

Empathy for Max

(Preliminary Project Report)

Christian Becker
Faculty of Technology
University of Bielefeld
33549 Bielefeld, Germany
cbecker@techfak.uni-bielefeld.de

Mitsuru Ishizuka
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

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

Ipke Wachsmuth
Faculty of Technology
University of Bielefeld
33549 Bielefeld, Germany
ipke@techfak.uni-bielefeld.de

Abstract

This paper first describes 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 [8, 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 [15]. Then, the integration of both systems is suggested for the purpose of realizing a highly believable agent with empathic qualities.

Keywords: Embodied Conversational Agents, Empathy, Emotion Recognition, Bio-signals, Emotion Simulation, Affect

1. Introduction and Motivation

In the growing field of Embodied Conversational Agents (in short, ECAs) [4, 16] techniques from artificial intelligence, computer animation and human-computer interaction are combined to create a synthetic communication partner. ECAs are computer-generated, humanoid characters, which are able to conduct a natural face-to-face dialogue with a human user (e.g. [5, 6]). The types of communication channels range from pure textual input to multi-modal speech-gesture interfaces [9].

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 that perceive and interpret 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.

This paper describes our efforts towards the integration of a highly expressive agent (Max) and an emotion recognition module that may serve to furnish Max with empathic behavior. Section 2 reports on Max and the emotion simulation system used by Max [2]. Section 3 summarizes the emotion recognition module and its use within the Empathic Companion application [15]. In Section 4, we outline our future work on integrating both components.

2. The ECA Max and his Emotion Simulation System

In the A.I. (Artificial Intelligence) Group of the Faculty of Technology at the University of Bielefeld, an embodied conversational agent called Max has been developed [8]. 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 differ-



Figure 1. The ECA Max in a VR setting.

ent 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 real-time on the basis of 21 simulated facial muscles (see Fig. 2).

On this technical basis, Max is controlled by a cognitively motivated architecture [8], which enables him to conduct deliberative as well as reactive behavior. Its deliberative component is realized as an extension of the Believe-Desire-Intention (BDI) framework that builds on [3]. As extension of this overall architecture, a concurrent emotion system has been developed and implemented, which is based on the dimensional theories of emotions as originally proposed by Wundt [18] (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.

3. Emotion Recognition from Bio-Signals and Empathic Response

In this section, we describe two modules of our Empathic Companion system [15]. The Empathic Companion (EC) is an animated 2D interface agent that accompanies a user in a virtual job interview scenario. The first module is responsible for recognizing emotions (inferred from biomet-

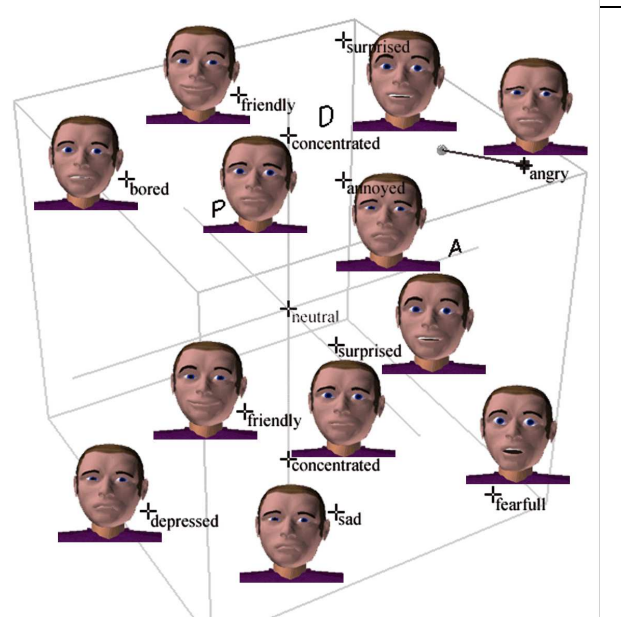


Figure 2. Named emotions with corresponding facial displays in the pleasure-arousal-dominance space.

ric measures) in real-time, whereas the second module selects an appropriate (empathic) response based on the user's emotion. Although both modules are independent of the job interview scenario, we will (for convenience) occasionally use examples pertaining to this 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 [10] 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 [11]. 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 are discussed in Andreassi [1] and applied e.g. in Picard [14].

