Proceedings of the 2nd Workshop

Emotion and Computing – Current Research and Future Impact

Dirk Reichardt, Paul Levi (Editors)

Osnabrück, Germany
September 10th, 2007

ISSN 1865-6374

Herausgeber:

Prof. Dr. D. Reichardt, Berufsakademie Stuttgart
Prof. Dr. P. Levi, Universität Stuttgart

Verlag/Bereitstellung (online):

Berufsakademie Stuttgart
- Staatliche Studienakademie -
Jägerstraße 56
70174 Stuttgart
Germany
http://www.ba-stuttgart.de

Emotion and Computing - current research and future impact

Workshop

The workshop focuses on the role of affect and emotion in computer systems including the three dimensions: emotion recognition, emotion generation and emotion modeling with special attention to AI specific problems and applications. Both shallow and deep models of emotion are in the focus of interest.

In recent years computer science research has shown increasing efforts in the field of software agents which incorporate emotion. The success of the first workshop on emotion and computing held at the KI 2006 in Bremen shows the interest within the scientific community. Several approaches have been made concerning emotion recognition, emotion modeling, 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.

A key question: Which commercially interesting applications would incorporate emotional aspects? One of the first interesting applications are dialogue systems which intend to generate natural human-like language and to react on emotional aspects of the utterances of the human partner adequately. Another field of application with significant AI influence is the field of computer games. An increasing interest in believable and intelligent computer characters can be recognized. As a third application type eLearning depends on the quality of learner and teacher interaction. Moreover, automatically recognizing the user's emotional state can help evaluating new products. In general many applications which integrate a user model often leave emotion and mood out of the adaptation loop. If we manage to cover this aspect, the quality of user adaptive systems can be increased significantly.

This workshop discusses the scientific methods considering their benefit for current and future applications. Especially when regarding the subject of emotion recognition, this also includes ethical aspects. The presented papers discuss theories, architectures and applications which are based upon emotional aspects of computing.

Organization and Scientific Committee

Organization

Prof. Dr. Dirk Reichardt, BA Stuttgart, University of Cooperative Education
http://www.ba-stuttgart.de/~reichard

Prof. Dr. Paul Levi, University of Stuttgart
http://www.ipvs.uni-stuttgart.de/ipvs/bv

Scientific Committee

Dipl.-Ing. Joscha Bach, University of Osnabrück
http://www.cognitive-agents.org

Dipl.-Inform. Christian Becker, University of Bielefeld
http://www.techfak.uni-bielefeld.de/~cbecker/

Prof. Dr. Nicola Henze, University of Hannover
http://www.kbs.uni-hannover.de/~henze/

Dr. Manfred Kerber, University of Birmingham
http://www.cs.bham.ac.uk/~mmk/

Prof. Dr. J.-J.Ch. Meyer, University of Utrecht
http://www.cs.uu.nl/staff/jj.html

Dipl.-Ing. Christian Peter, Fraunhofer-Institut für Graphische Datenverarbeitung, Rostock
http://www.igd-r.fraunhofer.de/private/cpeter

Dr. Götz Renner, DaimlerChrysler AG, Head of Customer Research Center, Berlin
http://www.daimlerchrysler.com

Prof. Dr. Michael M. Richter, University of Calgary
http://people.freenet.de/mmrichter/

Dr. Marc Schröder, DFKI Saarbrücken
http://www.dfki.de/~schroed/

Workshop Program

The workshop provides talks covering a variety of aspects of emotion and computing to set a basis for a subsequent moderated discussion. The presentations are documented by reviewed short research papers.

Session 1:	*Speech – Touch – Vision*
9:30 – 11:00	Welcome and Introduction
	Emotion Recognition in Speech using a Modified Version of the Median-Cut Algorithm
	Human Capabilities on Video-based Facial Expression Recognition
	Touch Perception and Emotional Appraisal for a Virtual Agent
Session 2:	*Empathy – Androids – Games*
11:30 – 13:00	Simulating Empathy for the Virtual Human Max
	Constructing Androids as Emotional Agents in Robot-Human Relationships
	Moods and Shallow Emotions for Balanced Speech Acts in Computer Games
	Interpretation of Intensity Variables for an Emotional Agent in the Public Goods Game
Session 3:	*Models – Applications*
14:30 – 16:00	Towards a Quantitative Model of Emotions for Intelligent Agents
	Using the Emotional Relation of Topics for Text Mining Based Recommendation Systems
	Application of emotion recognition methods in automotive research
Session 4:	*Discussion on Models*
16:30 – 18:00	True Feelings: Functionalist and Descriptionalist Modeling of Emotion
	Moderated Discussion
	Wrap up of Discussions and Workshop Summary

Table of Contents

Emotion Recognition in Speech using a Modified Version of the Median-Cut Algorithm

Emmanuel Lesser, Tim Dams, Maggy Goossens

University College of Antwerp, Dep. of Applied Engineering, Electronics-ICT
Paardenmarkt 92, B-2000 Antwerp, Belgium
lesser.emmanuel@student.ha.be, t.dams@ha.be, m.goossens@ha.be
http://www.ha.be, http://www.e-lab.be, http://www.avai-project.net

Abstract. Working towards natural human-machine interaction, it is essential for communication systems to be able to process speech as well as emotion. We propose a new approach for extracting emotion features out of speech signals. We combine general signal processing techniques with some novel methods on the transformed signal. Furthermore, we introduce two essential emotion recognition algorithms which are fundamental for accurate results. The research is motivated by the AVAI project, in which we aim to develop a set of databases and algorithms as a framework for the development of speech and emotion recognition and AI applications. The system we present is built around spectrograms and a modified median-cut image quantization algorithm. Finally, research results are discussed.

Keywords: speech, emotion recognition, spectrograms, median-cut, AI

1 Introduction

It is generally believed that accurate emotion recognition is easier achievable on multimodal input, since emotion can also be extracted from facial expressions [1]. Based on this proposition, an audio-visual emotion database has been developed by Martin et al. [2] that is freely available for research purposes. While multimodal systems may be more accurate, real-time processing of audio-visual input requires considerably more intensive computing than the processing of speech signals only. Since application-oriented emotion recognition is very likely to be combined with standard speech recognition, it is advisable to limit the necessary computing resources required by the emotion recognition system.

In this paper we introduce a new approach for developing an accurate emotion recognition system that requires only a minimum of computing power. This system uses speech signals as its exclusive input. The processing of the signals and the recognition of emotion is achieved by combining general signal processing techniques with a modified median-cut algorithm.

2 Architecture

It is necessary to gain insight into some specific theoretical aspects that motivate the fundamental build of the system. The benefits of this experimental approach will become clear when describing the various parts that make up the system.

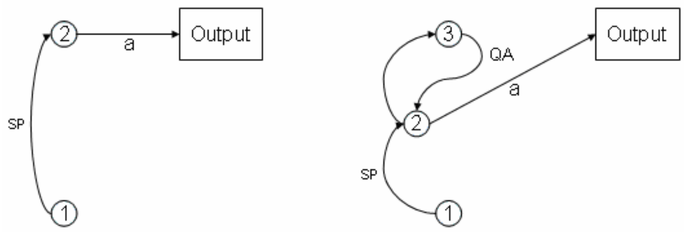
When developing speech and emotion systems, our primary goal is to achieve the best possible imitation of human neural networks. We want the system to act intelligently, responding naturally to the input signals. Consequently, we base the development of AI systems on research results of intelligent human behavior.

Several methods exist for the manual analysis of speech signals [3], [4]. These methods all visualize the signal in some way. This proves that accurate analysis of a signal is substantially easier for humans with the help of a visual aid.

Since we want our system to extract emotion by analyzing the input signals rather than to 'listen' to them, we will try to simulate human behavior by using visualization of the input.

Instead of working exclusively with signal processing algorithms to extract the necessary information from the signals, we introduce an additional step wherein the signal is transformed to the Fourier domain upon which we apply the modified median-cut algorithm for emotion extraction.

Figure 1(a) shows the classic, general process for extracting information out of a signal. In Figure 1(b) we illustrate the basic working of our proposed system. In contrast to the classical system, we extract information out of the input signals, by applying signal processing (SP) algorithms, which is then used in the next step (3) that depicts our novel method. By quantizing (QA) the extracted information in (3) we then return to the process (2).



(a) Classic system

(b) Proposed system

Fig. 1. Information extracting systems.

2.1 Signal Processing

We use spectrograms for the visualization of the input signals. Spectrograms are particularly interesting for emotion recognition since the various energy levels of the signal are represented in the image (related to the signal's amplitude). A Short-Time Fourier Transformation (STFT) is applied on the input signal, generating a spectrogram, followed by the generation of an image from the data that was obtained.

To enable continuous processing of the input, it is imperative that the spectrograms are generated at the same rate as the sampling frequency of the input signal. This can be achieved using a method proposed by Aming & Feng in [5], which essentially suggests to use Recursive Fourier Transforms (RFT's) using multiple data windows. Spectrogram images are temporarily stored to allow further processing.

2.2 Quantization Techniques & Additional Parts of the System

As described later, a modified version of the median-cut algorithm extracts information from the spectrograms and represents the core of the emotion recognition system.

The algorithm counts and sorts pixels from the spectrogram image according to color and color variations. Different colors in the spectrogram indicate different levels of energy in the signal, which in turn indicate different kinds of emotion or differences in gradation of the emotion. This method corresponds to reading a spectrogram (cf. human behavior), which fits the philosophy of this system, as described earlier.

In a final step the extracted information is matched against a reference database, implemented in SQL. No statistical models such as HMMs are necessary to obtain accurate recognition results. In addition to the use of a reference database, some efficient stand-alone recognition algorithms are applied on the output. The development of these algorithms, which allow user-independent recognition, is discussed in section 3.

3 Fundamental Issues Concerning the System and the Database

We discuss some fundamental aspects of the proposed system that are essential to the development of the emotion recognition algorithms.

First we have to consider which set of emotions will be recognized by our system. Two conditions have to be taken into account when stipulating our choice:

- The set of emotions has to match a common pattern of emotions.
- The system must be able to easily distinguish between these emotions.

Existing research on this topic [6] teaches us that sadness, anger and fear are emotions that best meet these requirements. Since positive emotions are equally important as negative ones in the context of our system, we chose to integrate the following set of emotions: sadness, anger, fear, happiness and joy.

3.1 Contents of the Database

We chose to fill the reference database with normalized values, independent of the original input and thus useful for recognition. This process is applied both when creating the reference database and during recognition. Furthermore, words and phrases are used to improve the accuracy of the results.

Since there is no strict definition of a certain emotion, we have to take so-called gradations in emotion (e.g. one can be 'happy' or 'happier') into account. It is practically impossible to record all kinds of gradation in the reference database. This issue has to be addressed in the recognition algorithm(s). We developed an algorithm that attempts to recognize a certain pattern in a signal that an emotion shows between its lowest and highest gradations. Applying this algorithm on the results after matching the signal to the reference database, followed by a second comparison with the database, has proven to generate satisfactory results.

3.2 User-Independent Recognition and voices

These aspects of the emotion recognition system are essential, since our approach does not include neural networks to be trained. In order to maintain a certain scientific standard, we opted not to use speech from random persons, but existing speech fragments instead. We carefully selected a set of 10 speech fragments per emotion, wherein the respective emotion is clearly expressed. Fragments were taken from various sources, such as film, radio and television. Wherever necessary, (background) noise or music was filtered by applying Fast Fourier Transforms (FFT) and/or specific Wavelet transformations on the signals. In addition, one professional voice-over sample was recorded per emotion.

To achieve user-independent recognition, we introduce a second recognition algorithm, which searches for predefined proportions between certain values in the spectrograms. Every specific proportion is attached to a certain emotion. Thus, we attempt to find a consistent sequence of proportions that matches one specific emotion as closely as possible. Naturally, this approach assumes that energy levels in speech are directly related to emotion in the same, permanently consistent way.

