Towards Facial Mimicry for a Virtual Human

Hana Boukricha and Ipke Wachsmuth

Faculty of Technology
Bielefeld University
33594 Bielefeld
Germany
{hboukric,ipke}@techfak.uni-bielefeld.de

Abstract. Mimicking others' facial expressions is believed to be important in making virtual humans as more natural and believable. As result of an empirical study conducted with a virtual human a large face repertoire of about 6000 faces arranged in Pleasure Arousal Dominance (PAD-) space with respect to two dominance values (dominant vs. submissive) was obtained. Each face in the face repertoire consists of different intensities of the virtual human's facial muscle actions called Action Units (AUs), modeled following the Facial Action Coding System (FACS). Using this face repertoire an approach towards realizing facial mimicry for a virtual human is topic of this paper. A preliminary evaluation of this first approach is realized with the basic emotions Happy and Angry.

1 Introduction

In his book *The Expression of the Emotions in Man and Animals* Darwin focuses on a more detailed and first scientific description of the meaning of different facial expressions as well as the facial muscles accompanying them. He also underlines the specific and functional role of facial expressions in expressing and communicating emotions [12]. Thus facial expressions play an important role in social interactions since detecting and understanding the facial expressions displayed by others allow an access to their intentional and affective states.

Nowadays the eventuality of being confronted with virtual characters embedded in computer related applications such as teaching or therapy applications [9], interactive museum guide applications [19], and movie-video applications, is increasing. Therefore features of human face-to-face interactions should be applied when designing human-computer interfaces, e.g., features underlying kinds of facial displays which play an essential role as a nonverbal communication channel [8]. Facial expressions are crucial not only in expressing and communicating emotions but also in mimicking the facial expressions of others. In social behavior mimicry has a necessary role in contributing to build bondings between humans. Mimicry acts as a 'social glue that binds humans together' since it contributes empathy, liking, rapport, and affiliation [10]. Bavelas et al. [2] argue for the role of mimicry as a communicative function in social interaction:

> By immediately displaying a reaction appropriate to the other's situation (e.g., a wince for the other's pain), the observer conveys, precisely and eloquently, both awareness of and involvement with the other's situation. (p. 278)

In human-computer interaction Brave et al. [5] and Prendinger et al. [23] found that agents showing involvement with their partner's situation through behaving empathically are judged by humans as more likeable, trustworthy and caring. In our work the definition of mimicry as empathy arousing mode introduced by Hoffman [18] is followed. Hoffman defines mimicry in terms of two sequential steps, namely imitation and feedback. That is, mimicry is the process involving the imitation of another's facial expression, voice, and posture, which triggers an afferent feedback eliciting the same feelings in oneself as those of the others.

The virtual human Emma (see Figure 1) has a face which replicates 44 Action Units (AUs) implemented inline with Ekman & Friesen's Facial Action Coding System (FACS) [13]. An empirical study consisting of human participants rating randomly generated facial expressions of the virtual human Emma with the bipolar adjectives from the "Semantic Differential Measures of Emotional State or Characteristic (Trait) Emotions" [21] (translated into German) has been conducted. Each facial expression was rated with 18 bipolar adjectives on a 1 to 7 Likert-Scale. Following [22] each group of 6 bipolar adjectives is used to represent one of the dimensions of pleasure, arousal, or dominance. As result of the empirical study a face repertoire of about 6000 faces arranged in Pleasure Arousal Dominance (PAD-) space with respect to two dominances values (dominant vs. submissive) was obtained. Each face in the face repertoire consists of different intensities of the virtual human Emma's AUs. A more comprehensive paper which also describes the study is to appear [4].

In this paper a first approach towards realizing facial mimicry for the virtual human Emma following the definition of mimicry introduced by Hoffman is presented. That is, using the face repertoire resulting from the empirical study, we are working towards enabling the virtual human Emma to imitate perceived emotional facial expression in terms of AUs and then to infer its related emotional state as a PAD-value. Since the virtual human Emma uses her own emotional facial expressions to infer the PAD-value related to a perceived facial expression, this can be considered as an internal simulation of the perceived facial expression, thus yielding a form of a facial-feedback-like approach. Therefore based on the face repertoire resulting from the empirical study, we are mainly interested in finding a backward mapping of AUs displaying emotional facial expressions on PAD-values and thus in exploring how the changes in the facial musculature of the virtual human Emma when imitating a facial expression can induce changes in her emotional state.

In the next section previous works on extending a virtual human's or robot's behavior to mimicking human's facial or multimodal expressions are outlined. A first investigation of an approach towards realizing facial mimicry for the virtual human Emma based on backward mapping AUs displaying emotional facial expressions on PAD-values is topic of Section 3. Finally a summary of the main conclusions and an outlook to future work are given.

