Empathy-Based Emotional Alignment for a Virtual Human: A Three-Step Approach

Hana Boukricha & Ipke Wachsmuth

Band 25 · Heft 3 · August 2011

KÜNSTLICHE INTELLIGENZ

Computational Assessment of Interest in Speech

Designing Emotions

Interviews with Rosalind Picard and Jonathan Gratch

Affective Computing Combined with Android Science

Emotion and Computing

Organ des Fachbereichs Künstliche Intelligenz der Gesellschaft für Informatik e.V.

🖉 Springer

🖉 Springer

Springer

FACHBEITRAG

Empathy-Based Emotional Alignment for a Virtual Human: A Three-Step Approach

Hana Boukricha · Ipke Wachsmuth

Abstract Allowing virtual humans to align to others' perceived emotions is believed to enhance their cooperative and communicative social skills. In our work, emotional alignment is realized by endowing a virtual human with the ability to empathize. Recent research shows that humans empathize with each other to different degrees depending on several factors including, among others, their mood, their personality, and their social relationships. Although providing virtual humans with features like affect, personality, and the ability to build social relationships, little attention has been devoted to the role of such features as factors modulating their empathic behavior. Supported by psychological models of empathy, we propose an approach to model empathy for the virtual human EMMA—an Empathic MultiModal Agent—consisting of three processing steps: First, the *Empathy Mechanism* by which an empathic emotion is produced. Second, the *Empathy Modulation* by which the empathic emotion is modulated. Third, the *Expression of Empathy* by which EMMA's multiple modalities are triggered through the modulated empathic emotion. The proposed model of empathy is illustrated in a conversational agent scenario involving the virtual humans MAX and EMMA.

Keywords Empathic virtual humans ·
Human-agent/agent-agent interaction · Internal simulation ·
Empathy

H. Boukricha (✉) · I. Wachsmuth
Faculty of Technology, Bielefeld University, P.O. Box 100131,
33501 Bielefeld, Germany
e-mail: hboukric@techfak.uni-bielefeld.de

1 Introduction and Motivation

Allowing virtual humans to align to others' perceived emotions is believed to enhance their cooperative and communicative social skills. In our work, emotional alignment is realized by endowing a virtual human with the ability to empathize with others. In human social interaction empathy plays a major role as a motivational basis of prosocial and cooperative behavior and as contributing to moral acts like helping, caring, and justice [12]. Recent neuropsychological findings [25] substantiate that empathic brain responses are prone to modulation and thus humans empathize with each other to different degrees depending on several modulation factors including, among others, their mood, their personality, and their social relationships.

Research on virtual humans exhibiting empathic behavior shows that they can reduce stress levels during job interview tasks [21] and that they can teach children to deal with frustration [5] and bullying [19]. Moreover, empirical evaluations show that empathic virtual humans are judged as being more likable, trustworthy, and caring [4]. However, an evaluation of the impact of a virtual human's empathic behavior in a competitive card game scenario [20] shows that displaying empathic emotions is perceived as significantly arousing and stress inducing and is thus inappropriate. Therefore, we believe that a modulation of a virtual human's empathic behavior through factors like its mood, personality, and relationship to its interaction partner will lead to a more appropriate behavior. Furthermore, although providing virtual humans with features like affect, personality, and the ability to build social relationships is the subject of increasing interest, little attention has been devoted to the role of such features as factors modulating their empathic behavior.

Supported by psychological models of empathy, we propose an approach to model empathy for the virtual human

EMMA—an Empathic MultiModal Agent—based on three processing steps:

1. The *Empathy Mechanism* consists of an internal simulation of perceived emotional facial expressions and results in an internal emotional feedback that represents the empathic emotion. Based on the results of an empirical study [3] the empathic emotion is represented by a Pleasure-Arousal-Dominance (PAD) value in PAD emotion space [1] of EMMA's affective architecture.

2. The *Empathy Modulation* consists of modulating the empathic emotion derived in step 1 through modulation factors such as EMMA's mood and relationship to the other, e.g., familiarity and liking.

3. The *Expression of Empathy* consists of triggering EMMA's multiple modalities through the modulated empathic emotion derived in step 2. EMMA's multiple modalities comprise a facial expression corresponding to the PAD value of the modulated empathic emotion, PAD-based prosody modulation of a verbal expression appropriate to the context, and eye blinking and breathing behaviors modulated by the arousal value of the modulated empathic emotion.

The proposed model of empathy is illustrated in a conversational agent scenario involving the virtual humans MAX [14] and EMMA [3]. In the conversational agent scenario [13] in which the virtual human MAX engages in a multimodal small talk with a human partner, we integrate EMMA as a third interaction partner. Within this scenario, the human partner can trigger the emotions of both virtual humans by either insult or kindness toward them. During the interaction of MAX with the human partner, EMMA follows the conversation and reacts to MAX's emotions empathically.

The paper is structured as follows: In Sect. 2, we present psychological models of empathy that are relevant for our work. In Sect. 3, we outline related works on existing empathic virtual human architectures. In Sect. 4, we introduce our three-step approach to model empathy for a virtual human. Subsequently in Sect. 5, we illustrate the proposed approach to model empathy in a conversational agent scenario. Finally in Sect. 6, we give a summary of the main aspects underlying our approach and an outlook on future work.

2 Psychological Models of Empathy

In this section we introduce psychological models of empathy that are relevant for our work.

2.1 Definitions of Empathy

The concept of empathy has no universal definition and its multiple definitions are subdivided into three major categories (cf. [10]): (a) empathy as an affective response to the other's emotions, (b) empathy as the cognitive understanding of the other's emotions, and (c) empathy as the combination of the above two definitions. In order to further refine the meaning of empathy, quite a number of researchers (cf. [10]) try to differentiate it from other related phenomena. Therefore, one important feature of empathy is the self-other distinction that differentiates it from the concept of emotional contagion.

Hoffman [12] defines empathy as an affective response more appropriate to another's situation than to one's own. Within his definition, Hoffman emphasizes that an empathic response does not need to be in a close match with the affect experienced by the other, but can be any emotional reaction compatible with the other's condition.