4 Modified Median-Cut Algorithm

Originating from the field of image quantization, the median-cut algorithm was first proposed by Heckbert in [7]. Essentially, this algorithm divides the original color space of an image in a certain amount of regions, allowing to work with palletized images. It proves useful for quickly extracting the necessary information out of spectrogram images, as required by our system.

Apart from counting the pixels of certain colors in the spectrogram, we need more information on the proportions between energy levels in the image, as explained in section 3.3. Since the median-cut algorithm counts pixels and calculates median values using color histograms, we can get relevant information for emotion recognition by only applying minor changes to the algorithm.

The algorithm executes the following steps (note that steps 1 through 4 are basically the general median-cut algorithm, except for one nuanced aspect):

1. Find the smallest box that contains all the colors in the image (in very exceptional cases, the whole image might be considered to be the smallest box).
2. Count all pixels that contain a certain primary color.
3. Calculate the median for this color and split the box into two regions at the median.

4. Assign a certain bit pattern to every region.
5. Repeat these steps on the result for all colors, in a predefined order.
6. Analyze the final bit pattern to evaluate the proportions between colors in the image. (This cannot be done by comparing results from step 2, since these pixels might also contain other colors, which could cause substantial inaccuracies, depending on which colors they are.)

The results of step 2 are directly compared to the reference database. The output of step 6 is passed on to the recognition algorithms. The mutual output is then retransferred to the recognition algorithms. As a final step these are matched a second time against the database, resulting in the recognized emotion.

5 EAT, Results & Research

We have developed a software tool named 'Emotion Analysis Tool' (EAT), which uses all our techniques to recognize emotions in speech. The EAT accepts live recordings as well as existing speech files. After going through 3 (highly customizable) steps, the program displays the percentual presence of each emotion in the signal to the user. Figure 2 shows the user interface of the EAT, displaying results for the speech fragment 'sad_female_child.wav'. (Note that the two operation modes of the EAT are 'Adult' and 'Infant'.)

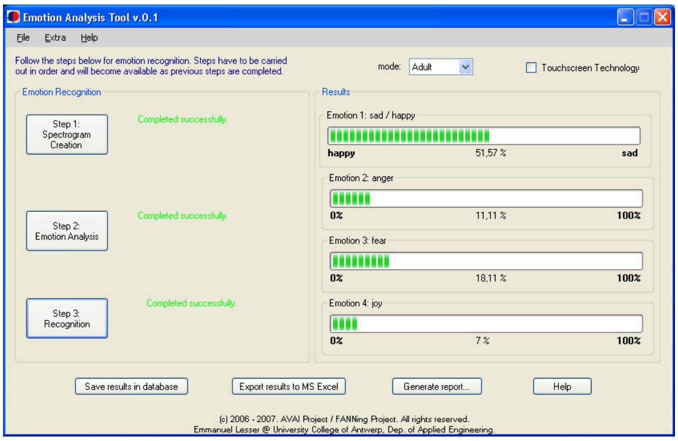


Fig. 2. Screenshot of results in the EAT for speech fragment 'sad_female_child.wav'

This speech fragment was recorded independently and is not part of the same set as the samples in the database.

The EAT has proven to work with an accuracy of about 60% for natural emotion. A substantial increase of database entries could lead to a much higher accuracy, depending on the quality of the samples.

Figure 3 contains four spectrograms originating form Sendlmeier et al.'s research [8]. The relationships between colors in the spectrograms and the respective emotions can clearly be seen, as well as the differences between the various spectrograms and the color proportions. Although in real-life-speech the differences may be more nuanced than those depicted, the use of efficient algorithms has proven to compensate this.

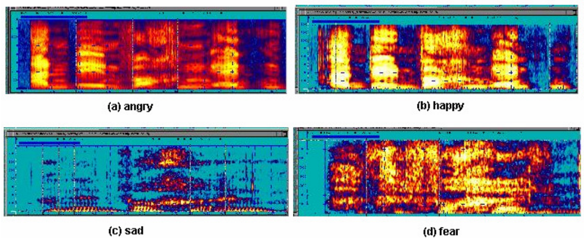


Fig. 3. Spectrograms from speech fragments containing different emotions

6 Conclusion

We have introduced a new approach for extracting emotion features from a speech signal. Although a little unorthodox considering the classical techniques, the proposed architecture achieves relatively accurate results, while using considerably less computing power than other systems. Even though the benefit of our system may not seem directly evident, the use of spectrograms improves the use of computing resources and reduces computing time. Additionally, it provides very substantial benefits for emotion recognition in noisy environments. Furthermore, our approach opens a lot of doors for new research, development and applications.

References

1. Kwang-Sub Byun, Chang-Hyun Park, Kwee-Bo Sim, "Emotion recognition from facial expression using hybrid-feature extraction", SICE, Vol. 3, 2483-2487 (2004)
2. Martin, O., Kotsia, I., Macq, B. and Pitas, I., "The eNTERFACE'05 Audio-Visual Emotion Database", Data Engineering Workshops Proc. (2006)
3. Alfaouri, M., "Digital analysis of signal intensities", TENCON 1996 Proc., Vol. 2 (1996)
4. Vivek, T., "Novel Speech Processing Techniques for Robust Speech Recognition", Swiss Federal Institute of Technology (2006)
5. Aming, M. G. and Kai Di Feng "Short-time Fourier transforms using cascade filter structures", Circuits and Systems II., Vol. 42, 631-641 (1995)
6. Sobin, C. and Alpert M., " Emotion in Speech: The Acoustic Attributes of Fear, Anger, Sadness, and Joy ", Journal of Psycholinguistic Research, Vol. 28, Nr. 4, 347-365 (1999)
7. Heckbert, P., "Color Image Quantization for Frame Buffer Display", SIGGRAPH Proc., Vol. 16, 297-307 (1982)
8. http://database.syntheticspeech.de/

Human Capabilities on Video-based Facial Expression Recognition

Matthias Wimmer[1], Ursula Zucker[2], and Bernd Radig[2]

[1] Faculty of Science and Engineering, Waseda University, Tokyo, Japan
[2] Institut für Informatik, Technische Universität München, Germany

Abstract. A lot of promising computer vision research has been conducted in order to automatically recognize facial expressions during the last decade. Some of them achieve high accuracy, however, it has not yet been investigated how accurately humans accomplish this task, which will introduce a comparable measure. Therefore, we conducted a survey on this issue and this paper evaluates the gathered information regarding the recognition rate and the confusion of facial expressions.

1 Introduction and Motivation

The psychologists Ekman and Friesen investigated social dependencies of facial expressions. They figure out six universal facial expressions [1] that are expressed and interpreted in the same way by humans all over the world, see Figure 1. Furthermore, they introduce the Facial Action Coding System (FACS) in order to precisely describe facial muscle activity [2]. Kanade et al. [3] gather a database (CKDB) of $I = 488$ short image sequences each showing one of the six universal facial expressions. Each sequence shows a neutral face at the beginning and then develops into one of the six universal expressions with peak activity.

Fig. 1. The six universal facial expressions as they occur in [3].

The computational task of facial expression interpretation is usually subdivided into three subordinate challenges, see Pantic et al. [5], detecting the face, extracting facial features, and inferring the facial expression. Michel et al. [4] train a Support Vector Machine (SVM) that determines the facial expression within video sequences of the CKDB by comparing the first frame with the neutral expression to the last frame with the peak expression. Schweiger et al. [6] compute the optical flow of several predefined regions within a human face in order to extract the facial features. Classification is conducted by neural networks.

However, the accuracy of these approaches cannot be stated realistically, because a comparable measure does not exist. Therefore, we conduct a comprehensive survey asking hundreds of persons to determine facial expressions. Similar to computer vision algorithms, humans are only provided visible information. The contribution of this paper is a realistic measure to state the accuracy of algorithms based on the accuracy of human beings recognizing facial expression.

2 Description of the Survey

We questioned a few hundred persons about the facial expressions visible in the CKDB. Note that this database only contains visible information and does not provide further communication channels or context information. The participants were shown randomly selected image sequences and they specified one of the six universal facial expressions or none in case they were not able to decide. Each participant annotated as many image sequences as he or she wanted.

3 Evaluation of the Survey's Results

This section investigates which facial expressions are recognized easily and which are more likely to be confused. We gathered $q = 5413$ annotations specified by $P = 250$ persons[3]. On average, each participant annotated $\frac{q}{P} \approx 22$ image sequences, which results in $\frac{q}{I} \approx 11$ annotations per sequence.

For each sequence of the CKDB, licensed FACS-experts manually specified Action Units, but, unfortunately, they do not directly relate to one of the six universal facial expressions, in most cases. Therefore, we cannot decide whether or not the participants' annotations are correct. This evaluation compares the annotations to one another, instead.

The set $\mathcal{E} = \{\text{happiness}, \text{sadness}, \text{disgust}, \text{fear}, \text{anger}, \text{surprise}, \text{none}\}$ contains the possible annotations. We subdivide the q annotations into the number of

[3] The entire set of annotations is available on request (matthias.wimmer@cs.tum.edu).

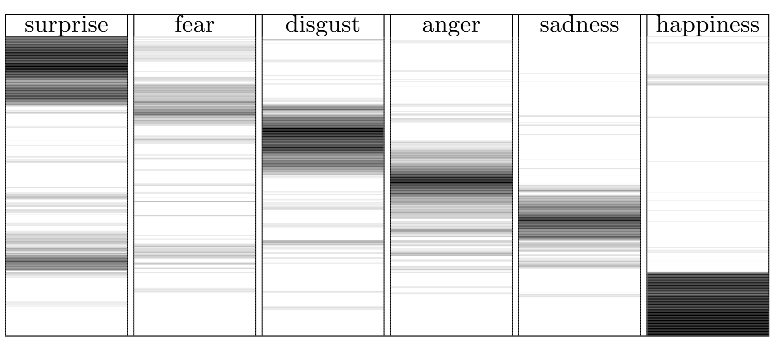


Fig. 2. The annotation rates for each facial expression of each image sequence sorted by similar annotation of the sequences. Darker regions denote a higher annotation rate.

annotations q_i for the image sequences i. Again, we subdivide each q_i into the number of annotations $q_{i,\epsilon}$ for the facial expression ϵ, see Equation 1. The annotation rate $r_{i,\epsilon}$ gives evidence about the number of annotations for the facial expression ϵ compared to the total amount of annotations for the sequence i.

$$q = \sum_{i=1}^{I} q_i, \qquad q_i = \sum_{\epsilon \in \mathcal{E}} q_{i,\epsilon}, \qquad r_{i,\epsilon} = \frac{q_{i,\epsilon}}{q_i} \tag{1}$$

Higher annotation rates indicate that the participants rather chose the same facial expression, which makes the annotation more reliable. Furthermore, $r_{\epsilon,i} \approx 0$ indicates that most participants did not specify ϵ for sequence i. Therefore, this image sequence does probably not show facial expression ϵ. Figure 2 illustrates the annotation rates for all 488 sequences. Every row denotes one image sequence i and indicates the annotation rate $r_{i,\epsilon}$ for all facial expressions $\epsilon \in \mathcal{E}$. We sort the rows of the table such that similarly specified image sequences are in adjacent rows. In this representation, the confusion of facial expressions is clearly visible. Happiness is best distinguished from other facial expressions. Sadness gets little confused with disgust or fear, but gets highly confused with anger or surprise. Fear is the hardest to tell apart.

Figure 3 shows the histograms of the annotation rates for all facial expressions. As mentioned above, well-recognized facial expressions have a lot of occurrences of $r_{i,\epsilon} \approx 0$ and of $r_{i,\epsilon} \approx 1$. Happiness is the most distinctive facial expression, because its histogram shows the most distinctive peaks for $r_{i,\epsilon} = 0$ and for $r_{i,\epsilon} = 1$.

