

Simulating the Emotion Dynamics of a Multimodal Conversational Agent

Christian Becker, Stefan Kopp, and Ipke Wachsmuth

Artificial Intelligence Group
Faculty of Technology, University of Bielefeld
D-33594 Bielefeld, Germany
{cbecker, skopp, ipke}@techfak.uni-bielefeld.de

Abstract. We describe an implemented system for the simulation and visualisation of the emotional state of a multimodal conversational agent called *Max*. The focus of the presented work lies on modeling a coherent course of emotions over time. The basic idea of the underlying emotion system is the linkage of two interrelated psychological concepts: an emotion axis – representing short-time system states – and an orthogonal mood axis that stands for an undirected, longer lasting system state. A third axis was added to realize a dimension of boredom. To enhance the believability and lifelikeness of *Max*, the emotion system has been integrated in the agent’s architecture. In result, *Max*’s facial expression, gesture, speech, and secondary behaviors as well as his cognitive functions are modulated by the emotional system that, in turn, is affected by information arising at various levels within the agent’s architecture.

1 Introduction

Natural and intuitive communication with a computer is a primary research goal in human-computer interaction. In recent times this goal has frequently led to the employment of humanoid agents as interlocutors that are able to both understand and use communicative means natural to human users, i.e., speech, gesture, gaze, and facial expression. In our lab an anthropomorphic humanoid agent named *Max* is being developed [7]. In order to enhance his lifelikeness and believability, we aim at creating the impression of coherent emotions, moods and personality. This requires, first, appraising internal or external events w.r.t. their influence on the agent’s emotion dynamics, secondly, modeling the emotions and moods of the agent and their course over time as well as their mutual interaction (*emotion dynamics*), and, finally, communicating the emotional state in identifiable ways. In this paper, we focus on the last two aspects. The first aspect – the appraisal of emotions – may be accomplished in at least two ways by other components of the overall architecture: (1) on a non-cognitive level by direct sensory feedback in a VR application which is not described here, and (2) on a cognitive level by a BDI interpreter which performs deliberative reasoning.

Based on original ideas of Wundt [18] and subsequent research, e.g., [10], the communication of affect can be seen as having three major dimensions of connotative

meaning: *pleasure*, *arousal*, and *dominance* (*PAD*). Further evidence has been found that emotions can be differentiated from moods by at least two aspects [8]: an emotion can usually be associated with its eliciting stimulus and is a short-lived psychological phenomenon, whereas a mood is a more diffuse and longer lasting phenomenon. Nevertheless it is assumed that moods are fortified and alleviated in some way by emotions [8] as well as that the elicitation of emotions is influenced by the actual mood in return; see [12]. Mood can also be understood as “an emotional state, perhaps of low intensity, capable of lasting for many minutes or several hours”; see [12].

We believe that both theoretical aspects, emotion and mood, along with their effects on each other, must be accounted for by a model that is to simulate coherent emotion dynamics for a human-like character. We thus developed a model that combines these two theories to realize an internal continuous dynamics and allows for symbolic output of categorical emotional terms. Furthermore, our model extends these theories by adding a concept of boredom that is understood as a state of relatively low arousal, which is attributed to an inadequately stimulating situation as proposed by [11]. This concept is especially useful as *Max* is situated in scenarios of real-time communication with a varying degree of interaction and the concept of boredom supports the emergence of proactive behavior in a natural way.

As far as the expression of emotions is concerned, two different kinds of information processing are distinguished: On the one hand, discrete emotional terms, modulated by a continuous intensity value, are used in order to trigger emotional expressions of the face as well as to influence deliberative reasoning in the cognitive architecture. On the other hand, involuntary facets of *Max*'s observable behavior – e.g., his simulated breathing rate, eye blink frequency, and speech pitch – are modulated by the continuous part of the emotional data like the intensity of arousal. Together with the capabilities of deliberative reasoning, nonverbal signals like gestures and facial expression, and response to various forms of user interaction (e.g., speech, gesture and gaze of the user), we expect *Max* to be perceived as a more creditable conversational partner.

