First ideas on the use of affective cues in an empathic computer-based companion

Nikolaus Bee
Institute of Computer Science
University of Augsburg, Germany
bee@informatik.uni-augsburg.de

Elisabeth André
Institute of Computer Science
University of Augsburg, Germany
andre@informatik.uni-augsburg.de

Thurid Vogt
Institute of Computer Science
University of Augsburg, Germany
vogt@informatik.uni-augsburg.de

Patrick Gebhard
DFKI
Embodied Agents Research Group, Germany
patrick.gebhard@dfki.de

ABSTRACT

This paper describes a system to enhance the interaction between humans and virtual characters with emotional mimicry and role-taking. Such system increases the believability of virtual agents. Mimicking necessitates a model of emotional intelligence to understand and display user's emotions. A more complex processing is however necessary for a reactive behavior, where the virtual agent reacts e.g. in an encouraging way which allows to actively change user's current state.

A virtual character with highly expressive capabilities was created to create a platform to figure out the differences in mimicking and role-taking. As we will concentrate on nonverbal behavior input from users, our agent will not be able to understand what, but how users are speaking.

1. INTRODUCTION

A virtual companion that stays for a long period of time with a user and that learns and knows about the preferences and wishes of its owner continuously, requires emotional intelligence that allows it to observe, to estimate and to manage its and the others emotions. It should be capable to detect users' affective state and respond appropriately to it in real-time [8]. In one situation, it might help if the agent shows that it feels with the user by simply mirroring the user's emotional state. This mimicry (i.e. parallel empathy) is the capability to display the user's emotion in a similar manner to the user's current emotional expression. In contrast, reactive behavior aims to understand the user's affective state and tries to alter or enhance it.

Assume, for example, a situation where your best friend fails an exam and tells this to you. This exam was very important for your friend as it decides about graduating. You might now react in two ways. One would be to completely empathize with your friend about the failed exam. Fortunately you are an attentive listener and finally your friend already feels better, true the motto: a problem shared is a problem halved. The other possibility for you to react, is to understand the current situation and you start to encourage your friend by, for instance, explaining that you also failed some exams but there is always a second chance. In cognitive science, theory of mind enables a person to infer from the users' verbal and non-verbal behavior what they intend to do, desire, think or belief.

While a lot of work has been done in creation of affective output for virtual characters, less work was done in combining the recognition of user's affective state with the affective display of current systems for embodied conversational agents.

Prendinger and colleagues [18] developed an empathic companion that accompanies a user in a virtual job interview. This system measures users' physiological state (skin conductance and electromyography) in real-time and interprets it as emotion. The virtual agent reacts with empathic feedback dependent on the users' current affective state. The reaction is calculated with a Bayesian net, that takes the physiological state and user's job interview answers into account. The Bayesian net is modeled after findings in literature. Gratch et al. [8] describe a system for rapport in human-machine dialogs. They detect speech and head orientation from the user to create continuing dialogs with their system. The speech detection is used for detecting backchannel opportunity points, disfluencies, questions and loudness and the head tracker detects head nods, shakes, gaze shifts and posture shifts. Their system is mainly meant to create contingent nonverbal behavior and neither takes the user's emotional state into account nor intends to react with affective behavior. McQuiggan et al. [11] created a system for empathic virtual agents. They compare parallel with reactive empathy behavior. Using a machine learning approach let's them automatically learn from the users' behavior. The system is trained by users' in-game behavior. Thus, it does not consider the real users' actual emotional state. Ochs et al. [15] focus on creating an empathic model for virtual agents with a combination of analyzing real human-machine dialogs and theoretical descriptions of emotions. This agent

Cite as: , , *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX.

helps the users to obtain empathic information about their emails. Their corpus was annotated taking vocal and semantic cues into account. Boukricha [2] proposes a computational model of affective theory of mind for empathic virtual agents. She defines affective theory of mind by sharing their emotions (mimicry) and by understanding the counterparts' emotions cognitively (role-taking). Detection of the users' current emotional state, which will be mapped to a virtual character, is planned by using facial feature detection. There is an ongoing project named SEMAINE [20] that deals with building Sensitive Artificial Listeners which are designed to sustain the communication with a person. Their system will recognize and generate non-verbal behavior in real-time.