3.1 Confusion Between Facial Expressions

Furthermore, we determine the level of confusion between different facial expressions. We consider two facial expressions to be confused if two different participants annotate the same image sequence with these expressions. $H(\epsilon_1 \wedge \epsilon_2)$ represents the number of sequences that are both annotated as ϵ_1 and as ϵ_2.

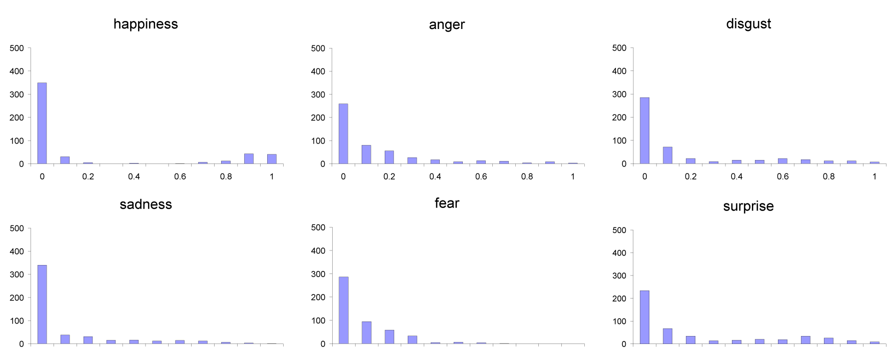


Fig. 3. Distribution of the annotation rate $r_{i,\epsilon}$ for each facial expression ϵ.

$\tau(\epsilon_1, \epsilon_2)$	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	100.0%	42.4%	23.8%	7.3%	43.2%	28.9%
Disgust	42.4%	100.0%	32.6%	6.2%	19.3%	24.6%
Fear	23.8%	32.6%	100.0%	11.0%	15.8%	**43.8%**
Happiness	7.3%	6.2%	11.0%	100.0%	6.7%	14.5%
Sadness	43.2%	19.3%	15.8%	6.7%	100.0%	28.9%
Surprise	28.9%	24.6%	**43.8%**	14.5%	28.9%	100.0%

Table 1. The amount of confusion $\tau(\epsilon_1, \epsilon_2)$ between the six universal facial expressions.

$H(\epsilon_1 \vee \epsilon_2)$ represents the number of sequences that are either annotated as ϵ_1 or as ϵ_2. The quotient $\tau(\epsilon_1, \epsilon_2) = \frac{H(\epsilon_1 \wedge \epsilon_2)}{H(\epsilon_1 \vee \epsilon_2)}$ determines the amount of confusion between two facial expressions ϵ_1, ϵ_2.

Table 1 illustrates the confusion of either pair of facial expressions. The participants confused fear and surprise most often. According to FACS, some Action Units are similar in these two facial expressions. Further expressions, which get easily confused because of some coinciding Action Units are anger and sadness, anger and disgust, fear and disgust, and anger and surprise. People confused least between happiness and disgust, and happiness and sadness.

4 Conclusion

The interpretation of the data gathered by our survey shows that humans are not as good in determining facial expressions as computer vision researchers would expect them to be. These poor recognition rates partly originate from the fact that the CKDB does not contain natural expressions but acted ones. Furthermore, this recording was conducted in a laboratory environment rather than in a real-world scene. In our opinion, the most decisive reason for the poor results is the consideration of video information only. We expect humans to be more accurate being provided further information as well, such as audio information and long-term context information. Therefore, we recommend integrating this information into facial expression interpretation algorithms as well.

References

1. P. Ekman. Universals and cultural differences in facial expressions of emotion. In *Nebraska Symposium on Motivation*, vol 19, pp 207–283, 1972. U of Nebraska Press.
2. P. Ekman. Facial expressions. In T. Dalgleish and M. Power, editors, *Handbook of Cognition and Emotion*, New York, 1999. John Wiley & Sons Ltd.
3. T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Int. Conf. on Automatic Face and Gesture Recognition*, pp 46–53, 2000.
4. P. Michel and R. El Kaliouby. Real time facial expression recognition in video using support vector machines. In *Int. Conf. on Multimodal Interfaces*, 2003.
5. M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE PAMI*, 22(12):pp 1424–1445, 2000.
6. R. Schweiger, P. Bayerl, and H. Neumann. Neural architecture for temporal emotion classification. In *Affective Dialogue Systems, LNAI 3068*, pp 49–52, 2004.

Touch Perception and Emotional Appraisal for a Virtual Agent

Nhung Nguyen, Ipke Wachsmuth, Stefan Kopp

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

Abstract. Building virtual agents that are able to perceive and appraise touch increase the lifelikeness and interaction possibilities in human computer interaction. In this paper we introduce work on how a sense of touch was realized for the virtual human Max and how he can emotionally react to it by appraising the different kinds of tactile sensation.

1 Introduction and Related Work

Virtual agents have found various attention in affective computing. For instance, agents have been built that can express emotions in their faces, e.g. [4] and voices, e.g. [3]. Also physical agents were developed that can perceive touch and are able to respond to tactile sensations in different ways, e.g. [9] and [10]. An interesting research question is how to develop virtual agents that have a 'sense of touch', in that they could perceive touch and emotionally appraise it. The context for our research is the virtual agent Max [6] that has been equipped with a dynamic emotion system, which can respond to various kinds of stimuli from verbal input, goal achievement or failure [2]. In a first experimental system, Max was equipped with a simple sense of touch that could evoke valenced reactions. In this scenario

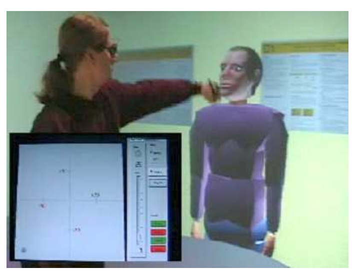


Fig. 1. When touched on his right cheek, Max reacts with negatively valenced emotional impulses.

human users can interact with the graphical representation of Max in a CAVE-like VR environment by means of motion tracking of their hands (Figure 1). The user's hand movements are tracked in order to detect the distance between the hands and the three dimensional geometries forming Max's face. In this first system, touching Max's left cheek gave a positive appraisal and touching Max's right cheek gave a negative appraisal. The stimuli thus elicit emotions that can fortify or alleviate Max's state of mood, which in turn causes Max to display corresponding facial (happy or annoyed) expressions.

In this first setting, the quality of touch could not be differentiated, that is, a touch of the cheek caused an undifferentiated 'all-or-nothing' reaction. In human interaction, however, touch influences emotions in many more subtle ways. Someone gently stroking our arm might evoke happiness in us, while getting beaten puts us immediately in a negative emotional state. On the other hand, touch can also bear a communicative meaning in that someone might want to convey his or her emotions that he or she likes us. Could such a distinguished touch perception and corresponding emotional appraisal also be possible for a virtual agent? In this paper, we describe how touch receptors were developed and technically realized for Max's virtual body. These receptors allow for differentiating between different qualities of tactile stimulation. That way Max can be enabled to extract and emotionally response to the affective content of a tactile stimulation, exceeding the simple all-or-nothing reactions that were possible before.

2 A Sense of Touch for Max

The virtual humanoid agent Max is a situated artificial communicator for modeling and researching communicative behavior in natural face-to-face interactions [7]. Findings from studies on the human tactile systems were incorporated to build an artificial sense of touch for Max, which is conceived not only for virtual but for artificial agents in general. When modeling touch, one important distinction to draw is between active and passive touch [5]. Passive touch is the mere sensation of being touched by some other object, whereas in active touch the sensing individual herself evokes the tactile sensation by actively controlling the stimulation. Here, we focus on the affective content of passive touch. That is, we do not care about the agent's attention, intention or motor control, but can focus on the kind of tactile stimuli passively applied to the agent's body.

In our work on modeling and realizing passive touch for Max's whole body [8], each tactile stimulation is associated with three characteristics, namely, *where* on Max's body it was applied, *what* kind of tactile stimulation it was, e.g. stroking or tapping, and *how* it is emotionally appraised. The realization and explanation of these issues are outlined in the following.

2.1 Where is Max Touched

Max has a segmented body, i.e. his virtual graphical embodiment consists of several geometry parts. Around every geometry representing a limb of Max's body,

Fig. 2. The virtual agent Max with the proximity aura (left) and without the proximity aura (right).

17 proximity geometries were added forming a "proximity aura" (see Figure 2). This allows us to make predictions, when an object in the VR environment is approaching Max's body. By means of the aura we are also able to identify the body part an object may be going to touch. Below the proximity aura, the surface of Max's body is covered with a virtual "skin". This skin consists of flat quadrangle geometries varying in size, each representing a single skin receptor (shown in Figure 3). The receptors are located on the body in neighbourhoods, which are represented in a somatotopic map (similar to the map in the human brain). This representation encodes the information which body limb a virtual skin receptor is attached to, and it allows to determine in a fine-grained way where Max is being touched. Depending on the location on the body, a tactile stimulation can thus be interpreted differently. For example, Max could be more ticklish under his arms than on his knees.

2.2 How is Max Touched

Instead of different kinds of skin receptors as in the human skin, we propose only one kind of virtual skin receptor for Max for it is sufficient to discriminate between different tactile stimulations. Every object that is graphically represented in our VR environment can cause tactile stimuli on Max's virtual skin. In addition, a motion-tracked human hand is a stimulus source. The simulation of touch is based on detecting collisions (using the V-Collide collision engine) between these two types of geometries, the virtual skin receptors and external objects of the environment. Each geometry's collision with a skin receptor is regarded as tactile stimulus. Specific stimulation patterns arise from the temporal and spatial changes connected to the stimulation. When a stimulus, e.g., is moving continuously over the skin, neighbouring receptors are responding successively over time (Figure 4). This temporal information along with the spatial information about each triggering receptor, extracted from the somatotopic map, allows to classify the stimulation as a continuous touch of the respective body

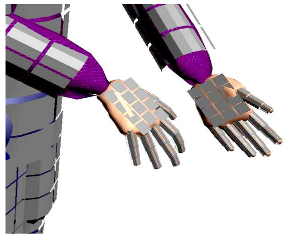


Fig. 3. Virtual skin receptors arranged on the surface of Max's hand. The back of the hand is arranged with 6-neighbourhood receptors, the palm with 8-neighbourhood receptors for higher tactile resolution.

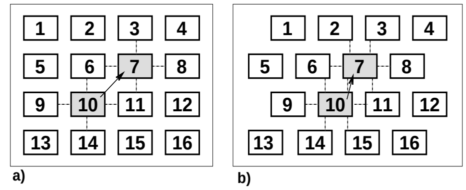


Fig. 4. Schematic depiction of a) eight and b) six neighbouring receptors. Arrows indicate stimuli moving directly from receptor 10 to receptor 7 (highlighted in grey), interpreted as moving tactile stimulus. Dashed lines indicate direct neighbours.

parts. The central component, that fuses these stimulations of the receptors into a coherent touch sensation forms our touch sensor.

For example if a tactile stimulus moves with no spatial interruptions over the agent's body, this can be regarded as light stroking. If a tactile stimulus is applied over a short period of time at one location, this can be interpreted as a short tap.

2.3 Emotional Appraisal

For each classified tactile stimulation we can associate an emotional appraisal. For example, when someone is quickly stroking the bottom part of Max's arm, this is classified as tickling and can be positively (or negatively) appraised. To this end, the touch sensor sends valenced impulses to the emotion system, which drives the emotion dynamics. That is, touch does not directly give rise to specific kinds of emotion, but only controls the number of impulses sent as well as the strength of their valence. This will lead to an increase or decrease of the agent's mood, such that a 'gentle stroke' applied several times can take a comforting effect on Max's mood state. Furher, it is conceivable that the touch sensor can also draw upon informations of the environment, e.g., about the velocity of an object touching Max's body. This would allow to appraise any impact of an object with a high velocity more negatively.