After discussing related work in the following section, the emotion system is described in detail in Section 3. The integration of the emotion system in the cognitive architecture and the expression of affect are presented in Section 4.

2 Related Work

Several architectures were proposed for human-computer interfaces that account for affective states. Some have been implemented in expressive characters which are able to show their affective states via different communication channels such as voice quality and facial expressions, and which include such states in their deliberation process for action planning. Most of the existing computational systems of emotions rely on appraisal theories and logical reasoning about the eliciting factors of emotions. The majority of these systems are based on the OCC model by Ortony, Clore and Collins [13], see [6, 14, 16, 1] for examples. They either try to simulate a different number of discrete internal emotional categories, or aim at the expressive aspect of

emotions and therefore simulate only a smaller number of the emotion categories than established by the OCC model (see [17]). In [4] a “personality system” is presented which includes “emotional states” based on the OCC model and also uses a concept called “mood” to mediate between the static personality and the dynamic emotions of a virtual human. The notion of an “emotional impulse” as introduced in [4] is appropriate for our emotion system as well, but in contrast to their approach, our emotion system does not need an associated emotion category as an input signal.

Especially the course of emotions in time as well as the mutual interaction of different emotion categories (i.e. the *emotion dynamics*) was found difficult to realize in purely cognitively motivated architectures of emotions (see [14] for details). De Rosi et al. [15] focus on the dynamics of their agent’s affective states in a conversational situation. In contrast to their approach, which concentrates on event-driven emotions of the OCC model, our system is not designed to represent a mixture of multiple emotions so far. But with respect to the dynamics of emotions and their response decay in the absence of new specific stimuli they deal with the same conceptual problems as in our case. Addressing these problems, it appears a promising approach to ground a model of emotions on dimensional theories, as theoretically founded by [18, 10]. This is true particularly when focusing on the expression of emotions and their course over time (see [2] for example) because they are self-contained and complete, even if their expressive power is in some way restricted.

3 Internal Structure of the Emotion System

The emotion system in our approach contains two conceptually different components (see Fig. 1). The first component is mainly concerned with the course of emotions and moods over time and the mutual interaction between emotions and moods, whereas in the second component the categorization on discrete emotional terms is accomplished after a mapping into the PAD space.

3.1 Emotions and Moods over Time and Their Mutual Interaction

The concept of emotions is linked to the concept of moods using a two-dimensional space defined by an x-axis of emotional valence and an orthogonal y-axis that represents the valence of moods (see Fig. 2). The system tends to hold both valences in absolute zero because this point is interpreted as the prevalent state of mental balance. Therefore two independent spiral springs are simulated, one for each axis, which create two reset forces F_x and F_y whenever the point of reference is displaced from the origin.

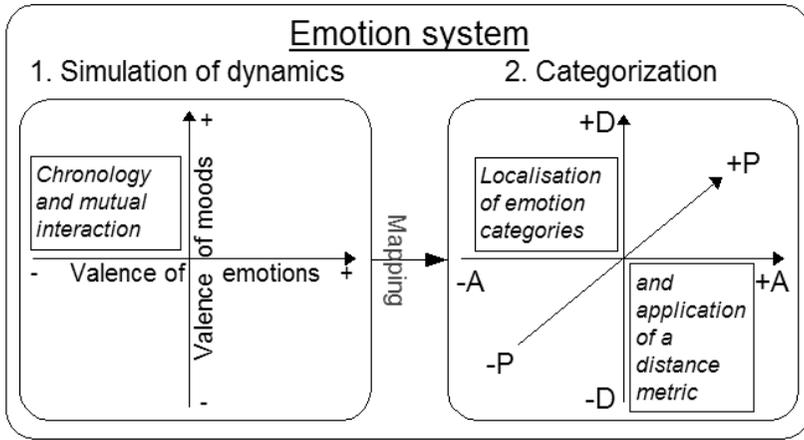


Fig. 1. Internal structure of the emotion system

The exerted forces are proportional to the value of the corresponding valences x and y just as if the simulated spiral springs were anchored in the origin and attached to the point of reference. The mass-spring model was chosen here mainly based on the heuristics that it better mimics the time course of emotions than linear and exponential decreasing models.