Our objective is the creation of an empathic listening agent that analyzes the user's emotive state and responds to it in real-time. To realize such an agent, we have been experimenting with several metaphors, such as that of a virtual pet or that of a virtual butler. While the virtual pet is not able to verbally respond to the user's state, the virtual butler gives both verbal as well as non-verbal feedback. Imagine the user has a rather bad day and is talking to the agent with a depressed voice. In the case of parallel empathy, the agent would simply mimicry the user's emotive state and show depression as well (see Fig. 1 left). In contrast to that, reactive empathy requires the agent to decide which emotion to display as a response to the user's emotive state. Here, the agent's emotions does not necessarily coincide with the user's emotions. That is the agent might, for example, decide to sheer the user up by showing encouragement (see Fig. 1 right). The agent is also able to provide simple verbal feedback, such as "I know how you feel!" or "That is really awful!".

2. THEORY OF MIND

Theory of mind is the cognitive ability to understand what others intend to do or think. It enables us to interpret the counterpart's behavior. Furthermore, it allows us to assume or predict what our counterpart intends, desires and believes. Such characteristic is essential for a virtual agent with emotional intelligence. As we are particularly interested in the interaction between real and virtual world (human-agent), a model that reflects the real and virtual world in an agent's mind is necessary. The attentive capability of our agent model will include both worlds. In contrast to the affective cognitive model, that will be limited to the real world and the user only, as our empathic listener is currently alone in its virtual world.

Children develop a theory of mind with 3-5 years. Typical tests for humans to detect the capability of theory of mind are the appearance-reality or false-belief task. A cognitive model that passes the latter task was, for example, implemented by Bringsjord [3]. As our virtual agents do not have to understand false-beliefs, we will simplify our theory of mind and build the cognitive ability of our virtual agent on an affective theory of mind for mirroring users' emotions and for being aware of users' emotions, which will allow us to react on users' emotions.

Boukricha [2] describes Affective Theory of Mind as a model that shares emotions (being able to mimic a person's emotion) and that understands a person's emotions (being able to alter a person's emotion). The first part of this model's behavioral pattern is innate and the expression of emotional feedback is involuntarily. Such feedback behavior does not need a high level of cognitive capabilities. Whereas, for the second part of the model for AToM, i.e. to react with pity or sympathy to the users' emotional state, our system needs a higher level of processing. The virtual agent shall understand in what emotional state the user is to react in an appropriate way. So it necessitates, when the input components detect e.g. sadness from the user that the current state in the emotional model of our virtual character shifts to something appropriate for 'pity for'. This lets our virtual character display the correct emotion for interacting with a user.

3. AFFECTIVE INTERACTION

There are many emotional models around, e.g. EMA [7]. We use ALMA because it describes how emotions evolve over time. We combine the component for affective sensory input with an emotional model for our empathic listener that allows it to act or react to the users' feelings. Although the feedback as listener is limited in a way, timing and understanding the user becomes crucial. Our system detects the users' emotions from voice via the tool EmoVoice [22]. Further, we use a realistic virtual character with highly expressive facial emotions to display appropriate facial expressions.

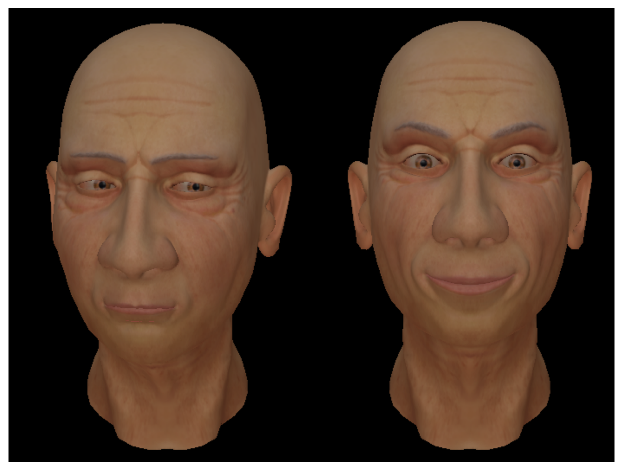

Figure 1: The virtual character Alfred is designed utilizing FACS to compose facial expressions.