The emotive system is, on the one hand, fed with external stimuli, such as tactile stimuli. On the other hand, the cognitive system exerts influence on the emotional state [1]. In turn, Max's behavior is influenced by his (simulated) emotions, determining as system parameters the way in which Max performs actions. Simulated facial muscles enable him to express emotional states. Max is also able to verbally utter his current emotional states ('I am angry now'). In the case of being hit by a (virtual) ball Max could thus also display or verbalize anger, or he could laugh in the case of someone tickling him.

3 Conclusion

In this paper we introduced an approach to simulate touch perception for the virtual agent Max, based on attaching a large number of virtual skin receptors to his body. Stimulations of these receptors by external objects are calculated in real-time by detecting collisions between the object and, first, Max's "proximity aura" and, then, each single receptor connected to the body part the object is approaching. This method enables a high degree of sensitivity as it was not possible before, neither for virtual agents nor for physical robots. We have presented a way how this perceptual capability can be utilized along with a present emotion simulation system to appraise the tactile stimulations and to accumulate them to determine the "affective content" of touch. Possible applications of this work include a virtual gaming scenario, in which touch perception increases the lifelikeness and interaction possibilities. Human players could touch Max in order to attract his attention or, e.g., could play ball with him. Depending on the quality of the tactile stimulation he could 'feel' a fast moving ball hitting his arm and show an angry face. Another important application is in active touch, where touch perception, goals, and emotional appraisal could be used for Max to develop a form of body awareness.

References

1. C. Becker, N. Lessmann, S. Kopp, and I. Wachsmuth. Connecting feelings and thoughts - modeling the interaction of emotion and cognition in embodied agents. In *Proceedings of the Seventh International Conference on Cognitive Modeling (ICCM-06)*, pages 32–37, Trieste, 2006. Edizioni Goliardiche.
2. C. Becker and I. Wachsmuth. Modeling primary and secondary emotions for a believable communication agent. In *International Workshop on Emotion and Computing, in conj. with the 29th annual German Conference on Artificial Intelligence (KI2006)*, pages 31–34, 2006.
3. C. Creed and R. Beale. Evaluating affective embodied agents over extended interactions. In P. L. D. Reichardt and J.-J. C. Meyer, editors, *Proceedings of the 1st Workshop on Emotion and Computing*.
4. F. de Rosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. de Carolis. From greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies. Special Issue on Applications of Affective Computing in HCI*, 59:81–118, 2003.
5. J. J. Gibson. The purple perils - the implications of active touch. March 1963.
6. S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth. Max - A Multimodal Assistant in Virtual Reality Construction. *KI-Künstliche Intelligenz*, 4/03:11–17, 2003.
7. N. Lessmann, S. Kopp, and I. Wachsmuth. Situated interaction with a virtual human - perception, action, and cognition. In G. Rickheit and I. Wachsmuth, editors, *Situated Communication*, pages 287–323. Mouton de Gruyter, Berlin, 2006.
8. N. Nguyen. Modeling and Simulation of Tactile Perception for Artificial Virtual Agents. Master's thesis, University of Bielefeld, Technical Faculty, Research Group Artificial Intelligence, 2007.

9. W. D. Stiehl and C. Breazeal. Affective touch for robotic companions. In *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction (ACII 2005)*, pages 747–754, Beijing, China, 2005.
10. S. Yohanan, M. Chan, J. Hopkins, H. Sun, and K. MacLean. Hapticat: exploration of affective touch. In *ICMI '05: Proceedings of the 7th international conference on Multimodal interfaces*, pages 222–229, New York, NY, USA, 2005. ACM Press.

Simulating Empathy for the Virtual Human Max

Hana Boukricha, Christian Becker, Ipke Wachsmuth

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

Abstract. Addressing user's emotions in human-computer interaction significantly enhances the believability and lifelikeness of virtual humans. Emotion recognition and interpretation is realized in our approach by integrating empathy as a designated process within the agent's cognitive architecture. In this paper we describe this empathy process which comprises of two interconnected components: a belief-desire-intention (BDI) based cognitive component and an affective component based on the emotion simulation system of the virtual human Max. The application and a preliminary evaluation of this empathy system are reported on in the context of a 3D competitive card game scenario.

1 Introduction and Related Work

Empathy plays a prominent role in human-human interaction because it represents a motivational basis of prosocial behaviour and contributes to moral acts like helping, caring and justice [7].

Brave et al. [5] accentuate that empathy is a fundamental and powerful means to manifest caring in humans. They investigate the psychological impact of affective agents, which are endowed with the ability to behave empathically. In their study [5] two conditions were evaluated in which the agent expresses self-oriented emotions and other-oriented, empathic emotions. Based on online questionnaires, they found that subjects judge the empathic agent as more likeable, trustworthy and caring as the self-emotional agent.

Prendiger et al. [9] investigate the impact of the affective virtual human Max upon humans within a 3D card game on the basis of human player's physiological responses in four different conditions: non-emotional condition, self-centered condition, negative-empathic condition and positive-empathic condition. It was found that, first, the absence of negative emotions within a competitive card game is stress-inducing and, secondly, the valence of the human's emotion is congruent with the valence of the emotion expressed by the agent.

These findings demonstrate that realizing empathic agents is important in human-computer interaction. The authors of these studies primarily investigate the impact of empathic agents upon humans. Their implementations of an agent's empathic reactions to human's emotions, however, are based on simple heuristics rather than a more detailed analysis of the internal processes involved in the elicitation of empathy between humans.

Definitions of empathy fall into one of two major categories (cf. [5], [7]): (1) empathy is defined as an affective reaction to the emotions of others, (2) empathy is viewed as cognitive understanding of another's emotions. In our approach we acknowledge both of these definitions by combining an affective component with a cognitive component in the cognitive architecture of our virtual human Max.

2 A First Approach for an Empathy System

We have experimentally implemented an empathy system for the virtual human Max [4]. It is based on the fundamental assumption that Max is quasi-egocentric. Quasi-egocentrism is observed in two-year-old children [7], who know that empathic emotions are to be ascribed to another individual although being experienced subjectively. However, due to cognitive limitations they do not yet understand that others have their own independent inner states and may appraise a given situation differently. These children appraise observed situations of others in the same way as if they were in that situation themselves.

The architecture of the virtual human Max can be divided into a cognition module and an emotion module (cf. [8], [3]). Following our understanding of empathy we also divide the empathy system into two interconnected components (see Figure 1). The cognitive component of the empathy system is characterized by the cognitive understanding of how emotions occur in humans. This component is implemented within the cognition module. The affective component of the empathy system then assures the simulation of a similar emotion dynamics for the human as it is generated for Max himself (cf. [1]). As an example scenario the interactive implementation of the card game Skip-Bo is chosen (cf. [2]). In this game the agent performs sequences of plans to reach his intended goal of winning the game. In our realization of empathy the agent generates a hypothesis about the emotional state of the human player by appraising the game situation for the human player in the cognitive component. This appraisal is based on the same mechanism that is used in the agent's own appraisal processes. Every human's action is analyzed by plans that generate the same emotional impulses (EI in Figure 1) as if Max would perform these actions himself. These impulses, however, are now driving the hypothesized emotion dynamics of the human player which is simulated within the emotion module of Max. In this appraisal processes the agent does not integrate any information concerning the cognitive state of the human player. Thereby, Max behaves quasi-egocentric and our approach follows the idea of the situational role-taking as described in [6].

In the affective component the course of emotions and moods over time and their mutual interaction as well as the mapping to pleasure-arousal-dominance-space are also modeled for the human player now. Reflecting the assumption that the agent is quasi-egocentric, we use identical parameters for the human player's emotion dynamics as we are using for our virtual human Max. The emotional state of the human player is represented by two additional reference points in

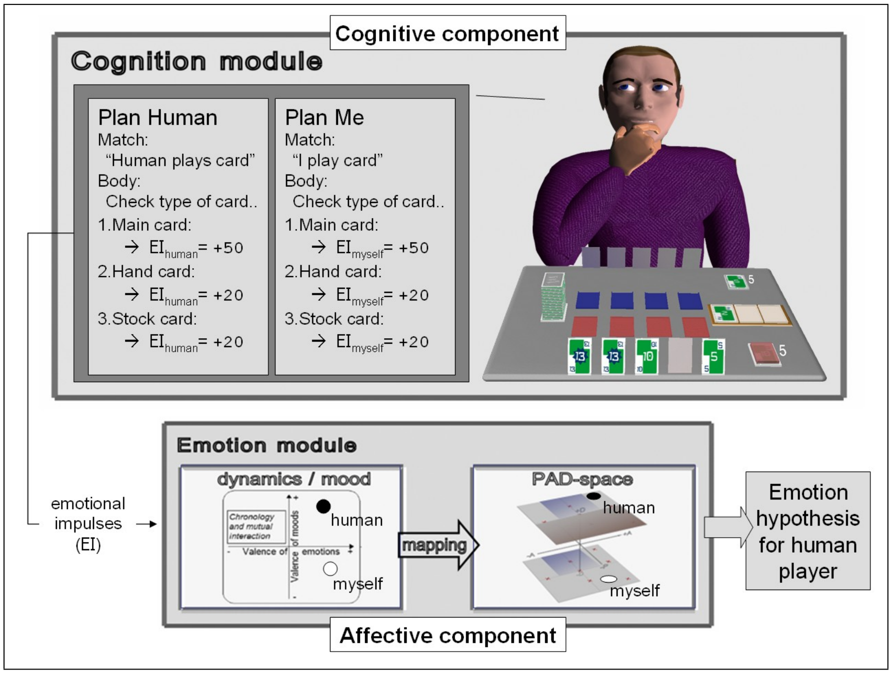


Fig. 1. Empathy simulation as a combination of cognitive appraisal and affect dynamics: The events "Human plays card" and "I play card" trigger appropriate "Plan Human" and "Plan Me" leading to emotional impulses (EI) for Max himself and the human player. These two impulses drive two independent emotion dynamics as represented in the "Emotion module" by their respective white and black circles. After mapping into pleasure-arousal-dominance-space (PAD-space) an emotion hypothesis for the human player is derived.

the emotion module as indicated by the two black circles in Figure 1. Thus, Max always distinguishes between his own and another one's emotional state. Even though he might experience fear himself at a given moment during the game, he might hypothesize at the same time that the human player is happy.

3 A Preliminary Evaluation of the Empathy System

To evaluate our approach to modeling empathy we are currently comparing the hypothetical arousal values generated by our empathy system with arousal values we reassigned to the emotion categories provided by the emotion recognition system of [9].

3.1 Procedure

So far, three of a total of 32 game sessions (sessions 24, 25 and 28) were replayed in realtime. During replay the empathy system generated an emotion hypothesis which was analyzed as follows. The emotion recognition system of [9] provided us with five discreet emotion categories. As these categories are based on a mapping in valence-arousal-space, we reassigned pleasure and arousal values as shown in table 1.

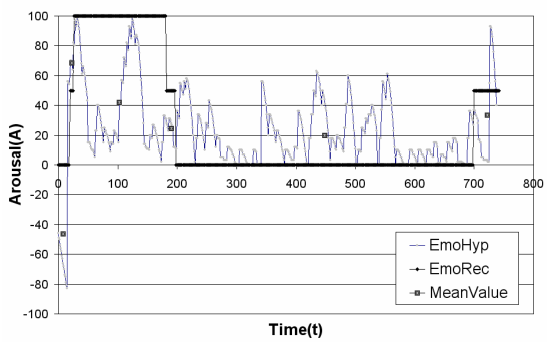
Emotion Cat.	**Pleasure**	**Arousal**
Exited	100	100
Joyful	100	50
Relaxed	0	0
Frustrated	-100	50
Fearful	-100	100

Table 1. The pleasure and arousal values reassigned to the emotion categories provided by the emotion recognition system of [9]. These values were chosen with respect to the valence-arousal-space introduced by Lang (cf. [9]).

