Proceedings of the 4th Workshop

Emotion and Computing - Current Research and Future Impact

Dirk Reichardt (Editor)

Paderborn, Germany

September 15th, 2009

ISSN 1865-6374

4th Workshop
Emotion & Computing
Current Research and Future Impact

D. Reichardt

Applied Computer Science and Information Technology
Baden-Württemberg Cooperative State University Stuttgart (DHBW)
Stuttgart, Germany
reichardt@dhbw-stuttgart.de

The workshop series "emotion and computing - current research and future impact" has been providing a platform for discussion of emotion related topics of computer science and AI since 2006. In recent years computer science research has shown increasing efforts in the field of software agents which incorporate emotion.

Several approaches have been made concerning emotion recognition, emotion modelling, generation of emotional user interfaces and dialogue systems as well as anthropomorphic communication agents. Motivations for emotional computing are manifold. From a scientific point of view, emotions play an essential role in decision making, as well as in perception and learning. Furthermore, emotions influence rational thinking and therefore should be part of rational agents as proposed by artificial intelligence research. Another focus is on human computer interfaces which include believable animations of interface agents. From a user perspective, emotional interfaces can significantly increase motivation and engagement which is of high relevance to the games and e-learning industry.

This workshop intends to discuss the scientific methods considering their benefit for current and future applications. Especially when regarding the subject of emotion recognition, this also includes ethical aspects.

This year we are proud to have 6 presentations of ongoing research work and a life demo within our workshop – which is a rather high number regarding the IVA and ACII Conference at Amsterdam at the same time. It has become a tradition to select a discussion topic and integrate an open discussion session on this. Last year we started to integrate web based discussions by providing a discussion (mind mapping) feature on the workshop website http://www.emotion-and-computing.de . We are looking forward to interesting presentations and fruitful discussions.

Dirk Reichardt

Scientific Committee:

Dipl.-Ing. Joscha Bach

Dr. Christian Becker-Asano, ATR, Kyoto, Japan

Prof. Dr. Nicola Henze, University of Hannover

Dr. Manfred Kerber, University of Birmingham

Dr. Michael Kipp, DFKI Saarbrücken

Prof. Dr. Paul Levi, University of Stuttgart

Prof. Dr. John-Jules Charles Meyer, University of Utrecht

Dipl.-Ing. Christian Peter, Fraunhofer IGD, Rostock

Dr. Götz Renner, Daimler AG, Customer Research Center, Stuttgart

Prof. Dr. Michael M. Richter, University of Calgary

Workshop Program:

Presentations

The happiness cube paradigm
Stylianos Giannoulis, Fons Verbeek

Dynamic Facial Expression Classification Based on Human Visual Cues Information
Zakia Hammal, Corentin Massot

Classifying Facial Pain Expressions
Michael Siebers , Miriam Kunz , Stefan Lautenbacher , and Ute Schmid

Towards Facial Mimicry for a Virtual Human
Hana Boukricha and Ipke Wachsmuth

The OCC Model Revisited
Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer

For Androids, Emotions are just in Time
Wolfgang Gessner

Demo

Emotion Recognition Using Optical Flow
Micheal Girgis, Dirk Reichardt

Scientific Research Papers

presented at the workshop

The happiness cube paradigm; eliciting happiness through sound, video, light and odor. Assessment of affective state with non-invasive techniques.

Stylianos Giannoulis[1], Dr. Fons Verbeek[1]

[1]Media Technology. Leiden Institute of Advanced Computer Science, Leiden University, Niels Bohrweg 1, 2333 CA, Leiden, the Netherlands
s.giannoulis@umail.leidenuniv.nl , fverbeek@liacs.nl

Abstract. Emotion elicitation and physiological responses are 2 fields that have been studied extensively the last 20 years. Based on the already existing research, a scientific experiment is described with the goal to elicit emotions of happiness to the participants by the use of video, sound and odorants. Contrary to most already existing research, the goal of this experiment is to elicit just one emotion -happiness-. Moreover, the expected multisensory experience is of great significance since most of the existing research on emotion elicitation is usually focusing only on one or two at most sensory modalities.

Keywords: emotion elicitation, happiness, physiological response, biofeedback

1 Introduction

As emotions play a crucial role in normal and abnormal functioning [1], it is not surprising that there is a growing interest and research performed the last 20 years in emotion elicitation and their physiological responses. Several different techniques have been used so far in order to induce certain emotions, including: a) images and sounds b) expressive behavior c) scripted and unscripted social interactions d) music [2] e) smell [3]. Moreover, indoor lightning is proved to have effect on mood and cognition according its luminance levels and color temperature [4].

2 Factors of Happiness

Although emotional response is a complex field of study because of certain characteristics that are not universally of the same volume like : a) the threshold for eliciting components of a particular emotion b) the peak and amplitude of the response c) the rise time to peak and d) recovery time [5]), research from different study fields like: a) Neuroscience [6], [7] b) Psychology [8],[9] c) Social Sciences [10], [11] has already put great effort on discovering what could make people happy, how the brain behaves in certain affective states and what are the physiological responses when experiencing certain emotions [12], [13].

From the perspective of social sciences, research shows that employment, marriage [14], income, education, gender, religion, social life and health [15] have a strong correlation to overall life-satisfaction.

From the neuroscience and endocrinology point of view, it is widely accepted that neurotransmitters like dopamine and serotonin play a great role in our current affective state [16],[17]. Additionally endorphin levels are strongly related with increase in positive affect [18]. Additionally, research showed that norepinephrine levels have great effect on our positive or negative affective state [19].

The smell of lavender is connected with an increase of alpha waves in the brain and consequently it is promoting relaxation [20]. Finally in a experiment performed in laboratory concerning mood elicitation through odorants, vanillin was rated by the participants as the most pleasant [3].

Moreover, there is evidence that placing a patient close to a window with a view outside could speed up the healing of wounds [21].

Listening to music produces changes in the autonomous nervous system which are associated with emotional states. In an experiment conducted by C. L. Krumhansl, [9], it was proved that music has significant effect on our physiological responses compared to the pre-music baserate levels: Cardiac interbeat interval (IBl), pulse transmission time to the finger (FPTT), Pulse transmission time to the ear (EPTT), Systolic blood pressure (SBP), diastolic blood pressure (DBP), and mean arterial pressure (MAP) increased, while Finger pulse amplitude (FPA), respiration intercycle interval (ICI), Respiration depth (RD) and respiration-sinus asynchrony (RSA), skin conductance level (SCL) and temperature on the finger (TEM) decreased. In another experiment [22] music not only managed to induce the desired moods of happiness, sadness and fear but Etzel [22] also agrees with Gabrielson and Lindstrom [23] that fast tempos can be associated with expressions of activity, excitement, happiness, joy, potency, surprise, anger and fear, as it was proven that the music to induce happiness had quick rhythm and melody while music used to induce sadness had much slower tempos.

Moreover from the research done on people's response to different colors, findings show that people describe as quiet and serene colors of short wavelength like blue or green, while colors of long wavelength like red and orange, are described as arousing and hot [24].

Last but not least, one of the most commonly used methods of emotion elicitation is the projection of pictures. This led to the development of the International Affective Picture System (IAPS), a large set of color photographs chosen specifically in order to evoke emotions that include pleasure, arousal, dominance to men and women [25]. The IAPS stimuli are standardized on the basis of ratings of pleasure and arousal experienced by the respondents [26] and experiments conducted in both sexes showed that pictures which include food, sports and adventure, were rated high from both sexes as pleasant and arousing while pictures of babies and nature were rated by both sexes as pleasant but not highly arousing. Men rated pictures of erotic scenes between people of the opposite sex also as highly arousing and pleasant, while women rated these pictures as less arousing and pleasant [26].

3 Methods for Emotion Measurements

The most commonly used non-computer based technique to measure emotional states are self-reports. There are 5 different and dominant affective measures in the mood literature[27]: the Mood Adjective Checklist or MACL, the Profile of Mood States or POMS the original and revised versions of the Multiple Affect Adjective Checklist or MAACL & MAACL-R, the Differential Emotion Scale or DES but it seems that the most commonly used standardized scale is currently the Positive Affective Negative Affective Scale or mostly known as PANAS scale, developed by Watson and Clark [28]. In the PANAS scale, respondents rate to what extend they have experienced each mood on a 5-point scale (1= very slightly or not at all, 5=extremely) in different time periods like: today, past few days, week, past few weeks, year and in general [27].

Although Watson and Clark report extensive reliability and validating data on the PANAS scales [27], self-report might not be always enough in order to come to safe conclusions as emotions belong to our personal sphere, and respondents have the tendency to present themselves in a socially desirable way if they perceive some topics or questions as threatening [29].

In neuroscience in order to indicate which parts of the brain are stimulated while experiencing specific emotions, Magnetic Resonance Imaging also known as MRI [30], functional Magnetic Resonance Imaging also known as fMRI [31], Positronic Emitting Tomographies also known as PET scans [6] or Electrocephalography also known as EEG [32] have been used in different experiments.

But when it comes to emotions and their physiological responses, sensors that measure: a) skin conductance level, b) heart rate levels c) blood pressure levels d) skin temperature levels e) muscle activity levels are used. The afore mentioned physiological responses can give safe indications of affective state physiological expressions and that's why they are the most commonly targeted responses.

The last 40 years, there is a lot of research concerning emotion and facial expression, which led to the development of a number of observer-based systems of facial expression measurements with the Facial Action Coding System or FACS, developed by Paul Ekman and Wallace Friesen in 1978 [8] being the most commonly used for facial expression recognition. FACS is based on the fact that all people, with a few exceptions of course, have the same facial muscles [33]. In facial actions, various facial muscles are used. Based on that fact and on observers that were capable of distinguishing the appearance changes resulting from the use of various facial muscle, Ekman and Friesen, managed to categorize the facial actions, which are called Action Units, and also describe which facial muscle or muscles are used in each Action Unit. In the new version of FACS, FACS 2002, 66 different Action Units are described [34]. It is important to explain that FACS each self is a system that describes which muscles are used in each Action Unit. It is not a system that describes emotional facial expression but there is a lot of research conducted the last 15 years [35], [36] on computer-based facial recognition of emotion, based on Ekman's FACS.

4 Happiness and Physiological Responses

The research has shown so far that there is no standardized bodily expression of happiness and of emotions in general, which is based on the fact that people experience and express emotions in different intensity [5]. For that reason we can't say so far that when happiness is experienced, there is indication of certain and specific levels of heart beat, skin conductance, blood pressure, respiration, or skin temperature, but physiological response of happiness is described in relation to physiological responses of other emotions like sadness, anger,[12] disgust [6] or fear [22].

5 The Happiness Cube

5.1 Introduction

Based on the afore mentioned scientific research on emotion elicitation and assessment, our goal is to design and construct an installation that would allow the participant to experience happiness in a multisensory way with the use of video, sound, light, smell combined with assessment in a basic level of the user's physiological response and affective state.

We believe that a 20 minute session would be enough to draw safe conclusions about the effectiveness of our emotion elicitation and assessment hypothesis. This session will be divided in 4 parts: a) self-report, during which the participants are going to fill in a short computer-based questionnaire b) the relaxation part (~7 minute duration) b) the happiness elicitation part (~13 minute duration) c) self report session, during which the participants are going to fill in a very short computer based questionnaire.

5.2 The Questionnaires

The questionnaire will help us assess the participants' current affective state before and after the 20 minute session. We think that filling in a computer-based questionnaire is a better evaluation method than a paper-based method, especially when it comes to disclosure of very personal data like emotions, as it seems that computer administration increases self-disclosure compared to other conventional methods [37].

Both questionnaires should be as short as possible, as emotions are characterized from high intensity but very short duration [38]. For that reason we decided that the first questionnaire shouldn't include more than 15 questions, providing also information about the participants' age, sex, marital status, work satisfaction, health, factors that as mentioned before play an important role in people's affective state [14], [15]. Moreover questions from the PANAS scales are also included, as well as other questions related to well-being and life-satisfaction, like the famous Cantril question, developed by Hadley Cantril in 1965:

> "Here is a picture of a ladder, representing the ladder of life. Suppose we say the top of the ladder (step 10) represents the best possible life for you, and the bottom (step 0) represents the worst possible life for you. Where on the ladder do you feel you personally stand at the present time?"

The questions that provide information about the participants' life-satisfaction and affective state will be rated by the participants in a 5 or a 7 at most point scale, as a 5 or 7 point scale has a clear mid point and there is no significant change in data interpretation, when analyzing the collected data from a bigger than a 9 point scale [39].

The questions that provide information about the participants' current affective state and were used in the questionnaire before the 20 minute affective state, are going to be answered again by the participants after the 20 minute session in order to check if there is a significant change in their current affective state. This questionnaire has to be even shorter as emotions and moods differentiate from each other in duration and intensity. Emotions are experienced in general in a highly intensive way, while moods are characterized by low intensity. Moreover emotions last for a relatively short time compared to moods [38].

5.3 The 7 minute relaxation session

After filling in the first questionnaire, the participants are entering a 2m x 2m x 2m cube. The outer part of the cube will be coated by black fabric in order to create an environment that prevents external factors like daylight to disturb the participant's experience. The goal is to create an environment that will also give the participant the impression of being closer to a more natural environment, as urbanization is growing rapidly all over the world, with a percentage of 49% of the world population to be urban by 2005, which was expected to rise up to 50% by 2008 and up to almost 60% by 2030 [40] leading to alienation from natural environment. For that reason the inner part of the cube will be coated with green fabric, a color soothing for the eye. And inside the cube green and blue light is going to create an even more relaxing environment.

Lavender odor will be emitted inside the cube, as the smell of lavender is strongly correlated with increase of alpha waves production in the brain, promoting relaxation [20].

Although there is a lot of research already performed on emotion elicitation through music, it seems that the researchers have not come to a final and unanimous conclusion of what kind of musical piece elicits feelings of happiness. Scientists have experimented with different kinds of music [9], [22] and the participants' interest in the music is an important factor for influencing emotional responses [41]. So as taste in music is subjective and what makes one participant happy could probably not elicit the same emotion to other participants, the best solution is to use music that would have an as universal effect as possible. For that reason we think that the best idea is to use natural sounds, based on our hypothesis that urban living is increasing rapidly and people are getting more and more alienated from natural environments.

5.4 The 14 minute happiness session

The session to elicit happy emotions is going to follow after the short relaxation session.

Based on the fact that specific picture content in the International Affective Picture System (IAPS), is eliciting arousal and pleasure in both men and women[26], we are going to project pictures of food, sports, adventures and intimacy between opposite sexes.

Research on emotion elicitation and odorants has shown that participants rated vanillin as the most pleasant odor among the rest [3]. Moreover consumption of chocolate is associated with different neurotransmitter levels in our body, like serotonin, dopamine and endorphins [17]. For this session lavender will be replaced by vanilla and chocolate odors.

Blue and green lighting could still be a good medium to elicit happiness as research in lighting has shown that participants describe blue and green light as pleasant [42],agreeable and relaxing[43].

Natural sounds will still be used and the use of headphones will be able to create a more immersive environment.

5.5 Measuring emotions

There are many different techniques in order to measure emotional changes in subjects and the most common involve: a) self-report b) use of hardware and software that analyze and measure physiological responses. Self-report methods can give indications of emotional changes in subjects, but as respondents could answer questions according to a desirable affective state and not according to their actual affective state, this method of measuring the emotional state of the participants might not be enough.

Moreover, applying sensors to measure heart rate, temperature, blood pressure or skin conductance, might create a feeling of insecurity or even stress to the participants, so it is important to find less invasive methods in order to measure physiological responses. The use of an infrared thermometer to measure body temperature and a pulse rate ring in order to monitor changes in the participants' heart beat seems currently as an interesting approach for measuring the participants' physiological responses.

Currently, the most dominant idea is to monitor the whole session with a camera and then analyze the participants' facial expressions and responses with software based on the afore mentioned FACS [8], [34]. In that way although the subjects will know that they are going to be monitored, they will not be surrounded by sensors and cables that might be distracting or even create stressful feelings to them.

6 Discussion

The happiness cube is a scientific experiment and it is already under construction and development. The testing is scheduled for the end of August 2009-beginning of September 2009.

From this experiment we hope to get valid results that support our hypothesis about creating an artificial environment that could elicit emotions of happiness with the use of video, sound, light and smell.

