Physiologically Interactive Gaming with the 3D Agent Max*

Christian Becker[1], Arturo Nakasone[2], Helmut Prendinger[3], Mitsuru Ishizuka[2], Ipke Wachsmuth[1]

[1] Faculty of Technology, University of Bielefeld, 33594 Bielefeld, Germany
[2] Dept. 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
[3] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

Abstract. Physiologically interactive (or affective) gaming refers to research on the evocation and detection of emotion during game play [21]. In this paper, we first describe the two building blocks of our approach to affective gaming. The building blocks correspond to two independently conducted research strands on affective human–computer interaction: one on an emotion simulation system for an expressive 3D humanoid agent called *Max*, which was designed at the University of Bielefeld [13, 2]; the other one on a real-time system for empathic (agent) feedback that is based on human emotional states derived from physiological information, and developed at the University of Tokyo and the National Institute of Informatics [19]. Then, the integration of both systems is motivated in the setting of a cards game called *Skip-Bo* that is played by a human game partner and Max. Physiological user information is used to enable empathic feedback through non-verbal behaviors of the humanoid agent Max. With regard to the new area of Conversational Informatics we discuss the measurement of human physiological activity in game interactions and non-verbal agent behavior.

1 Introduction and Motivation

In the growing field of Embodied Conversational Agents (in short, ECAs) [7, 20], techniques from AI (artificial intelligence), computer animation, and human–computer interaction are combined to create a synthetic communication partner. ECAs are computer-generated, humanoid characters that are able to conduct a natural face-to-face dialogue with a human user (e.g. [8, 9]). The types of communication channels range from pure textual input to multi-modal speech–gesture interfaces [14]. In addition to the different (standard) cognitive components, such an agent is often equipped with an emotional component in order to increase its believability. Moreover, research in the field of machine emotion recognition is offering initial results on the perception and interpretation of different kinds of user feedback on the basis of the very same psychological theories. Hence, by adjusting the agent's behavior with respect to both – its own as well as the interlocutor's emotional state – the agent may adapt to the cognitive and affective state of the human interlocutor and therefore may be experienced as a more sensible and trustworthy interaction partner.

Affective gaming is a very young research field that aims at creating a new type of gaming experience by adapting the game development to the human player's affective state. Gilleade et al. [11] propose three high-level design heuristics for affective games:

– *Assist me*: Games that detect user frustration and aim at providing assistance within the game course to avoid user frustration.
– *Challenge me*: Games that detect the user's arousal and adapt the difficulty level accordingly in order to prevent user boredom or an overcharged user state.
– *Emote me*: Games that track the user's emotional response to specific game content and adjust subsequent content to achieve a high level of 'dramatic' user experience.

From the perspective of emotion recognition technology, affective gaming is closely related to biofeedback systems, where people 'learn' to control physiological activity such as blood flow, muscular action, or brain waves by being provided with real-time graphical representations of their biometric state. Most prominently, the "Relax-to-Win" game is a competitive two player racing game where a player's level of arousal, derived from the player's skin conductance, determines the speed of a dragon figure [4]. The more the player relaxes, the faster the dragon will move, and thus win the race. The "Mind Balance" game uses brain waves (Visually Evoked Potentials, or VEPs) in order to assist a graphical tightrope walker to keep its balance [6]. Stevenson [21] uses pupil dilation (with co-recorded electrodermal activity) to measure player arousal during game play. On the other hand, some researchers argue that biometric

* This paper is an extended version of Becker et al. [3].

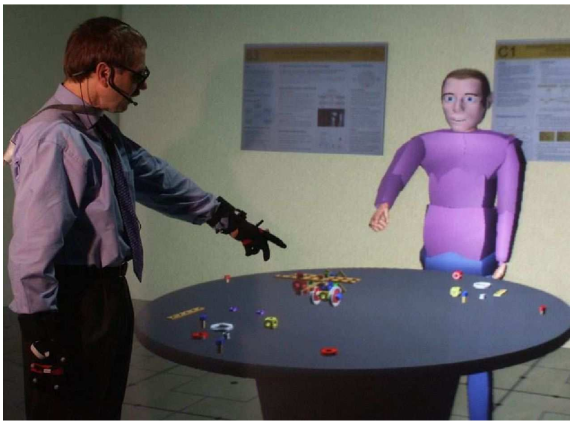

Fig. 1. The Embodied Conversational Agent Max in a Virtual Reality setting.