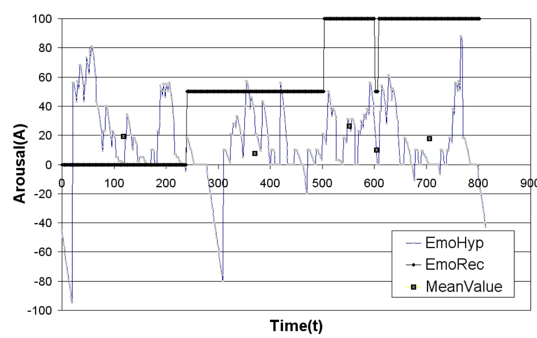
For our first analysis we concentrated on arousal values, because we assume the pleasure values derived from biometrical data (cf. [9]) to be much less reliable. We expected to find similarities between the arousal values generated by our empathy system on the one hand and reassigned to the emotion categories as shown in table 1 on the other hand. Figures 2(a), 2(b) and 2(c) show the courses of the two arousal values over time for the subjects 24, 25 and 28, respectively. The empathy system provides a more continuous course of arousal (EmoHyp) whereas the remapping of the categorical output of emotion recognition results in more discreet course of arousal (EmoRec) over time. In order to compare these values we decided to calculate the mean values (MeanValue) of the arousal values provided by our empathy system over all intervals, in which the reassigned arousal values (EmoRec) remained stable.

3.2 Results and Discussion

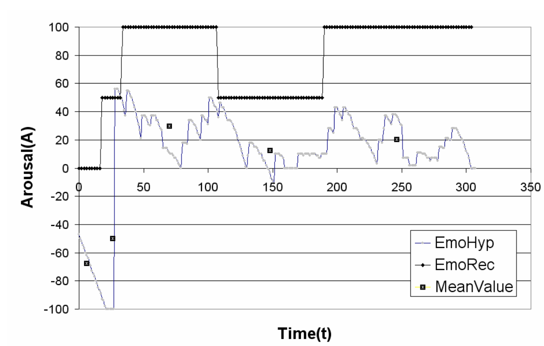
For a preliminary evaluation we compared the arousal mean values of our empathy system with the reassigned arousal values. For subject 28 (cf. Figure 2(c)) we found that the relative offsets of the mean values are comparable to the arousal values of emotion recognition. The mean values corresponding to each interval increase or decrease together with the reassigned arousal values. However, in case of the two other subjects (24 and 25) we found less similarities. For example, in the first interval of Figure 2(b) a mean value of 20 was calculated, while the emotion recognition value remained zero. During the following interval (240 to 502 seconds) the mean value decreases to 7 although the arousal value of emotion recognition increases to 50. Nevertheless, we consider this evaluation as acceptable because some similarities between the two arousal courses could be found. In order to gain statistically significant results we are currently analysing more gaming sessions. We also plan to compare the empathy arousal values to the raw arousal values derived from biometrical data used in the emotion recognition system of [9] thereby avoiding remapping of emotion categories.



(a) Subject 24



(b) Subject 25



(c) Subject 28

Fig. 2. The arousal courses (EmoHyp and EmoRec) and the mean values (MeanValue).

4 Conclusion

We presented a first realization of empathy for a virtual human Max based on the conceptual distinction between cognitive and affective empathy. Some aspects of cognitive empathy are captured by following the situational role-taking approach. The consideration of the course of emotions and moods over time and their mutual interaction in the simulation of the human's hypothesized emotions reflects aspects of the affective component of empathy. Our next step is to evaluate this first approach of an empathy system and to also integrate a cognitive model of the human player to realize an empathy process that follows the individual role-taking approach (cf. [6]).

5 Acknowledgements

This research has been partially supported by the Deutsche Forschungsgemeinschaft in the Collaborative Research Center SFB 673, and by the Japan Society for the Promotion of Science.

References

1. C. Becker, S. Kopp, and I. Wachsmuth. Simulating the emotion dynamics of a multimodal conversational agent. In *Workshop on Affective Dialogue Systems*, pages 154–165, 2004.
2. C. Becker, H. Prendinger, M. Ishizuka, and I. Wachsmuth. Evaluating affective feedback of the 3d agent max in a competitive cards game. In *Affective Computing and Intelligent Interaction*, pages 466–473. Springer, Berlin, 2005.
3. C. Becker and I. Wachsmuth. Modeling primary and secondary emotions for a believable communication agent. In D. Reichardt, P. Levi, and J.-J. C. Meyer, editors, *Proceedings of the 1st Workshop on Emotion and Computing*, pages 31–34, Bremen, 2006.
4. H. Boukricha. Affektive Partnermodellierung: Generierung von Emotionshypothesen durch einen interaktionsfähigen künstlichen Agenten. Master's thesis, University of Bielefeld, Faculty of Technology, 2007.
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. E. T. Higgins. Role-taking and social judgment: Alternative developmental perspectives and processes. In J. H. Flavell and L. Ross, editors, *Social cognitive development: Frontiers and possible futures*, pages 119–153. Cambridge University Press, Cambridge, England, 1981.
7. M. L. Hoffman. *Empathy and Moral Development*. Cambridge University Press, 2000.
8. N. Lessmann, S. Kopp, and I. Wachsmuth. Situated interaction with a virtual human - perception, action, and cognition. In G. Rickheit and I. Wachsmuth, editors, *Situated Communication*, pages 287–323. Mouton de Gruyter, Berlin, 2006.
9. 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.

Constructing Androids as Emotional Agents in Robot-Human Relationship

Wolfgang Gessner, Gesine Lenore Schiewer, Guerino Mazzola, Alex Ringenbach

Abstract. The integrative development of a structural theory of emotion and action results in a new communication medium (MMI) at the user interface between robots and humans which serves purposes already highly awaited. In our project human emotion provides a model for analyzing, reconstructing and implementing the complex interrelations of mimics, gestures, verbal communication and elicitation of action schemes. The results will be transformed into technical, but human-like tools and devices to be used in android robots.

Keywords: Androids, Emotional agents, Subjectivity, Action disposition, Emotional mimic expression, Gesture, Posture, Speech act theory

1. Developing believable and effective android behaviour based on mentalistic theories of emotions

The state of simulation of mechanical abilities in human-like robots has been pushed and enforced worldwide and has reached impressive results. In contrast to this, the communicative abilities of robots are still in need of further improvement. The central idea of our research[1]) is the amplification of the communicative abilities of robots based on the transfer of a special cognitive theory of emotions (Gessner 2004) to the behaviour of androids. A new kind of steering of android's behavior towards human beings will be achieved.

In order to construct human-like robots as autonomous agents they have to coincide with the special communicative abilities and instruments which evolution and cultural development has brought up exclusively in human beings. We accept and tolerate other beings primarily when their behaviour is predictable and comprehensible. Therefore only android robots with distinct 'personalities' will be believable and convincing robots (cf. Breazeal 2002). Successful robot-human interactions will be dependent on multimodal communication coordinating speech, facial expression (as a universal language) and gesture/posture that can simulate the complexity of interpersonal interaction.

---

[1] The Project 'Emotional Agents for Controlling Expression, Action, and Speech in Man-Machine Interaction' is supported by Hasler Stiftung Bern (MMI grant nr. 1978).

According to our approach the coordination of multimodal communication and action can be achieved by using human emotions as a model for android construction. Based on our cognitive emotion theory (Gessner 2004) the realization of robots as emotional agents which are able to situation-adequate, coordinated and individually appropriate responses (*qua* structural equivalents to human emotions) will be achievable. Up to now mindless and faceless devices (technically sophisticated as they are) have to be changed in believable and accepted *communicative* partners showing consistent behaviour (cf. Ortony 2003).

As a cognitive theory, our own theory of emotion can rely on earlier approaches, especially the well-known approach of Ortony, Clore and Collins (OCC) (1988) and the appraisal-theoretic approach developed especially by Scherer et al. (cf. Scherer, Schorr, & Johnstone 2001). In appraisal theories of emotions the role of a fixed number of criteria for appraisals of everyday situations in the elicitation of emotions is evaluated. Contrasting to this, we insist that a complete emotion theory has to rely on the specific kind of individual dispositions – i.e., the individual cognitive triggering, causing a specific subjective interpretation of an emotion-inducing situation, including subjectively erroneous representations of such situations (cf. Gessner 2004, p. 127). Therefore, our attitude-centered theory focuses on the analysis of the individual interpretation mechanisms regarding an emotion-inducing situation, whereby we use elaborate analytical instruments regarding the inner perspective of human beings and their interpretation of emotion-inducing situations.

2. Subjectivity-modeling approach and the conception of emotion in android behaviour

Central to our research is the kernel concept of ‘subjectivity’, which is able to represent the perspectivity which is characteristic for concrete situatedness and allows single emotion elicitation, emotion-related action and/or communicative behaviour to unfold. Subjects are ‘mediums’ representing structures, states and events of the objective world solely from a certain singular perspective and in different quantities and qualifications. They will store these elements in the form of representations, deal with these representations in various manners, draw conclusions and act in a (mostly) predictable manner based on these special assumptions of their own (Tuomela 1995).

A methodological device for the representation and articulation of subjectivity is given by the ‘language of mind’ (LOM) by using the natural means of people expressing their inner thoughts, which represent linguistic universals as categories in which their inner experiences are represented (cf. Fodor 1976, Wierzbicka 1999). The reconstruction of mental states and the ways we are reflecting about them and communicate them to others in searching self- and other-understanding shows up to be essential for reconstructing the inner organization of subjectivity (cf. Morton 2003). This trend (back) to rule-based and even mentalistic models has already reached robotics (cf. Xie 2003), who speaks of ‘mental syntax’ and of ‘imitating the mind’ as ‘the embodiment of an artificial mind’ in connection with ‘autonomous behaviour’ (p. 306 ff.).

In trying to explicate the concept of 'subjectivity', many attempts have been made (cf. Pauen 2001, the variants of Metzinger's 'representationalist' self-model theory of subjectivity (1993, 2003) and Zahevi's (2006) monograph on 'subjectivity and selfhood', which covers the whole field of such approaches). Beyond these conceptual refinements we have decided to follow a pragmatic line in using propositional attitudes to define a 'language of mind' (LOM), which can represent the intra- and intersubjective structures of mental attitudes in an abstract and handsome manner.

In the abridged form to be presented here (due to limited space, cf. Gessner 2004, ch. 6 and 7 for details), propositional attitudes together with the conventionally given logical operators (&, v, ¬), modal operators ($\Diamond$,$\Box$) and some special entailment operators ($\otimes\Rightarrow$ for implication by natural laws, $\bullet\Rightarrow$ for cognitive implication) and with some propositional categories (P for 'proposition', Δ for ‚event', HS resp. HS* for '(the actualization of an) action scheme', Ω for an existing opportunity to act)) build the core vocabulary of LOM:

Epistemic operators are B (....) (*belief*) and K (....) (*knowledge*), the operator of ability C (....) *(can do)* describes the ability to actualize an action scheme HS, i.e., HS*. Volitional operators are W (...) (*wish*) for the disposition to evaluate the actualization of a corresponding action scheme positively, I (...) for intention, i.e. the readiness to actualize this action scheme, HS*. A normative operator O (...) (*obligation*) stands for the existence of (moral or judicial) norms, i.e. imperatives steering actions.

All propositional attitudes are accompanied by the subscripts $_{\varepsilon,\ \alpha,\ \Pi}$ for *ego* (I, me) resp. *alter* (the other person) resp. 'persons in general' for the 'generalized person. Subscripted time indicators V, G, Z indicate the time span in which the attitude is given, with V → G or V → Z etc. indicating the transition of time spans over which the attitude is held. Finally, $\lceil$ indicates that the (complex) proposition before this sign is valid *only under the condition of* the truth of the proposition following after this sign.