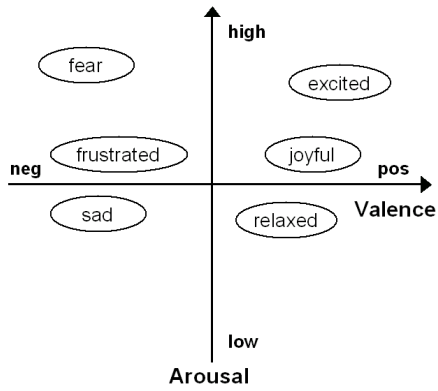


Figure 3. Some named emotions in the arousal–valence space.

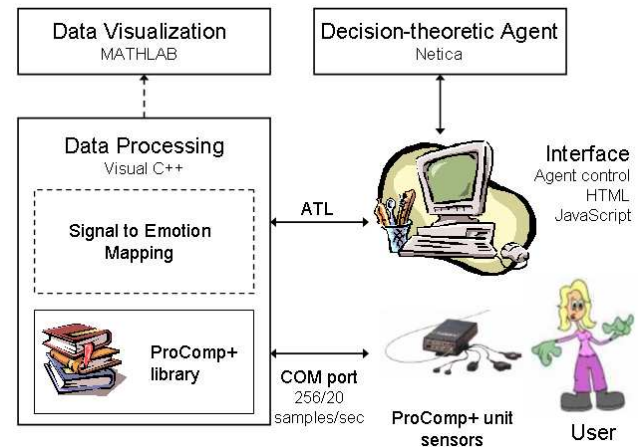


Figure 4. System architecture.

3.2. System Architecture for Real-time Emotion Recognition

Since the EC application assumes real-time emotion recognition, the system architecture depicted in Fig. 4 has been implemented on the Windows XP platform. Below, we will explain each of its components.

3.2.1. Data Capturing The user is attached to sensors of the ProComp+ unit from [17]. 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 at 20s/s. Data capturing is achieved by a module written in Visual C++ that employs the ProComp+ data capture library.

3.2.2. 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 a relaxation period preceding the interaction. The current mean value is derived from a segment of five seconds, the average duration of an emotion [11]. 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, but optionally, a “Signal to Emotion Mapping” module is available, if no complex decisions are required. The connection between the Data Processing component and the User Interface is established by the Active Template Library (ATL).

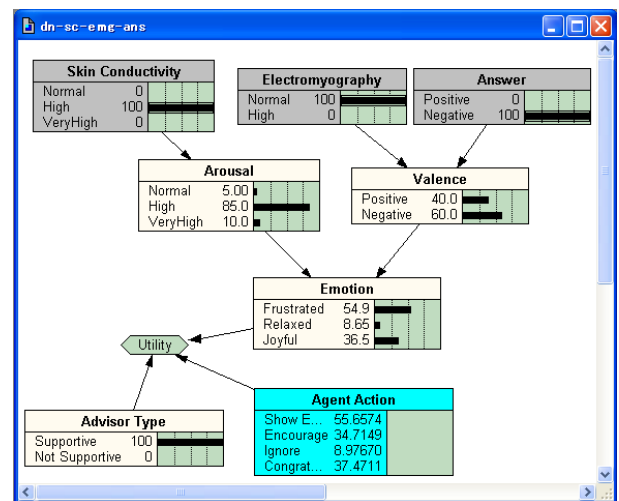


Figure 5. Simple decision network.

3.2.3. 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 [12] 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.

3.2.4. Decision-theoretic Agent A decision network (implemented with Netica from [13]) is used to combine bio-signals and other facts about the interaction, relate them to emotions, and suggest appropriate, e.g. empathic agent decisions.

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. 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 [10]. 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 decision node in Fig. 5 lists some possible actions. If the advisor type is supportive, the utility function is set to give priority to empathic responses.

- *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 the user is aroused in a positive way, it applauds the user ("Well done!", "Good job! You said the right thing", etc.).

"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. Future Work

Future joint work is directed towards combining the systems developed at the University of Bielefeld on the one hand, and the University of Tokyo and the National Institute

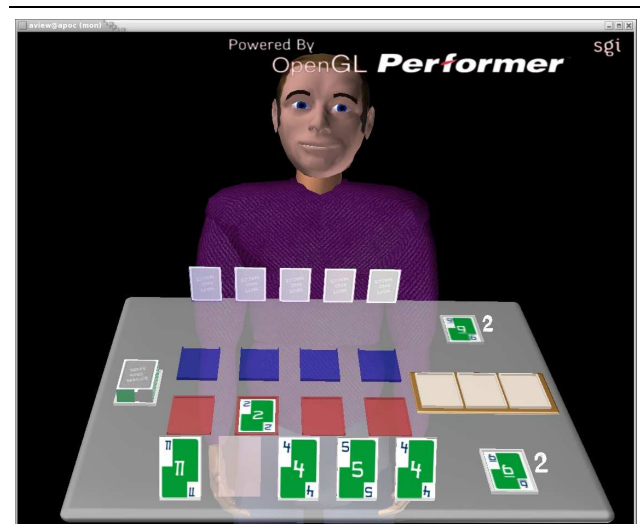


Figure 6. The card game "Skip-Bo" as an interaction scenario for an Empathic Max.