By adjusting the two spring constants d_x and d_y as well as the simulated inertial mass m of the point of reference, the course over time of both concepts can be biased intuitively. These parameters can also be construed as personality-related aspects of the emotion system.

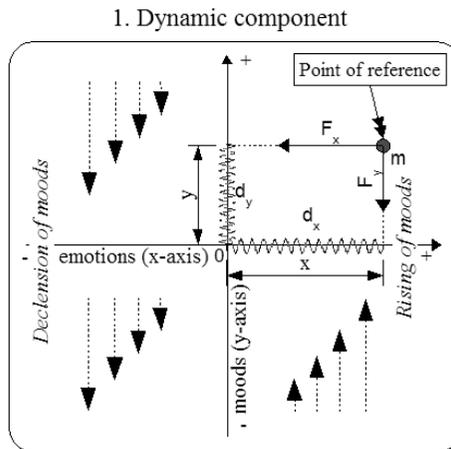


Fig. 2. The linkage of emotions and moods and their courses over time

In order to simulate the alleviating and fortifying effects of emotions on moods, the emotional valence is interpreted as a gradient for changing the valence of moods at every simulation step according to Eq. 1. The independent parameter a is a personality-related aspect of the character, with smaller values of a resulting in a more sluggish agent and a greater values of a leading to a more moody agent.

$$\frac{\Delta y}{\Delta t} = a \cdot x$$

Equation 1. The influence of emotions (x) on moods (y)

3.2 The Concept of Boredom

In addition to the emotion dynamics described above, a concept of boredom is added to the dynamic component as a third, orthogonal z-axis. Assuming that the absence of stimuli is responsible for the emergence of boredom (as proposed by [11]), the degree of boredom starts to increase linearly over time if the point of reference lies within an epsilon neighborhood of absolute zero (as given by ϵ_x and ϵ_y , see Fig. 3). Outside of this neighborhood the value of boredom is reset to zero per default. The co-domain of the boredom parameter is given by the interval $[-1, 0]$, so the agent is most bored if the value of negative one is reached.

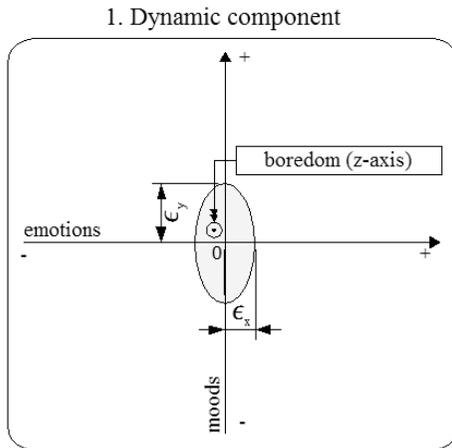


Fig. 3. The epsilon neighborhood

The linear increase of boredom can be described by the equation $z(t + 1) = z(t) - b$, where the parameter b is again a personality-related aspect of the emotion system.

3.3 Mapping into PAD-Space for Categorization

The dynamic component provides the following triple at any time step t :

$$D(t) = (x_t, y_t, z_t), \quad \text{with } x_t = [-1, 1], y_t = [-1, 1], z_t = [-1, 0]$$

Equation 2. The output triple of the dynamic component

The variable x_t denotes the emotional state, the variable y_t stands for the actual valence of the mood, and z_t represents the degree of boredom. Given this triple, the mapping into PAD space for categorization is implemented according to the function $K(x_t, y_t, z_t, t)$ as shown in Eq. 3. This mapping results in a triple consisting of the functions $p(x_t, y_t)$ for the calculation of the *pleasure* value, $a(x_t, z_t)$ for the *arousal* value and $d(t)$ for the *dominance* value.