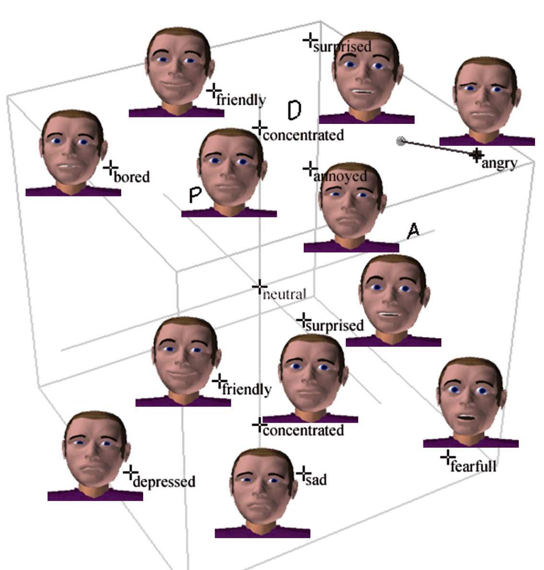

Fig. 2. Named emotions with corresponding facial displays in the pleasure-arousal-dominance space.

information is easily corrupted by artifacts and thus propose a non-physiological approach to emotion (arousal) detection, e.g. button pressure, game pad grip or tilt, and vibration of the game pad [22, 10].

In this paper, we will describe our own efforts towards the realization of affective gaming. Section 2 reports on Max and the emotion simulation system used by Max [2]. Section 3 describes the emotion recognition module and its use within the game application. In Section 4, we develop our ideas of a physiologically interactive game environment.

2 The ECA Max and his Emotion Simulation System

In the AI Group of the Faculty of Technology at the University of Bielefeld, an embodied conversational agent called Max has been developed [13]. Currently, Max has basic abilities for multi-modal interaction as he is capable of synchronized auditory speech and facial and bodily gestures [2]. He is situated in different kinds of interaction scenarios with varying degrees of user perception; starting with pure desktop point-and-click interaction (see Sect. 4, Fig. 6) up to fully immersive 3D (three-dimensional) gesture and speech recognition in a Virtual Reality application (see Fig. 1). In all cases the 3D computer animated visualization of Max is based on an underlying kinematic skeleton, so that he can point to objects in his virtual environment using inverse kinematics routines and his face is realized by morphing a 3D representation in realtime on the basis of 21 simulated facial muscles (see Fig. 2).

On this technical basis, Max is controlled by a cognitively motivated architecture [13], which enables him to conduct deliberative as well as reactive behavior. His deliberative component is realized as an extension of the Believe-Desire-Intention (BDI) framework that builds on [5]. As extension of this overall architecture, a concurrent emotion system has been developed and implemented [2], which is based on the dimensional theories of emotions as originally proposed by Wundt [24] (see Fig. 2). It primarily aims at increasing the believability of Max by influencing his rational reasoning and modulating his reactive behaviors like directing facial displays. As the emotion system employs a well-defined and transparent interface, it is triggered by the cognitive architecture on conceptually different levels. It is also designed flexibly enough to be used in various kinds of agent architectures including purely reactive ones.

Therefore the emotion simulation system provides Max with a course of emotions over time that is independent of the perceived emotional state of the human player. Within the affective gaming context it is now possible to experiment with different strategies of combining these concurrent information channels of intrinsic emotions and external emotion recognition. That way different types of behavior patterns of the competing players might be observable and the human player might get the impression of an emotional and empathic artificial opponent with his own feelings and dynamically changing playing strategies. An affective game scenario also provides clearly defined conflictive goals for the agent and the human player, i.e. to win the game. Thereby the agent might at every given state of the game be able to derive a power relationship between the human player and himself. This information enables the agent to distinguish between the emotion categories fear and anger by changing the dominance parameter of the emotion simulation system as described in [2] and again the facial expressions (see Fig. 2) as well as his non-verbal utterances will change accordingly.