Davis [6] introduces an organizational model of empathy. This is based on an inclusive definition of empathy as a set of constructs having to do with the responses of one individual to the experiences of another. He defines four related constructs: *Antecedents*, which refer to the characteristics of the observer, the other, or the situation; *processes*, which refer to the particular mechanism by which empathic outcomes are produced; *intrapersonal outcomes*, which refer to cognitive and affective responses produced in the observer which are not manifested in overt behavior toward the other; and *interpersonal outcomes*, which refer to behavioral responses toward the other.

For *intrapersonal outcomes* Davis distinguishes *parallel* from *reactive* outcomes. He defines *parallel outcomes* as affective responses similar to observed affect in others and *reactive outcomes* as affective responses different from observed affect in others. Thus, like Hoffman he stresses that an empathic response does not need to be in a close match with the affect experienced by the other, but can be any emotional reaction compatible with the other's condition.

2.2 Mechanisms of Empathy

Hoffman [12] argues that empathy is multidetermined and introduces five major mechanisms or processes as modes of empathic arousal: *mimicry*, *classical conditioning*, *direct association*, *mediated association*, and *role* or *perspective-taking*. Further, he stresses that these modes commonly operate in conjunction with one another. What determines which particular mode will operate is the nature of the situation, e.g., the perception of strong expressive cues will more likely foster *mimicry*. The *processes* introduced by Davis [6] (see Sect. 2.1) are very similar to the modes of empathic arousal introduced by Hoffman.

Recent neuropsychological studies (cf. [25]) try to investigate the neural basis of empathy. The results show that observing as well as internally simulating another person's emotional state automatically activates parts of the neural networks involved in processing that same state in oneself.

These findings suggest the existence of a *shared representational system*.

2.3 Modulation of Empathy

Vignemont and Singer [25] claim that empathy does not always arise automatically every time we observe others displaying emotions. They argue that empathic brain responses are prone to modulation and propose several modulation factors which they group into four categories: (a) Features of observed emotion—*valence*, *intensity*, *saliency*, *primary* vs. *secondary emotion*, (b) relation between empathizer and the observed other—*affective link* and *nurturance*, *familiarity* and *similarity*, *communicative intentions*, (c) situative context—*appraisal of the situation*, *display of multiple emotions*, and (d) features of the empathizer—*mood arousal*, *personality*, *gender* and *age*, *emotional repertoire*, *emotional regulation capacities*.

In their cognitive theory of emotion, Ortony, Clore, and Collins [18] introduce four factors modulating the intensity of an empathic emotion: (a) *Desirability-for-self* as the degree to which the desirable/undesirable event for the other is desirable/undesirable for oneself, (b) *desirability-for-other* as the degree to which the event is presumed to be desirable/undesirable for the other person, (c) *deservingness* as the degree to which the other person deserves/not deserves the event, and (4) *liking* as the degree to which the other person is liked/disliked.

The next important question is at what stage modulation occurs during empathic processing. Vignemont and Singer [25] propose two possible models: First, in the *late appraisal model* of empathy, the empathic response is directly and automatically activated by the perception of an emotional cue. As next, the empathic response is modulated or inhibited through different modulation factors. Second, in the *early appraisal model* of empathy, the empathic response is not directly and automatically activated by the perception of an emotional cue. The emotional cue is first processed and evaluated in the context of different modulation factors. Whether an empathic response arises depends on the outcome of evaluation.

The implications of the introduced psychological models of empathy for our work are detailed in Sect. 4.

3 Related Work

There are various attempts to endow virtual humans with the ability to empathize. McQuiggan et al. [16] propose an inductive framework for modeling parallel and reactive empathy in virtual agents. Their framework is called *CARE* (*Companion Assisted Reactive Empathizer*) and is based on learning empirically grounded models of empathy from observing human-agent social interactions. In a virtual training environment, users' situation data, affective states, physiological responses, and other characteristics are gathered during interaction with virtual characters. The training users are able to evaluate the virtual character's empathic reaction allowing it to learn models of empathy from "good" examples. An empirical evaluation of this approach shows that it generates appropriate empathic reactions for virtual agents.

Based on an empirical and theoretical approach Ochs et al. [17] propose a computational model of empathic emotions. The empirical part is based on analyzing human-machine dialogs in order to identify the characteristics of emotional dialog situations. The theoretical part is based on cognitive psychological theories and consists of determining the type and intensity of the empathic emotion. The results of an empirical evaluation of this approach highlights the positive impact of the empathic virtual agent on human-machine interaction.

Rodrigues et al. [22] propose a generic computational model of empathy between synthetic characters. The model is implemented into an affective agent architecture and the intensity of the empathic emotion is determined by the following modulation factors: *similarity*, *affective link*, *mood*, and *personality*. To evaluate their model a small scenario of four characters was defined. The scenario was implemented with respect to two conditions, without the empathy model and with the empathy model. The results suggest that their model has significant effects on subjects' perception of the empathic behaviors of the characters.

While significant advances have been made in modeling empathy for virtual humans, the modulation of the empathic emotion through factors like the empathizer's mood and relationship to the other is either missing or only the intensity of the empathic emotion is modulated. Hoffman [12] and Davis [6] emphasize that the empathic response to the other's emotion does not need to be in a close match with the affect experienced by the other, but can be any emotional reaction compatible with the other's condition (see Sect. 2.1). Accordingly, in our work the modulation factors not only affect the intensity of the empathic emotion but also its related type. That is, depending on the values of the modulation factors, EMMA's empathic emotion can be from different type and intensity to perceived emotion.

Moreover, research on emotion recognition mostly focuses on recognition of seven prototypic emotion categories which result from a cross-cultural study by Ekman and Friesen [7]. This categorical approach suffers from the fact that it is limited to a predefined number of prototypic emotions. In contrast, the dimensional approach is believed to be more convenient to model, analyze, and interpret the subtlety, complexity, and continuity of emotional expressions. However, little attention has been devoted to emotion recognition using a dimensional rather than a categorical approach, in particular regarding emotion recognition from fa-

cial expressions (see [11] for a review). Therefore, we introduce a new approach to infer PAD values from facial Action Units (AUs) [9] displaying emotions. Furthermore, the modulation of the type and intensity of the empathic emotion is realized in EMMA's PAD emotion space. These approaches are applied resp. in the first and second processing steps of the empathy model.