The next step could include: a) automation of the multisensory experience b) wireless measurements of physiological responses c) leave the initiative to the user in order for him to train a computer-based program, which understands his physiological responses and could adjust indoor lighting, temperature, available TV channels, music according to the user's current affective state and preferences.

References

1. Gross, J.J., Levenson, R.W.: Emotion Elicitation Using Films, Cognition and Emotion 9 (1), 87-108 (1995)
2. Rottenberg, J., Ray, R.D., Gross, J.J.: Emotion Elicitation Using Films. In: Coan J. A., Allen J.J.B., Handbook of Emotion Elicitation and Assessment, pp. 9-28. Oxford University Press, New York (2007)
3. Alaoui- Ismaili, O., Robin O., Rada H., Dittmar A., Vernet-Maury E.: Basic Emotion Evoked by Odorants: Comparison Between Autonomic Responses and Self-Evaluation, Physiology and Behavior 62, No.4, 713-720 (1997)
4. Knez, I.: Effects of Indoor Lighting on Mood and Cognition, Journal of environmental Psychology 15, 39-51 (1995)
5. Davidson, R.J.: Neuropsychological Perspectives on Affective Style and their Cognitive Consequences. In: Dagleish, T., Power, M.J., Handbook of Emotion and Cognition, pp.103-123. Wiley, Chichester, (1999)
6. Lane, R.D. , Reiman, E. M., Ahern, G. L., Schwartz, G.E., Davidson R.J.: Neuroanatomical Correlates of Happiness, Sadness and Disgust, the American Journal of Psychiatry 154, 926-933 (1997)
7. Davidson, R. J.: Well-being and Affective Style: Neural Substrates and Biobehavioural Correlates, Phil. Trans. R. Soc. Lond. B 359 , 1395–1411 (2004)
8. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, (1978)
9. Krumhansl, C.L.: An Exploratory Study of Musical Emotion and Psychopsychology, Canadian J. Exp. Psychol. 51, 336-352 (1997)
10. Bjornskov, C., Gupta, N.D., Pedersen, P.J.: Analyzing Trends in Subjective Well-Being in 15 European Countries, 1973–2002, Journal of Happiness Studies 9, is. 2, 317-330 (2007)
11. Veenhoven, R.: Is Happiness Relative, Social Indicators Research 24, 1-34 (1991)
12. Schwartz, G.E., Weinberger, D.A., Singer, J.A.: Cardiovascular Differentiation of

Happiness, Sadness, Anger and Fear Following Imagery and Exercise, Psychosomatic Medicine 43, No. 4, 343-364 (1981)
13. Collet, C., Vernet Maury, E., Delhomme, G., Dittmar A.: Autonomic Nervous System Response Patterns Specificity to Basic Emotions, Journal of the Autonomic Nervous System 62, 45-57 (1996)
14. Veenhoven, R.: Developments in Satisfaction Research, Social Indicators Research 1, 1-46 (1996)
15. Helliwell, J., Putnam, R.,D.: The Social Context of well-being. In: Huppert, F.A., Baylis, N., Keverne, B., The science of well-being, pp.435-459. Oxford University Press, New York (2005)
16. Cabioglu, M.T., Ergene, N., Tan, U.: Smoking Cessation after Acupuncture Treatment, The International Journal of Neuroscience 117, is.5, 571-578 (2007)
17. Parker, G., Parker, I., Brotchie, H.: Review: Mood State Effects of Chocolate, Journal of Affective Disorders 92, 149-159 (2006)
18. Harte, J. L., Eifert, J.H., Smith, R.: The Effects of Running and Meditation on Beta-endorphin, Corticotropin-releasing Hormone and Cortisol in Plasma, and on Mood, Biological Psychology 40, is 3, 251-265 (1995)
19. Harmer, C.J., Shelley, N. C., Cowen, P. J., Goodwin, G. M.: Increased Positive Versus Negative Affective Perception and Memory in Healthy Volunteers Following Selective Serotonin and Norepinephrine Reuptake Inhibition, Psychiatry: Interpersonal and Biological Processes 161, 1256-1263 (2004)
20. Fontaine, D.K., Briggs, L.P., Pope-Smith, B.: Designing Humanistic Critical Care Environments, Critical Care 24, is. 3, 21-34 (2001)
21. Ulrich, R.: View through a Window May Influence Recovery from Surgery. *Science*, *224*(4647), 420–421 (1984)
22. Etzel, J.A., Johnsen, E.L. Dickerson, J., Tranel, D., Adolphs, R.: Cardiovascular and Respiratory Responses during Musical Mood Induction, International Journal of Psychophysiology 61, 57-69 (2006)
23. Gabrielson, A., Lindstrom, E.: The Influence of Musical Structure on Emotional Expression. In: Juslin, P. N., Sloboda, J.A., Music and Emotion: Theory and research, Oxford University Press, Oxford, pp. 223-248 (2001)
24. Lawrence, L., Garber Jr, Hyatt, E.M., Boya, U.O.: The Mediating Effects of the Appearance of Nondurable Consumer Goods and their Packaging on Consumer Behavior. In: Schifferstein, H.N.J., Hekkert, P., Product Experience, pp. 581- 602. Elsevier, Amsterdam, (2008)
25. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual, Technical Report no. A-6, University of Florida, Gainesville, Fl. (2005)
26. Bradley, M.M., Lang, P.J.: The International Affective Picture System (IAPS) in the Study of Emotion and Attention. In: Coan, J.A., Allen, J.J.B., 2007, Handbook of Emotion Elicitation and Assessment, pp.29- 46. Oxford University Press, New York, (2007)
27. Gray E.K., Watson D.: Assessing Positive and Negative Affect via Self-Report. In: Coan, J.A., Allen, J.J.B., 2007, Handbook of Emotion Elicitation and Assessment, pp. 171-202. Oxford University Press, New York (2007)
28. Watson, D., Clark, L.A., Telegen, A.: Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scale. Journal of Personality and Social Psychology 54, 1063-1070 (1988)
29. Schaeffer, N.C.: Asking Questions About Threatening Topics: A Selective Overview. In: Ston, A.A., Turkkan, J.S., Bachrach, C.A., Jobe, J.B., Kurtzman H.S., Cain V. S., (2000), the Science of self-report, pp. 105-121. Lawrence Erlbaum Associates, New Jersey, (2000)
30. Bremner, J.D.: Does Stress Damage the Brain? Biol. Psychiat. 45, 797–805 (1999)
31. Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hyman, S.E., & Rosen, B. R.: Response and Habituation of the Human Amygdala during Visual Processing of Facial Expression. Neuron 17, 875–887 (1996)
32. Ahern, G. L., & Schwartz, G. E.: Differential Lateralization for Positive and Negative Emotion in the Human Brain: EEG spectral analysis. Neuropsychologia 23, 745–755 (1985)
33. Schmidt, K.L., Cohn J.F.: Human Facial Expressions as Adaptations: Evolutionary

Perspectives in Facial Expression Research. Yearbook of Physical Anthropology, 116, 8-24 (2001)
34. Ekman, P., Friesen, W., Hager, J. C.: Facial Action Coding System, Salt Lake City, UT: Research Nexus (2002)
35. Cohn, J.F.: Automated Analysis of the Configuration and Timing of Facial Expression. In: Ekman P., Rosenberg E.L., What the Face Reveals. Basic and Applied Studies on Spontaneous Expression Using the Facial Action Coding System (FACS), pp. 388-392. Oxford Univerisity Press, New York, (2005)
36. Bartlet, M.S., Movellan, J.R., Littlewort, G., Braathen, B., Frank, M.G., Sejnowski, T.J.: Toward Automatic Recognition of Spontaneous Facial Actions. In: Ekman P., Rosenberg E.L., What the Face Reveals. Basic and Applied Studies on Spontaneous Expression Using the Facial Action Coding System (FACS), pp. 393- 412. Oxford University Press, New York, (2005)
37. Anthony, J.C., Neumark, Y., Van Etten, M.L.: Do I Do what I Say? A Perspective of Self-report Methods in Drug Dependence Epidemiology. In: Stone A.A., Turkkan J.S., Bachrach C.A., Jobe J.B., Kurtzman H.S., Cain V. S., The Science of Self-report, pp.175-198. Lawrence Erlbaum Associates, New Jersey, (2000)
38. Scherer, K.R.: What are Emotions? And how They Can Be Measured?, Social Science Information , vol 44 (4), 695-729 (2005)
39. Haslam, S.A., McGarty, C.: Survey Design. In: Haslam, S.A., McGarty, C., Research Methods and Statistics in Psychology, pp. 102-129. Sage Publications, London (2003)
40. United Nations, Department of Economic and Social Affairs, Population Division . World Urbanisation Prospects: the 2005 Revision, Working Paper No. ESA/P/WP/200 (2006)
41. Aldridge, D.: An Overview of Music Therapy Research, Complementary Therapies in Medicine 2, 204-216 (1994)
42. Valdez, P., Mehrabian, A.: Effects of Colors on Emotions, Journal of Experimental Psychology 123, vol. 4, 394-409, (1994)
43. Mahnke, F.H.: Color, Environment and Human Response. Wiley, New York (1996)

Dynamic Facial Expression Classification Based on Human Visual Cues Information

Zakia Hammal[1], Corentin Massot[2]

[1] Universite de Montréal, [2] McGill University
zakia_hammal@yahoo.fr

Abstract. Compared to human performances automatic recognition of facial expressions in real life situations remains a big challenge for computer systems. Here, we aim at including new strategies into a previously developed model by Hammal *et al.* [1,2] that has been successfully used for the recognition of the six basic facial expressions. We extend this model based on the Transferable Belief Model (TBM) by, first, introducing the most important transient facial features used by human observers for facial expressions recognition. A new filtering-based method is introduced for the detection of transient facial features and the precise evaluation of their orientations. Second, a dynamic and progressive fusion process of the permanent facial features (such as eyes, eyebrows and mouth) and of the transient facial features (such as nasal root wrinkles and nasolabial furrows) is introduced and allows to deal with asynchronous facial features deformations. Experimental results show an increase by 12 % in average compared to the original model and compare favorably to human observers.

Keywords: Facial expressions, Classification, TBM, Uncertainty, Multiscale spatial filtering.

1 Introduction

Despite all the efforts in computer vision [6], human performances remain far better than any automatic facial expressions recognition system. Recent works in computer vision [2] compared their model to human behavior [5] to take advantage of their strategies for facial expressions classification. The current paper is in the continuity of these works where in addition to the permanent facial features the most useful transient facial features by human observers (such as nasal root winkles and nasolabial furrows) are added in the classification process. The two main contributions of the current paper are: First, compared to the canny based methods [1,12] where the obtained results are usually noisy and require a high threshold (to prevent false detection) leading to several missed detections, a new and robust method based on multiscale spatial filtering is proposed. The proposed method allows the automatic detection of wrinkles and the estimation of the corresponding orientation. Secondly, given that the temporal dynamics of the facial features is a critical factor [11] for the facial expressions interpretation, a dynamic fusion process of the permanents (eyes, eyebrows and mouth) as well as the transient facial features based on the temporal modeling of the Transferable Belief Model (TBM) is proposed. Recent works have been proposed to model the facial behavior dynamics as temporal segments of AUs activation in a consecutive predefined number of frames [7]. However, these methods are very sensitive to the frame size and do not allow the dynamic recognition of facial expression in the case of asynchronous facial features deformation. The proposed model deals with these considerations. Moreover, based on the dynamic TBM fusion process, the proposed model allows the dynamic recognition of pure facial expressions (Happiness, Surprise, Fear, Disgust, Sadness, Anger) and Neutral and explicitly models the doubt between expressions in the case of blends, combinations or uncertainty between two or several expressions [9].

2. Visual cues extraction

2.1. Permanent facial features

Assuming that the face is at a near-frontal position the fully automatic face and facial feature (eyes, eyebrows and mouth) segmentation is made [3] (see Fig. 1 (a)). Based on the work of Hammal *et al.*, five characteristic distances $D_i \; 1 \le i \le 5$ coding the motion of the selected facial feature points according to the neutral state are used (Fig. 1 (a) see [1, 2] for detailed explanation of this choice).

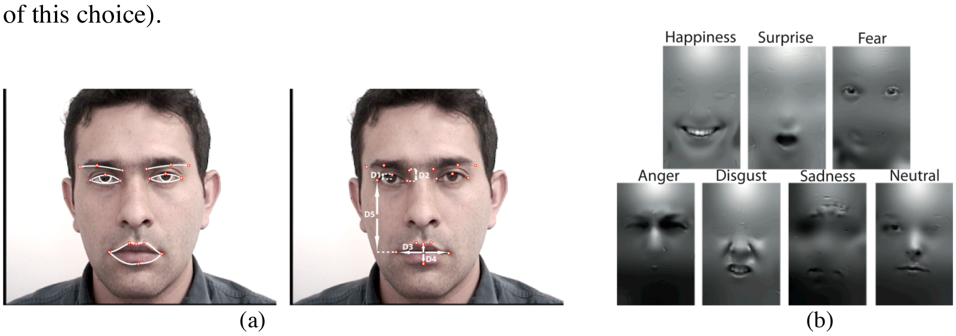
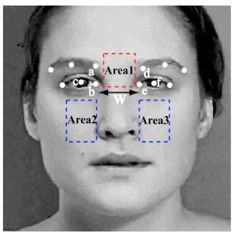


Figure 1: (a) Example of facial features segmentation and the corresponding characteristic distances [3], (b) Effective faces representing diagnostic filtering functions for the human observers [5].

However, the five characteristic distances are necessary but not sufficient to attaint the human visual system performances for facial expression recognition [2]. To understand how humans are so efficient at classifying facial expressions, Smith *et al.* [5] revealed the precise effective features for the categorization of the six basic facial expressions plus Neutral (Fig. 2 (b) summarizes their experiment results). Taking advantage of the Smith *et al.* outcome, the ”suboptimal” features (such as nasal root wrinkles (see Fig. 2 (b) Anger) and nasolabial furrow (see Fig. 2 (b) Disgust)) will be added to the permanent facial features for the dynamic classification process.

2.2 Transient features

The Nasal root wrinkles and the Nasolabial furrows are used to provide additional information to support the recognition of facial expressions. The transient feature areas are located using the position of the permanent facial features (Fig. 2).



Transient features	Transient feature regions description
Area 1:Nasal	Y11= a(y)-0.7*W; X11= b(x)+2
Root wrinkle	Y12= a(y)+0.4*W; X12= d(x)-2
Area 2:Left	Y21= Y12; X21= c(x)+R
Nasolabial Furrow	Y22= Y12+1.5*W; X22= X11
Area 3: Right	Y31= Y12; X31= f(x)+R
Nasolabial Furrow	Y32= Y12+1.5*W; X32= X12

Figure 2: Left: Example of detection of wrinkles regions. Right: Descriptions of transient feature regions. R is the eyes radius, and W the distance between eyes corners.

In the current paper we propose a filtering based-method that estimates the appearance of transient facial features, provides an estimation of the confidence in this estimation and accurately measures the orientation of the detected wrinkles when necessary. The proposed method is based on the filtering of the image by a bank of spatial filters. We chose Log-Normal filters which have the advantage of being easily tuned in frequency and orientation and separable, which make them well suited for detecting features at different scales and orientations [13]. They are defined as follow:

$$|G_{i,j}(f,\theta)|^2 = |G_i(f).G_j(\theta)|^2 = A.\frac{1}{f}.\exp\left(-\frac{1}{2}\left(\frac{\ln(f/f_i)}{\sigma_r}\right)^2\right).\exp\left(-\frac{1}{2}\left(\frac{\theta-\theta_j}{\sigma_\theta}\right)^2\right) \quad (1)$$

Where $G_{i,j}$ is the transfer function of the filter, $G_i(f)$ and $G_j(\theta)$ respectively represents the frequency and the orientation components of the filter; f_i is the central frequency, θ_j, the central orientation, σ_r, the frequency bandwidth and σ_θ, the orientation bandwidth and $\boldsymbol{A}$, a normalization factor.

Fig. 3 describes the different steps of the transient facial features detection and nasolabial furrows orientation estimation. A retinal prefiltering is applied on the input frame (Fig. 3 (b)) aiming at increasing contours contrast while removing lighting conditions [13]. Transient feature regions are extracted and a Hamming circular window is applied in order to remove image border information (Fig. 3 (c)). A bank of log-normal filters of 7 central frequencies and 15 orientations (Fig. 3 (d)) are then applied to the power spectra of each selected area. We measure the response of orientation bands that corresponds to the sum of the responses of all filters sharing the same central orientation at different spatial frequencies (Fig. 3 (d), grey area). This allows analyzing an oriented transient feature independently of its spatial frequency making the detection more robust to individual morphological differences and facial deformations accruing during facial expressions.