3 Emotion Recognition from Bio-Signals

In this section, we describe the module that is in charge of recognizing emotions inferred from biometric measures in real-time. The module is based on a module that was used in our Empathic Companion system [19]. Although the functionality of the new system is similar to the previously developed module, it was redesigned in order to account for the particular type of interaction between the user and the Max agent in the game scenario. We start with explaining how a user's physiological activity can be interpreted as emotional states.

3.1 Relating Physiological Signals to Emotions

Lang [15] claims that all emotions can be characterized in terms of judged valence (pleasant or unpleasant) and arousal (calm or aroused). Figure 3 shows some named emotions as coordinates in the arousal–valence space. The relation between physiological signals and arousal/valence is established in psychophysiology that argues that the activation of the autonomic nervous system (ANS) changes while emotions are elicited [16]. The following two signals have been chosen for their high reliability:

- 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.

Other signals (electrocardiogram, EEG, respiration, temperature, pupil dilation) are discussed in Andreassi [1] and applied e.g. in Picard [18].

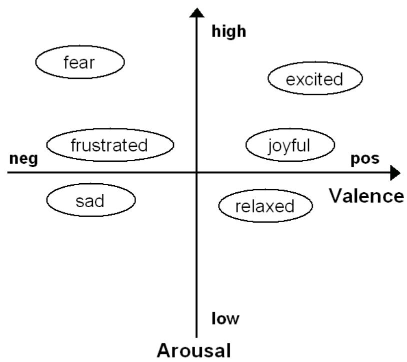

Fig. 3. Some named emotions in the arousal–valence space.

3.2 Real-time Emotion Recognition

The module architecture depicted in Fig. 4 shows how the system handles real-time emotion recognition. It has been implemented as an ActiveX component on a Windows XP platform and a proxy program, in order to allow data exchange between the emotion recognition component and the Max agent system. Each of the main components will be explained below.

Initialization Data File, Synchronization Layer, and Interface Layer The Initialization Data file contains parameter definitions (e.g. sampling rate and category definitions). The Synchronization Layer is used to initialize these parameters and to provide a framework for coordination between the other layers on data acquisition, use, and storage. The Interface layer provides the interface functions for communication between the Max agent and the component. It allows the Max agent to retrieve the user's current emotional state, his underlying valence and arousal values, update the current state of the game in the component, and control the sensor acquisition process.

Device Layer The user is attached to sensors of the ProComp Infiniti unit from Thought Technologies [23]. The ProComp Infiniti encoder is able to use input from up to eight sensors simultaneously. Currently, we only use galvanic skin response (GSR) and electromyography (EMG) sensors. Input from the sensors is digitally sampled at the rate of 20 samples/sec. In order to perform the data acquisition, this layer makes use of the ProComp Infiniti data capture library, known as TTLAPI. The data acquisition process is performed periodically once it has been activated, and it maintains queues for the retrieved data of each sensor, storing new values and discarding old ones. The size of these queues can also be configured using the Initialization Data file.

Signal Categorization Layer When prompted by the Max application through the interface of the emotion recognition component, this layer evaluates the data currently stored in the Device Layer queues. 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 of three minutes preceding the interaction.