4 A Three-Step Approach for Empathy-Based Emotional Alignment

The Empathic MultiModal Agent—EMMA has a face which replicates 44 Action Units (AUs) implemented in line with Ekman and Friesen's Facial Action Coding System (FACS) [9]. EMMA's AUs include nine upper face units and 25 lower face units. The remaining AUs represent head and eye units. In an empirical study [3], human participants rated randomly generated facial expressions of EMMA with 18 bipolar adjectives on a one to seven Likert-Scale. Each group of six bipolar adjectives represents one of the dimensions of pleasure, arousal, and dominance. A two dimensional non-linear regression analysis of AU activation with pleasure and arousal was calculated for those faces that had been rated as showing high dominance and those that had been rated as showing low dominance. As a result three dimensional non-linear regression planes for each AU in pleasure-arousal spaces of high and low dominance were obtained (e.g., see Fig. 1). By combining all planes of all AUs the facial expression corresponding to each point in Pleasure-Arousal-Dominance (PAD) space is recomposed and a face repertoire is reconstructed. The face repertoire comprises faces arranged in PAD space with respect to two dominance values (dominant vs. submissive). A more comprehensive description of the empirical study is given in [3].

Similar to the virtual human MAX, EMMA has a cognitive architecture composed of an emotion simulation module [1] and a Belief-Desire-Intention (BDI) module [15]. The emotion simulation module consists of two components: First, the dynamics/mood component for the calculation of the course of emotions and moods over time and their mutual interaction. Second, the PAD space in which primary and secondary emotions are located and their intensity values can be calculated. At each point in time, the emotion simulation module outputs values of pleasure, arousal, and one of two possible values of dominance (dominant vs. submissive) as well as intensity values of primary and secondary emotions.

The restriction of the face repertoire, which results from the empirical study, to two dominance values (dominant vs. submissive) was done in order to adapt it to the emotion simulation module. Using this face repertoire, each PAD value output by the emotion simulation module over time is expressed by its corresponding face in the face repertoire.

The above outlined components constitute the framework used to realize our approach for modeling empathy for EMMA. Our approach is based on three processing steps: *Empathy Mechanism*, *Empathy Modulation*, and *Expression of Empathy* (see Fig. 2) introduced in the following.

4.1 Empathy Mechanism

Facial expressions are crucial in expressing and communicating emotions [8]. Thus, in our model of empathy, the mechanism by which an empathic outcome is produced—the *Empathy Mechanism*—is based on an internal simulation of perceived emotional facial expressions. It results in an emotional feedback that represents the empathic emotion.

Both Hoffman and Davis introduce *mimicry* as an *Empathy Mechanism* (see Sect. 2.2). Hoffman defines *facial mimicry* as the process involving the *imitation* of another's facial expressions, which triggers an afferent *feedback* eliciting the same feelings in oneself as those of the other. Following this, in our *Empathy Mechanism*, the internal simulation of perceived emotional facial expressions consists of an *internal imitation* of these expressions and results in an *emotional feedback* that represents the perceived emotional state. This is based on the use of a *shared representational system*. That is, our virtual human uses the same face repertoire to express its own emotions as well as to understand emotions from perceived facial expressions.

Accordingly, using her own AUs and their activation functions (regression planes) in PAD space, EMMA internally simulates a perceived facial expression by first mapping it to AUs with corresponding activation values (*internal imitation*) and by subsequently inferring its related emotional state as a PAD value (*emotional feedback*); (see Fig. 2). In order to determine the PAD value from AUs' activation values corresponding to a perceived facial expression, the following two methods are used and represented by (1) and (2). First, in (1), a weighted summation of the AUs' activation functions is calculated. The weights $w_i(t)$ represent the AUs' activation values at each point t in time. The p, a, d value corresponding to the maximum of the weighted summation is returned as $(p, a, d)_{t,hint}$. Since most of the AUs' activation functions have their maximum values at the boundaries of the PAD space (e.g., see Fig. 1), the resulting p, a, d value could also lie at the boundaries of this space. Thus, this method is only used to determine in which quadrant of the PAD space the p, a, d value to determine could lie. Accordingly, the domain of the AUs' activation functions is restricted to that quadrant of the PAD space. Second, in (2), the AUs' activation functions with restricted domain are translated to $w_i(t)$ and the sum of their absolute values is calculated. The p, a, d value resulting from the minimum of that sum denoted by $(p, a, d)_{t,final}$ represents the determined PAD value at timestamp t.

AU12 Lip Corner Puller
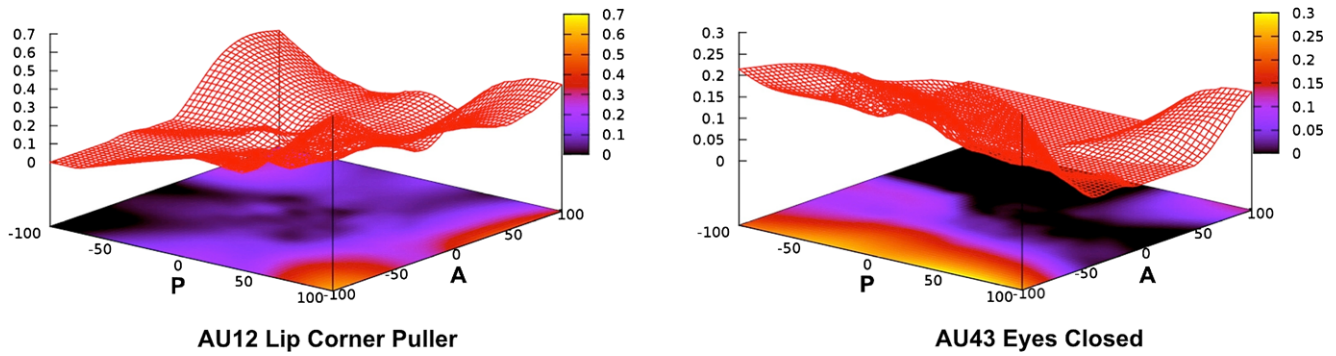
AU43 Eyes Closed

Fig. 1 The activation functions (regression planes) of AU12 (Lip Corner Puller) and AU43 (Eyes Closed) in PA space of high dominance. The labels P and A denote resp. pleasure and arousal. The *vertical axis* represents the AU's activation values. The activation function of AU12 shows a high activation with respect to positive pleasure independent of the arousal value and the activation function of AU43 shows a high activation with respect to negative arousal independent of the pleasure value