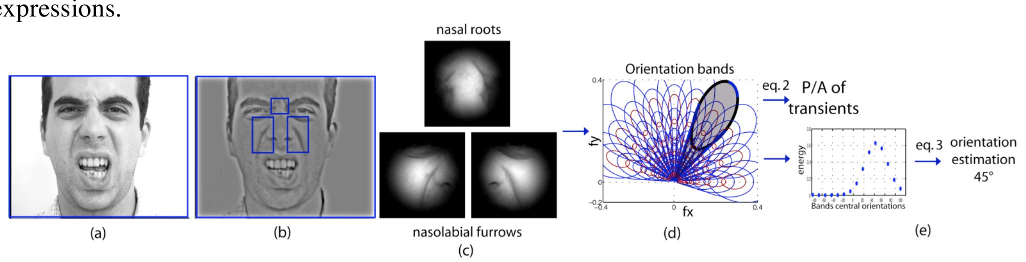


Figure 3: Transient features detection and orientation estimation.

Nasal root wrinkles and nasolabial furrows are defined as “present” or “absent” based on a confidence measure (Fig. 3 (e)). This latter is computed based on two variables (Eq. 2, Eq. 3): first the amount of energy in each orientation band and the variance of the distribution of the energy across orientation bands. The confidence is then maximal when a contour is present at a given orientation. Moreover, it is obtained independently on each frame. Threshold values on the confidence measure and on the nasolabial furrows angles are obtained after learning process over all test databases in order to build BBA models associated to each transient feature (see section 3.1).

$$C_t = \frac{\sigma_t}{(1/2-\mu_t)} \quad (2) \qquad \theta_t = \sum_{j=1}^{15} B_j * \theta_j \quad (3)$$

Eq. 2 gives a measure of the confidence of the presence of a feature C_t on the frame $\boldsymbol{t}$, μ_t and σ_t are the average and the standard deviation of the orientation band responses; Eq. 3 gives the final orientation estimation θ_t where B_j is the response of the orientation band j at the central orientation θ_j. Tested on three benchmark databases [10][14][16] the proposed method reaches a precision of 89%.

Once the nasolabial furrows detected, their orientation is measured by linear combination of the orientation bands responses (Eq. 3, Fig 3(e)). The orientation corresponds to the angle between them and the horizontal plan of the corresponding area. Fig. 4 shows examples of detection of nasolabial furrows and nasal roots during sequences of Disgust and Happiness expressions respectively. One can observe that nasolabial orientations are different according to the expression.

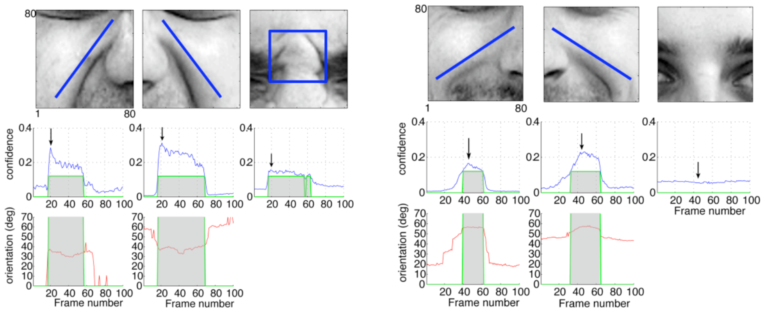


Figure 4: Example of nasolabial furrows and nasal roots detection during Disgust and Happiness sequence; gray temporal windows indicate the presence of the transient features based on the confidence threshold (0.12 validated on three databases); third rows display the measured angles of the nasolabial furrows (around 45 for Disgust and around 60 for Happiness).

2.2 Numerical to symbolic conversion

Once the permanent and transient facial features detected, a numerical to symbolic conversion is carried out in order to convert the measured distances, transient features and the corresponding angles into symbolic states reflecting their behavior. First, the value of each characteristic distance D_i is coded with five symbolic states (see Table1 and [1,2]) reflecting the magnitude of the corresponding deformations: S_i if the current distance is roughly equal to its corresponding value in the Neutral expression, C_i^+ (vs. C_i^-) if the current distance is significantly higher (vs. lower) than its corresponding value in the Neutral expression, and $S_i \cup C_i^+$ (vs. $S_i \cup C_i^-$) if the current distance is neither sufficiently higher (vs. lower) to be in C_i^+ (vs. C_i^-), nor sufficiently stable to be in S_i. Secondly, Nasal root and nasolabial furrows information are coded with two states: "present" P_j or "absent" A_j $1 \leq j \leq 2$ according to the detection confidence measure as described previously. The explicit doubt of their state $P_j \cup A_j$ (P_j *or* A_j) is also modeled and allows modeling the uncertainty of their detection (see section 3.1 for the modeling of the corresponding doubt).

Table 1: Rules table defining the visual cues states corresponding to each facial expression

	D_1	D_2	D_3	D_4	D_5	TF_1	TF_2	An
Happiness	C_1^-	$S_2 \cup C_2^-$	C_3^+	C_4^+	C_5^-	A_1	P_2	Op
Surprise	C_1^+	C_2^+	C_3^-	C_4^+	C_5^+	A_1	A_2	-
Disgust	C_1^-	C_2^-	$S_3 \cup C_3^+$	C_4^+	S_5	P_1	P_2	Cl
Anger	C_1^-	C_2^-	S_3	$S_4 \cup C_4^-$	S_5	P_1	P_2	$Op \cup Cl$
Sadness	C_1^-	C_2^+	S_3	C_4^+	S_5	A_1	A_2	-
Fear	C_1^+	$S_2 \cup C_2^+$	$S_3 \cup C_3^-$	$S_4 \cup C_4^+$	$S_5 \cup C_5^+$	A_1	A_2	-
Neutral	S_1	S_2	S_3	S_4	S_5	A_1	A_2	-

Finally, the nasolabial furrows angles can take two states: when present they can be "opened" Op or "closed" Cl if the value of the corresponding angle is higher or lower respectively than a predefined value. As for the wrinkles detection a doubt state $Op \cup Cl$ is also introduced to model the uncertainty of the measured angles (see section 3.1).

Table 1 defines the characteristic distances, the transient feature and the nasolabial furrow angle states according to each facial expression. However, a logic system is not sufficient to model the facial expression. Indeed, an automatic facial expression model should explicitly model the doubt and uncertainty of the sensors (such as $P_j \cup A_j$) generating its conclusion with confidence that reflects uncertainty of the sensors detection and tracking. For this reason, we use the Transferable Belief Model.

3. Visual cues fusion by the Transferable Belief Model

The TBM [8] considers the definition of the frame of discernment $\Omega = \{H_1,...,H_N\}$ of N exclusive and exhaustive hypotheses characterizing the six basic facial expressions as well as Neutral Ω = {*Happiness* (E_1), *Surprise* (E_2), *Disgust* (E_3), *Fear* (E_4), *Anger* (E_5), *Sadness* (E_6), *Neutral* (E_7)}. The TBM approach requires the definition of the Basic Belief Assignment (BBA) (the belief in each state) associated to each independent source of information (permanent facial feature and transient facial feature states).

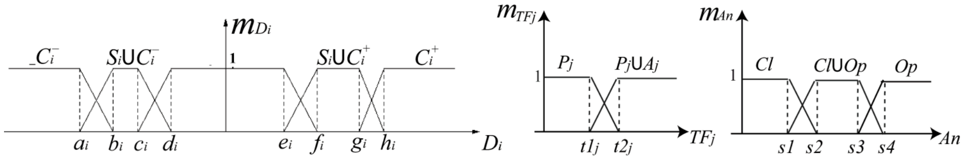


Figure 5: Left: model of BBAs for the characteristic distances (for more detail see [1]); middle: model of BBAs for the transient features detection; right: model of BBAs of the Nasolabial furrow angles. The thresholds have been derived by statistical analysis on benchmark databases.

3.1. Belief Modeling

The permanent facial features (characteristic distances) model is based on the work of Hammal *et al.* [1]. The Basic Belief Assignment (BBA) $m_{Di}^{\Omega Di}$ of each characteristic distance state D_i is defined as:

$$m_{Di}^{\Omega Di} : 2^{\Omega Di} \rightarrow [0,\ 1] \qquad A^{\Omega Di} \rightarrow m_{Di}^{\Omega Di}(A), \ \sum_{A \in 2^{\Omega Di}} m_{Di}^{\Omega Di} = 1 \qquad 1 \le i \le 5 \tag{4}$$

Where $\Omega_{D_i} = \{C_i^+, C_i^-, S_i\}$ is the power set, $2^{\Omega_{D_i}} = \{\{C_i^+\},\{C_i^-\},\{S_i\},\{S_i,C_i^+\},\{S_i,C_i^-\},\{S_i,C_i^+,C_i^-\}\}$, the frame of discernment, $\{S_i,C_i^+\}$ (vs. $\{S_i,C_i^-\}$) the doubt state between C_i^+ (vs. C_i^-) and S_i, $m_{D_i}^{\Omega D_i}(A)$, the belief in the proposition $A \in 2^{\Omega_{D_i}}$ without favoring any proposition of A in case of doubt proposition. This is the main difference with the Bayesian model, which implies equiprobability of the propositions of A. The piece of evidence $m_{Di}^{\Omega Di}$ associated with each symbolic state given the value of the characteristic distance D_i is obtained by the model depicted in Fig. 5 left.

The BBA $m_{TF_j}^{\Omega TFj}$ of the states of each transient feature TF_i is defined as:

$$m_{TF_j}^{\Omega TFj} : 2^{\Omega TFj} \rightarrow [0,\ 1] \qquad B^{\Omega TFj} \rightarrow m_{TF_j}^{\Omega TFj}(B), \ \sum_{B \in 2^{\Omega TFj}} m_{TF_j}^{\Omega TFj} = 1 \qquad 1 \le j \le 2 \tag{5}$$

where TF_1 means the *nasal root wrinkles*, TF_2, the *nasolabial furrow*, $\Omega_{TF_j} = \{P_j, A_j\}$, the power set, $2^{\Omega_{TF_j}} = \{\{P_j\},\{A_j\},\{P_j,A_j\}\}$, the frame of discernment. P_j means the presence of transient features and A_j, their absence. From the frame of discernment only the states P_j (the wrinkles are present without any doubt) and the state $\{P_j, A_j\}$ (there is a doubt in their detection and noted $P_j \cup A_j$) are considered. Then if the wrinkles are detected as present (the confidence threshold is higher than the defined value) the corresponding state is P_j if not the corresponding state is $P_j \cup A_j$. The piece of evidence $m_{TF_j}^{\Omega TFj}$ of each state is derived according to the model depicted in Fig. 5 middle.

a. Nasal root

The nasal root wrinkles are used as a refinement process and are associated to Disgust and Anger (without favoring any of them) with the piece of evidence: $m_{TF_1}^{\Omega_{TF_1}}(P_1) = m_{TF_1}^{\Omega_{TF_1}}(Anger \cup Disgust) = 1$. If they are not present: the current expression is one of the 7 expressions with the piece of evidence: $m_{TF_1}^{\Omega_{TF_1}}(P_1 \cup A_1) = m_{TF_1}^{\Omega_{TF_1}}(Anger \cup Disgust \cup Happiness \cup Surprise \cup Fear \cup Sadness \cup Neutral) = 1$

b. Nasolabial furrows
The nasolabial furrows are associated to Happiness, Disgust and Anger expressions with the piece of evidence: $m_{TF_2}^{\Omega_{TF_2}}(P_2) = m_{TF_2}^{\Omega_{TF_2}}(Happiness \cup Anger \cup Disgust) = 1$. If they are absent: the current expression is one of the 7 expressions with the piece of evidence $m_{TF_2}^{\Omega_{TF_2}}(P_2 \cup A_2) = m_{TF_2}^{\Omega_{TF_2}}(Anger \cup Disgust \cup Happiness \cup Surprise \cup Fear \cup Sadness \cup Neutral) = 1$

c. Orientation
The basic belief assignment of the nasolabial furrow angles allows assigning a belief on the state of the detected angles as:

$$m_{An}^{\Omega_{An}} : 2^{\Omega_{An}} \to [0,\ 1]\ ,\ C^{\Omega_{An}} \to m_{An}^{\Omega_{An}}(C),\quad \sum_{C \in 2^{\Omega_{An}}} m_{An}^{\Omega_{An}} = 1 \qquad (6)$$

Where An is the angle, $\Omega_{An} = \{Op, Cl\}$, the power set $2^{\Omega_{An}} = \{\{Op\},\{Cl\},\{Op,Cl\}\}$, the frame of discernment, Op and Cl mean opened and closed angles (see section 2.2) $\{Op,Cl\}$ means Op or Cl and corresponds to the doubt between Op and Cl (noted $Op \cup Cl$). The pieces of evidence associated to the states of the detected angles are defined using the model proposed in Fig. 6 right. The thresholds associated to each state are validated by statistical analysis on 3 benchmark databases [10,14,16].
The piece of evidence associated to each one of the 3 expressions Happiness, Disgust and Anger are then computed according to the BBAs of the angles using the fuzzy-like model of Fig. 5 right as:

$$m_{An}^{\Omega_{An}}(An \le s1) = m_{An}^{\Omega_{An}}(Cl) = m_{An}^{\Omega_{An}}(Disgust) = 1$$
$$m_{An}^{\Omega_{An}}(An \ge s4) = m_{An}^{\Omega_{An}}(Op) = m_{An}^{\Omega_{An}}(Happiness) = 1$$
$$m_{An}^{\Omega_{An}}(s2 \le An \le s3) = m_{An}^{\Omega_{An}}(Op \cup Cl) = m_{An}^{\Omega_{An}}(Happiness \cup Disgust \cup Anger) = 1$$

In the other cases the piece of evidence of the expression or subset of expressions corresponds to the projection of the value of the corresponding angles using the model of Fig. 6 right.

4. Dynamic fusion

The dynamic and asynchronous behavior of the facial features is introduced by combining at each time *t* their previous deformations from the *beginning* until the *end* of the sequence to take a decision. The analysis of the facial feature states is made inside an increasing temporal window Δt (Fig. 6 (a)). The size of the window Δt increases progressively at each time from the *beginning* until the *end* of the expression. Then, at each time *t* inside the window Δt, the current state of each facial feature (e.g. characteristic distances and transient features) is selected based on the combination of their current state at time *t* and of the whole set of their past states since the beginning of the expression which then takes into account asynchronous facial feature deformations (Fig. 6(b)).
The dynamic fusion of the BBAs (see Fig6 (b)) is made according to the number of appearance of each symbolic states noted $\underset{\Delta t}{Nb}(state)$ and their integral (sum) of plausibility noted $\underset{\Delta t}{Pl}(state)$ computed inside the temporal window Δt. For instance, for a characteristic distance D_j and for the $state = C^-$:

$$K_t(C^+) = \begin{cases} 1 \ if & m_{D_j}(C^-) \ne 0 \\ 0 \ otherwise & 1 \le t \le \Delta t \end{cases} \qquad \underset{\Delta t}{Nb}(C^-) = \sum_{t=1}^{\Delta t} K_t(C^-)\ (8)\ ; \qquad \underset{\Delta t}{Pl}(C^-) = \sum_{t=1}^{\Delta t} (m_{D_j}(C^-) + m_{D_j}(S \cup C^-)) \qquad (9)$$

From the two parameters $\underset{\Delta t}{Nb}(state)$ and $\underset{\Delta t}{Pl}(state)$, the selected states of each visual cues at each time *t* inside the temporal window Δt are chosen as:

$$\underset{\Delta t}{State}(D_i) = \max(\underset{\Delta t}{Pl}(state_{D_i}) / \underset{\Delta t}{Nb}(state_{D_i})),\ state_{D_i} \in \left\{C_i^+, C_i^-, S_i \cup C_i^+, S_i \cup C_i^-\right\}\ 1 \le i \le 5 \qquad (10)$$

$$State_{\Delta t}(TR_j) = \max_{\Delta t}(Pl(state_{TRj}) / Nb_{\Delta t}(state_{TRj}))\ ,\ state_{TR_j} \in \{P_j,\ P_j \cup A_j\},\ 1 \le j \le 2 \qquad (11)$$

$$State_{\Delta t}(An) = \max_{\Delta t}(Pl(state_{An}) / Nb_{\Delta t}(state_{An}))\ ,\ state_{An} \in \{Op,\ Cl,\ Op \cup Cl\} \qquad (12)$$

Fig. 6 (c) shows an example of the temporal evolution of the states of the characteristic distance D_2. One can see the correction of the false detection state (C^+) by the temporal fusion process (Eq. 10).
The piece of evidence associated to each chosen state corresponds to its maximum piece of evidence inside the current temporal increasing window as:

$$m_{\underset{\Delta t}{State}(Cues),\Delta t} = \max(m_{Cues_{i,1 \ldots \Delta t}})\ ,\ Cues \in \{D_i, TF_j, An\},\ 1 \le i \le 5,\ 1 \le j \le 2 \quad (13)$$

Then at time t between the *beginning* and the *end* of the expression sequence, once the basic belief assignments of all the visual cues are defined, the corresponding expression is selected according to the rules Table 1.