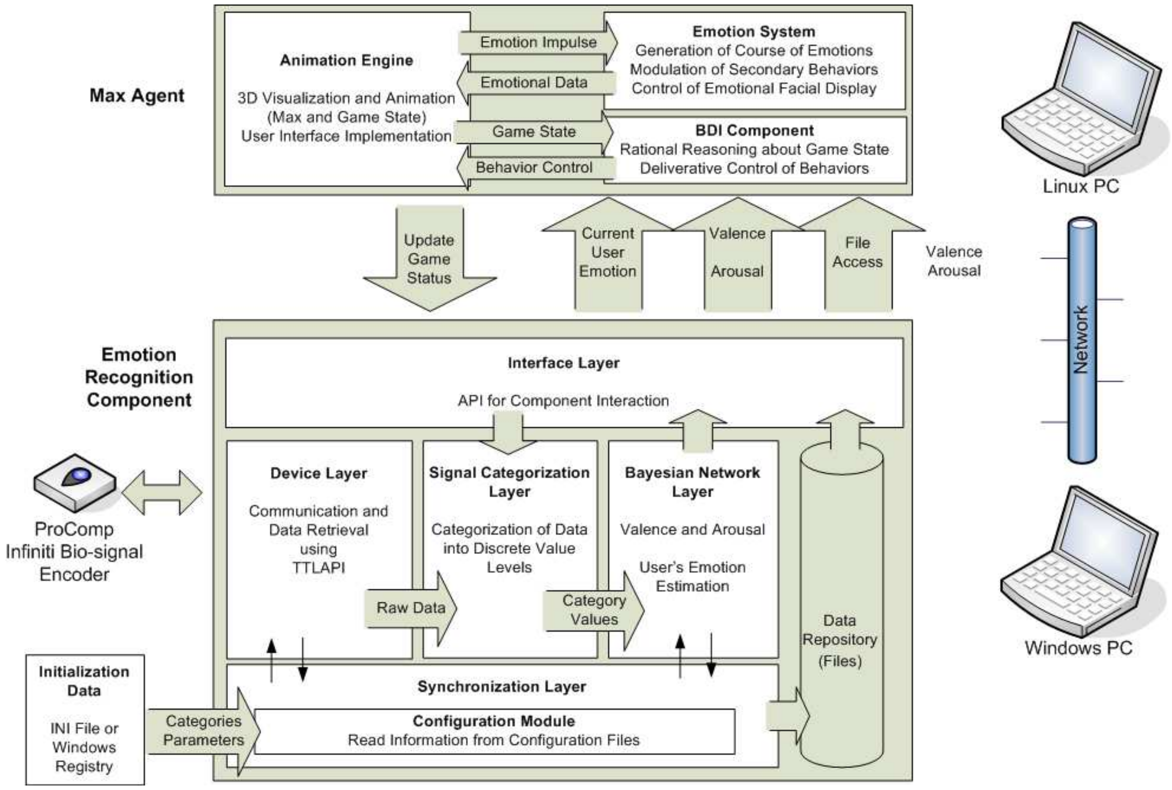

Fig. 4. System architecture for real-time emotion recognition and the Max agent.

The current mean value is derived from a segment of five seconds, the average duration of an emotion [16]. 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". The current categorization is determined by the Bayesian network model used in the respective layer discussed below. However, the design of this layer allows us to more flexibly define categories based on the information stored in the Initialization Data file.

Bayesian Network Layer Once the raw data from the sensors has been categorized, a Bayesian network (implemented with Netica [17]) is used to combine the categorized information from the bio-signals and other facts about the interaction and determine the user's emotion based on these values. This network is shown in Fig. 5.

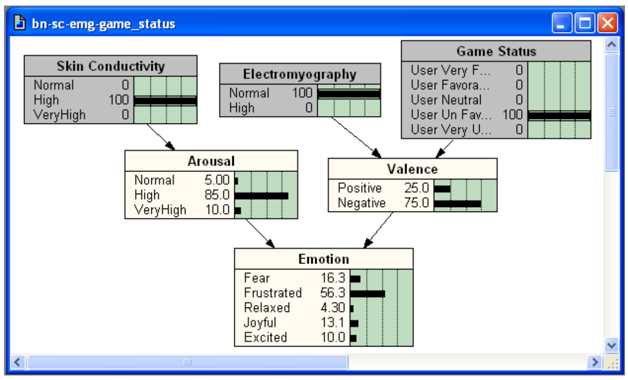

Fig. 5. Simple Bayesian network to determine a user's emotional state from bio-signals and game status.

Specifically, the Bayesian network is used to derive the user's emotional state by first relating skin conductance to arousal, and EMG and the current state of the game from the user's perspective to valence, and then inferring the user's emotional state by applying the model of [15]. The probabilities have been set in accord with the literature (whereby the concrete numbers are made up). Some examples are: "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; "Frustrated" is defined by increased arousal and negative valence.

The node "Game Status" represents situations where the game is one of the following states: very favorable for the user, favorable (for the user), neutral, unfavorable, or very unfavorable. This ('non-physiological') node was included to the network in order to more easily hypothesize the user's positive or negative appraisal of the current situation of the game, as the user's EMG value changes are often too small to evaluate valence. EMG activity is typically seen for strong emotions only.