Using this nomenclature, which in the same time expresses mental structures and makes them describable, one can formulate mental states of different order and complexity:

Simple (non-iterated) propositional attitudes as *intrasubjective* mental states could be assumptions like $B_{\varepsilon,G}$ (P), wishes like $W_{\varepsilon,V}$ (Δ) or intentions like $I_{\varepsilon,G}$ (HS* $(\Delta)_{\varepsilon,G\rightarrow Z}$. But intrasubjective attitudes can also be complex, i.e. iterated propositional attitudes like $K_{\varepsilon,G}$ ($B_{\varepsilon,V}$ ($\neg\,\Delta$ T Δ I $\neg\,\Delta$)) $\lceil$ HS* $(\Delta)_{\varepsilon,\ G\rightarrow Z}$, which indicates a higher order assumption about changes in the objective world under a special condition, with ($\neg\,\Delta$ T Δ I $\neg\,\Delta$) signifying the transition of a certain event into its contradiction instead of continuing without the intervention of an additional causal factor. As another example, the estimation of opportunities of getting future knowledge could be written as $\neg\ B_{\varepsilon,G\rightarrow Z}$ ($C_{\Pi,G\rightarrow Z}$ ($K_{\Pi,G\rightarrow Z}$ (P $\vee$ $\neg$ P))), expressing a general sceptical belief about such opportunities.

Turning to *intersubjective* attitudes (i.e., attitudes representing or reflecting not only personal states, but the mental state(s) of at least one other person simultaneously), the following formula describes the inner structure of a changing personal assumption concerning the intentions of another person: $B_{\varepsilon,V\rightarrow G}$ ($K_{\varepsilon,V\rightarrow Z}$

$(I_{\alpha,G}$ (HS* (Δ))) & $B_{\varepsilon,G\rightarrow Z}$ (¬ $I_{\alpha,G}$ (HS* (Δ))). In determining the options given in a potential action situation somebody could have arrived at the following description of his mental state: $B_{\varepsilon,V\rightarrow G}$ ($C_{\alpha,\ G\rightarrow Z}$ (HS* (Δ))) & $\Omega_{\alpha,G\rightarrow Z}$ (HS* (Δ)) & $K_{\varepsilon,A}$ (¬ $O_{\Pi,G\rightarrow Z}$ (¬ HS* (Δ))). The subjective state described here expresses some complex attitudes concerning abilities and opportunities to act, coexisting with an assumption on the non-intervention of act-relevant norms.

The examples given demonstrate that by introducing LOM a conceptual framework is set which enables real people to express their subjective states in arbitrarily gradable complexity and at the same time serves researches to describe these subjective mental states in a precise and comprehensible manner.

This tool will serve as a means to describe the elicitation conditions of emotions, too. In this form of application topical and focal elements have to be distinguished: *Topical elements* as attitudes which are already given in a subject, i.e. the (mostly value-related or norm-related) elements which are the background of emotions to originate, and *focal* elements as the (mostly epistemic) elements which are developed in representing the actual situation which gives rise to the 'triggering' of the corresponding emotion. Based on this an emotion like 'annoyance/irritation (the german 'Ärger') can be understood as individually instantiated reaction to a 'standard situation' which is describable in an abstract, i.e. not situation-specific manner by using LOM as follows:

(1)	$K_{\varepsilon,G}$ (Δ_G)	1st focal element
(2)	& $K_{\varepsilon,V\rightarrow G}$ ($O_{\alpha,V\rightarrow G}$ (HS* (Δ) $_{V\rightarrow G}$))	1st topical element
(3)	& $B_{\varepsilon,V\rightarrow G}$ ($K_{\alpha,V\rightarrow G}$ ($O_{\alpha,V\rightarrow G}$ (HS* (Δ) $_G$)))	2nd topical element
(4)	& $B_{\varepsilon,G}$ ($I_{\alpha,V\rightarrow G}$ ¬ (HS* $(\Delta)_G$))	2nd focal element

•⇒ **ANNOYANCE / IRRITATION** Resulting emotion

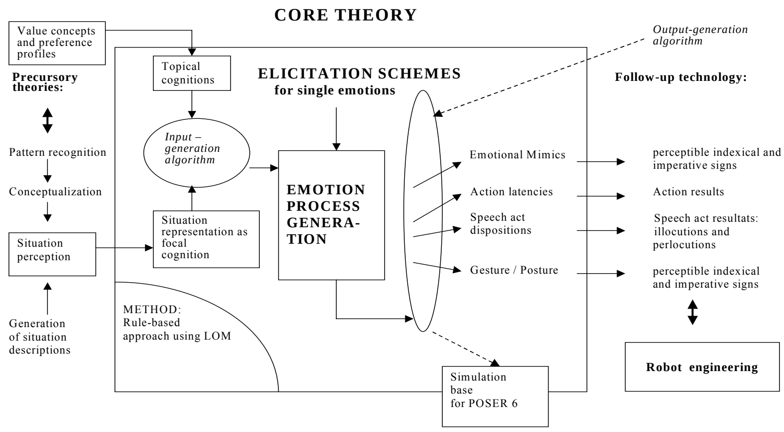
⌈ { $K_{\varepsilon,V\rightarrow G}$ (HS* (Δ) $_{V\rightarrow G}$) ⊗⇒ ¬ Δ_G) } Background cognition

In conjunction the conditions formulated above define the complete elicitation condition for annoyance/irritation. Moreover, LOM can be used to formulate the corresponding dimensions of this (and all other) cognitive emotion(s) in the following way:

1. In utilizing the fact that emotion elicitation and the buildup of action latencies are referring to the same situation-type, the parallels and interactions of unfolding emotions and dispositions to take specific actions can be analysed.
2. Accordingly, the constitutive rules for speech-acts (Searle & Vanderveken 1985) as cognitive patterns describable in LOM are 'tested' by the subject's language-generating mechanisms on 'matching' to the same situation and will show up to predict the appearance of specific speech-acts in reaction to this situation, which in turn will interact with real actions and the emotion proper having emerged in this situation.

3. The emoting subject's emotional mimics indicates inner states of this subject and (at the same time) has imperative functions on other persons receiving these 'signallings'. LOM can be used to reformulate these inner states and allows the reformulation of the relation between these specific subjective states and the semiotic apparatus of facial expression in its different dimensions.
4. Finally, LOM can be used in defining production rules for describing restrictions on the corpus of gestures and postures which could in principle interact with specific inner states of subjects and the triggering of core emotions by matching the appropriate elicitation condition of an emotion.

The following scheme depicts the principal inner relations of the core theory of emotion described here. Besides this it shows that the systematic development of situation representations (as a *precursory* theory) is not part of this project but will be left to the further development of existing approaches. Correspondingly, the project is not involved in real robot engineering (as *follow-up* technology) but tries to develop some steering mechanism of possible and meaningful android behaviour and (as a first step) to generate simulations of these behaviour in appropriate media of presentation.



3. Implementation

The LOM formalism developed has been restated in terms of the Backus Naur Form (BNF) as a context-free language type. This gives a foundation for the design and implementation of a software operationalising the LOM semantics and is also capable of yielding an output for multimedia rendering of emotional states and processes, which includes animated 3D models and acoustic utterances. The overall architecture of this design includes three components: the two input units for terms of focal and topical cognition and the internal process unit which is in turn divided into three units: (1) the cognitive elicitation schemata, (2) the logical processing machine, and (3) the multimedia interpretation schemata. The output unit must then interpret the process unit's result (generated by the multimedia interpretation schemata) in terms of a specialized software for multimedia rendering, such as the human animation software POSER 6. The latter may be driven by Python commands (an object-oriented interpreter language used in POSER 6), which allows for a code-driven processing of POSER 6 graphics and audio data. First 3D-implementations using splines and facial-action-points for reconstructing mimic and movements have been made and will be coordinated with the other dimensions of artificial emotion.

In dealing with android design we choose an elaborate task to be solved. Another possible task consists in the design of subjectivity-based emotional interfaces. Such interfaces, acting between e.g. sophisticated cars and their drivers or IT-operating systems and their intricate users could give the term 'smart interface' a new and exciting meaning.

Bibliography

Breazeal, C. (2002). Designing sociable robots. Cambridge, Mass.: The MIT Press ...

Fodor, J. A. (1976). The language of thought. Hassocks: The Harvester Press.

Gessner, W. (2004). Die kognitive Emergenz von Emotionen. Paderborn: Mentis.

Metzinger, Thomas (1993). Subjekt und Selbstmodell. Paderborn: Schöningh.

Metzinger, Thomas (2003). Being no one: the self-model theory of subjectivity. Cambridge, Mass.: The MIT Press.

Morton, A. (2003). The importance of being understood. Folk psychology as ethics. London: Routledge.

Ortony, A. (2003). On making believable emotional agents believable. In: Trappl, Petta & Payr (Eds.) Emotions in humans and artifacts. Cambridge, Mass.: The MIT Press, 189-211.

Ortony, A., Clore, G. L., & Collins, A. (1988). The cognitive structure of emotions. Cambridge: Cambridge University Press.

Pauen, Michael (2001). Grundprobleme der Philosophie des Geistes. Eine Einführung. Frankfurt a.M., Fischer.

Scherer, K.R., Schorr, A. & Johnstone, T. (Eds.). (2001). Appraisal processes in emotions. Theory, methods, research. Oxford: Oxford University Press.

Searle, John & Daniel Vanderveken (1985). Foundations of illocutionary logic. Cambridge: Cambridge University Press.

Tuomela, R. (1995). The importance of us. Stanford: Stanford University Press.

Wierzbicka, Anna (1999). Emotions across languages and cultures. Diversity and universals. Cambridge: Cambridge University Press.

Xie, M. (2003). Fundamentals of robotics. Linking perception to action. New Jersey: World Scientific.

Zahavi, Dan (2006). Subjectivity and selfhood: Investigating the first-person perspective. Cambridge, Mass.: The MIT Press.

Moods and Shallow Emotions for Balanced Speech Acts in Computer Games

Mireille Rojas Romero and Manfred Kerber

Computer Science, University of Birmingham, Birmingham, England

Abstract. In most state-of-the-art Role-Playing Games (RPG) speech acts are formed by pre-written dialogues. While it is possible to convey any dialogue in this way and it may be adequate in certain situations, it typically looks unnatural when the situation is repeated in a different context and/or when the agent should be in a different emotional state – for instance when an agent is repeatedly asked the same question. In order to make the speech acts more realistic, we add to the usual representation of agents by *goals* (what the agent wants to achieve), *standards*, *attitudes*, and *personality* (capabilities which the agent has and determine what it can and would do), firstly a *social status*, and secondly *emotional state*, which typically changes as the game advances.

1 Introduction

Many Non-Player Characters (NPC) in computer games make very good first impressions, but on repeated encounters it becomes clear that they follow quite simple patterns and do not show the sophistication which a human player would. This is partly due to the lack of emotional state, which can show in different ways, mainly, in action selection [1], facial expression [13], vocal effects [11] or language chosen for the conversation [2]. If we want to improve this situation, we need an explicit representation of emotional state. We will focus on this representation with the emotionally based language generation as our main motivation. We distinguish less specific mood from more specific (directed) emotions.

In this work we propose a way how to extend agents in role playing games so that they firstly change their emotional state, and that secondly the emotional state influences the way how language is generated. Furthermore we consider the niceness/grumpiness of agents and their social status. A nice agent would tend toward polite expressions while a grumpy one would tend toward rude expressions. If somebody addresses someone who is socially above himself he may use expressions which are more polite than those he usually uses (likewise a person who is in a good mood) [14].

In the next section we discuss how agents are modelled, in section 3 an emotional model, and in section 4 we use this model to adjust speech acts.

2 How to Model an Agent

There are many different ways how to represent agents and we will briefly introduce one. In most aspects we could as well have chosen a different approach and still have extended it by a way to represent moods and emotions as proposed in the following. We represent agents as follows and will explain the representation of mood and emotion in more detail in the next section:

Goals: The states of affairs that an agent wants to obtain (e.g., staying alive).
Attitudes: The disposition to like or dislike objects or aspects of objects and/or other agents (e.g., positive towards friends).
Standards: Beliefs about how things should be (e.g., heal people).
Personality: Personality is a set of consistent behaviour, emotion and thought patterns that characterize a person.
Mood: Mood is an undirected kind of emotion.
Emotion: Emotions are valenced reactions to either (consequences of) events, (actions of) agents or (aspects of) objects.
Social Status: represents the relative social position of agents.