Consider the intensity values of AU12 and AU43 (see Fig. 1) equal to resp. 40 and 20 at time t. By applying (1), $(p, a, d)_{t,hint}$ is equal to $(100, -100, 100)$. By applying (2), $(p, a, d)_{t,final}$ is equal to $(68, -75, 100)$. Increasing the intensity of AU12 increases the value of positive pleasure. Increasing the intensity of AU43 increases the value of negative arousal. That is, smiling (AU12) with eyes closed (AU43) is mapped to positive pleasure and negative arousal. Thus, a PAD value is determined as a hypothesis about the emotional state related to a perceived facial expression.

$$\sum_{i=1}^{n} AU_i(p, a, d) \times w_i(t)$$

$$= AU_{\sum}(p, a, d, t)$$

$$(p, a, d)_{t,hint} = AU_{max}(t) \qquad (1)$$

$$\sum_{i=1}^{n} \left| AU_i(p, a, d) - w_i(t) \right|$$

$$= AU_{\sum}(p, a, d, t)$$

$$(p, a, d)_{t,final} = AU_{min}(t) \qquad (2)$$

The inferred PAD value is represented in EMMA's PAD emotion space. Therefore, its related primary emotion as well as its corresponding intensity value can also be inferred. Note that secondary emotions generated by EMMA's emotion simulation module will be relevant in future work (see Sect. 6). Since empathy implies a self-other distinction (see Sect. 2.1), the inferred PAD value is represented in EMMA's PAD emotion space by an additional reference point. Thus, EMMA always distinguishes between her own and the other's perceived emotional state.

The elicitation of an empathic emotion is caused by detecting the occurrence of a desirable or a not desirable event for others [18]. Therefore, in our approach the empathic emotion is elicited after detecting a fast and at the same time salient change in the other's emotional state that indicates the occurrence of an emotional event or if the other's emotional state is perceived as salient. That is, with respect to a predetermined short time interval T, the difference between inferred PAD values corresponding to the timestamps t_{k-1} and t_k, with $t_k - t_{k-1} \leq T$, is calculated as $|PAD_{t_k} - PAD_{t_{k-1}}|$. If this exceeds a predefined saliency threshold $TH1$ or if $|PAD_{t_k}|$ exceeds a predefined saliency threshold $TH2$, then the emotional state PAD_{t_k} and its related primary emotion represent the empathic emotion (see Fig. 2). Otherwise, no empathic emotion is elicited and the next processing steps, *Empathy Modulation* and *Expression of Empathy*, are not triggered (see Fig. 2). The predefined thresholds can be interpreted as representing EMMA's responsiveness to the other's situation. The empathic emotion is asserted as belief about perceived emotional state in EMMA's BDI module (see Fig. 2).

4.2 Empathy Modulation

Our model of empathy follows the *late appraisal model* of empathy (see Sect. 2.3), in that the *Empathy Mechanism* resulting in an empathic emotion takes place before the *Empathy Modulation* can occur. Therefore, the *Empathy Modulation* represents the second processing step of our approach. It consists of modulating the empathic emotion resulting from the first processing step through different modulation factors. Since the *mood*, the *emotional repertoire*, the *emotional regulation* of the empathizer as well as *desirability-for-self* are factors modulating the empathic emotion (see Sect. 2.3), the modulation of the empathic emotion is integrated into EMMA's affective architecture.

The empathic emotion as well as EMMA's emotional state are represented by PAD values. Thus, the modulation takes place in EMMA's PAD emotion space. This is realized
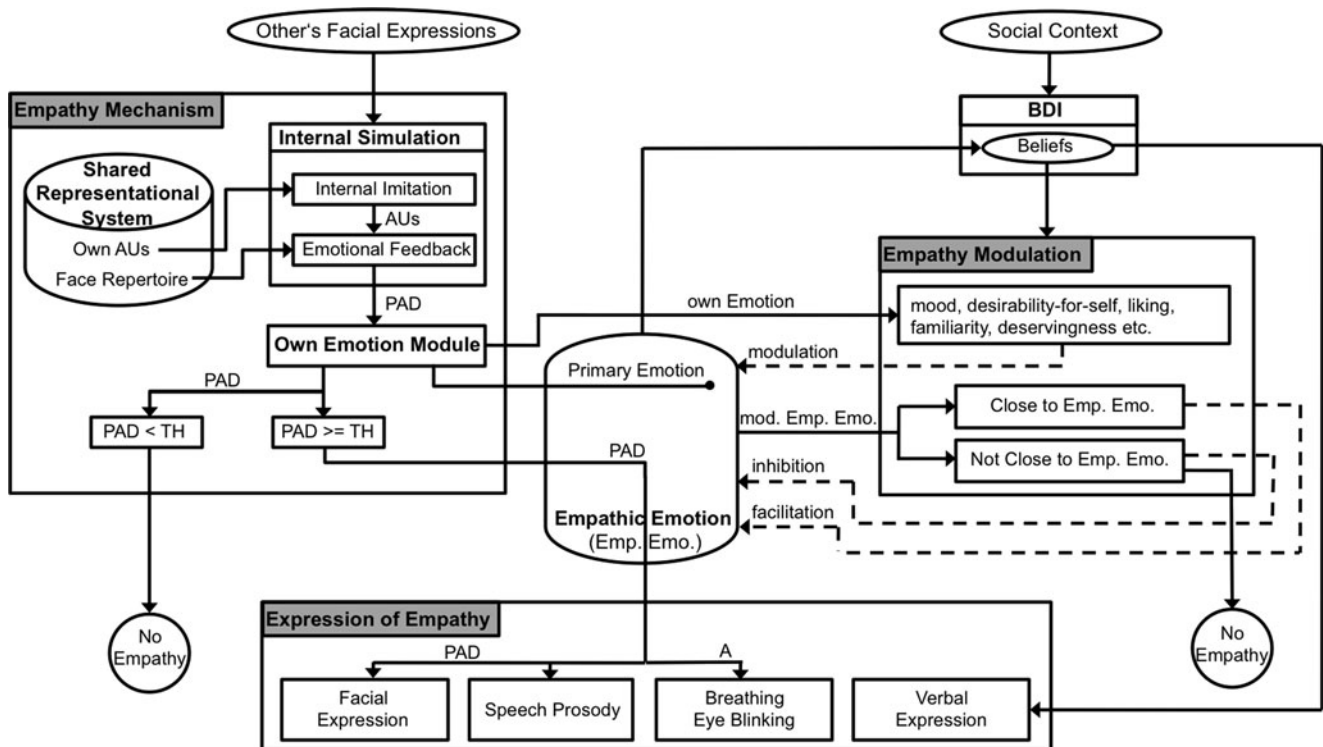
Fig. 2 The empathy model based on the *late appraisal model* of empathy [25] and consists of three processing steps: *Empathy Mechanism*, *Empathy Modulation*, and *Expression of Empathy*