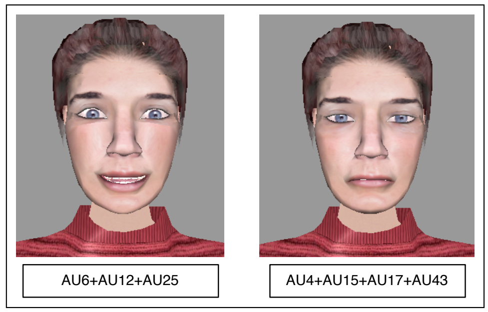

Fig. 1. Emma's face with two example facial expressions: Happy and Sad.

2 Related Work

There are various attempts in extending an agent's or robot's behavior to mimicking human's facial or multimodal expressions. The works of Caridakis et al. [7] and Courgeon et al. [11] consist of perceiving, interpreting, planning, and then animating the multimodal expression of the human. In [7] video recorded human's facial expressions and gestures are processed and analyzed. From an expression recognition module, Facial Animation Parameters (FAPs) are derived and expressed by the agent's face (the gesture's symbolic name is not being derived from the expression recognition module thus the gesture is manually communicated to the agent). Five expressivity parameters related to the movement's spatial volume, speed, energy, fluidity, and repetitivity are extracted from analyzing the image data and used to affect the quality of the agent's expressive behavior. In [11] from user's action on a 3D device (Joystick), a modulated target in PAD-space is computed and integrated with the output of a facial expression recognition module. The facial expression mirrored by the agent correspond to a blend of emotions derived from modulated target in PAD-space and from combining facial expression recognition rates of seven basic emotions. Breazeal et al. [6] primarily concentrate on the imitation task related to mimicry. They explore how imitation as a social learning and teaching process contributes to building socially intelligent robots. The robot identifies one of the basic emotions as emotion related to the imitated facial expression and uses this information to link new facial expressions with emotion labels.

A common characteristic among these works is that they mainly investigate the function of how the agent or robot is better enabled to learn reproducing or mirroring humans' facial expressions. In contrast, in our work we are not only interested in the imitation task of facial mimicry but also in the facial feedback task of facial mimicry. That means, given a perceived emotional facial expression in terms of AUs and based on the rich face repertoire provided by the empirical study, we are working on developing a system of backward mapping AUs displaying emotional facial expressions on PAD-values since the intensity of an emotion as well as comparing different emotions is better measured by real numbers. Therefore we aim at exploring how the changes in the facial musculature of the agent when imitating a facial expression can induce changes in its emotional state by investigating how altering the intensities of perceived AUs can impact the infered PAD-value.

3 Towards Facial Mimicry for a Virtual Human

New approaches in facial expression analysis [1] [15] [20] [25] attempt at recognizing AUs from a human face since detection of AUs allows a more flexible and versatile interpretation function of facial expressions. That is, the interpretation is not restricted to recognizing the emotional states related to a facial expression, but also the related mental cognitive states can be recognized. By these approaches laborious facial expression imitation learning methods for reproducing and mirroring humans' facial expressions can be avoided when the agent's or robot's face is modeled following FACS.

[1] and [17] are developing a system of facial expression analysis with AU recognition in spontaneous expressions since spontaneous expressions occur more frequently in everyday interaction. Following neuropsychological studies (cf. [16]) Bartlett et al. [1] state the importance of analyzing spontaneous facial expressions as they differ from posed facial expressions in their dynamics and in which muscles are moved. Spontaneous (involuntary) facial expressions are initiated subcortically and are characterized by fast and smooth onsets with different facial muscles (AUs) peaking simultaneously, while posed (voluntary) facial expressions are initiated cortically and are characterized by slow and jerky onsets with different facial muscles more often not peaking simultaneously.

Because currently we do not have data at hand from a system of facial expression analysis as described above, the starting point of our conception to realize facial mimicry is a vector of AUs' intensities available from simulating AUs expressing emotion with the virtual human Emma's face. In a first investigation of the idea of developing a system of backward mapping AUs displaying an emotional facial expression on PAD-values, we start up with some assumptions in order to reduce the complexity of this task. First we assume that the simulated facial expression has the same characteristics as a spontaneous facial expression. That is, the facial expression has a fast and smooth onset with different AUs peaking simultaneously. And second, Pleasure Arousal (PA-) courses related to facial expression onset are output by the system. In this paper only PA-courses for different patterns (AU combinations) of the basic emotions Happy and Angry are presented.