4 Game Playing with Max Based on User Physiological Information

As our affective gaming scenario, the classical cards game "Skip-Bo" has been implemented as a face-to-face interaction scenario between a human player and the Max agent (see Fig. 6). This setup has the potential to greatly increase the affective bandwidth of future games (as also claimed in [11]) due to its bi-directionality: the game not only provokes emotional reactions in players (both in the human player and Max), but the agent also actively refers to the dynamically changing emotional state of the user when giving empathic feedback by means of non-verbal utterances.

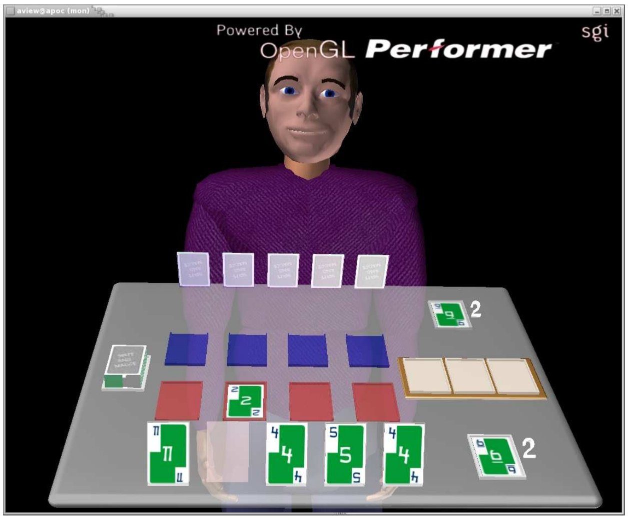

Fig. 6. The card game "Skip-Bo" as an interaction scenario for an empathic Max.

We are currently preparing a study in which we want to evaluate the user's game experience. A notorious difficulty in testing games is that users have different motivations to play a game as well as different experience levels with a game. They might be challenge-oriented or relaxation-oriented, novice or expert. Hence, one has to be cautious not to decrease the difficulty level for an advanced player who derives satisfaction from completing difficult challenges, or not to keep the difficulty level of a repeatedly loosing player who merely wants to play a relaxing game. Nevertheless, most game researchers would agree that future games should adapt the game course development to the user's emotional state [10].

For the sake of controllability, the study will confine human player motivation to challenge-oriented which is supported by rewarding winning players. Note that in the Skip-Bo game scenario the Max agent is not supposed to speak to the human player as this kind of feedback would complicate the intended intercultural comparative studies concerning the subjective experience of the agent. However, Max uses diverse facial expressions as well as dynamically generated bodily gestures to give direct non-verbal feedback. Indirect feedback will be given by modulating the eye blink frequency as well as the breath frequency according to his emotional state. The experimental conditions are as follows:

1. *Non-affective condition*: The Max agent plays the game challenge-oriented, and neither shows emotional behavior nor is he aware of the user's emotional state. Nevertheless the emotion recognition data as well as the emotion simulation data are recorded for later analysis.
2. *Affective Self-emotional condition*: The Max agent plays challenge-oriented and shows affective behavior that is evoked only by his own actions. The human player's actions have no effect on his own emotional state and it is not aware of the user's emotional state.
3. *Negative Empathic condition*: The Max agent plays in competitive (egoistic) style and shows affective behavior depending on his internal emotional state. Now also the user's actions are influencing Max's emotional state and he is aware of the user's affective state and responds accordingly. E.g. when the user shows frustration, Max will display *Schadenfreude* ("joy about the user's distress"). On the other hand, when the user dominates the game and is recognized to be in a positively valenced state, Max will express ignorance by looking aside.
4. *Positive Empathic condition*: Here, the Max agent adapts his playing strategy depending on the user's affective state hypothesized from biometric information and game state. When the user is recognizably frustrated, Max will alter his playing strategy from egoistic to defensive or altruistic. The latter strategy can be considered as an instance of (social) empathic behavior where Max plays in an amiable way. This condition corresponds most closely to an embodied agent in the role of an encouraging parent towards a child, or in the role of a teacher to a student. (One characterization of empathy is "as I understand it, this is what you are going through. This is how I would *feel in your*

shoes" [12]. By (slight) contrast, sympathy is characterized as "this is how *I* feel *about* what you're going through. This is how I would feel *in your shoes*" [12].)