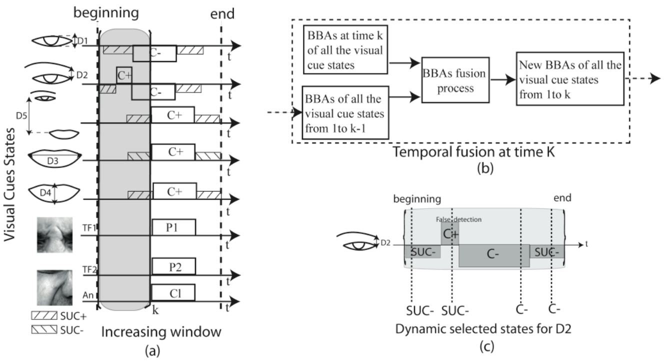


Figure 6. (a) Example of the increasing temporal window during a sequence of Disgust expression; (b) BBAs selection definition process at time k based on the current BBAs and the whole set of the past BBAs for all the visual cues; (c) Example of the selection process of the characteristic distances states inside the increasing temporal window where the state C^- is chosen.

5. Belief fusion and decision

The fusion process of all the visual cue states is done at each time (Fig. 6 (a)) using the conjunctive combination rule (Eq. 14) [15] and results in m^{Ω} the BBA of the corresponding expression or subset of expressions:

$$m^{\Omega} = \oplus m^{\Omega}_{Cues} \qquad (14)$$

From Table 1 and the BBAs of the states of each sensor: the characteristic distance states $m^{\Omega Di}_{D_i}$, the transient feature states $m^{\Omega TFi}_{TF_i}$ and the angles states $m^{\Omega An}_{An}$, a set of BBAs on facial expressions is derived for each sensor as: $m^{\Omega}_{D_i}$, m^{Ω}_{TFi} and m^{Ω}_{An}. The fusion process of the BBAs $m^{\Omega}_{D_i}$, m^{Ω}_{TFi} and m^{Ω}_{An} is performed successively using the conjunctive combination rule (Eq. 14). For example, if we consider two characteristic distances D_i and D_j with two BBAs $m^{\Omega}_{D_i}$ and $m^{\Omega}_{D_j}$ derived on the same frame of discernment, the joint BBA $m^{\Omega}_{D_{i,j}}$ is given using the conjunctive combination as:

$$m^{\Omega}_{D_{i,j}}(A) = (m^{\Omega}_{D_i} \oplus m^{\Omega}_{D_j})(A) = \sum_{E \cap F = A} m^{\Omega}_{D_i}(E) * m^{\Omega}_{D_j}(F) \qquad (15)$$

The results of the combination of the characteristic distances are then combined to the BBAs of the transient feature states by the conjunctive combination as:

$$m^{\Omega}_{D_i,TF_j}(G) = (m^{\Omega}_{D_i} \oplus m^{\Omega}_{TFj})(G) = \sum_{A \cap B = G} m^{\Omega}_{D_i}(A) * m^{\Omega}_{TFj}(B) \qquad (16)$$

The obtained results are finally combined to the BBAs of the angle states as:

$$m^{\Omega}_{D,TF_j,An}(H) = (m^{\Omega}_{D_i,TFj} \oplus m^{\Omega}_{An})(H) = \sum_{G \cap C = H} m^{\Omega}_{D_i,TF_j}(G) * m^{\Omega}_{An}(C) \qquad (17)$$

Where A, B, E, F, G, H, C denote propositions and $E \cap F$, $A \cap B$, $G \cap C$ the conjunction (intersection) between the corresponding propositions. This leads to propositions with a lower number of elements and with more accurate pieces of evidence.

For example, if the BBAs resulted from Eq. 16 is the doubt between Happiness and Disgust as: $m^{\Omega}_{D_i,TF_j} = m^{\Omega}_{D_i,TF_j}(Happiness \cup Disgust)$ and the BBAs resulted from the nasal root angles is Happiness as: $m^{\Omega_{An}}_{An}(Op) = m^{\Omega_{An}}_{An}(Happiness)$ the combination of these two BBAs according to the Eq. 17 leads to the selection of Happiness as : $m^{\Omega}_{D_i,TF_j}(Disgust \cup Happiness) \oplus m^{\Omega}_{An}(Happiness) = Happiness$

The decision is the ultimate step of the classification process. It consists in making a choice between various hypotheses E_e and their possible combinations. Making a decision is associated with a risk except if the result is sure ($m(E_e) = 1$). Several decision criteria can be used [8, 15]. In this paper the decision was made using the credibility as:

$$Bel : 2^{\Omega} \rightarrow [0,\ 1], \qquad I \rightarrow Bel(I) = \sum_{B \subseteq I, B \neq 0} m^{\Omega}(B), \forall\ I \in \Omega \tag{18}$$

6. Results

The classification results were performed on all the six basic facial expression and on three benchmark databases Cohn-Kanade [10], STOIC [4] (a total of 182 videos) and CAFE [16] (80 images). Recall and Precision are used for the evaluation of the proposed method. Finally we use the F-measure which combines evenly *Recall* and *Precision* as $F = 2 * \mathrm{Re}call * \mathrm{Pr}ecision / (\mathrm{Re}call + \mathrm{Pr}ecision)$. The obtained performances are reported in Table 2 left. The best performances are obtained for Happiness and Anger. The lowest performances are obtained for Fear and Sadness expressions. These results are explained by two doubt states that appear frequently: the doubt between Fear and Surprise, and the doubt between Sadness and Anger expressions. Interestingly, these expressions are also notoriously difficult to discriminate for human observers [4]. Compared to the previous model [1] the introduction of the temporal modeling of all the facial expressions leads to an increase of performances of 12% in average. In order to compare our results with human performances we validate part of the used videos by human observers. Table 2 right reports the human performances. 15 human observers discriminated between the six basic facial expressions on 80 videos randomly interleaved in 4 separate blocks. Performances obtained by our system and by humans are not significantly different (two-way ANOVA, $P > 0.33$).

Table 2: Recall and Precision measures (in %). Left, model performances; right, human performances

Exp	*Recall*	*Precision*	*F*
E_1	92	90	91
E_2	77	96	85
E_3	83	95	89
E_4	70	68	69
E_5	100	88	94
E_6	74	70	72

Exp	Recall	Precision	F
E_1	100	89	94
E_2	88	90	89
E_3	75	86	80
E_4	98	72	83
E_5	88	97	92
E_6	88	92	90

7. Conclusions

We have presented a model for the recognition of the six basic facial expression sequences using a dynamic modeling of the TBM. It deals with uncertainty of facial features segmentation and doubt between expected facial expressions. We introduced a new method based on multiscale filtering for the detection of the presence and the orientation of the transient facial features. Then, we proposed a temporal modeling process for the dynamic and progressive fusion of the permanent and transient facial features behavior from the beginning to the end of the sequence dealing with asynchronous facial feature deformations. Compared to the previous results the obtained performances increased by 12% and compare favorably to human observers. Besides the use of the most important human visual cues for facial expressions classification future work will introduce another possible human strategy consisting in weighting each visual cue according to its importance for the expected facial expression.

References

1. Hammal Z., Couvreur L., Caplier A., Rombaut M., Facial expressions classification: A new approach based on transferable belief model. International Journal of Approximate Reasoning , 46(3):542-567 (2007).
2. Hammal Z., Arguin M., Gosselin F.. Comparing a transferable belief model capable of recognizing facial expressions with human data. International Journal of Vision. 9/2/22/, doi:10.1167/9.2.22. (2009).
3. Hammal Z., Eveno N., Caplier A., Coulon, P-Y. Parametric models for facial features segmentation, Signal processing, 86 :399-413, 2006.
4. Roy, S., Roy, C., Fortin, I., Either-Majcher, C., Belin, P., & Gosselin, F. A dynamic facial expression database. *Proceedings of the Vision Sciences Society,* Sarasota, Florida (2007).
5. Smith M., Cottrell G., Gosselin F. Schyns P.G, Transmitting and decoding facial expressions of emotions, Psychological Science, 16184–189 (2005)
6. Pantic M., Rothkrantz, L.J.M. Automatic analysis of facial expressions: The state of the art. IEEE Transactions On Pattern Analysis and Machine Intelligence, 22(12):1424–1445 (2000)
7. Pantic M., Patras I., Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences, IEEE Transactions on Systems, Man and Cybernetics - Part B, 36(2):433-449 (2006)
8. Smets P., Kruse R., The transferable belief model, Artificial Intelligence, 66:191–234 (1994)
9. Youn A.W., Rowland D., Calder A.J., Etcoff N.L., Seth A., Perrett D.I., Facial expression megamix: Tests of dimensional and category accounts of emotion recognition, Cognition, 63, 271–313 (1997)
10. Cohn-Kanade database. 2000. http://vasc.ri.cmu.edu/idb/html/face/facial expression.
11. Cohn J.F.., Schmidt K.L., The timing of facial motion in posed and spontaneous smiles. Journal Wavelets,Multiresolution & Information Processing, 2(2) (2004)
12. Tian Y., Kanade T., Cohn J.F., Recognizing Action Units for Facial Expression Analysis, IEEE Trans. Pattern Analysis and Machine Intelligence, 23(2) :97-115 (2001)
13. Massot C., Herault J., Model of Frequency Analysis in the Visual Cortex and the Shape from Texture Problem, International Journal of Computer Vision, 76(2) (2008)
14. Hammal-Caplier database, 2003. http://viscog.psy.umontreal.ca/ zakia/index.htm.
15. Denoeux T., Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence, Artificial Intelligence, 172:234–264 (2008)
16. Dailey M., Cottrell G.W., Reilly J.. California facial expressions (cafe). unpublished digital images, University of California, SanDiego (2001)

Classifying Facial Pain Expressions
Individual Classifiers vs. Global Classifiers

Michael Siebers[1], Miriam Kunz[2], Stefan Lautenbacher[2], and Ute Schmid[1]

[1] Faculty Information Systems and Applied Computer Science, University of Bamberg, Germany
[2] Physiological Psychology, University of Bamberg, Germany

Abstract. Pain is a highly affective state that is accompanied by a facial expression. In this paper we compare different classifiers on their possibilities to classify pain from facial point data. Furthermore we investigate the need for training classifiers on each subject's data individually. We show that most used classifiers are suitable for the classifying facial pain expressions. The second issue cannot be finally decided.

1 Motivation and Introduction

Pain is an highly unpleasant feeling everybody is familiar with. Humans have an urge to communicate pain to others. This is done both verbally—the typical "ouch"—and nonverbally. Other persons react on those pain messages, e.g. with pity by providing help. Important for nonverbal communicating pain is the facial expression. Especially for those individuals that are not able to communicate pain verbally (e.g. patients with dementia) the facial expression of pain can serve as a valid communication channel of pain [1, 2]. This finding can be valuable for the clinical context where individualized classifiers for pain can support relatives and nursing staff in recognizing pain states of patients[3].

This pilot study tries to find possibilities for deciding whether a person experiences pain using facial images. To be precise, this study tries to answer two questions. First, which classifiers are suitable to predict pain by facial features. Second, is a global classifier, trained on different persons, sufficient or must individualized classifiers be trained for each person as psychological studies [1, 2] suggest.

Since this is a pilot study we are only handling a limited number of data records. The work reported in this paper is mainly explorative concerning the general investigation of suitable learning approaches, appropriate feature selection, and efficient preprocessing of data. Most importantly, we aimed at getting first support for the psychological hypothesis that pain classification must be dealt with individual classifiers and that general classifiers—as often used for classification of emotions as happiness [4]—are significantly less reliable.

2 Data Acquisition

In order to learn classifiers we needed pain-annotated facial data. These were derived from videos obtained during a previously conducted psychophysiological study and stills of non-painful and painful events were extracted. On those images facial features were annotated. Based on these features relational measures were calculated.

2.1 Psychophysiological Study

In this study, pressure stimuli of various intensities were applied to the upper edge of the trapezius muscle using a pressure Algometer (SOMETIC). Each subject received 40 stimuli, half-and-half painful and not painful. As the first stimulus a pressure of 5 kg was applied. This stimulus was repeated with increasing weight until the subject showed a pain face or the maximum of 8 kg was reached. This target intensity was used as the painful stimulus, whereas 1 kg was used to induce non-painful pressure sensations.

A total of 30 subjects took part in the study. Each of them was video taped. In the background a LED light lit up while pressure was applied. In addition to the video recordings, the subjective pain ratings were assessed on a Visual Analogue Scale[3]. Furthermore, the pressure intensity and expert's ratings on whether subjects displayed pain-typical facial expressions (this was done by M. Kunz who is a certified coder for facial expressions) were recorded for each stimulus application.

As this study was not initially conducted to be used for automated facial pain detection, it is no surprise that we came across some problems concerning image quality. The main one was that some subjects' hair was hiding parts of their faces.

2.2 Still Extraction

Since we wanted to evaluate the classification performances of different classifiers, we decided to use ideal data only. Therefore among all subjects only those showing the most prototypic pain faces were chosen for further consideration. Six person showed appropriate mimics. From their videos ten stills per stimulus were extracted. Stills were taken from all 20 non-pain stimuli and all pain stimuli showing a pain face. All stills were taken from within the first second after facial expression onset. This resulted in a total of 1200 non-pain and 880 pain images.

From further study one subject was excluded. He showed a pain face only in six of the 20 pain stimuli—others at least twice as much. Additionally he was the only male among the remaining subjects.

---

[3] A visual analogue scale (VAS) is a scale where values are stated on a continuum—ranging from one extreme to another. In this study it is implemented as percentage value.

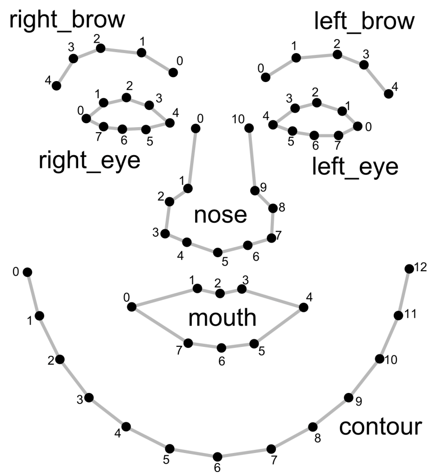


Fig. 1. Annotated feature points and shapes.

2.3 Facial Feature Annotation

On the images we marked 58 points as seen in Fig. 1. Those formed the face's contour (13 points), mouth (8 points), nose (11 points), eyes (8 points each) and eye brows (5 points each). We chose those points with FACS [5] in mind. However, we didn't use FACS directly, as we are in doubt whether it is suited to map facial pain expressions. For the annotation process we used the AM Tools of Tim Cootes [6].

As the images hardly displayed variation only the first three images of each stimulus were used. Few stimuli had to be excluded as the face was not completely on the images. So the final data set consisted of 534 records (294 non-pain; 240 pain) of five female subjects.

To eliminate the size and the position of the face within the image the x- and y-coordinates of the points were standardized to a range of $[0, 1]$.

2.4 Relational Measures

To the pure data record additional attributes were added. Distances considered as relevant were calculated and added to the data record.

- width and height of the mouth
- widths and heights of both eyes
- span from the tip to the root of the nose
- distance between the eye brows
- distances between eye and eye brow
- distance between mouth and nose

All distances can be seen in Fig. 2. Moreover we added the angle between the mouth corners and the mouth center.

