While a user's goals, beliefs and competence received considerable attention in the human-computer interface literature, there is still few work to date that discusses 'extra-rational' factors of communication, such as affective states and attitudes. This section reports on the physiological basics of emotions, by first describing features of physiological signals and then relating them to emotions.

2.1 Physiological Signals

An essential function of emotion is organization. Emotions can mobilize and arrange disparate response systems of the human body in order to cope with environmental events that menace survival. Emotions are a product of Darwinian evolution as a means to coordinate fast response in a life-threatening situation. Since emotions have been mainly targeted at flight or fight situations, negative feelings can be seen as more 'natural' while the function of (the small set of) positive emotions is primarily to reset the body after negative emotion mobilization.

Psychophysiology argues that the activation of the autonomic nervous system (ANS) changes while emotions are elicited (Levenson [8]). In the following, we will discuss some of the relevant physiological signals.

Galvanic skin response (GSR) is an indicator of skin conductivity. Under certain circumstances, the glands in the skin produce ionic sweat which change the electrical resistance. By passing small voltage across two electrodes, the conductance can be measured between them. The electrodes can be attached to two fingers. Skin conductance increases with a person's overall level of arousal (Lang [7]).

Electromyography (EMG) measures muscle activity by detecting surface voltage that occurs when the tiny muscle fibers are contracted by means of electrical impulses. Since all the muscle fibers in the area contract, the signal is composed of a constantly varying difference of potentials between its positive and negative electrodes. In order to minimize the variations, a sensor is attached to each head of the muscle, and the ground electrode is placed on the muscle belly. The sensors can be attached to the masseter muscle (see Fig. 1), where the voltage increases when the teeth are clenched, e.g. when a person is angry, but also with laughter and other facial movements.

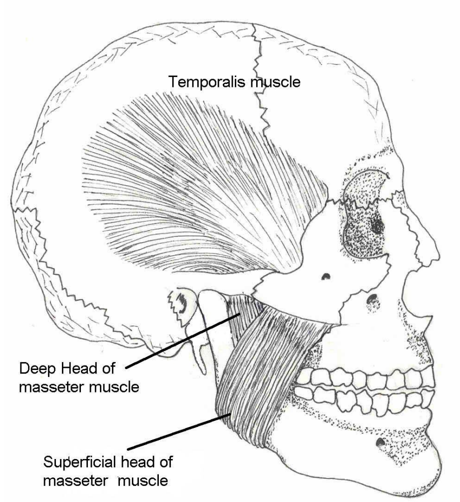

Fig. 1 Masseter muscle (figure borrowed from [10]).

Mean facial muscle activity has been shown to correlate with negatively valenced emotions (Lang [7], Healey [6]).

Blood volume pulse (BVP) indicates blood flow. The BVP sensor is a device that emits light and measures the amount of light reflected back. The sensor can be attached to the finger tip, but reliability is low since the data assessment is very sensitive to movement. Blood volume pulse and heart rate, which can be computed from BVP, have been shown to increase with negatively valenced emotions, such as anxiety and fear (Healey [6]).

Electrocardiography (ECG) is used in medicine as a as method to diagnose heart diseases. The ECG signal measures electrical impulses of the heart muscle by detecting voltage on the surface of the upper body. By employing appropriate filters [6], the heart rate can be obtained. Therefore, similar to BVP, the ECG signal correlates with negative valenced emotions.

Electroencephalogram (EEG) detects small electrical voltages generated by neurons. In order to measure overall brain activity, 128 electrodes have to be placed on the surface of the skull. EEG is a good indicator of a person's arousal.

Table 1 Sensors, physiological signals and some of their features.

Sensor	GSR	EMG	BVP	ECG	EEG
Sensor placement	finger	masseter muscle (among other places)	finger	upper body	skull
Reliability of sensor	high	medium-high	low (sensitive to movements)	high	medium (artefacts)
Measured signal	skin conductivity	electrical pulses of the muscle	heart rate	electrical impulses of the heart	electrical impulses of the brain
Interpretation of signal	simple	simple	simple	filtering necessary to retrieve heart rate	difficult
Signal influenced by	arousal	valence	valence	valence	arousal and valence

Respiration is sensed by using a stretchable belt worn around a person's chest. It measures the expansion and contraction of the rib cage. The rate of respiration is influenced by arousal (besides physical activity).