$$K(x_t, y_t, z_t, t) = (p(x_t, y_t), a(x_t, y_t), d(t)),$$

$$\text{with } p(x_t, y_t) = \frac{1}{2} \bullet (x_t + y_t) \quad \text{and} \quad a(x_t, z_t) = |x_t| + z_t$$

Equation 3. Mapping from dynamic component into categorization component

Pleasure is assumed to be the overall valence information in PAD space and therefore calculated as the standardized sum of both the actual emotional valence as represented by x_t and the valence of the actual mood as given by y_t . That way, the agent will feel a maximum of joy when his emotion as well as his mood is most positive and a maximum of reluctance in the contrary case.

The agent's arousal ranges from sleepiness to a maximum of mental awareness and physiological exertion. As it is assumed that any kind of emotion is characterized by high arousal, in contrast to moods, only the absolute value of *emotional* valence is considered in the function $a(x_t, z_t)$. The addition of the (negatively signed) value of boredom reflects its relation to the mental state of inactivity.

The independent parameter of dominance (or, in the other extreme, submissiveness) cannot be derived from the dynamic component. In human terms, this parameter describes the agent's feelings of control and influence over situations and events versus feelings of being controlled and influenced by external circumstances [10]. By introducing this parameter it is possible to distinguish between anger and fear as well as between sadness and annoyance. Anger and annoyance come along with the feeling of control over the situation whereas fear and sadness are characterized by a feeling of being controlled by external circumstances ([9] gives a theoretical foundation). Therefore, it is in principle not possible to derive such information from the dynamic component. The BDI interpreter of the cognitive architecture of *Max*, however, is capable of controlling the state of dominance in an adequate way.

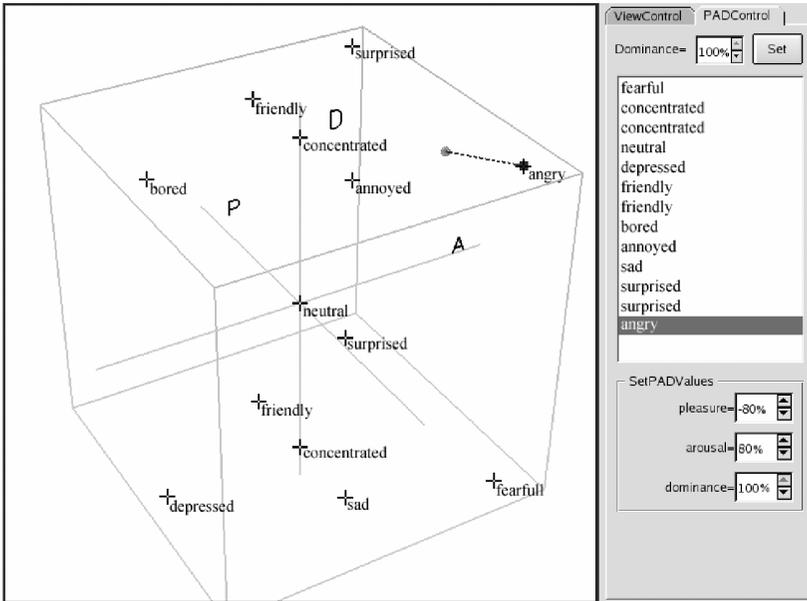


Fig. 4. The emotion categories in PAD space

Several emotion categories have been anchored in PAD space by defining adequate PAD triples (see Fig. 4). Some categories exist twice because it is assumed unnecessary to distinguish between a dominant and a submissive case for these emotion categories. The aforementioned point of reference is presented in Fig. 4 as the grey sphere currently linked to the emotion category “angry” (indicated by the dotted line to give a visual feedback of the activated emotion category).