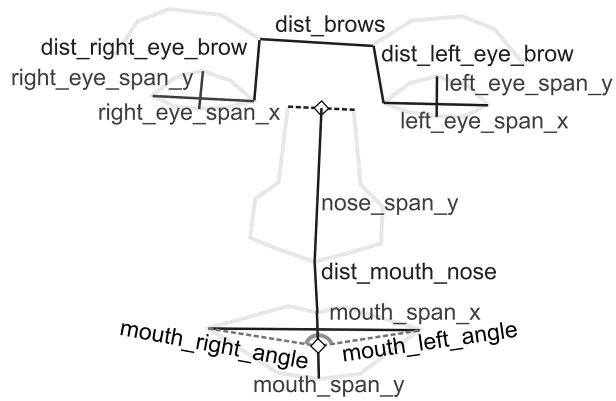


Fig. 2. Calculated distances and angles

It might be that these distances scale proportionally, or it might not be. To survey this we included proportions into the data set.

- ratio of mouth height to mouth width
- ratio of eye height to eye width (both sides) [called: eye ratio]
- ratio of left eye width to right eye width
- ratio of left eye height to right eye height
- ratio of left eye ratio to right eye ratio

It is understood that the pain face is not the usual face but a distortion of the neutral face. Taking this into account we averaged the distances, angles and ratios of the non-pain data records and added for every record the ratios of those attribute to their means. As the deformation of the face is potentially highly individual we also added the ratios of those attributes to their means of the considered subject's non-pain data.

Of course all mentioned relational measures are implied in the point data alone. However they would be inaccessible for some classifiers, e.g. Decision Trees (see Sect. 3.1). Hence we decided to include those measures in an explicit form.

3 Experiments

Based on the data set with 534 entries and 178 attributes (including image id, person number and class) classifiers were trained.

3.1 Classifiers

In the following the used classifier are briefly described. For further reading see [7] or [8](Support Vector Machines).

Decision Trees classify an example according to a set of tree structured if-then-rules. Each inner node of the tree denotes an attribute and branches according to its values. The *Decision Tree* classifier is the only symbolic classifier we used. We considered using it as the constructed decision model is human understandable and thus has an explaining element. For learning this trees we used ID3Numerical, a modification of Quinlan's ID3 [9] which can handle numerical attributes.

Naive Bayes is a statistical classifier that is based on Bayes' Law. It estimates the probability $P(c \mid \boldsymbol{x})$. For this purpose it assumes that all attributes are statistical independent (*thus naive*).

Support Vector Machines try to find a hyperplane – probably in a higher space – that separates the classes. For classification only the data records closest to the hyperplane are used – the *Support Vectors*.

k-Nearest Neighbours (kNN) considers each data record as points in $\mathbb{R}^n$. An unseen instance is now classified searching the k nearest points in the training example space and assigning their mode class. All training examples are kept in the feature space. Therefor *kNN* is called a *lazy learner*.

The Perceptron considers each data record as a vector of real-valued inputs $(x_1, \ldots, x_n)$. Those are weighted with $(w_0, \ldots, w_n)$. For classification it calculates the linear combination $\sum_{i=0}^{n} w_i x_i$, where x_0 is always 1. If the result is greater than 0 it returns *positive*, *negative* otherwise.

Neural Networks use a graph of sigmoid units to calculate the desired class. Sigmoid units are similar to perceptrons except that they use a sigmoidal output. This output is serving as input for the next unit. In this study we used a linear and acyclic network.

Classification by Regression trains a regression model for each class. The class with the higher predicted value is then selected. As base regression model we used linear regression. Linear regression tries to find the linear function that predicts the examples best.

3.2 Procedure

For each classifier the learning generally consisted of two phases. First its optimal parameters were obtained then its performance was measured in an cross-validation. As *kNN* suffers from curse of dimensionality, the attributes were weighted for this classifier before parameter optimization.

Cross-Validation was done using 10 partitions. The partitions were drawn according to single data records by means of stratified sampling[4].

[4] Stratified exampling means randomly assigning data records to folds while trying to presume class distribution.

Parameters were optimized via a systematic grid search. Parameter combinations were systematically evaluated using 10-folded cross-validations. The combination which performed best was selected. For the *Neural Network* no systematic approach was chosen. Taking the long training times and the size of the parameter space into account we decided on an evolutionary technique.

Attribute Weighting was performed using forward weighting. Initially every attribute was assigned a weight of 0. Each attribute was then independently weighted using a linear search.

Attribute Selection was used to overcome *Naive Bayes'* assumptions that attributes are statistical independent. Carried out as forward selection, an initial population is created – one individual per attribute. Then further attributes are added to the best ones as long as performance increases.

This proceedings were carried out with the whole data set (global) and once for each subject – using only its data (individual classifiers). We ran the experiments with RapidMiner[5] on a Fujitsu Siemens Computers LIFEBOOK T Series (Intel©Core™2 Duo P8400, 2 GB) and a Fujitsu Siemens Computers Esprimo (Intel®Pentium®4 3.00 GHz, 1 GB).

4 Results

The results of the different classifiers are shown in Tab. 1. Obviously most classifiers are suitable for this task.

Table 1. Experiments results

	global	individual
Decision Tree	0.955	0.970
Support Vector Machine	0.983	0.981
Regression	0.968	0.972
Perceptron	0.559[a]	0.474[b]
Neural Network	0.450[c]	0.514[b]
Naive Bayes	0.923	0.993
k-nearest Neighbours	0.989	0.999

[a] Parameter Optimization canceled after 19 days. Best interim result is displayed.
[b] Parameters not optimized due to time constraints.
[c] Parameter Optimization canceled after 10 days. Using not optimal parameters.

[5] www.rapid-i.com

4.1 Classifier Performance

Apparently *Support Vector Machines* and *k-nearest Neighbours* perform best. Unfortunately it is not possible to build an exact ranking. Due to the number of subjects, no significance tests were made. Decision Tree, Support Vector Machine, Regression, Naive Bayes and *k*-nearest Neighbours perform similar as adults classifying clinical pain videos[10].

No detailed statement about the *Neural Network* can be made. The global parameter optimization crashed after 10 days. Since no interim results were available performance estimation was done with manually chosen parameters. A repetition of the parameter optimization was not possible due to time limitations.

The detail plot in Fig. 3 shows that the data is probably not linear separable. This explains the bad performance of the *Perceptron* classifier. It performs even worse than guessing[6].

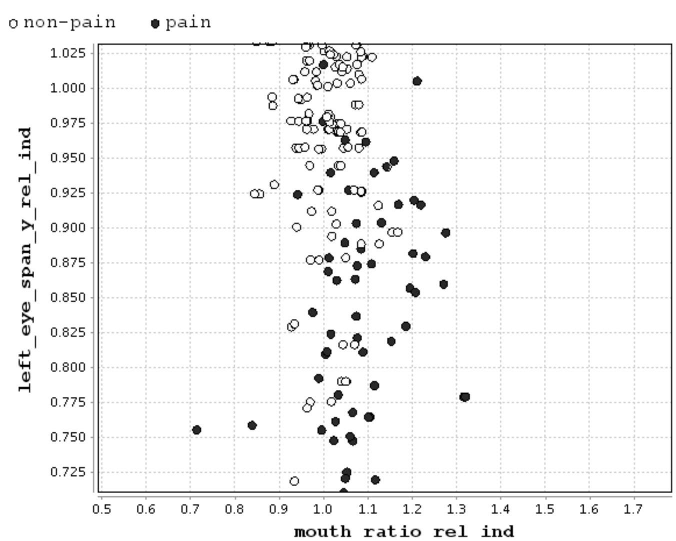


Fig. 3. Detail plot of the used data set. The topmost attributes of the global decision tree model were used as axis.

4.2 Global vs. Individual Classification

For most classifiers the individual approach seems to deliver better results. But—due to the small number of subjects—also here no detailed comparison is possible.

[6] Guessing probability: 0.550 for global classification, 0.583 for individual classification

5 Conclusion and Future Work

We showed that many classifiers are suitable for predicting pain by facial expression. Global and individual classifiers perform nearly equally. But due to the low subject number and since we only selected those subjects that showed prototypical facial pain displays[11] it is by now questionable if these findings can be generalized.

Currently we are able to decide if a shown face is prototypical for pain or not. However, the true question is whether the person whose face is depicted is experiencing pain. Therefore we will deal with predicting the self-reported VAS values in further work. Additionally we will tackle the issue of mixing up different emotions, first trying to distinguish between pain and disgust. Therefore we will use more subjects for the initial psychological study.

Further studies should also address a broader variety of subjects—for example regarding age or gender.

Acknowledgements We thank Simone Burkhardt and Ina Schulz for the data collection and picture extraction. The study was supported by the Dr. Werner Jackstädt-Stiftung.

References

1. Kunz, M., Mylius, V., Scharmann, S., Schepelman, K., Lautenbacher, S.: Influence of dementia on multiple components of pain. European Journal of Pain **13** (2009) 317–325
2. Kunz, M., Scharmann, S., Hemmeter, U., Schepelman, K., Lautenbacher, S.: The facial expression of pain in patients with dementia. PAIN **133** (2007) 221–228
3. Salomon, P.E., Prkachin, K.M., Farewell, V.: Enhancing sensitivity to facial expression of pain. PAIN **71** (1997) 279–284
4. Strupp, S., Schmitz, N., Berns, K.: Visual-Based Emotion Detection for Natural Man-Machine Interaction. In Dengel, A.R., Berns, K., Breuel, T.M., Bomarius, F., Roth-Berghofer, T.R., eds.: KI 2008: advances in artificial intelligence: 31st Annual German Conference on AI, KI 2008, Kaiserslautern, Germany, September 23 - 26, 2008 ; proceedings. Volume 5243 of LNAI., Berlin, Springer (2008) 356–363
5. Ekman, P., Friesen, W.V.: Facial action coding system. Consulting Psychologists Press, Palo Alto, Calif. (1978)
6. Cootes, T.F.: Am Tools. http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/software/am_tools_doc/index.html
7. Mitchell, T.M.: Machine learning. McGraw-Hill, Boston, Mass. (1997)
8. Bishop, C.M.: Pattern Recognition and Machine Learning. 1. edn. Springer Science + Business Media LLC, New York, NY (2006)
9. Quinlan, J.R.: Induction of decision trees. Machine Learning **1**(1) (1986) 81–106
10. Deyo, K.S., Prkachin, K.M., Mercer, S.R.: Development of sensitivity to facial expression of pain. PAIN **107**(1-2) (2004/1) 16–21
11. Prkachin, K.M.: The consistency of facial expressions of pain: a comparison across modalities. PAIN **51** (1992) 297–306

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.

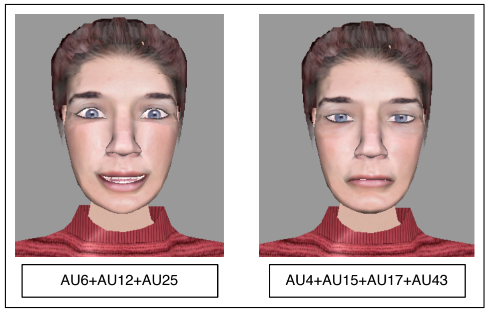


Fig. 1. Emma's face with two example facial expressions: Happy and Sad.

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.

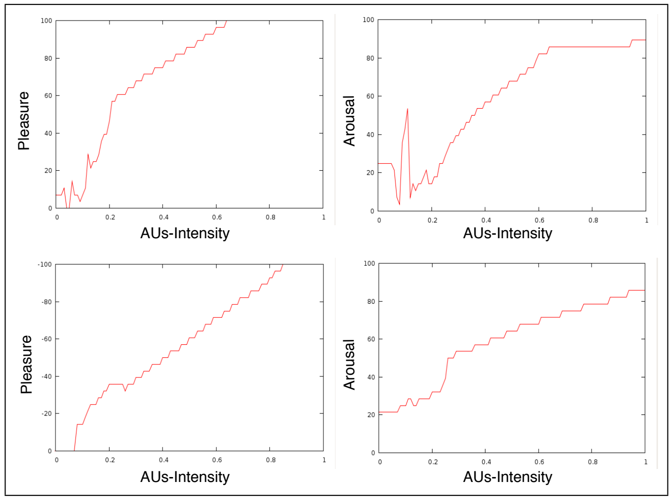


Fig. 2. Plots of Pleasure and Arousal (PA) courses, showing pleasure over AU intensity (left) and Arousal over AU intensity (right). The upper plots show PA-courses corresponding to AU6+AU12. The lower plots show PA-courses corresponding to AU4+AU5+AU7+AU10. The AUs of each facial expression are activated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

The OCC Model Revisited

Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer

Department of Information and Computing Sciences,
Utrecht University, The Netherlands
{bass,mehdi,jj}@cs.uu.nl

Abstract. Although popular among computer scientists, the OCC model of emotions contains a number of ambiguities that stand in the way of a truthful formalization or implementation. This paper aims to identify and clarify several of these ambiguities. Furthermore, a new inheritance-based view of the logical structure of emotions of the OCC model is proposed and discussed.

1 Introduction

In their book "The Cognitive Structure of Emotions" [1], Ortony, Clore & Collins have proposed a very interesting model of emotions that provides a clear and convincing structure of the eliciting conditions of emotions and the variables that affect their intensities. This psychological model is popular among computer scientists that are building systems that reason about emotions or incorporate emotions in artificial characters.

However, although many ad hoc or simplified implementations of "the OCC model" have been made, there have been fewer attempts at formalizing the complete, logical structure of the proposed emotion model (e.g., [2–4]). We are attempting to do so, using a formal logic containing constructs to reason about agents, their beliefs and actions, objects, and events. We are currently formalizing the eliciting conditions of emotions in this logic, trying to stay as close as possible to the book. Unfortunately, we have run into several ambiguities that stand in the way of a truthful formalization. We realize that satisfying logicians was not the primary concern of the book, but given their aim to provide a "computationally tractable account" (quoted from the back cover), it is important that computer scientists have an unambiguous structure of emotions to work with.

The contribution of this paper is to provide clarifications of the logical structure underlying the OCC model. After identifying several issues, we propose an inheritance-based view of the OCC model, supported by a new logical structure (figure 2) and new emotion type specifications (table 2). The structure that we propose aims to resolve several ambiguities in the OCC model, while our emotion type specifications have a stronger correspondence with the logical structure.

It should be emphasized that the current authors are not psychologists. We certainly do not claim to have more knowledge of emotions than OCC. We are computer scientists and logicians, and we have studied the OCC model from this perspective. Our critique only concerns the logical structure underlying the OCC model. Nevertheless, it is entirely possible that our interpretation of the OCC model is colored by our computer scientist's way of thinking. However, we do not find this possibility problematic, because the aim of our research into emotions is to make emotions *computable* (and make useful systems with them, of course).

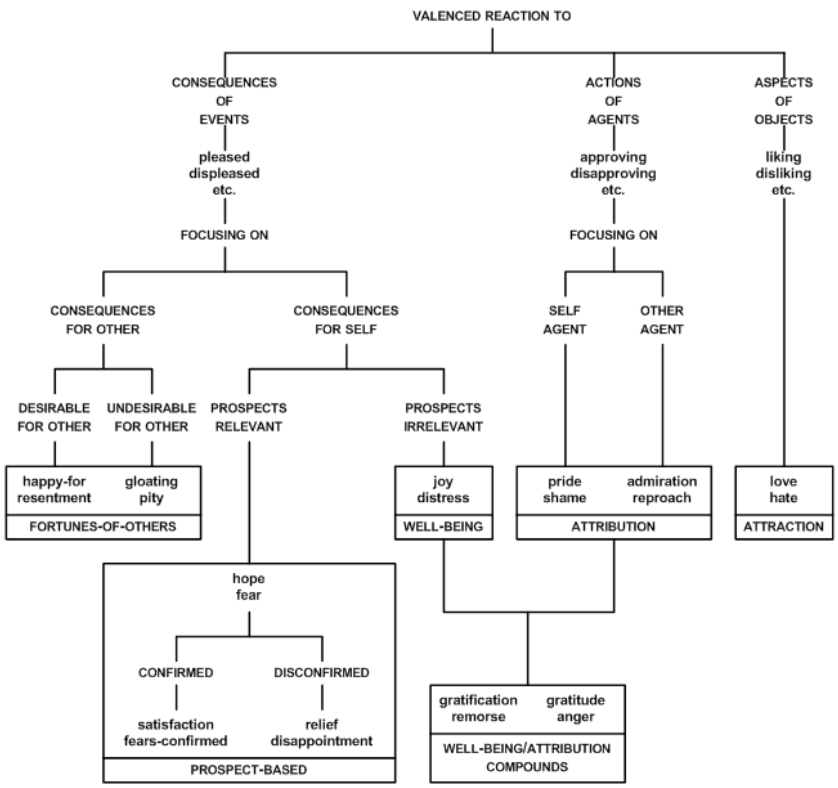


Fig. 1. The original structure of emotions of the OCC model, copied from page 19 [1].