by applying the following equation each time t an empathic emotion is elicited:

$$empEmo_{t,mod} = ownEmo_t + (empEmo_t - ownEmo_t)$$

$$\times \left(\sum_{i=1}^{n} p_{i,t} \times w_i \right) \Big/ \left(\sum_{i=1}^{n} w_i \right) \qquad (3)$$

The value $empEmo_{t,mod}$ represents the modulated empathic emotion. The value $ownEmo_t$ represents EMMA's emotional state. The value $empEmo_t$ represents the non-modulated empathic emotion resulting from the previous processing step. The values $p_{i,t}$ represent a set of predefined parameters discussed below.

Following the psychological background introduced in Sect. 2.3, in our approach the modulation factors *empathizer's mood* and *desirability-for-self* are represented by $ownEmo_t$. The parameters $p_{i,t}$ represent arbitrary predefined modulation factors with values ranging in [0, 1] such as *liking* and *familiarity*. *Familiarity* can be represented by values ranging in [0, 1] from *non-familiar* to *most-familiar*. Based on [18], *liking* can be represented by values ranging in [−1, 1] from *disliked* to *most-liked*. The value 0 represents neither *liked* nor *disliked*. Only the impact of positive values of $p_{i,t}$ is discussed in this paper. Thus, only positive values of *liking* are considered. The values of the modulation factors $p_{i,t}$ are either predefined or are determined from

the context at each timestamp t. These are asserted as beliefs in EMMA's BDI module (see Fig. 2).

By applying (3), $empEmo_{t,mod}$ lies on the straight line spanned by $ownEmo_t$ and $empEmo_t$ (see Fig. 3). Accordingly, we define the *degree of empathy* as the distance between $empEmo_{t,mod}$ and $empEmo_t$. That is, the closer $empEmo_{t,mod}$ to $empEmo_t$, the higher the *degree of empathy*. The less close $empEmo_{t,mod}$ to $empEmo_t$, the lower the *degree of empathy*. The *degree of empathy* is impacted by the values of the modulation factors at each point in time t. The impact of the modulation factors $p_{i,t}$ is calculated through a weighted mean of their current values at timestamp t. For example, *liking* can be defined as having more impact on the *degree of empathy* than *familiarity* and thus can be weighted higher.

In [22] the impact of the modulation factor *empathizer's mood* on the intensity of the empathic emotion is defined as follows: A negative mood increases the potential of a negative empathic emotion and decreases the potential of a positive one. In contrast, a positive mood increases the potential of a positive empathic emotion and decreases the potential of a negative one. Similar to this approach, EMMA is more sensitive to the empathic emotion when her emotional state is more similar to the empathic emotion. EMMA is more resistant to the empathic emotion when her emotional state is less similar to the empathic emotion. That is, the closer
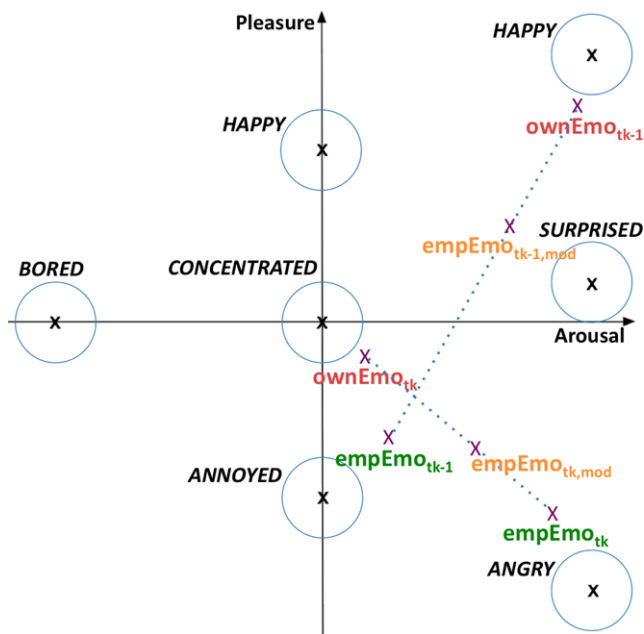
Fig. 3 The PA space of high dominance of EMMA's emotion simulation module. The primary emotions *happy*, *surprised*, *angry*, *annoyed*, *bored*, and the neutral state *concentrated* are located at different PA values. The reference points $ownEmo_{t_{k-1}}$ and $ownEmo_{t_k}$ represent EMMA's emotional state at timestamps t_{k-1} and t_k. The reference points $empEmo_{t_{k-1}}$ and $empEmo_{t_k}$ represent the non-modulated empathic emotion at timestamps t_{k-1} and t_k. The reference points $empEmo_{t_{k-1},mod}$ and $empEmo_{t_k,mod}$ represent the modulated empathic emotion at timestamps t_{k-1} and t_k