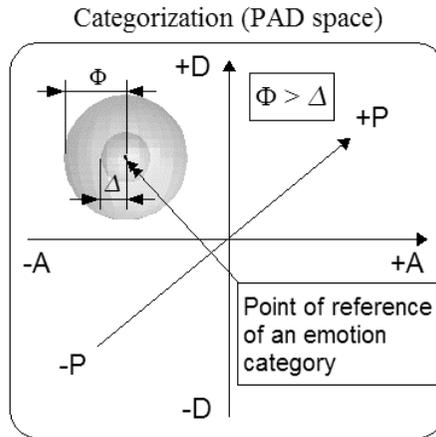


Fig. 5. Thresholds Φ and Δ for each emotion category

An emotion category is activated if the point of reference is getting closer than Φ units to that particular category and if there is no other emotion category within a

smaller distance (see Fig. 5). Until the distance d gets below Δ units, the weight of the activated emotion category is calculated by the following equation:

$$w = 1 - \frac{d - \Delta}{\Phi - \Delta}$$

Equation 4. Calculation of weight w for each emotion category

The weight w is set to 1, if the distance gets below Δ . In Eq. 4, Φ can be interpreted as the activation threshold and Δ as the saturation threshold, which are both global constants of the emotion system and valid for every emotion category. If no emotion category fulfills the above conditions for activation, the cognitive state of “confusion” is activated in the emotion system. This may occur in the case of a strongly positive valence of moods together with a highly negative emotional valence, or vice versa.

4 Integration and Application

4.1 Integration in the Agent Architecture

Max is employed as a presentation agent in the Heinz-Nixdorf MuseumsForum (*HNF*; Paderborn, Germany). In this environment, the agent’s task is to conduct multimodal smalltalk dialogs with visitors as well as to give explanations about the exhibition he is part of. *Max* is provided with typed user input as well as camera input and responds using synthetic speech, gesture, and facial expression. The agent may also leave the presentation screen when very annoyed, and reenter the screen after “calming down”.

The system’s overall architecture (shown in Fig. 6) is similar to those commonly applied in embodied conversational agents (e.g. *FXPAL* [3]). It exhibits a two-level structure of concurrent reactive and deliberative processing, the latter being responsible for the agent’s conversational capabilities. The emotion system has been added to this architecture as a separate module that incessantly receives input from and sends data to several other components as indicated by the solid arrows in Fig. 6.

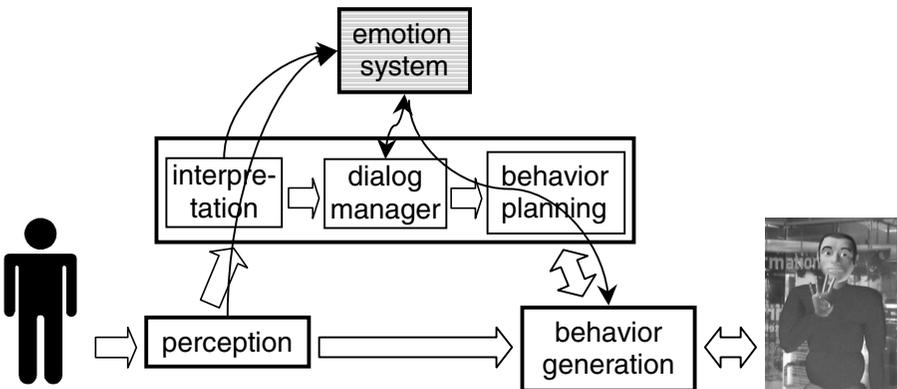


Fig. 6. Integration of the emotion system in the agent’s architecture

4.2 The Input and Output Data of the Emotion System

The components of the cognitive architecture of *Max* essentially feed the emotion system with emotional valence information. These positive or negative valences always originate from deliberative processes (*interpretation* and *dialog manager*) or as direct reactions to a positive or negative stimulus (*perception*). For example, the *interpretation* module sends a positive valence value if the user's utterance has been understood as a compliment. Likewise, the achievement of a desired discourse goal, e.g., coming to know the user's age after having asked for it, causes the *dialog manager* to send positive input to the emotion system. At the reactive level, the *perception* module delivers a positive impulse each time a user has been visually perceived¹. Finally, a significant boredom value in the emotion system, which only arises in cases of no interaction with visitors, leads to the random execution of "displacement activities" (e.g., yawning, stretching, looking around) in the *behavior generation* component. The execution of such behaviors also reduces the boredom value in the emotion system.