2 The OCC Model

The OCC model describes a hierarchy that classifies 22 emotion types (see figure 1). The hierarchy contains three branches, namely emotions concerning consequences of events (e.g., joy and pity), actions of agents (e.g., pride and reproach), and aspects of objects (e.g., love and hate). Additionally, some branches combine to form a group of compound emotions, namely emotions concerning consequences of events *caused* by actions of agents (e.g., gratitude and anger). Because these notions (i.e. events, actions, and objects) are also commonly used in agent models, this makes the OCC model suitable for use in artificial agents.

Throughout the book, specifications are given for each of the 22 emotion types. For example, below is the specification of the class of emotions labeled as 'fear' in the OCC model (copied from page 112 [1]):

FEAR EMOTIONS
TYPE SPECIFICATION: (displeased about) the prospect of an undesirable event
TOKENS: apprehensive, anxious, cowering, dread, fear, fright, nervous, petrified, scared, terrified, timid, worried, etc.
VARIABLES AFFECTING INTENSITY:
(1) the degree to which the event is undesirable
(2) the likelihood of the event
EXAMPLE: The employee, suspecting he was no longer needed, feared that he would be fired.

Joy:	(pleased about) a desirable event
Distress:	(displeased about) an undesirable event
Happy-for:	(pleased about) an event presumed to be desirable for someone else
Pity:	(displeased about) an event presumed to be undesirable for someone else
Gloating:	(pleased about) an event presumed to be undesirable for someone else
Resentment:	(displeased about) an event presumed to be desirable for someone else
Hope:	(pleased about) the prospect of a desirable event
Fear:	(displeased about) the prospect of an undesirable event
Satisfaction:	(pleased about) the confirmation of the prospect of a desirable event
Fears-confirmed:	(displeased about) the confirmation of the prospect of an undesirable event
Relief:	(pleased about) the disconfirmation of the prospect of an undesirable event
Disappointment:	(displeased about) the disconfirmation of the prospect of a desirable event
Pride:	(approving of) one's own praiseworthy action
Shame:	(disapproving of) one's own blameworthy action
Admiration:	(approving of) someone else's praiseworthy action
Reproach:	(disapproving of) someone else's blameworthy action
Gratification:	(approving of) one's own praiseworthy action and (being pleased about) the related desirable event
Remorse:	(disapproving of) one's own blameworthy action and (being displeased about) the related undesirable event
Gratitude:	(approving of) someone else's praiseworthy action and (being pleased about) the related desirable event
Anger:	(disapproving of) someone else's blameworthy action and (being displeased about) the related undesirable event
Love:	(liking) an appealing object
Hate:	(disliking) an unappealing object

Table 1. The emotion type specifications of the OCC model, copied from the book [1].

Besides an example at the bottom, these specifications have three elements:

(1) The *type specification* provides, in a concise sentence, the conditions that elicit an emotion of the type in question.

(2) A list of *tokens* is provided, showing which emotion words can be classified as belonging to the emotion type in question. For example, 'fright', 'scared', and 'terrified' are all types of fear. (Of course, 'fear' is also a type of fear.)

(3) For each emotion type, a list of *variables affecting intensity* is provided. These variables are local to the emotion type in question, i.e., global variables (such as arousal) that affect all emotions are not included. The idea is that higher values for these variables result in higher emotional intensities.

In table 1 we have summarized the type specifications of all 22 emotion types.

3 Ambiguities and Clarifications

The presentation of the structure of the OCC model (see figure 1) is sure to make many computer scientists (including us) think of an inheritance diagram; that is, each emotion type is like its parent type plus some specialization. For example, displeased is a negatively valenced reaction to a consequence of an event; distress is displeased *plus* a focus on the self; regret is distress *plus* the constraint that the consequence in question signals a loss of opportunity, etc. Although it may not have been the authors' intention to present the structure of emotions as an inheritance hierarchy, we think it is actually useful to think of it that way and pursue this direction further. However, this interpretation reveals several ambiguities in the logical structure underlying the OCC model. In this section, we will list a number of such issues, together with our proposed clarifications.

(1) In table 1, the phrase "desirable event" is used many times. However, events are actually always appraised with respect to their *consequences*. For example, an earthquake in itself does not have a valence; only the consequences of this event (e.g., valuable lessons for seismologists, property damage, loss of life) are appraised as being

desirable or undesirable. Because desirability only applies to consequences of events, every instance of the phrase "desirable event" should actually be read as a shorthand for "desirable consequence of an event." Furthermore, the term "prospect" (used in, e.g., hope and fear) is intentionally ambiguous: it is used to refer to both *future* events and *uncertain* (past or current) events. Many formalizations appear to use OCC's notion of prospect in only one of these senses. For example, Adam [4] and Gratch & Marsella [2] only used uncertain prospects when formalizing hope and fear, whereas Steunebrink, Dastani & Meyer [3] only used future prospects.

(2) Looking closely at figure 1, we see that each of the three branches is headed by a set of emotional words; namely, pleased/displeased for event-based emotions, approving/disapproving for action-based emotions, and liking/disliking for object-based emotions. These are supposed to be the most *undifferentiated* emotion types. For example, "being pleased" represents the situation where one has appraised a consequence of an event as being desirable, but it says nothing about, e.g., whether that consequence is viewed as prospective or actual (i.e., one cannot say whether it is hope or joy), or whether that consequence is presumed to be desirable or undesirable for someone else (i.e., one cannot say whether it is happy-for or gloating). When the structure as shown in figure 1 is regarded as an inheritance hierarchy, pleased/displeased, approving/disapproving and liking/disliking then become *generalized* emotion types, from which all emotion types below them are derived. Indeed, in OCC's structure of *variables affecting intensity*, each emotion type inherits all variables from its parents. For example, 'gratification' inherits all variables from 'joy' and 'pride' (this can even be seen as a kind of *multiple inheritance*).

(3) As can be seen in figure 1, 'joy' is classified as an emotion type arising from positively appraising the consequences of an event where the focus is on consequences for the self and prospects are irrelevant. However, the type specification of 'joy' is given as "(pleased about) a desirable event" (see table 1), with no mention of a focus on the self or the irrelevance of prospects. So either the type specification of 'joy' is incomplete, or a focus on the self and a disregard of prospects is implicitly assumed to be the default. (In personal communication, OCC have acknowledged the latter to be the case.) A downside of assuming defaults, however, is that the type specification of 'joy' then conflates with 'pleased'; just as 'joy', 'pleased' is a valenced reaction to a desirable event. The difference is only implicit; namely, 'pleased' is undifferentiated, whereas 'joy' has a default differentiation on the self and a disregard of prospects.

(4) As can be seen in table 1, 'joy' is specified as "(pleased about) a desirable event" and 'distress' is specified as "(displeased about) an undesirable event." Crossing these specifications, one may wonder whether there is such a thing as "being pleased about an undesirable event" or "being displeased about a desirable event." The same holds for other pairs of opposite emotion types (i.e., pride/shame, love/hate, etc.): almost all type specifications contain a duplicate positivity or negativity. Often this is not problematic; for example, if one approves of an action, it must be praiseworthy, and vice versa. However, in some cases this duplication can introduce ambiguity; for example, in the type specification of 'hope', does "being pleased" enforce the phrase "desirable event," or does the pleasure apply to having the "prospect" thereof? In other cases, pleased/displeased and desirable/undesirable are indeed crossed; for example, compare

the type specifications of ‘relief’ and ‘disappointment’ in table 1. In the next section we will propose a way to remove these duplications and ambiguities.

(5) In figure 1, the six prospect-based emotion types are grouped in a framed structure with an inner hierarchy. How should this nesting be interpreted? On page 19 [1], OCC explicitly state that the figure represents a *logical* structure of emotions, not a *temporal* one. This would mean that both satisfaction/fears-confirmed and relief/disappointment are more differentiated types of hope/fear, or, to continue our inheritance-based perspective, they are *specializations* of hope/fear. In particular, disappointment would then have to be a specialization of fear. However, this cannot be, because disappointment does not inherit anything from fear; it inherits its *variables affecting intensity* from hope, but its valence is reversed (hope is positive and disappointment is negative). The relation between hope and disappointment appears to be more of a temporal kind. For example, first Bob hopes Alice will show up for their date, but when she does not, his hope turns into disappointment. Thus satisfaction, fears-confirmed, relief, and disappointment are not special kinds of hope or fear, but more like continuations of hope or fear, counting from the point when an event has been perceived that signals the confirmation or disconfirmation of the thing hoped for or feared. In other words, these four types are emotions in response to *actual* consequences of events, namely consequences signaling the confirmation or disconfirmation of a *previously prospective* consequence. In the next section we will propose to move these four emotion types from under hope/fear to become specializations of joy/distress.

(6) Consider the group of fortunes-of-others emotion types (happy-for, resentment, gloating, pity) and their type specifications in table 1. For there to be a ‘happy-for’ emotion, the consequence that is desirable for the other must also be desirable for oneself to some degree (probably because it satisfies an interest goal; page 94 [1]). But if a consequence of an event is appraised as being desirable for oneself, the conditions for ‘joy’ are satisfied (see table 1). So logically speaking, happy-for/gloating implies joy, and resentment/pity implies distress. If the specifications of joy and distress are subsets of those of the fortunes-of-others emotions, then joy and distress are generalizations of happy-for, resentment, gloating, and pity and should thus be their parents in the hierarchy.

(7) The structure of emotions as illustrated in figure 1 contains the phrases “focusing on” and “prospects (ir)relevant” several times. We believe this terminology weakens the logical structure, or at least the inheritance view of it. It may indeed be intuitive to say that ‘joy’ is an emotion type where one is “focusing on” a consequence for the self and that ‘happy-for’ is an emotion type where one is “focusing on” a consequence for someone else, but as we have just shown above, ‘happy-for’ logically implies ‘joy’. Moreover, we have also shown in (5) that the connections from hope/fear to the emotion types below them are not purely logical but also temporal. So there are relations between emotion types in the structure that are not captured by presenting it as in figure 1. In the next section, this will be resolved by abandoning the phrases “focusing on” and “prospects (ir)relevant” and adopting a stricter inheritance-based view.

(8) Finally, the branch of attraction-based emotions (i.e., liking/disliking and love/hate) looks a bit awkward to us, because there are no conditions to distinguish love/hate from its generalization liking/disliking. In other words, love/hate does not seem to ex-

tend its parent type, at least not judging from figure 1. However, whereas liking/disliking has appealingness as its only *variable affecting intensity*, love/hate extends this by adding the variable *familiarity*. According to OCC, the more familiar one is with an appealing object, the more one will love it, and the more familiar one is with an unappealing object, the more one will hate it. Interestingly, OCC have chosen not to differentiate based on familiarity. In contrast, below approving/disapproving, the variable *strength of cognitive unit* is used to differentiate between pride/shame on the one hand (i.e., the acting agent is in a cognitive unit with the experiencing agent) and admiration/reproach on the other hand (i.e., the acting agent is distinct from the experiencing agent). Likewise, below pleased/displeased, the variable *likelihood* is used to differentiate between hope/fear on the one hand (i.e., an event is possible but not certain) and joy/distress on the other hand (i.e., an event has actually happened). Analogously, one could use the variable *familiarity* to differentiate between love/hate on the one hand (i.e., for familiar objects) and interest/disgust on the other hand (i.e., for unfamiliar objects).

4 The OCC Model Revisited

The structure of emotions as proposed by OCC and illustrated in figure 1 may be intuitive, but it falls short of capturing the logical structure underlying the OCC model. Taking the points of the previous section into consideration, we have constructed a new hierarchy, as shown in figure 2. It may not be as beautiful as the original figure; e.g., the naming of groups of emotion types ("attraction," "attribution," etc.) has been lost. Although it does not contain anything really new (we are not psychologists, after all), the structure has been altered in several ways to bring it more in line with our computer scientists' inheritance-based way of thinking. Below we elaborate on the differences.

(1) Figure 2 represents the inheritance structure explicitly, because it has labels at every point in the hierarchy, and the conditions of every child node are a superset of those of its parent node(s).

(2) Along each connection are written (in small caps) the additional condition(s) that make an emotion type a specialization of its parent type(s). Ambiguous terms are avoided; for example, we have omitted or replaced the phrases "focusing on" and "prospects (ir)relevant." In distinguishing pride/shame from admiration/reproach, the action in question must have been performed by either the self or another agent, so the phrase "focusing on" is redundant. In distinguishing hope/fear from joy/distress, we consider the consequence in question to be perceived as either prospective or actual. In contrast, the phrase "prospects irrelevant," as used in figure 1, suggests that there is no differentiation on prospects for joy/distress, which would mean that joy/distress is a generalization of hope/fear, which would contradict figure 1 because joy/distress is depicted next to hope/fear instead of above it. To avoid such ambiguities we have strived to use clearer terms.

(3) Satisfaction, fears-confirmed, relief, and disappointment have been moved from under hope/fear to become specializations of joy/distress. This is because we regard a confirmation or disconfirmation to be an *actual* consequence (of an event). According to OCC, these four emotion types require an "attendant" hope or fear emotion; however, we interpret this as meaning that one emotion requires the *presence* of another emotion

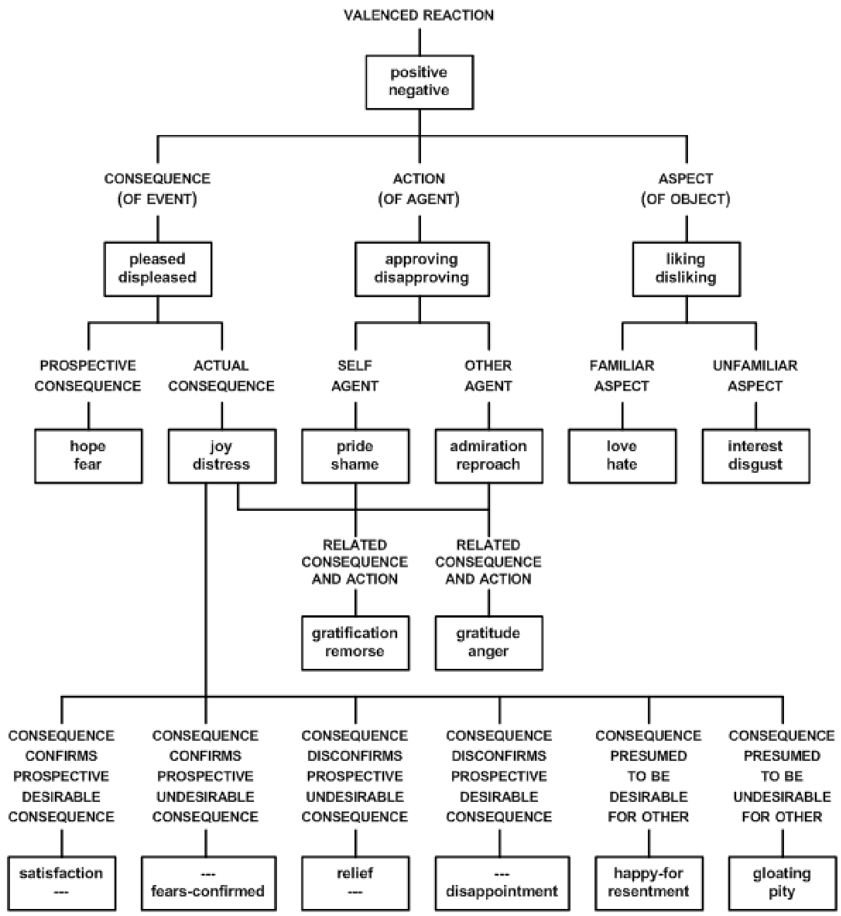


Fig. 2. A disambiguated, inheritance-based hierarchy of emotions of the OCC model.

to be elicited (e.g., relief requires that something is or was feared), not that one emotion is a specialization of another emotion (e.g., relief is not a special type of fear). The conditions that we have put on the connections in figure 2 capture this interpretation as follows: the phrase "prospective desirable consequence" matches the type specification for hope, and "prospective undesirable consequence" matches with fear. Thus the conditions of satisfaction, fears-confirmed, relief, and disappointment require the existence of an attendant hope or fear emotion, in line with OCC's specifications.