$ownEmo_t$ to $empEmo_t$, the higher the *degree of empathy* and the less the modulation factors $p_{i,t}$ can impact the *degree of empathy*. The less close $ownEmo_t$ to $empEmo_t$, the lower the *degree of empathy* and the more the modulation factors $p_{i,t}$ can impact the *degree of empathy*. Considering the modulation factors $p_{i,t}$, the higher their value of weighted mean, the closer $empEmo_{t,mod}$ to $empEmo_t$ and the higher the *degree of empathy*. The lower their value of weighted mean, the less close $empEmo_{t,mod}$ to $empEmo_t$, the lower the *degree of empathy*, and the closer $empEmo_{t,mod}$ to $ownEmo_t$ (see Fig. 3).

Hoffman [12] and Davis [6] emphasize that the empathic response to the other's emotion should be more appropriate to the other's situation than to one's own and can be any emotional reaction compatible with the other's condition (see Sect. 2.1). Thus, $empEmo_{t,mod}$ is facilitated only if its related primary emotion is defined as close enough to that of $empEmo_t$. Otherwise, $empEmo_{t,mod}$ is inhibited and the *Expression of Empathy*, is not triggered (see Fig. 2). That is, for each primary emotion located in PAD space, a distance threshold to close primary emotions in PAD space is defined. The distance threshold consists of threshold values in pleasure, arousal, and dominance dimensions. Primary emotions defined as close to $empEmo_t$'s primary emotion should

represent emotional reactions that are compatible with the other's condition.

For example, Fig. 3 shows the PA space of high dominance of EMMA's emotion simulation module. At time t_{k-1}, $ownEmo_{t_{k-1}}$ has as related primary emotion *happy*, $empEmo_{t_{k-1}}$ has as related primary emotion *annoyed*, and the values of $p_{i,t_{k-1}}$ are set to the value 0.4. $empEmo_{t_{k-1},mod}$ has as related primary emotion *surprised* which is defined as not close enough to *annoyed*. At this stage $empEmo_{t_{k-1},mod}$ is inhibited and EMMA's expression of empathy is not triggered. At time t_k, $ownEmo_{t_k}$ is the neutral state *concentrated*, $empEmo_{t_k}$ has as related primary emotion *angry*, and the values of p_{i,t_k} are set to the value 0.6. $empEmo_{t_k,mod}$ has as related primary emotion *annoyed* which is defined as close enough to *angry*. At this stage $empEmo_{t_k,mod}$ is facilitated and EMMA's expression of empathy is triggered.

4.3 Expression of Empathy

The *Expression of Empathy* (see Fig. 2) represents the third processing step of our approach. It consists of triggering EMMA's expression of empathy by the modulated empathic emotion derived in the previous processing step. Based on EMMA's face repertoire, the PAD value of the modulated empathic emotion triggers EMMA's corresponding facial expression. In the *EmoSpeak* module of the German text-to-speech system MARY (Modular Architecture for Research on speech sYnthesis), PAD based emotional prosody rules are specified [24] [23]. Using *EmoSpeak*, EMMA's speech prosody is modulated by the PAD value of the modulated empathic emotion. The frequencies of EMMA's eye blinking and breathing are modulated by the arousal value of the modulated empathic emotion. That is, the higher the arousal value, the higher the frequencies of EMMA's eye blinking and breathing. Triggering other modalities like verbal utterances or spatial actions as expressions of empathy depends on the scenario's context.

5 Example Scenario

As mentioned in Sect. 1, we choose a conversational agent scenario in which the virtual human MAX acts as a museum guide [13]. Within this scenario, MAX engages with human partners in a natural face-to-face conversation and conducts multimodal small talk dialogs using speech, gestures, and facial behaviors. The emotions of MAX can be triggered in different ways. For example MAX's emotions are triggered positively when a new person enters his field of view or when the human partner's verbal expression is interpreted as a compliment. MAX's emotions are triggered negatively when the human partner's verbal expression is interpreted as obscene or politically incorrect.
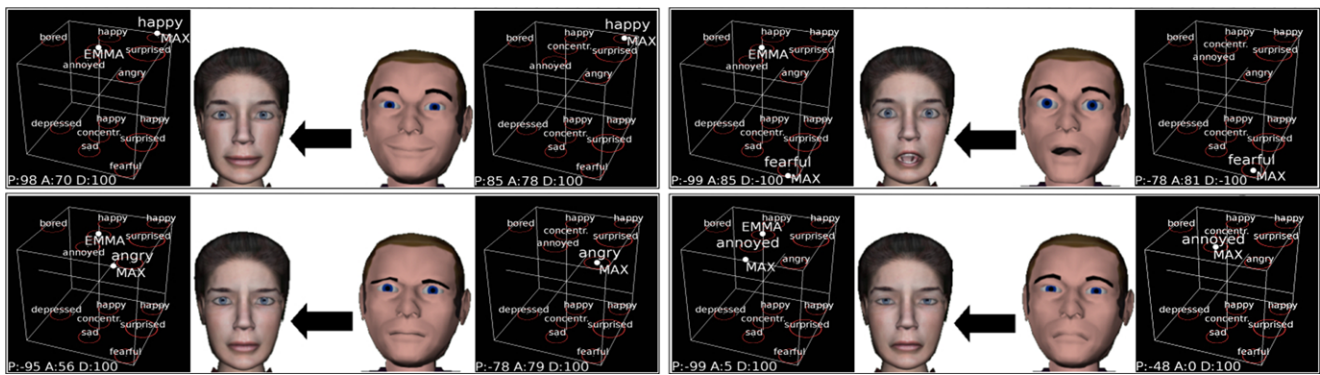
Fig. 4 In each framed illustration: *On the right*, MAX's PAD emotion space and MAX's facial expressions corresponding to the primary emotions *fearful*, *annoyed*, *happy*, and *angry*. The reference point labeled with MAX represents its current PAD value. *On the left*, the facial expressions internally simulated by EMMA and their mapping to PAD values represented in EMMA's PAD emotion space. The derived PAD values are represented by the reference points labeled with MAX

In this scenario, we integrate EMMA as a third interaction partner. The human partner can also engage in a small talk dialog with EMMA and her emotions can be triggered in the same way as those of MAX. When the human partner is interacting with MAX, EMMA follows the conversation by directing her attention to the speaking agent. When attending to MAX, EMMA's empathy process is triggered. In the following, the three processing steps of the empathy model are illustrated within this scenario.