Temperature is a slow signal that measures peripheral skin temperature. Increased body temperature can be associated with arousal.

Signals and their features most relevant for emotion recognition are summarized in Table 1.

2.2 From Signals to Emotions

Given that we obtained meaningful physiological signals from a person, the question is how to relate them to that person's affective state. We currently use the emotion model of Lang [7] which characterizes emotions in terms of valence and arousal and thus allows to derive emotions from physiological variables. Figure 2 shows some named emotions as coordinates in the valence-arousal space.

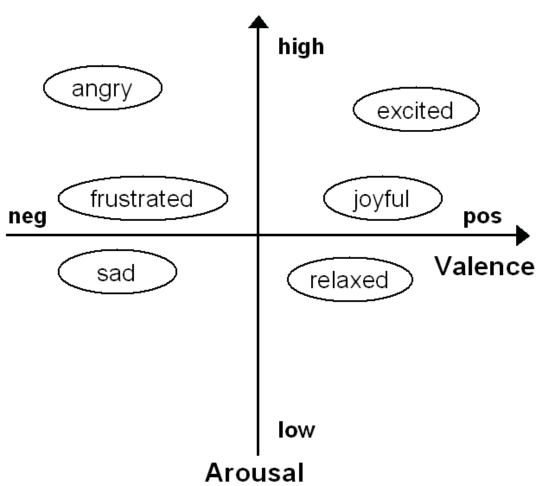

Fig. 2 Some named emotions in the valence-arousal space.

Advanced learning methods for feature-based recognition of emotional states from physiological signals can be found in the work of Picard et al. [14] and Lisetti et al. [9].

3. The Emotion Mirror

This section describes the Emotion Mirror, a character-based application aimed at training interpersonal communication skills known as *emotional intelligence* (Goleman [5]), specifically the abilities to be aware of and to regulate one's emotions. A job interview situation is one example where emotional intelligence is beneficial, as the interviewee has to manage his or her emotions when confronted with unpleasant and probing questions of the interviewer.† Since physiological manifestations of stress may reflect negatively on the interviewer's impression of the interviewee, a virtual job interview alerting the user (as interviewee) about his or her arousal level might serve as a valuable preparatory training environment. The Emotion Mirror application assumes that users are biased to conceive life-like characters as veritable social actors (the 'Media Equation' [17]), and hence *actually* get aroused when interviewed by a virtual agent.††

The job interview scenario features two life-like characters, the interviewer to the left and the 'Emotion Mirror Agent' to the right (see Fig. 3). Users in the role of interviewees are attached to the physiological sensors and the emotions recognized during the interview are shown to them by the Emotion Mirror agent.

Below, we will first discuss the system architecture of the Emotion Mirror and then describe an interaction session.

3.1 System Architecture

In order to realize online emotion recognition, the following architecture has been implemented on Windows

†A job interview scenario featuring an 'affective mirror' has been suggested by Picard [13, p. 86], but to our knowledge, it was never implemented.

††It is certainly true that an online interview cannot induce the stress level of a face-to-face or phone interview.