The mood and personality traits of an agent are characterized by firstly a set of default values $\mathbf{D}$ and secondly by the emotional sensitivity $\mathbf{S}$, which describes how strongly an agent is influenced in its mood by external events. All these entities go with a certainty value between -1 and 1 to represent the degree (as in the expert system *Mycin*). Our agents differ from rational agents in so far as they have an emotional state. This allows non-playing character agents to influence their speech acts.

Note that while it is our intention to make agents this way more *human-like*, we do *not* claim that the simple representation of emotions is a cognitively adequate representation of human emotions. To the contrary, we are well aware that the emotions as we describe and employ them in this work are *shallow*, and can at best offer an approximation of the *appearance* of human emotions.

3 How to Represent Emotions

We do not investigate which emotions humans have (and agents should have), whether they can be classified in different categories and if yes how. Emotions have been studied extensively in psychology (and made there way into psychology textbooks, see for instance [5, p.142–163]). There are different ways how they can be characterized. The attempts to classify them have a long history and go back at least to Wundt who wanted in 1896 to reduce all emotions to a combination of three dimensions pleasantness/unpleasantness, calm/excitement and relaxation/tension. In addition, some authors claim the existence of a set of basic emotions, Frijda [4], for instance, suggests that there are six basic emotions (desire, happiness, interest, surprise, wonder, and sorrow), while Plutchik [16] claims that they are eight (acceptance, anger, anticipation, disgust, joy, fear, sadness, and surprise).

We describe agents by personality traits and by a set of moods. However, there is no generally accepted definition of mood, emotion or personality. In [10] personality is defined to name those reaction tendencies that are consistent over situations and time. Personality is also described in terms of traits, some theories claim that there are five of these traits (openness to experience, conscientiousness, extroversion, agreeableness and neuroticism) to completely describe personality [9]. Others [3] claim that there should be only two (neuroticism and extroversion). Moods and emotions on the other hand are referred to states of the mind, which are only momentary in the case of emotions, but prolongated in the case of mood and due to accumulative effect of emotions [8]. According to Velásquez [20] mood and emotions are only differentiated in terms of level of arousal. In addition, Moffat [10] suggests that personality and emotions are basically the same mechanism only differentiated by time and duration.

In Ortony, Clore and Collins' theory of emotions [12] (also known as the OCC model), 22 *emotion types* are distinguished. According to this model emotions result from focusing on one of three significant aspects of the world, *events* and their consequences, *agents* and their actions, and *objects*. So the emotions are differentiated according to the three main categories: event-based emotions, action-based emotions, and aspect-based emotions, and they are triggered by the *goals*, *standards*, and *attitudes* of agents. We relate our work to OCC, and pair the 22 emotions in OCC to the eleven pairs: *Joy/Distress*, *Hope/Fear*, *Satisfaction/FearsConfirmed*, *Relief/Disappointment*, *HappyFor/Resents*, *SorryFor/Gloats*, *Grateful/Angry*, *Gratification/Remorse*, *Proud/Shame*, *Admire/Reproach*, and *Love/Hate*.

In our model while some emotions are directed, e.g. $loves(A, B, \alpha)$, A loves (likes) B with degree α, others such as $joy(A, \alpha)$ are not, and eventually integrate the mood of an agent. The mood comprises: *Joy/Distress*, *Hope/Fear*, *Grateful/Angry*, and *Relief/Disappointment*, which are represented for each agent by the certainty value α. Moods along with another personality feature, *niceness/grumpiness*, form a characteristic personality vector which is the basis for an agent's way of language utterances. In our model, *Niceness/Grumpiness* is closely related to the *agreeable* trait in the five factor model of personality [9], which refers a person's disposition to be kind, warm or sympathetic. We show this personality trait by polite or rude utterances accordingly. We take the 5 values *Niceness/Grumpiness*, *Joy/Distress*, *Hope/Fear*, *Grateful/Angry*, and *Relief/Disappointment* to form a personality vector. The mood has direct influence on facial expressions, on action selection and verbal expressions: Joy goes hand in hand with high action level (as opposed to distress) and politeness, hope with a high action level, while anger is linked to impoliteness. Relief will result in sighs as will disappointment (albeit of a different kind). For instance, for a tough character called *Herban* from the game Neverwinter Nights, we might have:

$$\mathbf{D}_{Herban} = \begin{pmatrix} \mathit{Niceness/Grumpiness} \\ \mathit{Joy/Distress} \\ \mathit{Hope/Fear} \\ \mathit{Grateful/Angry} \\ \mathit{Relief/Disappointment} \end{pmatrix} = \begin{pmatrix} -0.9 \\ 0.0 \\ -0.4 \\ -0.8 \\ 0.0 \end{pmatrix}$$

In addition, each character has associated an emotional sensitivity. So we would select $\mathbf{S}_{Herban}$ as 2 throughout. That is, Herban would get much angrier than another agent when a particular event occurs.

Let us look more closely at the mood "Joy/Distress", which is supposed to be changed when something happens that an agent wants to happen.

$$wants(A, F, \alpha) \wedge happens(F) \longrightarrow joy(A, \alpha) \tag{1}$$

That is, when an agent wants something and this actually happens then the agent will increase its joy value by $\mathbf{S}_{\text{joy}} \odot \alpha$, where $\mathbf{S}_{\text{joy}}$ is the agent's emotional sensitivity for joy. E.g., for $x \in \mathbb{R}$ and $\alpha > 0$ we define: $x \odot \alpha = 1 - (1 - \alpha)^x$.

The update of a mood (and of niceness) is done by combining the current value a for the mood with the update value $b = \mathbf{S} \odot \alpha$ according to the standard combination of certainty values used in *Mycin* [17].

There is an old objection against the usage of certainty values: Where do the numerical values come from? Why does the α propagate in this particular way? There has never been a very good answer to this. Tests in *Mycin* have, however, shown that as long as you choose the values reasonable you get quite reasonable results and the overall behaviour of the system is quite robust as to the exact values chosen. We assume that a methodology which has successfully been applied in a much more critical application area will be appropriate here as well. We are aware, however, that further investigation is needed.

Note that there are many ways how to compute the intensity of an emotion based on the intensities associated with the preconditions (e.g., logarithmic combination [15], additive [2]). In our approach we employ *Mycin*'s combination method, that is, we take the minimum from all the values of the preconditions in a rule and multiply it with the strength of the rule [6].

Also note that we get a vector space this way and can combine moods in a simple computational way (we have not the space to give an explicit proof here). A mood value a decays over time and returns to its initial state by a decay factor. More advanced emotions such as love are updated by rules and do not decay. They are assumed to remain stable unless some events trigger rules for an explicity change. In addition to *joy* we can specify other emotions such as *hope* by update rules, which are triggered by the game. If an agent A wants F with certainty value α, but F is missing, then he hopes for it with α:

$$\mathit{wants}(A, F, \alpha) \wedge \neg \mathit{happens}(F) \longrightarrow \mathit{hopes}(A, \alpha) \tag{2}$$

4 Templates to express emotional state

Natural language generation (NLG) is the process of deliberately building some kind of natural language output (speech or text) from a non-linguistic representation in order to meet some specified communicative goals [7]. There is no general agreement about the input and the components of the language generation process, but it usually involves to determine the communicative goal of the system, determine the content of the system prompt based on the communicative goal, and produces text that is syntactically and morphologically correct. An input to NLG can be as simple as:

$$Departure(train_{306}, location_{abdn}, time_{1000}) \tag{3}$$

Template-based and *plan-based* are the two current approaches to NLG. A template is a static sequence of words with gaps to fill in, e.g.

[train] *is leaving* [town] now.

This approach is sometimes employed in conversational agents [18] and it is the one that initially we will use for our approach. Template-based systems map their non-linguistic input directly (without any intermediate representations) to the linguistic surface structure [19]. So, in this approach the semantic representation $Departure(train_{306}, location_{abdn}, time_{1000})$ might be directly associated with the template (3).

Our templates are based on a niceness value. How nice some character is towards another depends first on its basic niceness (as defined in the personality of the agent), second on its social rank, third its *Joy/Distress*, and finally on particular emotional attachment of the agent toward another *Love/Hate* (if given). For instance, when a rude agent A (default niceness -0.9) is interacting with another agent B who is socially superior ($\mathit{socialstatus}(A) < \mathit{socialstatus}(B)$) then he will not be as rude as usual, but he will try to be more polite, the same rude character A will be less rude if he is in a joyful emotional state, or if he loves (likes) B. Two specific rules are applicable to adjust the niceness toward another agent (from initially the overall niceness value of the agent) with respect to the respective social status and whether or not A likes B.

$$\textbf{Rule NiceSocial} \qquad \text{IF } \mathit{socialstatus}(A) < \mathit{socialstatus}(B) \text{ THEN}$$
$$NiceTo(A, B) := NiceTo(A, B) \oplus (\mathit{socialstatus}(B) \ominus \mathit{socialstatus}(A)) \tag{4}$$

$$\textbf{Rule NiceLikes} \qquad \text{IF } \mathit{loves}(A, B, \alpha)\text{THEN}$$
$$NiceTo(A, B) := NiceTo(A, B) \oplus \alpha \tag{5}$$

Based on the niceness value and the social status of a character, we define both an *associated semantic representation* and a *template*. So, for instance, the semantic representation in (6) has associated the template "[exp] I'm going to [v] you [sth]", which will be triggered by the system when an utterance request of *introduce his purpose* is demanded.

$$\text{[exp] I'm going to [v] you [sth]} \longleftarrow \text{Template} \tag{6}$$

The gaps in this template, i.e., [exp], [v] and [sth], can take the following values: [exp] = "I'm very pleased to tell you that" *(very polite)*, "I'm happy to tell you that" *(mannerly)*, "well" *(nice)*, "[ ]" *(neutral)*, "now, listen" *(crude)*, "now, listen you body" *(rude)*, "now listen you worthless hide" *(ill-mannered)*.
[v]= "teach", and [sth]= "some tips on combat techniques"

This give us a wide linguistic spectrum of possibilities of text utterances. Then, for example, a well-mannered character might introduce his purpose like: "*I'm very pleased to tell you that I'm going to teach you some tips on combat techniques*", in contrast with a rude character which might say instead "*now listen you worthless hide I'm going to teach you some tips on combat techniques*".

5 Summary

We have presented briefly the motivations of our work to represent emotional agents with mental states. Agents have emotions, which influence their utterances, for instance, if the agent is in a polite and joyful mood then its utterances will be much more polite than when it is in an impolite mood. Using a relatively simple template-based approach it is possible to improve the naturalness of language considerably. More sophisticated utterances can be generated by planning methods, which remain as future work. Our work can be extended to produce actions and postures which are modified by the agent's mood and emotions.