Our architecture provides currently components to sense emotional states from the user using a microphone, a component to process affective states for mimicry or role taking and a component to display affect with virtual characters.

3.1 Affect Sensing

For sensing affect from users' voice, we use EmoVoice. It is a framework that provides support for the acquisition of emotional speech corpora and the training of classifiers. Furthermore, it is suitable both for offline as well as online vocal emotion recognition. The framework is intended to be used by non-experts and therefore comes with an interface to create an own personal or application specific emotion recognizer [22].

3.2 Affect Model - ALMA Bio

For the affect simulation in real-time, we rely on *ALMA Bio* and extended version of the computational model ALMA

[6]. ALMA Bio allows processing of bio signals. In this context bio signals are treated as unspecific emotions that contains individual (measured) pleasure, arousal and dominance values. Bio signal emotions are used as input to change or intensify the current mood.

ALMA provides three affect types as they occur in human beings: (1) *emotions* reflect short-term affect that decays after a short period of time; (2) *moods* reflect medium-term affect, which is generally not related to a concrete event, action or object; and (3) *personality* reflects individual differences in mental characteristics and affective dispositions.

ALMA implements the cognitive model of emotions developed by Ortony, Clore and Collins (OCC) [16] combined with the *BigFive* model of personality [10] and a simulation of mood based on the PAD model [12]. The relations between the different affect types is an central part of the affect simulation:

- *A given personality* defines a default mood and influences the intensities of different emotions.
- *The current mood* amplifies or dampens the intensities of emotions.
- *Emotions* as short term affective events influence the longer-term mood.

Elicited emotions influence an individual's mood. The higher the intensity of an emotion, the higher the particular mood change. Emotions usually do not last forever. Over a specific period the intensity of emotions decays and the influence on the current mood fades.

The current mood also influences the intensity of emotions [13]. This simulates, for example, the intensity increase of *joy* and the intensity decrease of *distress*, when a individual is in an *exuberant* mood. Mood is represented by a triple of the mood traits pleasure (P), arousal (A) and dominance (D). The mood's trait values define the mood class. If, for example, every trait value is positive (+P,+A,+D), the mood is *exuberant*.

3.3 Alfred – Affect Display

The affective display of our virtual agent consists of an enriched set of facial animations. "Alfed", Our realistic looking virtual character is able to talk with a text-to-speech (TTS) system or by playing prerecorded audio files.

3.3.1 Facial Expression

Ekman and Friesen developed the Facial Action Coding System (FACS) to classify human facial expressions [4]. FACS divides the face into action units (AU) to describe the different expressions a face can display (e.g. inner brow raiser, nose wrinkler, or cheek puffer). Although FACS was originally designed to analyze natural facial expressions, it turned out to be usable as a standard for production purposes too. That is why FACS based coding systems are used with the generation of facial expressions displayed by virtual characters, like Kong in Peter Jackson's King Kong [19]. But the usage of FACS is not only limited to virtual characters in movies. The gaming industry with Half-Life 2 by Valve, also utilizes the FACS system to produce the facial expressions of their characters [21].

Alfred (see Fig. 1), a butler-like character, uses these action units to synthesize an unlimited set of different facial expressions. The action units were designed using morph targets and thus gives the designer the full power in defining the facial expression outlook. The system includes a tool to control the single action units. The tool allows to store the result in a XML file for later usage in our agent system [1].

We chose the FACS-based approach for our facial animation system, because of the Facial Expression Repertoire (FER) [5], which maps over 150 emotional expressions to the action units of FACS. Not only does it explain in detail, which action unit must be activated for certain facial expressions, it further provides a rich dataset of videos which show how the action units ought to be designed.

Alfred's mesh has a resolution of about 21.000 triangles. For displaying more detailed wrinkles in the face, normal maps baked from a high-resolution mesh are used [14]. The morph targets for the action units are modeled using the actor's templates from the FER. For rendering the character and its animations the Horde3D GameEngine [9] is used.