The emotion system in turn supplies the cognitive architecture of *Max* with the following data:

1. the mood valence and the degree of boredom of the dynamic component
2. the corresponding PAD triple
3. the emotion category and its intensity if one is activated or "confused" otherwise

The first two kinds of information are non-cognitive information types. They are used in the *behavior generation* module to trigger secondary actions and to modulate involuntary facets of *Max's* observable behavior, namely, the rate of his simulated breathing, the frequency of eye blink, and the pitch as well as the rate of his speech.

The third kind of information is mainly used within the *dialog manager* at the cognitive level of *Max's* architecture (see [1] for details). In general, deliberative reasoning is realized by a BDI interpreter that operates on the agent's *beliefs*, on *desires* representing persistent goals and a library of plans – each having preconditions, context conditions, an effect and a utility function – to formulate *intentions*. The interpreter continually pursues the applicable plan with the highest utility value as an intention.

The categorical output of the emotion system is incessantly asserted as belief of the agent. That way, the agent's plan selection is influenced by his current affective state which he can also verbalize. In addition, the emotion is used as precondition and context condition of plans to choose among alternative actions or even to trigger actions when becoming "aware" of a certain emotion (by asserting an according belief). Finally, based on the mappings proposed by Ekman [5], the current emotion category along with its intensity value is directly reflected in *Max's* facial expressions. This facial expression is then superposed on possible conversational behaviors like smiling.

¹ This case shows how non-cognitive emergence of emotions can be modeled. Another example – realized in a VR application of *Max* not described here – is positive valence due to the human interlocutor petting *Max's* cheeks.

4.3 Example

Instead of demonstrating the well-known expression of basic emotions on the agent's face, we show here an example situation in which the current emotional state of *Max*, being engaged in a conversation with a visitor, arises from the previous discourse and significantly influences the agent's behavior. After being offended several times by verbal input of the visitor, the accumulation of the respective impulses in the emotion system results in increasingly negative emotions that become available to the agent's deliberative processes. When first becoming "angry" the agent says "Now I'm getting angry" with a low pitch and rate of his voice as well as an appropriate facial expression of anger (see Fig. 7 left). Further negative impulses result in the emotional state of "annoyance" together with a bad mood. In effect, a plan is triggered which causes the agent to leave the display (Fig. 7 right) and to stay away until the emotion system has returned into balanced mood. The period of absence can either be shortened by complimenting *Max* or extended by insulting him again.



Fig. 7. *Max* is getting angry and leaves the display

5 Conclusion

We presented an emotion system that has been integrated in a cognitive architecture of a multimodal conversational agent called *Max*. The underlying theories of emotions and moods, extended by a concept of boredom, are successfully combined to achieve a coherent long-time behavior as well as convincing spontaneous reactions. The additivity of the emotion dynamics and the bidirectional connection between emotions and moods support the traceability of the expressed emotions. By modulating the involuntary facets of *Max*'s observable behavior, controlling his facial expressions and influencing the deliberation process, *Max* is able to express his emotional state by a wide range of different communication channels.

Furthermore the emotion system was devised open enough to be applied in any agent framework. It provides both fluent non-categorical data and weighted discrete emotion categories to be used by other technical components in order to meet the diverse requirements of each component.