In this first investigation, the task of deriving the PA-courses related to onset of simulated AUs expressing emotion is modeled as a coarse nearest neighbor search problem in multiple dimensions. Since each face in the face repertoire is a combination of different AUs with different intensities, each face can be represented as a multidimensional vector of AUs' intensities. Using a Euclidean metrical distance function the face vector including the most similar AUs' intensities to given AUs' intensities is extracted and the PA-values related to this face vector are returned as the predicted PA-values. That is, given a vector of intensities of simulated AUs: $f_{sim} =< i_{sim}(AU_{i_1}), i_{sim}(AU_{i_2}), ..., i_{sim}(AU_{i_k}) >\in R^k$, R is the set of real numbers, $\{i_1, i_2, ..., i_k\} \subseteq AI$: AI is the set of overall AU Identifiers and a repertoire of faces arranged in PA-space: $FR = \{f_1 f_2 ... f_m\}$ with $f_{fr} = < i_{fr}(AU_{j_1}), i_{fr}(AU_{j_2}), ..., i_{fr}(AU_{j_l}) >\in R^l$, R is the set of real numbers, $1 \leq fr \leq m$, $\{j_1, j_2, ..., j_l\} = AI$, the function of returning the face from face repertoire (FR) including the most similar AUs' intensities to the given AUs' intensities can be described as

$$argmin_{f_{fr} \in FR}\{dist(f_{sim}, f_{fr})\} = f_{min} \tag{1}$$

with $f_{min} \in FR$ and f_{min} including the most similar AUs' intensities to the given AUs' intensities. The function *dist* is defined as follows

$$dist(f_{sim}, f_{fr}) = \sum_{e \in ID} \sqrt{(i_{sim}(AU_e) - i_{fr}(AU_e))^2} \tag{2}$$

During activation of the AUs simulated with the virtual human Emma's face the values of increasing AUs' intensities are sequentially processed with the function *argmin* (1) thus getting the PA-courses related to facial expression onset.

In order to reduce the dimension of the search space, only faces from face repertoire arranged in PA-space of highest dominance with values of positive pleasure and high arousal, and negative pleasure and high arousal are considered to respectively calculate the PA-courses related to the onsets of the facial expressions Happy and Angry since the emotions Happy and Angry correlate with respectively positive and negative pleasure values, high arousal values, and positive dominance values (cf. [24]).

In this first investigation with the coarse nearest neighbor search function, an overall increase in the values of pleasure and arousal is recorded from onsets of different patterns (cf. [14]) of the facial expression Happy, (AU6, AU12), (AU6, AU12, AU25), and (AU12, AU25). An overall decrease in pleasure and increase in arousal is recorded from onsets of different patterns (cf. [14]) of the facial expression Angry, (AU4, AU5, AU7, AU10) and (AU4, AU5, AU7, AU10, AU27), (e.g., see Figure 2). The PA-courses show more jerky patterns in the interval [0, 0.3] of increasing intensities. This is due to the coarse nearest neighbor classification that returns exactly one nearest neighbor. A smoother course of PA-values can be recorded by searching for the k-nearest neighbor with more adequate PA-values.

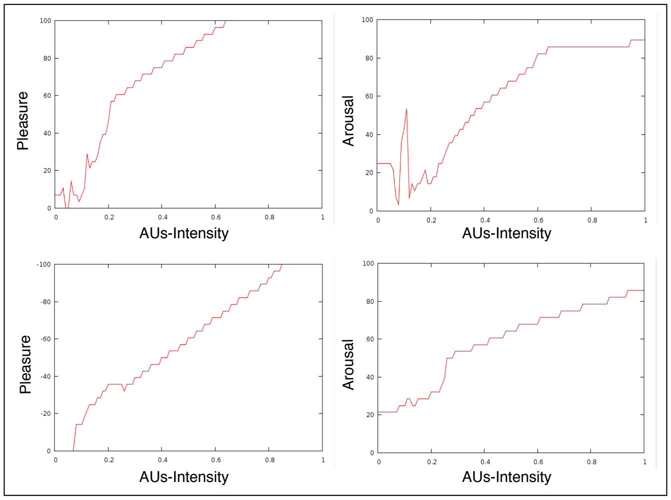

Fig. 2. Plots of Pleasure and Arousal (PA) courses, showing pleasure over AU intensity (left) and Arousal over AU intensity (right). The upper plots show PA-courses corresponding to AU6+AU12. The lower plots show PA-courses corresponding to AU4+AU5+AU7+AU10. The AUs of each facial expression are activated with the same intensity values.

4 Conclusion and Future Work

Based on a face repertoire resulting from an empirical study and consisting of faces arranged in PAD-space, a first investigation of backward mapping AUs displaying emotion to PAD-values using a nearest neighbor search function was introduced. Since the AUs in each considered facial expression of the emotions Happy and Angry are activated with the same intensity values thus having the same values of apex, as a next step we aim at altering these values of apex in order to better investigate the impact of each AU on the PA-courses. Furthermore the PA-courses of different AU combinations of additional basic emotions such as Sad and Fearful will be investigated.