3.3.2 Speech

The system interfaces the Microsoft Speech API to synchronize the audio output with the lip movements. This allows us to use any text-to-speech that supports SAPI 5. As the quality of common TTS systems may not be satisfactory, we integrated a module to synchronize prerecorded audio speech files with the lip movements of the virtual character. This allows us to use highly emotional sentences or affect bursts to be spoken through a virtual character. As FACS defines several action units involving mouth muscles (e.g. lip funneler, lip tightener, mouth stretch), we utilize the FACS system for lip movements. The approach is similar to displaying facial expressions. The output from the editor to modify the single action units is stored in a XML file. Reusing the FACS approach for visemes enables Alfred to display facial expressions and lip movements parallel.

4. EMPATHIC FEEDBACK

We designed an empathic model which is responsible to interact with mimicry and to react with reactive empathy. While the model for parallel empathy goes along with the emotional model in ALMA, the model for reactive empathy needs some further discussion. Both systems generate the facial expression by picking up the according facial expression dependent on the current emotional state from an emotional model.

4.1 Parallel Empathy

For mimicry (empathy or emotional contagion) our system (see Fig. 2) does not need a complex model to understand and interpret users' emotions. It is sufficient to map users' recognized emotions directly to the affective display of a virtual character. We created an emotional model in ALMA to represent the user's emotion in our system. We simply take the current state in the model and display it with the virtual character to mirror the user's emotional state. EmoVoice maps the classification result directly into ALMA. The same mapping between PAD values and the facial animation system is used to display emotional expression with Alfred, our virtual character. While in parallel empathy, we use ALMA to describe how emotions evolve over time and simply use it to "store" the current user's emotional state.

4.2 Reactive Empathy

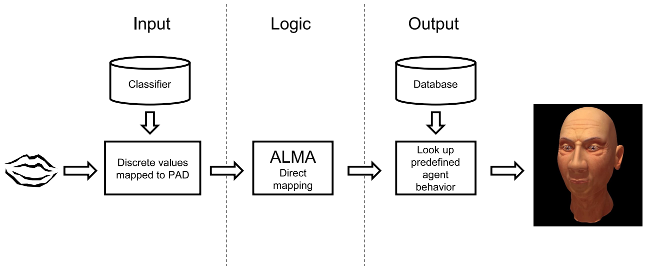

Figure 2: Model for mirroring

Here, the agent does not display emotions it feels, but tries to express deliberately emotions that might help to put the user into a better emotional state. Reactive empathy requires the agent to put himself into the shoes of the user in order to decide on an appropriate emotional display. For example, if the user is afraid of failing in a exam, the agent might tell the user that he will manage due to his good preparation. That is the agent praises the user which is perceived by the user a positive event. In addition, the agent utters a statement which decreases for the user the likelihood of the negative event (namely failing in the exam). Further, in reactive mode, it is crucial for the virtual agents to be able to utter vocally. Especially if its emotional display does not coincide with the user's emotion. For example, if the user is sad and the agent is smiling, the user might not perceive this as encouragement, but as gloating (see Fig. 1 right).

5. CONCLUSION

While developing the current architecture, we came up with some issues that need to be further discussed.

Reactive empathy cannot be modeled so far by ALMA Bio. While ALMA is at least capable to analyze the user's state. For example, is the user sad, is it an unpleasant event. Is the user the agents friend, will the agent feel "pity for". But is the user the agents enemy, the agent will feel gloating. In this case the reaction would not deal with empathy.

Another not trivial issue is the decision when a virtual agent with the ability of empathic reasoning should respond parallel or reactive [11]. Our system provides the possibility to test different behavior variants to figure out when a parallel empathic reaction or a reactive empathic reaction for a virtual character in a human-machine interaction is appropriate.