(4) Happy-for, resentment, gloating, and pity have been moved from the left of figure 1 to the bottom right of figure 2. This is due to the reasoning of point (6) in the previous section.

(5) In line with point (8) in the previous section, we have added interest/digust as an additional (i.e., besides love/hate) specialization of liking/disliking, based on the familiarity with the object in question.

(6) One thing that we find particularly attractive about figure 2 is that type specifications now follow *immediately* from the diagram. Descriptions can easily be formed by

positive and *negative*	are valenced reactions (to "something")
pleased	is being *positive* about a consequence (of an event)
displeased	is being *negative* about a consequence (of an event)
hope	is being *pleased* about a prospective consequence (of an event)
fear	is being *displeased* about a prospective consequence (of an event)
joy	is being *pleased* about an actual consequence (of an event)
distress	is being *displeased* about an actual consequence (of an event)
satisfaction	is *joy* about the confirmation of a prospective desirable consequence
fears-confirmed	is *distress* about the confirmation of a prospective undesirable consequence
relief	is *joy* about the disconfirmation of a prospective undesirable consequence
disappointment	is *distress* about the disconfirmation of a prospective desirable consequence
happy-for	is *joy* about a consequence (of an event) presumed to be desirable for someone else
resentment	is *distress* about a consequence (of an event) presumed to be desirable for someone else
gloating	is *joy* about a consequence (of an event) presumed to be undesirable for someone else
pity	is *distress* about a consequence (of an event) presumed to be undesirable for someone else
approving	is being *positive* about an action (of an agent)
disapproving	is being *negative* about an action (of an agent)
pride	is *approving* of one's own action
shame	is *disapproving* of one's own action
admiration	is *approving* of someone else's action
reproach	is *disapproving* of someone else's action
gratification	is *pride* about an action and *joy* about a related consequence
remorse	is *shame* about an action and *distress* about a related consequence
gratitude	is *admiration* about an action and *joy* about a related consequence
anger	is *reproach* about an action and *distress* about a related consequence
liking	is being *positive* about an aspect (of an object)
disliking	is being *negative* about an aspect (of an object)
love	is *liking* a familiar aspect (of an object)
hate	is *disliking* a familiar aspect (of an object)
interest	is *liking* an unfamiliar aspect (of an object)
disgust	is *disliking* an unfamiliar aspect (of an object)

Table 2. These emotion type specifications correspond directly to figure 2.

following any link from child to parent node, inserting the text on the link. The resulting type specifications are displayed in table 2 (*cf.* table 1).

5 Conclusion and Future Work

In this paper we have identified a number of ambiguities in the logical structure underlying the popular OCC model of emotions [1]. If these ambiguities are not resolved, computer scientists wishing to formalize or implement emotions may come up with conflicting interpretations of the psychological model. Therefore we have proposed a new inheritance-based view of the structure of emotions, which addresses the ambiguities that we have identified. This effort is part of our goal to formalize emotions in logic in a way which is both truthful to the OCC model and useful for Artificial Intelligence.

References

1. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, UK (1988)
2. Gratch, J., Marsella, S.: A domain-independent framework for modeling emotions. Journal of Cognitive Systems Research **5**(4) (2004) 269–306
3. Steunebrink, B.R., Dastani, M., Meyer, J.-J.Ch.: A logic of emotions for intelligent agents. In: Proceedings of the 22nd Conference on Artificial Intelligence (AAAI'07), AAAI Press (2007)
4. Adam, C.: The Emotions: From Psychological Theories to Logical Formalization and Implementation in a BDI Agent. PhD thesis, Institut National Polytechnique de Toulouse (2007)

For Androids, Emotions are just in Time.
Redefining the Need for Mentality, Process Structures and Situation Representation in order to develop real interactive Possibilities for Androids

Wolfgang Gessner

gessner@collegium.ethz.ch, wolfgang.gessner@fhnw.ch

Abstract. Emotions as primarily intersubjective processes have to be analyzed along the dimension in which they unfold: Time. So process philosophy will be the best fitting approach to get a grasp of the development of these special mental events and their individual and social functions. The analysis and construction of algorithms proves to be the appropriate methodology to describe the prerequisites for the elicitation of a single emotion, as well as its dependent modular dimensions (mimics, gestures, action, speech). Already single emotions (and the more sequences of different emotions or emotions mutually reacting on the other's emotion) are time-dependent series of mental events, which are, as processes, embedded in changing situations (i.e., subjectively represented objective events). The reconstruction of human emotions along these lines will open new perspectives for the construction of android emotions which more than just imitate the surface structures of stable pictures of human emotions. Robot emotions have to be appropriate to the generation of human emotions in order to achieve a parallelized functionality which can give rise to the possibility of interaction, communication and cooperation between androids and men. So the task will be twofold: (i) construction of an android mental system for representing situations and coping with them by emotions, and (ii) testing these new abilities in model worlds which are manageable and calculable in order to solve the 'hard problem' of situation representation for androids.

Keywords: Situation Representation, Emotion Elicitation, Cognitive Emotions, Subjectivity and Intersubjectivity, Functions of Emotion, Model Worlds, Process Philosophy.

Introduction

The attempts to provide a scientifically appropriate and intellectually satisfying understanding of emotions in different disciplines such as philosophy, psychology, (cognitive) ethology, biology and, *lbnl*, neurosciences are copious. Criss-crossing to these disciplinary (and methodological) varieties are the differences and discrepancies which discriminate between the approaches of single authors or, if established, mainstream schools of thought. Some scientists trying to give an overview have spoken of the number of emotion theories as numerous as the number of authors contributing to these fields, or even greater, because some authors have changed their minds during their scientific biography, leaving more than one theory on this battlefield of never ending controversies (for an exemplary insight see (among many others) Strongman's descriptions of these varieties, especially in the different editions of his work *The Psychology of Emotions* (1973, 1980, 1987, 1996) and Solomon (2004). My own theory [1]) has contributed to this plethora of theories and approaches by giving a radically cognitive approach to emotions (Gessner 2004, Gessner et al. 2007, Gessner & Schiewer 2008a and 2008b, Gessner, Schiewer and Ringenbach 2009 a), which nevertheless has some crosslinks and interconnections with the well known 'appraisal theories' (Scherer et al. 2001), with the OCC-model (Ortony, Clore & Collins 1990) and with Solomon's 'judgement theory' of emotions (1988, 1993, 2004). At the same time we have tried to clarify the battlefield by proposing definite distinctions of the phenomena at stake, i.e. the varieties of speaking about 'feelings' of different sorts in different domains (Gessner 2007). In our opinion this typological discrimination and determination of phenomena will give rise to a clarification of the explanation claims of several theories in relation to this typology of phenomena. Conversely, emotions as multi-modal phenomena will make it difficult to give clear and selective differentiations without accepting interferences and overlaps.

1. Emotions: subjective or intersubjective Phenomena?

In our original approach we have analyzed emotions primarily as singular, person-centered events which at the same time have a schematic character and a dynamic component of '(de-)activation', connected with typical dependent dimensions, namely emotional expression, gesture and posture, action latencies and the triggering of certain speech-acts respectively. These 'dependent' dimensions were based on (1) the individual's situation representation, thereby distinguishing between *focal*, actuality-centered perceptions and *topical*, already given thoughts and attitudes [2]) and (2) a certain set of 'elicitation schemes' for definite emotions which determine the internal states of a generalized person in connection with an abstract description of external states of the objective world as being defined as a 'situation' by this person. So we imagined a certain 'reaction-cycle', in which, based on the person's situation representation and its stored emotion elicitation schemes, a pattern-matching process of these two 'input-variables' would generate the four dependent dimensions as 'outputs', as different but stable structures typically connected with each single emotion.

Meanwhile we have learned that emotions have to be seen as an intersubjective process, in developmental perspective (Holodynski 2005) and in their actual unfolding as well. There is no quasi-autistic 'Robinson Crusoe' on his solitary island

[1] This work on theory of emotion and Human-Robot Interaction HRI ('Emotional Agents') has been funded by Haslerstiftung (Berne), grant-nr. MMI 1978.

[2] The necessity of conceiving the *subjective perception of* situations (and not objective situations or their features whatsoever) to build the basis for triggering emotions – in a line with the so-called THOMAS-theorem – we have argued for in Gessner 2004, ch. 4.1.

interacting with himself, perhaps looking to his emotional expressions with the aid of a mirror to learn about his inner states which are expressed by his mimical movements. Instead, there must be at least one second person observing these mimics in order to make sense of this game, maybe his mute servant 'Man Friday', but better still a second person actually able to have emotions itself and being able to express them, too. So let us take this example of two men (R and F) on their isolated island, both reacting to changing, more or less threatening situations (S), i.e. that the proverbial saber-toothed tiger is out of sight, comes closer or is ready to attack. Let us assume further that both men are, at least for the moment. not (or not yet) able to react or to speak, their situation-dependent emotions just expressing themselves in emotional mimics EM). Then there are certain fixed dimensions in which the decision which 'script' [3]) should be activated will be organized:

1st (*intra*subjective) level for Robinson:

- Perception of the situation S by R
- Interpretation of his perception of the situation S by R
- Perception of F's emotional mimics by R
- Interpretation of his perception of F's emotional mimics by R

1st (*intra*subjective) level for Friday:

- Perception of the situation S by F
- Interpretation of his perception of the situation S by F
- Perception of R's emotional mimics by F
- Interpretation of his perception of R's emotional mimics by F

In this first time-segment the buildup of a first assessment of (a) the situation itself and the assessment of the situation as it is mirrored by the emotional mimics of the respective other one will occur. Of course, these assessments can be different: If F already has noticed the approaching tiger, while R has not, then the expression of fear in the face of F (when perceived by R) will give rise for R to check the objective situation anew, maybe with the consequence of now altering his own assessment of S. In the following time-segments this first mutual assessments will iterate in more or less broad intervals of time, depending on the (eventually changing) estimation of the basic probability of tigers being around. In any case, after some iterations there will be a sort of implicit or even explicit understanding between R and F about the risk and peril of the situation given at the moment of the last iteration, which results in an implicit or explicit decision to act in a certain way: Fleeing the predator, if both faces are signalling the disability to cope with the situation or otherwise to stay together against the tiger by struggling against the beast and hoping to win or at least being able to dispel it. As a result, an intersubjective state between R and F will arise:

---

[3] A general description of rule-based action-scenarios has been given by Schank & Abelson 1977 in their classical development of time-structured descriptions of social interactive practices and processes.

2nd (*inter*subjective) level for both Robinson and Friday:

- Decision to flee or to fight, according to prior iterations of R's and F's perceptions of the eventually changing objective situation and of their assessment of their mutual perceptions of their emotional mimics, mirroring the (stable or changing) assessments of the objective situation by the respective other one.

Despite the relative simplicity of this situation and its variants, it will be easy to imagine numerous variable sequences of interaction between R and F, more or less prestructured as cooperative scripts, in order to be able to cope with the different variants of this situation(s), as seen concordantly or differently by both men. Also it will be plausible to imagine that the emotional mimics of these persons gives rise and hints to their respective assumptions about the mental states of the respective other person. In sum, a process of cooperation between both persons (if necessary) will be enormously facilitated by this twofold interaction between their mentally steered displays – emotional mimics – in coping cooperatively with a certain situation [4]). So we have identified emotional mimics as a functional entity for organizing elementary social processes, as a means for coordination or confrontation, but intersubjective in nature in both of these cases.

2. Emotions as unfolding in Time in an algorithmic Structure of Sequences

Emotions as primarily intersubjective processes have to be analyzed along the dimension in which they unfold: Time. An algorithm – in its most general sense to be adopted here [5]) – is a sequence of normatively steered steps to be executed over time, beginning in a certain starting state and finally terminating in an end-state, usually determined by some criterion of solving a determinate problem. In consequence, the first challenge for theorizing about emotions proper is to reconstruct the algorithmic structures along which the relevant emotions will be developed step-by-step in the temporal dimension, guided by rules, which are steering this rule-based unfolding over time dependant on situational inputs. Rules as normative structures [6]) of defining

---

[4] This possibility of interaction is not limited to (potentially) cooperative or collaborative interactions, it will also be useful to optimize one's own aims and interests in (potentially) competitive or confrontative situations as well – or the possibility of displaying emotional mimics will contribute to the decision about a continuation of this situation as a cooperative or as a confrontative one.

[5] Different conceptions of 'algorithm', which show some sort of family resemblance, accentuate different features as recursion and iteration, logical structure, deterministic or probabilistic structures, different forms of exactness and variance, and so on. All of these features will eventually be useful in the reconstruction-task of time-structures in (series of) emotions.

[6] The normativity of these rule-structures makes it impossible to reduce their essence to some sort of material entities or structures (brain structures, for example) but conversely opens the way to implement emotional structures to non-human hardware, i.e. materially realized program structures in android 'brains'.

steps to be taken under the condition of defined circumstances are essentially time-structuring entities, at least by the if-then-structure of execution they *describe* and *prescribe* at the same time. So our task as emotion theorists will be reconstruction, giving, at first, an idiographic analysis of single cases (dt. *Einzelfallanalyse*), which later on should be generalized to fixed 'scripts' in an algorithmic form, each one being in its typicality a general description of a certain emotion. The nomological base of a certain emotion type will then be no more and no less than the definition of an adequate algorithm. To be able to 'emote' adequately will mean to possess the competence to execute this algorithm, comparable to constructing a certain type of sentence, i.e. an interrogative, declarative or imperative sentence. Comparing the concrete, situation-based execution of emotional algorithms to the competence of speaking a language (as a cultural phenomenon) will be the right kind [7]) of perception, description and reconstruction of this type of mental phenomena.

3. Emotion in Time: the Contribution of Process Philosophy

We have thus tried to demonstrate that the analysis and construction of algorithms as rule-based systems will be the appropriate methodology to describe the prerequisites for the elicitation of a single emotion as well as its dependent modular dimensions (mimics, gestures, action, speech). This will be a *temporally structured* analysis of *processes*, which finds its counterpart in the pivotal ontological principles of 'process philosophy', as they are (among others) defined by Rescher (1996):

- That time and change are among the principal categories of metaphysical understanding
- That process is a principal category of ontological description
- That processes – and the force, energy, and power that they make manifest – are more fundamental, or at any rate not less fundamental, than things for the purposes of ontological theory (…)
- That contingency, emergence, novelty and creativity are among the fundamental categories of metaphysical understanding (op. cit., 31).

This conception is in best congruence with the view of emotions as a certain functional means of coping with (ever changing) situations and as systematically structured schemes for handling typologically definable situations [8]). The mainstream

[7] Consequently, all ‚dimensional' theories of emotions, and all approaches using statistical methods will generate (at best) only indirect contributions to this appropriate scientific approach to emotions, which is inevitable connected with the normative and time-structured view of defining algorithms, *which reflects the time-structured course of emotion-generating mental processes themselves*.

[8] Situations of 'fear' or of 'anger' are, in a certain abstract sense, identical despite all the differences they may have concerning the concrete content of these situations. Emotions are reactions to 'standard situations' which can and must be described in an abstract manner in order to avoid the 'misplaced concreteness' of direct descriptions of the situations at stake,

conception of emotions as (passively experienced) *passions* is abandoned in this view in favor of viewing them as a certain means of coping and as a steering instrument as well, embedded in the processes of interaction between individual persons which organize their (cooperative or competitive, singular or collaborative) plans and actions in everyday life or in some sort of professional activity. To deal adequately with these *dynamic, time-structured* interactions makes it necessary to center the theorist's attention on the categories of interactive relatedness, wholeness of sequences, functional units worked out in process structures of given tasks and their solutions (problem-solving), and in general to the activity, fluidity and evanescence which characterizes the course and flow of processes [9]). So process philosophy with its emphasis on the inescapable temporal dimension of processes in general and especially of interactive processes will be the best fitting methodological background approach to understand the development and execution of emotions as very special mental events and their individual and social functions.