Based on empathy arousing mechanisms our long term objective is to enable our virtual human Emma to address others' emotions by adjusting her subsequent behavior during interaction. One empathy arousing mechanism is facial mimicry and is introduced in this paper. Another empathy arousing mechanism is role-taking [18]. Metaphorically, role-taking is described as the ability of "seeing the world through another's eyes" or "putting yourself in another's shoes". Higgins [6] distinguishes two aspects of the role-taking process: situational role-taking (an example implementation is given in [3]) vs. individual role-taking. Situational role-taking refers to inferring that the other's viewpoint would be the same as our's in the same circumstances, whereas in individual role-taking the additional implications of the other's characteristics are considered. Thus role-taking allows including context related information since context information is crucial for judgment of facial expressions in terms of emotional states [26].

5 Acknowledgement

This research is partially supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center 673.

References

1. M. S. Bartlett, G. C. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1:22–35, 2006.
2. J. B. Bavelas, A. Black, N. Chovil, C. R. Lemery, and J. Mullett. Form and function in motor mimicry topographic evidence that the primary function is communicative. *Human Communication Research*, 14:275–299, 2006.
3. H. Boukricha, C. Becker, and I. Wachsmuth. 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*, pages 22–27, 2007.
4. H. Boukricha, I. Wachsmuth, A. Hofstätter, and K. Grammer. Pleasure-arousal-dominance driven facial expression simulation. To appear in Proceedings ACII 2009, IEEE.
5. S. Brave, C. Nass, and K. Hutchinson. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International journal of human-computer studies*, 62:162–178, 2005.
6. C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, 11:31 – 62, 2005.
7. G. Caridakis, A. Raouzaiou, E. Bevacqua, M. Mancini, K. Karpouzis, L. Malatesta, and C. Pelachaud. Virtual agent multimodal mimicry of humans. *Language Ressources and Evaluation*, 41:367–388, 2007.
8. J. Cassell. Nudge nudge wink wink: Elements of face-to face conversation for embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 1–27. The MIT Press, 2000.
9. J. Cassell. Towards a model of technology and literacy development: Story listening systems. *Applied Developmental Psychology*, 25:75–105, 2004.
10. T. L. Chartland, W. W. Maddux, and J. L. Lakin. Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. Hassin, J. Uleman, and J. Bargh, editors, *The new unconscious*, pages 334–361. New York: Oxford University Press, 2005.
11. M. Courgeon, J.-C. Martin, and C. Jacquemin. User's gestural exploration of different virtual agent's expressive profiles. *Proceedings of the 7th international conference on Autonomous agents and multiagent systems*, 3:1237–1240, 2008.
12. C. Darwin. *The Expression of the Emotions in Man And Animals.* The University of Chicago Press, 1969.
13. P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, Palo Alto, CA, 1978.

14. P. Ekman, W. V. Friesen, and J. C. Hager. *Facial Action Coding System: Investigator's Guide*. Research Nexus, a subsidiary of Network Information Research Corporation, Salt Lake City UT, USA, 2002.
15. R. ElKaliouby and P. Robinson. Mind reading machines: Automated inference of cognitive mental states from video. In *IEEE International Conference on Systems, Man and Cybernetics*, 2004.
16. M. Frank and P. Ekman. Not all smiles are created equal: the differences between enjoyment and nonenjoyment smiles. In *Humor: The International Journal for Research in Humor*, volume 6, pages 9–26, 1993.
17. A. Goneid and R. ElKaliouby. Facial feature analysis of spontaneous facial expression. In *The 10th International AI Applications Conference*, 2002.
18. M. L. Hoffman. *Empathy and Moral Development*. Cambridge University Press, 2000.
19. S. Kopp, L. Gesellensetter, N. Kraemer, and I. Wachsmuth. A conversational agent as museum guide – design and evaluation of a real-world application. In *Intelligent Virtual Agents*, pages 329–343. Springer, 2005.
20. Y. li Tian, T. Kanade, and J. F. Cohn. Recognizing action units for facial expression analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 23, pages 97–115, 2001.
21. A. Mehrabian and J. A. Russel. *An Approach to Environmental Psychology*. Cambridge, MIT Press, 1974.
22. Osgood, Suci, and Tannenbaum. *The Measurement of Meaning*. University of Illinois Press, Champaign, 1957.
23. H. Prendinger, C. Becker, and M. Ishizuka. A study in users' physiological response to an empathic interface agent. *International Journal of Humanoid Robotics*, 3(3):371–391, September 2006.
24. J. A. Russel and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11:273–294, 1977.
25. M. Valstar and M. Pantic. Biologically vs. logic inspired encoding of facial actions and emotions in video. In *IEEE International Conference on Multimedia and Expo*, pages 325–328, 2006.
26. H. G. Wallbott. Faces in context: The relative importance of facial expression and context information in determining emotion attributions. In *Facets of Emotion*. Lawrence Erlbaum Associates, Inc., 1988.