When attending to MAX, by means of the *Empathy Mechanism* introduced in Sect. 4.1 EMMA internally simulates MAX's facial expression. This is done by getting the values of his face muscles at each point in time, by mapping them to her own AUs, and by inferring their related PAD value as a hypothesis about MAX's emotional state. The inferred emotional state is represented by a second reference point in EMMA's PAD emotion space thus allowing EMMA to distinguish between her own and MAX's emotional states. This is illustrated in Fig. 4 by some example facial expressions of MAX corresponding to the emotions *fearful*, *annoyed*, *happy*, and *angry* at the timestamps of their highest intensity values. Figure 4 shows that the inferred PAD values are quite accurate compared to MAX's PAD values. MAX's facial expressions corresponding to *concentrated* and *surprised* are also mapped to quite accurate PAD values. The emotions *sad* and *depressed* are represented by the same facial expression as *annoyed*, thus EMMA maps them to the same PAD value as that for *annoyed* which results in a less accurate hypothesis about MAX's emotional state. At this point, further implication of context information is assumed to be relevant. This makes the scenario of EMMA interacting with MAX a more realistic, but also a challenging one. If MAX's perceived emotional state exceeds a predefined saliency threshold or if a salient change in MAX's emotional state is detected (see Sect. 4.1), the perceived emotional state is labeled as the empathic emotion and is asserted as belief in EMMA's BDI module.

Based on the *Empathy Modulation* introduced in Sect. 4.2, an elicited empathic emotion is modulated by means of the following two factors: First, EMMA's *mood* which changes dynamically over the interaction when the human partner triggers EMMA's emotions negatively or positively. Second, EMMA's relationship to MAX represented by EMMA's *liking* toward MAX and EMMA's *familiarity* with MAX. The values of *liking* and *familiarity* are predefined and do not change dynamically over the interaction. Thus, the impact of the *mood* factor as dynamically changing over the interaction can be better perceived in this scenario. For example, when EMMA's emotions are triggered by the human partner negatively, EMMA's empathy with MAX's positive emotional states is either low or is not triggered at all. Depending on the values and weights of *liking* and *familiarity* different profiles of EMMA's modulated empathy with MAX can be defined. The higher the values of *liking* and *familiarity*, the higher EMMA empathizes with MAX.

Once the empathic emotion is modulated, EMMA's multiple modalities are triggered by means of the *Expression of Empathy* introduced in Sect. 4.3. EMMA's verbal utterance is triggered as follows: In this scenario, MAX's emotions can be triggered negatively or positively by the human partner. Consequently, this is reflected by a negative or a positive change in MAX's pleasure value. By calculating the difference of the pleasure values of MAX's perceived emotional state, $P_{t_k} - P_{t_{k-1}}$, at timestamps t_{k-1} and t_k, EMMA can detect the changes in MAX's pleasure value. A positive change in pleasure triggers a verbal expression of EMMA that encourages the human partner to continue being kind to MAX. A positive change that results in a positive pleasure value triggers verbal expressions such as "That's great, you are so kind to MAX!". A positive change in the negative space of pleasure triggers verbal expressions such as "This

is not enough! You have to be kinder!". A negative change in pleasure triggers a verbal expression of EMMA that advises the human partner not to be unfriendly to MAX. A negative change in the space of positive pleasure triggers verbal expressions such as "Why are you saying that to MAX? This is nasty!". A negative change that results in a negative pleasure value triggers verbal expressions such as "Better think about what you are saying! This is really nasty!".

In this scenario we illustrated how by means of the three-step approach introduced in Sect. 4, EMMA can empathize with MAX and thus align to MAX's emotions depending on her mood and current relationship to him.

6 Conclusion and Future Work

Supported by psychological models of empathy, in this paper we introduced an approach to model empathy for EMMA based on three processing steps: *Empathy Mechanism*, *Empathy Modulation*, and *Expression of Empathy*. The *Empathy Mechanism* is based on using a *shared representational system*. That is, using her own AUs and their activation functions in PAD space, EMMA infers PAD values from perceived AUs displaying emotions. Therefore, our approach to infer emotions from perceived facial expressions is not limited to a predefined number of emotion categories. By means of the PAD values inferred from perceived facial expressions, EMMA can detect fast and salient changes in the other's emotional state, e.g., a fast and salient change from *happy* to *neutral* is perceived as a negative one. However, inferring emotions from other's perceived facial expressions without integrating contextual information may lead to false assumptions. Thus, EMMA is likely to falsely interpret, e.g., a facial expression accompanying the secondary emotion of relief as the primary emotion of happiness. In order to make the interpretation process more adequate (e.g., to distinguish between happiness and relief) we aim to consider *situational role-taking* (cf. [2]) as an additional *Empathy Mechanism*.

Further, the *Empathy Modulation* do not only affect the intensity of the empathic emotion but also its related type. This is in agreement with Hoffman's as well as Davis' emphasis that an empathic response does not need to be in a close match with the affect experienced by the other, but can be any emotional reaction compatible with the other's condition (see Sect. 2.1). The *Empathy Modulation* can be extended to further modulation factors like *deservingness* and *similarity* (see Sect. 2.3). EMMA's personality is modeled by means of parameters defined in her emotion simulation module. These parameters impact how emotional EMMA is (e.g., temperamental vs. lethargic) [1]. Which parameters these are and how they impact EMMA's degree of empathy with her interaction partner will be discussed in future work.