In a first smalltalk scenario the visual and auditory effects of the emotion system were experienced as beneficial for the believability of *Max* as a humanoid conversational agent. Especially the direct feedback by the facial expressions of *Max* was judged to be adequate and persuasive by a large number of visitors. Together with the described modulating effects of the emotion system a harmonious overall appearance was achieved at this stage.

There are still some open questions to be addressed in future work. Mainly the control of the dominance dimension by the surrounding components has to be conceptualized and implemented in further detail to better fulfill the requirements of the underlying psychological theories. In contrast to the cognitive models of emotions, higher-level emotion categories (e.g., *shame*) are not representable in our approach. However, with the experiences gained in our work it seems a promising challenge to investigate whether purely cognitive emotion models (like OCC) can be extended concerning the course of emotions and their mutual interactions.

References

1. André, E., Klesen, M., Gebhard, P., Allen, S. & Rist, T.: Integrating models of personality and emotions into lifelike characters. In: A. Paiva and C. Martinho (eds.): Proceedings of the workshop on Affect in Interactions. Siena. (1999) 136-149
2. Breazeal, C.: Kismet Homepage. <http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html>. (2003)
3. Churchill, E. F., Cook, L., Hodgson, P., Prevost, S. & Sullivan, J. W.: May I Help You?: Designing Embodied Conversational Agent Allies. In: Cassell et al. (eds.): Embodied Conversational Agents. MIT Press. (2000) Cambridge (MA). 64-94
4. Egges, A., Zhang, X., Kshirsagar, S. & Magnenat-Thalmann, N.: Emotional communication with virtual humans. In: Proceedings of the 9th International Conference on Multi-Media Modeling. Taiwan. (2003)
5. Ekman, P., Friesen, W.V. & Ancoli, S.: Facial Signs of emotional experience. In: Journal of Personality and Social Psychology. Vol. 29. (1980) 1125-1134
6. Elliot, C.: The Affective Reasoner: A Process Model of Emotions in a Multiagent System. Ph.D. Thesis. Northwestern University. (1992)
7. Kopp, S., Jung, B., Lessmann, N., Wachsmuth, I.: Max – A Multimodal Assistant in Virtual Reality Construction. KI-Künstliche Intelligenz 4/03. (2003) 11-17
8. Kruse, O.: Emotionsdynamik und Psychotherapie. (1985)
9. Lerner, J.S. & Keltner, D.: Beyond valence: Toward a model of emotion-specific influences on judgement and choice. In: Cognition and Emotion. Vol. 14. (2000) 473-493
10. Mehrabian, A.: Framework for a Comprehensive Description and Measurement of Emotional States. In: Genetic, Social, and General Psychology Monographs. Vol. 121. (1995) 339-361
11. Mikulas, W. L. & Vodanovich, S. J.: The essence of boredom. In: The Psychological Record. Vol. 43 (1993) 3-12
12. Oatley, K.: Towards a Cognitive Theory of Emotions. In: Cognition & Emotion. Vol. 1(1). (1987) 29-50
13. Ortony, A., Clore, G. L. & Collins, A.: The cognitive structure of emotions. Cambridge. (1988)
14. Reilly, W.S.N.: Believable Social and Emotional Agents. Ph.D. Thesis. Carnegie Mellon University. Pittsburg. (1996)

15. de Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V. & De Carolis, B.: From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. In: *International Journal of Human-Computer Studies*. Vol. 59. (2003) 81-118
16. Staller, A. & Petta, P.: Towards a Tractable Appraisal-Based Architecture for Situated Cognizers. In: Cañamero, D., Numaoka, C., & Petta, P., (eds.): *Workshop: Grounding Emotions in Adaptive Systems*. Vienna. (1998) 56-61
17. Velásquez, J.D.: Modeling Emotion-Based Decision Making. In: Cañamero, D. (eds.): *Emotional and Intelligent: The Tangled Knot of Cognition*. (1998) 164-169
18. Wundt, W.: *Vorlesung über die Menschen und Tierseele*. Siebente und Achte Auflage (1922)