Conversational feedback is missing in our approach. Nevertheless, it is essential in human-human interaction to keep a conversation running and to show understanding to what somebody says. A virtual agent that is not able to show that it is following the conversation with giving appropriate feedback to the users will hardly be accepted as an empathic listener. Engender rapport with a virtual agent can be as effective as a human listener [8]. Feedback without understanding the content of what is said is called envelope feedback. Although our agent will not understand what users say, a method for responding in an appropriate manner is necessary to keep a conversation running. Such signals must transmit engagement, interest, understanding, agreement and of course emotional feedback [17].

6. REFERENCES

[1] N. Bee, B. Falk, and E. André. Simplified facial animation control utilizing novel input devices: A comparative study. In *International Conference on Intelligent User Interfaces (IUI '09)*, pages 197–206, 2009.

[2] H. Boukricha. A first approach for simulating affective theory of mind through mimicry and role-taking. In *The Third International Conference on Cognitive Science, Symposium: Emotional Computer Systems and Interfaces*, 2008.

[3] S. Bringsjord, A. Shilliday, M. Clark, D. Werner, J. Taylor, A. Bringsjord, and E. Charpentie. Toward logic-based cognitively robust synthetic characters in digital environments. In *Proceedings of the First Conference on Artificial General Intelligence (AGI-08)*, 2008.

[4] P. Ekman and W. Friesen. *Unmasking the Face.* Prentice Hall, 1975.

[5] Facial Expression Repertoire. Filmakademie Baden-Württemberg. `http://research.animationsinstitut.de/`.

[6] P. Gebhard. Alma - a layered model of affect. In *Proc. of the 4th Int. Joint Conference on Autonomous Agents and Multiagent Systems*, pages 29–36. ACM, June 2005.

[7] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, December 2004.

[8] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In *Intelligent Virtual Agents (IVA 2007)*, pages 125–138, 2007.

[9] Horde3D GameEngine. University of Augsburg. `http://mm-werkstatt.informatik.uni-augsburg.de/projects/GameEngine/`.

[10] R. McCrae and O. John. An introduction to the five-factor model and its applications. *Journal of Personality*, 60:175–215, 1992.

[11] S. W. McQuiggan, J. L. Robison, R. Phillips, and J. C. Lester. Modelling parallel and reactive empathy in virtual agents: An inductive approach. In *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 167–174, 2008.

[12] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology: Developmental, Learning, Personality, Social*, 14:261–292, 1996.

[13] R. Neumann, B. Seibt, and F. Strack. The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Journal of Cognition & Emotion*, 15(6):725–747, 2001.

[14] C. Oat. Animated wrinkle maps. In *SIGGRAPH '07: ACM SIGGRAPH 2007 courses*, pages 33–37, New York, NY, USA, 2007. ACM.

[15] M. Ochs, C. Pelachaud, and D. Sadek. An empathic virtual dialog agent to improve human-machine interaction. In *7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 89–96, 2008.

[16] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive*

Structure of Emotions. Cambridge University Press, Cambridge, MA, 1988.

[17] C. Peters, C. Pelachaud, E. Bevacqua, M. Ochs, N. E. Chafai, and M. Mancini. Social capabilities for autonomous virtual characters. In *International Digital Games Conference*, pages 37–48, 2006.

[18] H. Prendinger, H. Dohi, H. Wang, S. Mayer, and M. Ishizuka. Empathic embodied interfaces: Addressing usersŠ affective state. In *In Proceedings Tutorial and Research Workshop on Affective Dialogue Systems, LNAI 3068*, pages 53–64, 2004.

[19] M. Sagar. Facial performance capture and expressive translation for king kong. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Sketches*, page 26, New York, NY, USA, 2006. ACM.

[20] M. Schröder, R. Cowie, D. Heylen, M. Pantic, C. Pelachaud, and B. Schuller. Towards responsive sensitive artificial listeners. In *Fourth International Workshop on Human-Computer Conversation*, Bellagio, Italy, 2008.

[21] Valve. Facial Expressions Primer from Half-Life 2 by Valve. `http://developer.valvesoftware.com/wiki/Facial_Expressions_Primer`.

[22] T. Vogt, E. André, and N. Bee. EmoVoice - a framework for online recognition of emotions from voice. In *Proceedings of Workshop on Perception and Interactive Technologies*, 2008.