We are currently conducting a study and hope to report on its results in the near future.

Acknowledgements

Christian Becker is supported by a JSPS Pre-Doctoral Fellowship. The research is partly supported by the JSPS Research Grant (1999-2003) for the Future Program.

References

1. J. L. Andreassi. *Psychophysiology. Human Behavior & Physiological Response.* Lawrence Erlbaum Associates, Mahwah, NJ, 4 edition, 2000.
2. C. Becker, S. Kopp, and I. Wachsmuth. Simulating the emotion dynamics of a multimodal conversational agent. In *Proceedings Tutorial and Research Workshop on Affective Dialogue Systems (ADS-04)*, LNAI 3068, pages 154–165, Berlin Heidelberg, 2004. Springer.
3. C. Becker, H. Prendinger, M. Ishizuka, and I. Wachsmuth. Empathy for Max (Preliminary project report). In *The 2005 International Conference on Active Media Technology (AMT-05)*, 2005. In press.
4. D. Bersak, G. McDarby, N. Augenblick, P. McDarby, D. McDonnell, B. McDonald, and R. Karkun. Intelligent biofeedback using an immersive competitive environment. In *Online Proceedings of the Ubicomp 2001 Workshop on Designing Ubiquitous Computer Games*, 2001.
5. W. Bratman, D. Israeal, and M. Pollack. Plans and resource bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.
6. R. Burke. Breathing new life into accessibility. *CodeZone Magazine*, 2003.
7. J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. Human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 29–63. The MIT Press, Cambridge, MA, 2000.
8. J. Cassell and K. R. Thórisson. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4–5):519–538, 1999.
9. J. Cassell, H. Vilhjálmsson, and T. Bickmore. BEAT: the Behavior Expression Animation Toolkit. In *Proceedings of SIGGRAPH-01*, pages 477–486, 2001.
10. K. M. Gilleade and A. Dix. Using frustration in the design of adaptive videogames. In *Proceedings of Advances in Computer Entertainment Technology (ACE-04).* ACM Press, 2004.
11. K. M. Gilleade, A. Dix, and J. Allanson. Affective videogames and modes of affective gaming: Assist me, challenge me, emote me ACE. In *The 2005 International Conference on Changing Views: Worlds in Play*, 2005.
12. J. Klein, Y. Moon, and R. Picard. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14:119–140, 2002.
13. S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth. Max – a multimodal assistant in virtual reality construction. *KI Zeitschift (German Journal of Artificial Intelligence), Special Issue on Embodied Conversational Agents*, 2003.
14. S. Kopp, T. Sowa, and I. Wachsmuth. Imitation games with an artificial agent: From mimicking to understanding shape-related iconic gestures. In *Proceedings 5th International Workshop on Gesture and Sign Language based Human–Computer Interaction*, 2003.
15. P. J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.
16. R. 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.
17. Netica. Norsys Software Corp., 2003. URL: `http://www.norsys.com`.
18. R. W. Picard. *Affective Computing.* The MIT Press, Cambridge, MA, 1997.
19. H. Prendinger, H. Dohi, H. Wang, S. Mayer, and M. Ishizuka. Empathic embodied interfaces: Addressing users' affective state. In *Proceedings Tutorial and Research Workshop on Affective Dialogue Systems*, LNAI 3068, pages 53–64, Berlin Heidelberg, 2004. Springer.
20. H. Prendinger and M. Ishizuka, editors. *Life-Like Characters. Tools, Affective Functions, and Applications.* Cognitive Technologies. Springer Verlag, Berlin Heidelberg, 2004.
21. C. Stevenson. Affective gaming: Using pupil dilation to determine player arousal. In *The 2005 International Conference on Changing Views: Worlds in Play*, 2005.
22. J. Sykes and S. Brown. Affective gaming. Measuring emotion through the gamepad. In *Proceedings of CHI-03*, pages 732–733, 2003.
23. Thought Technology Ltd., 2002. URL: `http://www.thoughttechnology.com`.
24. W. Wundt. *Vorlesung über die Menschen und Tierseele (Lecture on Humans and the Soul of Animals).* Voss Verlag, Leipzig, 7. and 8. edition, 1922.