We further aim at empirically evaluating the empathic behavior of EMMA within the scenario introduced in Sect. 5.

In particular we will focus on the impact of EMMA's mood, as a modulation factor that dynamically changes over the interaction, on human subjects' perception of EMMA's empathic behavior. Based on our approach to model empathy, we aim to modulate a virtual human's spatial helping actions in a cooperative spatial interaction task by its degree of empathy with its interaction partner. As *Empathy Mechanism*, we will consider *situational role-taking* (cf. [2]). As modulation factors we will consider the virtual human's *mood*, *liking*, and *deservingness* and show how the values of *liking*, and *deservingness* can change dynamically over the interaction. Furthermore, a challenging work is to model 'negative empathic' behavior by extending the values of the modulation factors *liking* and *deservingness* to negative ones.

References

1. Becker-Asano C, Wachsmuth I (2009) Affective computing with primary and secondary emotions in a virtual human. Auton Agents Multi-Agent Syst. doi:10.1007/s10458-009-9094-9
2. Boukricha H, Becker C, Wachsmuth I (2007) Simulating empathy for the virtual human max. In: 2nd international workshop on emotion and computing, in conj with the German conference on artificial intelligence (KI2007), Osnabrück, Germany, pp 22–27
3. Boukricha H, Wachsmuth I, Hofstätter A, Grammer K (2009) Pleasure-arousal-dominance driven facial expression simulation. In: 3rd international conference on affective computing and intelligent interaction and workshops (ACII), Amsterdam, Netherlands. IEEE Press, New York, pp 119–125
4. Brave S, Nass C, Hutchinson K (2005) Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int J Hum-Comput Stud 62:162–178
5. Burleson W, Picard R (2007) Affective learning companions. Educ Technol 47(1):28–32
6. Davis MH (1994) Empathy: a social psychological approach. Brown and Benchmark Publishers
7. Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. J Pers Soc Psychol 17(2):124–129
8. Ekman P, Friesen WV, Ellsworth P (1972) Emotion in the human face: guidelines for research and an integration of findings. Pergamon, Elmsford
9. Ekman P, Friesen WV, Hager JC (2002) Facial action coding system: investigator's guide. Research nexus, a subsidiary of network information research corporation, Salt Lake City UT, USA
10. Goldstein AP, Michaels GY (1985) Empathy development, training and consequences. Erlbaum, Elmsford
11. Gunes H, Pantic M (2010) Automatic dimensional and continuous emotion recognition. Int J Synth Emot 1(1):68–99
12. Hoffman ML (2000) Empathy and moral development. Cambridge University Press, Cambridge

13. Kopp S, Gesellensetter L, Kraemer N, Wachsmuth I (2005) A conversational agent as museum guide—design and evaluation of a real-world application. In: Intelligent virtual agents. Springer, Berlin, pp 329–343

14. Kopp S, Jung B, Lessmann N, Wachsmuth I (2003) Max— a multimodal assistant in virtual reality construction. Künstl Intell 4/03:11–17

15. Lessmann N, Kopp S, Wachsmuth I (2006) Situated interaction with a virtual human—perception, action, and cognition. In: Rickheit G, Wachsmuth I (eds) Situated communication. Mouton de Gruyter, Berlin, pp 287–323

16. McQuiggan S, Robison J, Phillips R, Lester J (2008) Modeling parallel and reactive empathy in virtual agents: An inductive approach. In: Padgham A, Parkes A, Müller A, Parson A (eds) Proc of 7th int conf on autonomous agents and multiagent systems (AAMAS 2008). Estoril, Portugal, pp 167–174

17. Ochs M, Pelachaud C, Sadek D (2008) An empathic virtual dialog agent to improve human-machine interaction. In: Padgham A, Parkers A, Müller A, Parson A (eds) Proc of 7th int conf on autonomous agents and multiagent systems (AAMAS 2008). Estoril, Portugal, pp 89–96

18. Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge

19. Paiva A, Dias J, Sobral D, Aylett R, Woods S, Hall L, Zoll C (2005) Learning by feeling: evoking empathy with synthetic characters. Appl Artif Intell 19:235–266. doi:10.1080/08839510590910165

20. Prendinger H, Becker C, Ishizuka M (2006) A study in users' physiological response to an empathic interface agent. Int J Humanoid Robot 3(3):371–391

21. Prendinger H, Ishizuka M (2005) The empathic companion: a character-based interface that addresses users' affective states. Appl Artif Intell 19:267–285. doi:10.1080/08839510590910174

22. Rodrigues SH, Mascarenhas S, Dias J, Paiva A (2009) I can feel it too!: Emergent empathic reactions between synthetic characters. In: 3rd international conference on affective computing and intelligent interaction and workshops (ACII), Amsterdam, Netherlands. IEEE Press, New York

23. Schröder M (2004) Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions. In: Affective dialogue systems, vol 3068. Springer, Berlin, pp 209–220

24. Schröder M, Trouvain J (2003) The german text-to-speech system Mary: a tool for research development and teaching. Int J Speech Technol 6(4):365–377. doi:10.1023/A:1025708916924

25. de Vignemont F, Singer T (2006) The empathic brain: how when and why? Trends Cogn Sci 10(10):435–441. doi:10.1016/j.tics.2006.08.008
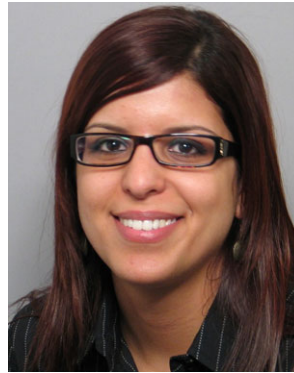
Hana Boukricha is a Ph.D. student and research assistent at the Collaborative Research Center 'Alignment in Communication' (CRC 673) in project A1—Modelling Partners. Her main research interests are cognitive modeling and empathic agent architectures.

Ipke Wachsmuth is chair of Artificial Intelligence at Bielefeld University. He is co-initiator and current coordinator of the Collaborative Research Center 'Alignment in Communication' (CRC 673).