of Informatics on the other – in order to achieve a highly believable and expressive empathic interlocutor. An integrated system is desirable for two reasons:

- Max has highly advanced abilities to express his emotions but limited access to the user's affective state;
- The Empathic Companion can recognize the user's affective state but the expression of its empathic feedback is rather simple (predefined 2D animation sequences).

The resulting interface agent will be able to react to the user's emotional state appropriately by modulating its behavior and facial expressions. Furthermore, the deliberate reasoning of the agent will be influenced according to the emotional state of the user.

As a face-to-face interaction scenario the classical card game "Skip-Bo" is currently being implemented (see Fig. 6). Most importantly, this game is expected to provoke diverse emotional reactions in human players who may choose between an offensive or defensive playing strategy. The game can easily be implemented as an interactive computer game with two to four players, the agent being an emotionally aware and responsive participant of the game.

Max will show empathy – often understood as awareness of the interlocutor's affective state and appropriate response [7] – in situations where e.g. the human player is frustrated about the development of the game. Hence, rather than being an opponent only, Max will also be able to act in the role of a teacher to the human player.

Within this scenario the agent is not supposed to speak to the human player as this kind of feedback would further

complicate the intended intercultural comparative studies concerning the subjective experience of the empathic agent. Diverse facial expressions as well as dynamically generated bodily gestures will be used to give direct non-verbal feedback and indirect feedback will be given by modulating the agent's eye blink frequency as well as its breath frequency according to its emotional state.

In the near future, two specific project goals will be realized. *First*, based on the biometric data and the actual game state, a heuristic will be developed that triggers the agent's emotion system, so that its emotional reactions as well as its playing strategies are altered by its emotions in a believable and comprehensible way. For instance, if the assumed emotional state of the human player is negatively valenced and the agent itself is leading the game, it might begin feeling dominant and alter its playing strategies to be defensive. If the emotional state of frustration is derived from the biometrical data the agent might even begin to play in an altruistic way. Thus the agent shows empathy by playing a defensive strategy on the one hand, and on the other hand its emotion dynamics would be biased towards the experience of anger instead of fear by internally changing to a dominant emotional stance.

Second, by taking into account the actual game state, the emotion detection system will be extended, so that the number of distinguishable emotion types increases. This extension follows the basic ideas of the three-dimensional emotion space, which also forms the basis of the emotion dynamics component of the ECA Max.

We also plan to qualitatively evaluate the believability of the combined system by conducting user experiments.

Acknowledgements

The first author is supported by a JSPS Pre-Doctoral Fellowship. The second and third author are 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] W. Bratman, D. Israeal, and M. Pollack. Plans and resource bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.
- [4] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsón, and H. Yan. Human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sulivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 29–63. The MIT Press, Cambridge, MA, 2000.
- [5] 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.
- [6] J. Cassell, H. Vilhjálmsón, and T. Bickmore. BEAT: the Behavior Expression Animation Toolkit. In *Proceedings of SIGGRAPH-01*, pages 477–486, 2001.
- [7] J. Klein, Y. Moon, and R. Picard. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14:119–140, 2002.
- [8] S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth. Max – a multimodal assistant in virtual reality construction. *KI Zeitschrift (German Journal of Artificial Intelligence), Special Issue on Embodied Conversational Agents*, 2003.
- [9] 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.
- [10] P. J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.
- [11] 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.
- [12] Microsoft. *Developing for Microsoft Agent*. Microsoft Press, Redmond, WA, 1998.
- [13] Netica. Norsys Software Corp., 2003. URL: <http://www.norsys.com>.
- [14] R. W. Picard. *Affective Computing*. The MIT Press, Cambridge, MA, 1997.
- [15] 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.
- [16] H. Prendinger and M. Ishizuka, editors. *Life-Like Characters. Tools, Affective Functions, and Applications*. Cognitive Technologies. Springer Verlag, Berlin Heidelberg, 2004.
- [17] Thought Technology Ltd., 2002. URL: <http://www.thoughttechnology.com>.
- [18] W. Wundt. *Vorlesung über die Menschen und Tierseele (Lecture on Humans and the Soul of Animals)*. Voss Verlag, Leipzig, 7. and 8. edition, 1922.