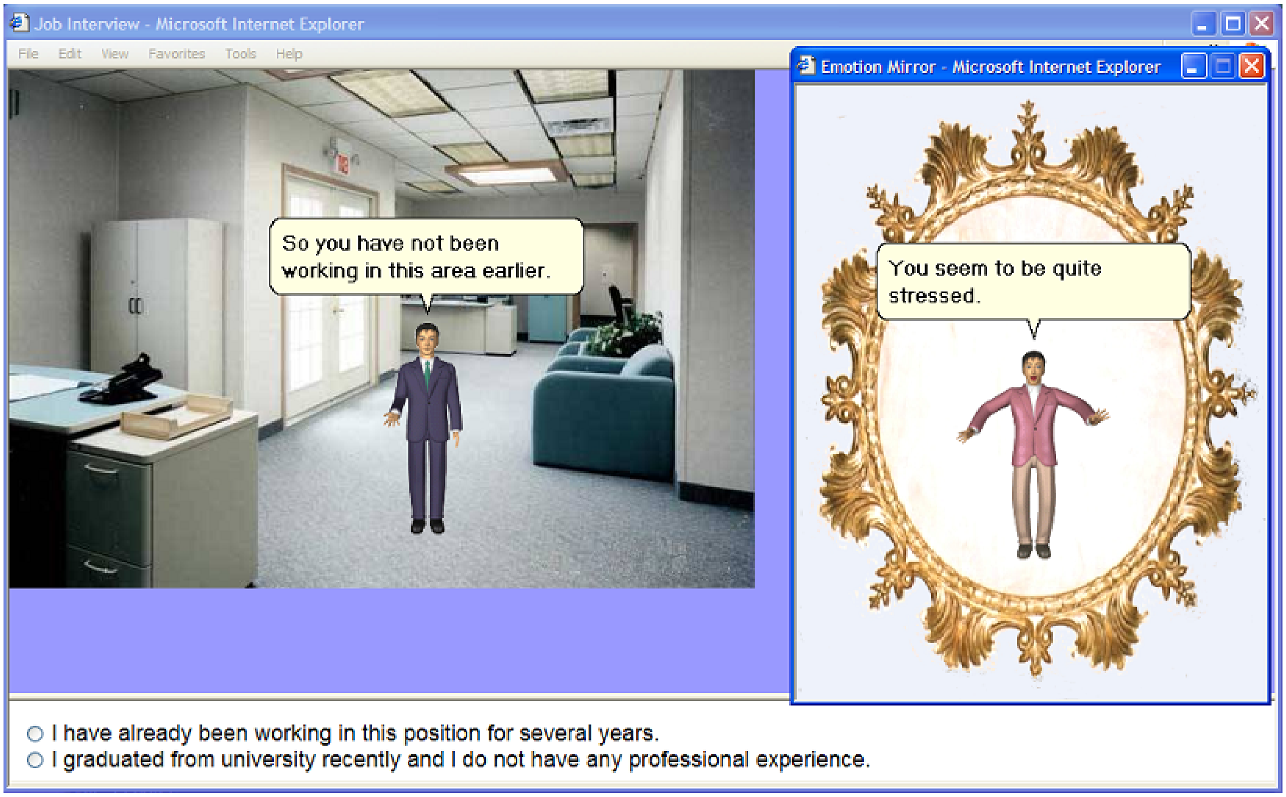

Fig. 3 Job Interview Scenario with Emotion Mirror.

XP and Windows Professional platforms (see Fig. 4).

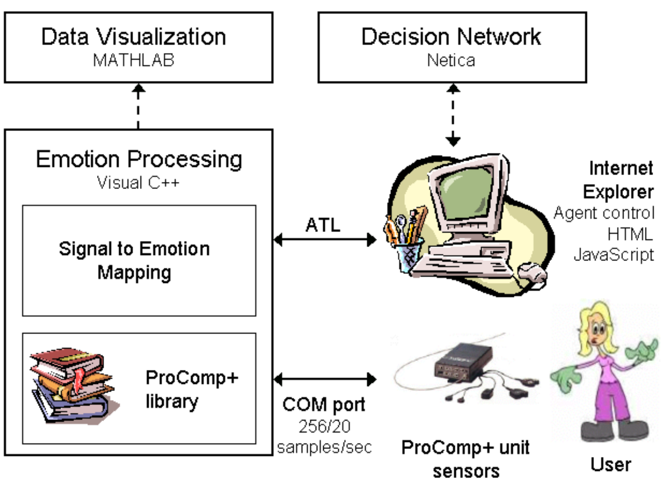

Fig. 4 System Architecture.

Data Capturing The user is attached to sensors of the ProComp+ unit from Thought Technology Ltd. [19]. 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/sec, GSR and EMG signals allow for a much lower rate, at 20 samples/sec.

Emotion Processing The emotion processing component retrieves new data every 50 milliseconds, evaluates them, and outputs an emotion. Data capturing is achieved by a module written in Visual C++ that employs the the ProComp+ data capture library. Data evaluation uses a simple Lang [7] mapping from physiological signal change to emotions. GSR (indicating arousal) and EMG (indicating valence) are used to derive the five emotions 'angry', 'frustrated', 'relaxed', 'joyful', and (positively) 'excited' as follows (see Table 2):

Table 2 Signal to emotion mapping.

Current average SC related to average SC of baseline	Current average EMG related to average EMG of baseline	Emotion
>130%	>300%	angry
>115% and <130%	>300%	frustrated
<115%	<300%	relaxed
>115% and <130%	<300%	joyful
>130%	<300%	excited

The connection between the Emotion Processing component and the User Interface is established 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). Hence, data capturing is initialized by events from the user interface, which will be described next.