4. Robot Emotions: Necessity of Similarity with human Emotions

The constructors of androids have learnt that robots in order to be able to get along with the artificial world of infrastructures surrounding us must have bodies similar to our own bodies, simply because this artificial world has been built in order to fit with the possibilities and restrictions of our own bodies. In parallel, being similar to the mind of humans is a precondition for androids in being accepted as natural interactors with humans, too. Future robot constructors, aiming at the future androids' competencies for communicative and cooperative interactions with us, must learn that androids have to develop mental systems which are (at least) similar to the mental competencies we have acquired during the long way of cultural evolution. This must include the ability to build, to store and to communicate attitudes, which can be expressed as propositional attitudes as a common core in different languages [10]). Based on this, at least the 'higher', language-dependent cognitive emotions will function and are explicable, understandable in principle and explainable when appearing in special, appropriate situations as well.

In (re-)constructing robot emotions for androids (in their future roles as service roboters or in any 'social role' whatsoever), these maxims has to be maintained. Androids' emotions have to be appropriate to the generation of human emotions in order to achieve a parallelized functionality which can give rise to the possibility of interaction, communication and cooperation between androids and men. Starting with our first circumscription of 45 human emotions (Gessner 2004) we have chosen 24

which is shorthand for an understanding of the *principal nature* or *essential functional mechanism* of the respective emotion (Cf. Gessner 2004, ch. 7).

[9] It can be left open for the moment if the contemporary conception of physical causality which has invaded psychology is sufficient for an adequate comprehension of psychic processes, or if a recurrence to purposive and/or teleological forms of explanation would be the better choice.

[10] Cf. Gessner 2004, Gessner et al. 2009 for a differentiated exposition and justification of this view, which also develops the possibilities which are given by this capable and highly productive instrument.

emotions seemingly being appropriate and useful for androids. Four criteria have been developed in guiding this selection:

1. Communicative functionality (→ Aim: efficiency of signalization)
2. Potential contribution to coordinate (→ Aim: successful steering of cooperation partners (CP) or competitors and opponents respectively (CO) [11]))
3. Believability of the emotion considering the android's status (→ Aim: Acceptance by cooperation partner or competitor)
4. Coherence and functional closure of the android's emotional system (→ Aim: Compatibility of the single emotions as an effective system of regulative abilities)

The following list of six emotions (standing substitutionally) gives some short justification for their functional adequacy:

ANNOYANCE / IRRITATION

(for CP): Makes it clear to the other one (eventually to the cooperation-partner himself too) that he has behaved wrongly, which has damaged interests of the cooperating person, or carries out an unwanted action (according to the cooperation aims) or is just starting it.

(for CO): Makes it clear to competitor that one has disapproved of one of his current actions and is asking him for its interruption or repurchase.

SUSPICION

(for CP): Shows communicatively to CP that one is not sure of the putative intentions of a third person or suspects a forthcoming deliberate injury of interests of one's own. This can draw a check of the intentions of this person or propose to conduct such a checking.

(for CO): Shows to the CO (or a third, watching person) that he has detected conditionally a 'suspicious' state or event or that one is prepared that such an event can enter. In addition, sanction readiness can be signalled to the potentially injuring party.

JOY

(for CP): Can signal the positive course of action or interaction and target achievement and is able to motivate perhaps also other persons involved by communicating a 'feeling of success' to them.

(for CO): Shows to the CO that one feels satisfaction about a damage, failure or loss of CO. Paradigmatic for the special case of joy which is called 'malicious joy'.

FEAR

(for CP): Can draw one's attention to a threatening negative change in the given or a future situation which the interaction partner perhaps has not noticed. Will signal the necessity of 'coping' with this danger to the CP as a possible mastering of this danger.

(for CO): Can signal to the CO that one is in an inferior position, that one cannot fend off an attack or that one checks the situation as not controllable by oneself.

---

[11] ‚Competitors' or ‚opponents' encompasses persons which stand in an antagonistic relation with the android, ranging from different interests to situations of conflict or even hostility – in equivalence to ‚negative' relations between humans.

SORROW / WORRY

(for CP): Can point out the possibility of an immediately impending or in the long run unfolding negative development to the CP or other persons and steer his or their attention at this relevant development. The necessity of providing coping potential and of the further observation of the situation is signalled.

(for CO): Can signal to the CO that one assesses the present condition and/or the further course of the situation rather negative.

SURPRISE

(for CP): Signals an unexpected situation attentively and transfers the necessity of reorientation to the CP in this situation with respect to his assumptions and his own scales of values newly (and perhaps also newly defines the task-formulation for the CP).

(for CO): Ditto, and furthermore the signalisation of a disappointment of expectation and/or at first the necessity of newly orientating oneself about the situation or newly interpreting it.

It is easy to imagine that emotions of this sort should not be conceived as monological mental events (at least not primarily), but are highly appropriate and effective in coordination and cooperation with partners and/or providing better chances of one's own interests against competitors of all sort. In any case, the reconstruction of human emotions along these lines will open new perspectives for the construction of android emotions which will more than just imitate the surface structures of human emotions.

5. Building Model Worlds for Human-Robot-Interaction

We have shown that the first task to build robot emotions will be the construction of an android mental system to represent situations and to cope with them by appropriate emotions. The best way of testing these new abilities will be the construction of model worlds which are manageable and calculable in order to solve the 'hard problem' of situation representation for androids.

In order to achieve this goal, we have developed a model world of chess (MWC), enacting a game situation between a human player and an android, which uses the capacities of a chess computer (Levy & Newborn 1991) as his intellectual capacities, but will show the emotional interaction competencies we want him to have in order to be a believable partner in this model world (Gessner 2009 b). The background for this model are the (internal and boundary) rules of chess, which at the same time define the 'meaning' of certain actions (Habermas 2009). So both partners will interact in different (normal and irregular) ways in just the same way as two real human players would do, admitting to the human partner to make (intentional or nonintentional) faults, breaking some rules, trying to deceive or in general dismiss of the normative requirements given in situations of such kind.

It will be easy to accept, in spite of the fact it cannot be argued for here, that the selection of emotions given above is easily transferable to situation types like MWC as well. Presumably the greatest problem there will be the identification of 'cues', indicators, features and hallmarks in these MWC-situations which are unambiguously allocatable to the conditions which will be responsible for the triggering of definite emotions. However the relative clarity of the situation chosen as a model world will facilitate this task. Providing this as a starting point, the android's ‚emotional world' has at least two sources of situational variance:

- The 'score' of the game itself as a result of the moves of both players up to now, which is evaluated by the chess computer and given as a feedback to the android, resulting in meta-information in form of appraisals as 'things are going well' or 'the game threats being lost'.

- The actual state of the normatively determined boundary conditions, which can be violated or not, i.e. by the moving back of a figure already beaten during the game, transgression of time or the breaking of other rules defining an orderly game of chess.

These sources will be supplemented by a third (and perhaps most valuable) source, namely the perception of the actions and of the emotional mimics of the human player by the android. He has his moves in the game, which have already been checked and processed by the chess computer, but moreover he has his irregular moves eventually violating the rules of chess or some boundary rules of the game. This will be accompanied by his recognition of the emotional mimics of his partner (and maybe some of his gestures, too), which can also be 'read out' and interpreted by appropriate pattern recognition systems. Given all these sources of information, the third source will generate the relevant surplus of information which will make it possible for the android (at least in a preliminary sense) to constitute elementary forms of intersubjectivity with his human opponent. In interpreting the actual mental states of the human out of his emotional mimics, the android will be able to comment on this 'information' by his own emotional mimics and, finally, a process of mutual reference on the situation itself and on his human partner's reactions will arise – an iterative process of 'understanding' and coordination of the same kind we have circumscribed in the 'Robinson/Friday-example given above.

Based on these three sources of information, the elicitation of emotions in the android can be based on the same pattern-matching process of situation representations and elicitation schemes which give rise to the triggering of emotions in the human itself. As a result, we can generate the ‚parallelism' or equivalence in the elicititation of emotion in both human and android which will – at least in the long run – make androids into believable partners in communicative, cooperative or competitive interactions with humans. Real intersubjectivity, defined as mutual reference between two actors regarding changing situations and to their (more or less congruent) situation interpretation as well, would be accomplishable.

6. Conclusion: Results of building Model Worlds

The whole development of the game situation must be admitted as *in general realitstic* because the real and potential variability of possible progressions of situations is enclosed in this MWC. Moreover it is *differentially realistic* insofar as players with different competencies in playing chess (and in betrayals …) compete with an android, whose chess brain can be adapted to different playing abilities, which will lead to an ever different course of events in any single game – just as in games between two humans.

In summarizing the main ideas of emotion generation in this MWC it can be ascertained that there is reference to the changing game situations on both sides, on the process of the game itself and its normative backing as well as its evaluation, and, *lbnl*, mutual reference on the players themselves, which are also subject to implicit and explicit evaluations. The whole interactive process generates continuously new events, which are at the same time subject to coping processes by emotions on both sides: the human and the robotic partner. So the whole process generates 'lifelike' situations with a high amount of adequacy in bringing about believable and comprehensible android emotions, well adapted to the situation and to the expectations of its human partner as well.

Defined in advance, the aim of the construction of MWC has been to create the 'elbow room' for a variety of emotional reactions of the android despite the rather restricted model surroundings. This aim seems attainable due to the following points:

- The subjective structures of the android are built up from the knowledge of objective situation elements and are converted in a rule-controlled generation process to his emotional reactions.

- The subjective structures of the human player including his facial expressions (as being a subset of the relevant subjectivity at work) are fed in the processing of the complete situation in MWC by reconnaissance of his mimics.

- All dimensions of a real situation are included variably in MWC only by epistemic or knowledge conditions, but also the abilities of all the protagonists, their intentions, their results and preferences, by mirroring them against the background of a normative structuring of the complete situation which is given by the validity of norms and rules.

- A motive for winning and various order motives which shall safeguard the regulated attainment of the aims of the game are not to insinuate only at the human teammate's side, but are also presumed as a kind of motivation structure for the android.

- All relevant elements of MWC are included completely and are made accessible in a symbolisation for further information processing. This 'operationalisation' of the model's states and surroundings is accomplished by its reformulation in propositional attitudes (as language of the mind, LOM).

- Taken together, subjectivity is realized by what will develop in a removable minimal form of realized intersubjectivity in the interplay of the references taken by a human teammate and the android.

The created situation is thus immune to the objection to be 'barely a simulation' of a real game-process. A high degree of reality is given to this interaction both to the human team mate and to the android, which both realize parallel structures of situation perception, situation hermeneutics and the interpretation of this model world by their complementary and parallel emotional reactions, which also 'work' according to the same complex algorithms. So the operating android should be able to pass the Turing-Test at least from the perspective of observers of the complete situation.

Looking deeper into future, the next steps would be the successive integration of further, qualitatively different model worlds into the android's mental frame and its behavioral scope and latitude. This could lead to a step-by-step growing enhancement of these elementary worlds, and will eventually finish in the android's asymptotic approach to the breadth and depth of real people's view of the world, which equally had to be developed by humans in a long-range process of phylogenetic cultural development and an ever renewed ontogenetic development of each individual within its life-span as well.

References

1. Gessner, W., G. L. Schiewer & A. Ringenbach: Why Androids Will Have Emotions: Constructing Human-Like Actors and Communicators Based on Exact Sciences of the Mind. In: Lalanne, D. & Kohlas, J. (Eds.): Human Machine Interaction. Research Results of the MMI Program. Berlin/Heidelberg: Springer (Lecture Notes in Computer Science, Subseries: Programming and Software Engineering, Vol. 5440), 133-163 (2009 a)
2. Gessner, W.: Skizze zu einer Modellumgebung ‚Schach' als intersubjektiver Situation zwischen einem Androiden und einem menschlichen Mitspieler (Revidierte Fassung). Zürich: Collegium Helveticum (Unpublished manuscript) (2009 b)
3. Gessner, W. & G. L. Schiewer: Identifying the common Ground of Emotion and Speech Acts in View of Human-Robot Interaction (HRI). In: Reichardt, D. (Ed.): Proceedings of the 3rd Workshop on Emotion and Computing - Current Research and Future Impact', KI 2008, TU Kaiserslautern (2008 a)
4. Gessner, W. & G. L. Schiewer: Prospects for Subjectivity and Personhood in Androids [Regular Paper]. In: Proceedings of IEEE RO-MAN 2008, 'From Personal Robots … towards Robots with Personalities?' The 17th International Symposium on Robot and Human Interactive Communication, TUM München (2008 b)
5. Gessner, W.: Talking about ‚Feelings'. A Solution by analytical Division of the Phenomena. Contributions from the Haslerstiftung Project Nr. MMI 1978, Working Paper nr. 31 (2007)
6. Gessner, W., G. L. Schiewer, G. Mazzola & A. Ringenbach: Constructing Androids as Emotional Agents in Robot-Human Relationships. In: Reichardt, D. & P.Levi (Eds.): Proceedings of the 2nd Workshop on Emotion and Computing - Current Research and Future Impact', KI 2007, Universität Osnabrück (2007)
7. Gessner, W.: Die kognitive Emergenz von Emotionen. Paderborn: mentis (2004)
8. Habermas, J.: Sprachtheoretische Grundlegung der Soziologie. Kap. 1: Vorlesungen zu einer sprachtheoretischen Grundlegung der Soziologie, S. 29-156. Frankfurt / M.: Suhrkamp (2009)
9. Holodynski, M.: Development of Emotions and Emotion Regulation. New York: Springer (2005)
10. Levy, D. & M. Newborn: How Computers play Chess. New York: Computer Science Press (1991)
11. Ortony, A., G. L. Clore & A. Collins: The cognitive Structure of Emotions. Cambridge: Cambridge University Press (1990)

12. Rescher, N.: Process Metaphysics: An Introduction to Process Philosophy. Albany: State University of New York Press (1996)
13. Schank, R. C. & R. P. Abelson: Scripts, Plans, Goals and Understanding: An Inquiry into human Knowledge Structures. Hillsdale N. J.: Lawrence Erlbaum (1977)
14. Scherer, K. R. et al.: Appraisal Processes in Emotions: Theory, Methods and Research. New York: Oxford University Press (2001)
15. Solomon, R. C.: On Emotions as Judgements. American Philosophical Quarterly, 25 / 2, repr. in Solomon, R. C. (2003). Not Passion's Slave. Emotions and Choice. Oxford: Oxford University Press, 92-114 (1988)
16. Solomon, R. C.: The Passions. Emotions and the Meaning of Life. Indianapolis: Hackett Publishing Company (1993)
17. Solomon, R. C.: Thinking about Feeling. Contemporary Philosophers of Emotion. Oxford: Oxford University Press (2004)
18. Strongman, K. T.: The Psychology of Emotion. Chichester, London: John Wiley (1973, 1980, 1887, 1996)

Demonstrations

presented at the workshop

Emotion Recognition Using Optical Flow

Michael Girgis[1], Dirk Reichardt[2]

1
[1]German University Cairo (GUC), Egypt
micheal.girgis@student.guc.edu.eg
[2]Duale Hochschule Baden-Württemberg Stuttgart, Germany
reichardt@dhbw-stuttgart.de

Emotion recognition in faces is based either on features or movements of muscles. In our experiment we applied optical flow on video sequences and determined typical muscle movements in order to detect Ekman's Action Units. The classifier for the eye region is trained by a genetic algorithm and results are tested against the Cohn-Kanade Database. The first experiments with eyebrow movement detection are promising.

The experiment is part of a emotion recognition framework which developed at the DHBW Stuttgart. The framework uses the OpenCV library and is implemented in C++. The head detection and tracking as well as the positioning of so called regions of interest is provided by the framework. In our experiment additional ROI-boxes are placed on the eyebrow region. In this region we place elements to track for an optical flow analysis. The region is divided in three vertical sections. The flow direction and strength is then calculated for each section. To compensate head motion, the flow values of reference points outside the eyebrow region are measured to compute relative motion. The classification of action units [2] AU1, AU2 and AU4 is done by a finite state machine. The state transition thresholds are determined by a genetic algorithm which uses 20 labeled samples from the Cohn-Kanade database [1] as a reference for the fitness function. The approach works quite nicely in sequences with low head motion already. Results and life testing can be shown at the workshop.

References

1. Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.
2. Paul Ekman, Wallace V. Friesen: Facial action coding system: A technique for the measurement of facial movement. Palo Alto, Calif.: Consulting Psychologists Press, 1978