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/JavaScript and utilizes the Microsoft Agent package [11] to control the behavior of animated agents including the text-to-speech engine.

Data Visualization In order to visualize the physiological signals obtained during an interview session, the data can be (optionally) exported to MATHLAB.

Decision Network Instead of computing emotions in the Emotion Processing component, a decision network allows for advanced (probabilistic) forms of combining signals and relate them to emotions as well as agent decisions. We will describe this component in more detail in the section on the emphatic companion.

3.2 Interacting with the Emotion Mirror

In an interaction session with the Emotion Mirror, the user is seated in front of a computer, with the GSR sensors attached to two fingers of the non-dominant hand, and the EMG sensors attached to the masseter muscle. The baseline for subsequent bio-signal changes is obtained during an initial relaxation period of 40 secs. where the user listens to music from Cafe del Mar (Vol. 9), as the average of GSR and EMG values.

An interview episode consists of four segments:

(i) The interviewer character asks a question;
(ii) The user selects an answer from a set of given options (the lower part in Fig. 3);
(iii) The interviewer responds to the user's answer;
(iv) the Emotion Mirror agent displays the user's emotion calculated from the data gathered during segment (iii).†

More precisely, signal values are taken every 50 milliseconds, for a period of 5 seconds. The psychophysiological literature, e.g. Levenson [8, p. 30], suggests 0.5–4 secs. as an approximation for the duration of an emotion.

†The assumption here is that the Emotion Mirror's response to the user's answer would elicit the user's emotion. However, it might well be the case that signals are better taken in segment (i), as an 'anticipatory' form of emotion.

When the (average) GSR signal is 15–30% above the baseline, the user's arousal level is assumed as 'high' (see also Table 2). If the signal is on average higher than 30%, the user is assumed to be very aroused. In case the (average) EMG signal is more than three times higher than average, the user can be assumed to experience a negative emotion.

For instance, if the recognized emotion is 'anger', the Emotion Mirror Agent would display a gesture expressing anxiety and utter, e.g.,"You seem to be quite stressed". Obviously, 'anger' and 'fear' are quite different emotions. However, in our model, they have the same features, high arousal and negative valence.†† Consequently, the emotion expressed by the agent refers to the interaction context, which makes a 'fear' emotion more likely than an 'anger' emotion.

Currently, we did not gain enough experiences with the Emotion Mirror in order to make sound propositions about its usability. However, it could be shown that people *do* get aroused for some questions. On the other hand, changes in valence are rare, possibly the threshold of an increase by more than 300% is set too high.

4. The Emphatic Companion

An agent that simply reflects the physiological state of the user will be too restrictive for most applications. Ideally, the agent recognizes the user's emotions and responds depending on the recognized emtion *and* in accord to some higher level communication goal, such as increased user motivation, improved learning, or also an improved interaction experience with the computer. An application where the interface *adapts* to the user according to some communication goal requires more sophisticated reasoning capabilities of the agent.

In this section, we will briefly introduce the Emphatic Companion agent that has an explicit representation of the utility of its actions. This type of agent is often called decision theoretic agent (Russell and Norvig [18]). A decision theoretic agent selects actions that maximize the outcome in terms of some utility function. Formally, the reasoning of decision theoretic agents is based on decision networks (also called 'influence diagrams'), which are an extension of Bayesian networks (Pearl [12]). 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 network depicted in Fig. 5 represents the decision problem of the Emphatic Companion agent.

Chance nodes 'Skin Conductivity', 'Electromyogra-

††Ekman et al. [4] show, however, show that skin temperature is a discriminating factor between anger and fear.

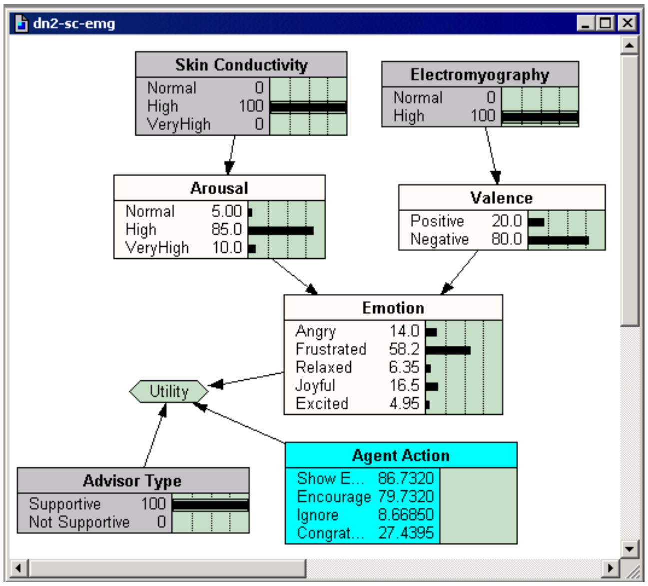

Fig. 5 Simple decision network.

phy', 'Arousal', 'Valence', 'Emotion', and 'Advisor Type'††† are chance nodes. The values for GSR and EMG are retrieved from the Emotion Recognition component. The advisor type can be set to either "supportive" or "not supportive".

Decision nodes The agent's can either show empathy, encourage the user, ignore the user's emotion, or congratulate.

Utility node This node defines the utility of performing an action, given advisor type and user emotion.

Let us assume GSR and EMG are both high (see Fig. 5). Then the most likely user emotion is 'frustration'. Since the advisor type is supportive (an emphatic companion), the action with maximal expected utility is to show empathy. On the other hand, if the advisor type is not supportive, then ignoring the user's emotion will maximize the actions expected utility.

5. Conclusions

This paper describes some possibilities to consider users' autonomic nervous system activity in character-based interfaces. After reporting on the meaning of various physiological signals, an architecture for real-time emotion recognition in web-based environments is explained and illustrated by two preliminary demonstrator system, the Emotion Mirror and the Emphatic Companion. While the Emotion Mirror agent simply reflects the user's emotions, the Emphatic Companion adapts its behavior depending on the user's emotions and its value function.

†††This is a deterministic node rather than a chance node.

Acknowledgement

This research is supported by the JSPS Research Grant (1999-2003) for the Future Program ("Mirai Kaitaku").

References

[1] Elisabeth Andrè and Martin E. Müller. Learning affective behavior. In *Proceedings HCI International 2003 (Vol.4)*, pages 512–516, 2003.
[2] Timothy Bickmore. *Relational Agents: Effecting Change through Human-Computer Relationships.* PhD thesis, Massachusetts Institute of Technology, 2003.
[3] Cristina Conati. Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16:555–575, 2002.
[4] Paul Ekman, Robert W. Levenson, and Wallace V. Friesen. Autonomic nervous system activity distinguishes among emotions. *Science*, 221:1208–1210, 1983.
[5] Daniel Goleman. *Emotional Intelligence.* Bantam Books, New York, 1995.
[6] Jennifer A. Healey. *Wearable and Automotive Systems for Affect Recognition from Physiology.* PhD thesis, Massachusetts Institute of Technology, 2000.
[7] Peter J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.
[8] Robert W. Levenson. 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, 1988.
[9] C. Lisetti, F. Nasoz, C. LeRouge, O. Ozyer, and K. Alvarez. Developing multimodal intelligent affective interfaces for tele-home health care, 2003. International Journal of Human-Computer Studies.
[10] URL: `home.teleport.com/~bobh/masseter.htm` (July '03).
[11] Microsoft. *Developing for Microsoft Agent.* Microsoft Press, 1998.
[12] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, Santa Mateo, 1988.
[13] Rosalind W. Picard. *Affective Computing.* The MIT Press, 1997.
[14] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001.
[15] Helmut Prendinger and Mitsuru Ishizuka, editors. *Life-like Characters. Tools, Affective Functions and Applications.* Cognitive Technologies. Springer Verlag, 2003.
[16] Helmut Prendinger, Sonja Mayer, Junichiro Mori, and Mitsuru Ishizuka. Persona effect revisited. Using bio-signals to measure and reflect the impact of character-based interfaces. In *Fourth International Working Conference on Intelligent Virtual Agents (IVA-03)*, 2003.
[17] Byron Reeves and Clifford 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.
[18] Stuart J. Russell and Peter Norvig. *Artificial Intelligence. A Modern Approach.* Prentice Hall, Inc., Upper Saddle River, New Jersey, 1995.
[19] Thought Technology Ltd., 2002.
URL: `http://www.thoughttechnology.com`.