References

1. D. J. Chaplin and A. El Rhalibi. IPD for emotional NPC societies in games. In *ACE '04: Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 51–60. ACM Press, 2004.
2. A. Egges, S. Kshirsagar, and N. Magnenat-Thalmann. Generic personality and emotion simulation for conversational agents. *Journal of Visualization and Computer Animation*, 15(1):1–13, 2004.
3. H. J. Eysenck. *Handbook of personality: Theory and research*, chapter Biological dimensions of personality, pages 244–276. New York: Guilford, 1990.
4. N. Frijda. *Emotions.* Cambridge University Press, 1987.
5. R. D. Gross. *Psychology – The Science of Mind and Behaviour.* Hodder & Stoughton, London, 1992.
6. P. Jackson. *Introduction to Expert Systems.* Adisson-Wesley, 2nd Edition, 1990.
7. D. Jurafsky and J. H. Martin. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* Prentice-Hall, 2000.
8. S. Kshirsagar. A multilayer personality model. In *SMARTGRAPH '02: Proceedings of the 2nd international symposium on Smart graphics*, pages 107–115. ACM Press, 2002.
9. R. R. McCrae and O. P. John. An Introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215, 1992.
10. D. Moffat. Personality Parameters and Programs. In Paolo Petta and Robert Trappl, editor, *Creating Personalities for Synthetic Actors, Towards Autonomous Personality Agents*, LNCS 1195, pages 120–165. Springer, 1997.
11. I. R. Murray and J. L. Arnott. Implementation and testing of a system for producing emotion- by-rule in synthetic speech. *specom*, pages 369–390, 1995.
12. A. Ortony, G. Clore, and A. Collins. *The Cognitive Structure of Emotions.* Cambridge University Press, 1988.
13. C. Pelachaud, N. I. Badler, and M. Steedman. Generating Facial Expressions for Speech. *Cognitive Science*, 20(1):1–46, 1996.
14. H. Prendiner and M. Ishizuka. Affective and social behaviour in animated agents. In *IFSA World Congress and 20th NAFIPS International Conference*, volume 5, pages 2918–2923, 2001.
15. H. Prendinger and M. Ishizuka. SCREAM: scripting emotion-based agent minds. In *AAMAS*, pages 350–351. ACM, 2002.
16. R. Plutchik & H. Kellerman, editor. *Emotion: Theory, research, and experience*, volume 1, chapter A general psychoevolutionary theory of emotion, pages 3–33. New York: Academic, 1980.
17. E. H. Shortliffe. *Computer-Based Medical Consultations: MYCIN.* American Elsevier, 1976.
18. M. Stone and D. DeCarlo. Crafting the illusion of meaning: Template-based specification of embodied conversational behavior. In *CASA*, pages 11–16. IEEE Computer Society, 2003.
19. K. Van Deemter, M. Theune, and E. Krahmer. Real versus template-based natural language generation: A false opposition? *Computational Linguistics*, 31(1), 2005.
20. J. D. Velásquez. Modeling Emotions and Other Motivations in Synthetic Agents. In *AAAI/IAAI*, pages 10–15, 1997.

Interpretation of Intensity Variables for an Emotional Agent in the Public Goods Game

Dirk M. Reichardt

BA Stuttgart - University of Cooperative Education
D-70180 Stuttgart, Germany
reichardt@ba-stuttgart.de

Abstract: The public goods game with punishment option is chosen as a scenario for an emotional agent. Experiments show that human behavior is based on emotional decisions in this scenario. In this paper, the task of building an emotional human like agent is approached by adapting the OCC model to the game situation. This paper focuses on the intensity variables and the rules for generating appropriate values.

Introduction

The public goods game (PGG) with punishment option is chosen as an adequate scenario for an emotional agent [1]. Two main reasons make it a valuable test environment: a) its simple structure allows a reasoning component of manageable complexity and b) experiments show that human decision making in this scenario is based upon emotions rather than being purely rational [2].

We chose the model by Ortony, Clore and Colins (OCC) [3] as a basis for an emotional agent which plays that game. The OCC model is based on a (rational) cognitive evaluation of the situation which then affects the emotional state of the agent. The model is enhanced by a personality model to represent different types of players. The main adaptations to the scenario are:

1. providing the knowledge about rules of the game and for reasoning on strategies, as well as on behaviors of others in an adequate way,
2. defining the derivation of the values for the intensity variables in the OCC model from the given facts in the PGG scenario and
3. identifying the parameters of the personality model affecting the generation of emotions and their effect on the decision making process.

In this paper we focus on the second point and discuss intensity variables of OCC.

Applying the OCC Model to the Public Goods Game

A good scenario for emotional behaviour can be provided by game theory which actually presumes rational agents. The idea of the public goods game is the following: Four players are in the game. Each player gets a certain amount of money. The

scenario contains a public project which is bearing s > 0 percent interest. It's up to the player to decide how much to invest in the public project. The sum of all contributions plus interest is then distributed among the players in equal shares. No matter how much was paid by the individual player, the share is the same. Therefore, the rational decision is not to contribute at least this is the consequence of game theory. If everyone behaves that way, the players get much less than they could. A major extension of the scenario is the punishment option. Each player may punish a defector. In an iterated game this may change the defector's attitude. However, punishing is not free of cost but the punishment option leads to higher average payments. In fact, there is no direct benefit for the punisher, he actually reduces his outcome. In order to remove this rational considerations, Fehr and Gächter [2] repeat the game with the condition that the group composition changes from period to period. Now, only others may benefit from the punishment, the punishment decision therefore becomes *altruistic* and the experiments showed a significant correlation between emotion and the punishment decisions. In further experiments, the *reputation* of a player is included [4]. The concepts of punishment and reputation give incentives for a rational behaviour which leads to a higher benefit for the group.
However, it is still not rational to punish, so there will be no 'punisher' reputation for any rational agent. Nevertheless, in all experiments human behaviour differs from this rational model.

This human behaviour is assumed to be grounded on emotion an the emotional agent should show similar behaviour. The OCC model provides the framework we chose to model an agent for this scenario. This model specifies 22 categories of in the three dimensions: *goal relevant events*, *actions of agents* and *aspects of objects*. In our approach, the OCC model is combined with a personality model which provides non-linear transfer functions to map current mood and elicitor intensity to emotion intensity. In this paper we analyze the use of intensity variables of the OCC model as a basis for the adaptation on the PGG scenario.

Interpretation of Intensity Variables

The OCC model defines the dimensions *consequences of events*, *actions of agents* and *aspects of objects*. Each of these dimensions includes emotion classes which appear in different intensity depending on the situation. In order to formally represent this, the OCC model introduces global and local intensity variables. The latter are local wrt. one of the dimensions mentioned above. How do these intensity variables get appropriate values in the PGG scenario? This issue will be discussed in the following.

Events, actions and aspects of objects in the PGG

Events: the two central events in PGG are the publication of payments and the publication of punishments. Those are considered as events since only the final result for the agent is considered rather than the individual actions of the other agents.

Actions: agents in PGG perform only two actions: payment (public project) and punishment. Note that agents maintain a history record storing all the actions. This information is considered in decisions as well.

Aspects of Objects: another agent is considered an object in this situation. As aspects, the current account balance and the payment history (given and received) are taken into account.

Central Intensity Variables

Ortony, Clore and Collins define desirability, praiseworthiness and appealingness as central intensity variables. In the following we provide an interpretation of these variables in the PGG scenario.

Desirability: generally speaking, an event is desirable if it contributes to the achievement of a goal. Therefore, we need to define the goals of the agent, first. OCC distinguishes a-goals (active pursuit), i-goals (interest) and r-goals (replenishment). Which goals can be identified within the PGG? The two main goals are *profit maximisation* and the *relative profit* compared to other agents. A high contribution of the others is certainly highly desirable for the agent. The event of being punished is undesirable, therefore the decision on the own contribution has to take a *prediction* of this event into account. How about r-goals? We'd propose to disregard this category for the PGG scenario. A far-fetched interpretation may be the following: if the agent is in a situation where some other agent does not comply and nobody does anything against that behaviour, it may be desirable to punish this agent *from time to time*.

Praiseworthiness: this variable is based upon the actions of other agents and it depends on standards and the agent's view of its own role. Certainly, in the PGG it may be considered praiseworthy if another agent puts a lot of money in the common project or if he or she punishes free riders – unless the agent itself is affected. What if the agent did not consider its own payment as punishable? The 'standard' is left to one's own discretion. Consider the situation that agent A pays 15 and the other agents pay 10. Now compare to the situation in which they pay 20. The praiseworthiness of the *identical* action of agent A varies with the context. Therefore our model contains absolute and relative components for the measurement of praiseworthiness.

Appealingness: this intensity variable refers to an attitude of the agent. Because every punishment costs the agent, it is not appealing. Note that punishing can serve a *higher goal* (to force the other agent to comply) and can therefore be considered desirable.

Mapping Global Intensity Variables

Sense of reality: since the PGG is not a real world scenario, the expected intensity of emotions is rather low. Improvements for the simulation can be achieved (with human players) in case real money is used.

Proximity: emotion intensity is highest directly after the elicitation. The model then includes a *fading* intensity value. On the other hand this implies that the emotion elicited in the previous round still exists and therefore the simulation is not functional.

Unexpectedness: since this refers to an event which could not be predicted at all, it cannot reasonably be included in the simulation because this means the event is not within the defined scenario and therefore technically impossible.

Arousal: the agent is not ‘embodied’ and there is no other reason for arousal since the simulation is limited to the defined actions. This does not make this aspect irrelevant for the simulation but a direct link to the (unspecific) arousal ‘caused’ by the basic emotions of the OCC model is derived.

Mapping Local Intensity Variables

Events

Likelihood: the expected value of payments and punishments is computed based upon a locally stored action history for every other agent. If there is no knowledge, the attitude of the agent (part of the personality model) is taken into account with a larger influence.

Effort: a sequence of actions is considered. If the agent tries for a longer time to persuade others to comply (in vain), the anger is stronger than in case he or she accepts his or her fate right away. Which one of these two behaviours is dominant, depends on the personality model.

Desirability for other: a model for the goals of the other agent is needed for that. Since there are only very few possible goals (see above) this is viable in the PGG.

Liking: it seems possible that an agent builds a certain sentiment for another (unknown) agent depending on the shown behavior. Nevertheless, since there are no contacts outside the game, the effects of linking are considered fairly low.

Deservingness: is the punishment for another agent justified? Or, on the other hand, does another agent deserve the earned money? The deservingness depends on the subjective standards of the agent.

Actions

Strength of cognitive unit: since the groups are changed after 12 rounds, the relevance of this aspect is considered rather low. The only emotional reaction, which does not directly lead to an action, is the comparison of the group results.

Expectation-deviation: an agent that always pays the full amount suddenly contributes nothing. The strength of the emotional reaction of others may be higher considering pride, shame, admiration and reproach.

Aspects of objects

Familiarity: the only relevant objects in PGG are other agents. The agent keeps a record on the behavior of others. Therefore the reactions may differ in case two agents meet again in another group.

Implementation

The 'rational' part of the agent provides basic information on statistics like the average payments in the group, the last payments, the average payments of an agent, the average punishment for free riders, and punishment effects on agents. This contributes to the *likelihood* computation of an action or event. Moreover, the agent tries to compute a classification. Other agents in the same group are mapped on stereotypes and the classification carries a confidence factor. If the other agent behaves differently, an *expectation deviation* is sensed and influences the emotional state. The desirability is computed as consisting of two aspects: relative and absolute. Both deliver a satisfaction value between 0 and 1 and the final desirability is concluded as a weighted average. The weights are part of the personality model, defining the goal preference (profit maximization vs. relative profit) as a bipolar scale. Considering desirability for others, their expected goal preference is used.

Conclusion and Future Work

The first implementation of a core model is tested in a simulation environment. Besides functional tests, some initial plausibility tests have been performed. The current agent implements a "monitoring", analytical emotional component. The emotional state can be seen and can be evaluated by an observer, yet, the actions of the agents will be coupled as a next step in the work in progress.

References

[1] D. Reichardt, "Will Artificial Emotional Agents Show Altruistic Punishment In The Public Goods Game?", in Proceedings of the 1st Workshop Emotion and Computing Current Research and Future Impact, Reichardt, D., Levi P.,Meyer, J.-J.Chr. (Editors), 29th Annual German Conference on Artificial Intelligence, Bremen, 2006, ISBN 3-88722-664-X

[2] Fehr, E., Gächter, S., "Altruistic Punishment in Humans", Nature, Volume 415, pp. 137-140., 2002

[3] Ortony, A., Clore, G., Collins, A., "The Cognitive Structure of Emotions", Cambridge University Press, 1988

[4] Brandt, H., Hauert, Ch. & Sigmund, K., "Punishment and reputation in spatial public goods games", Proc. R. Soc. Lond B **270**, pp. 1099-1104.

Towards a Quantitative Model of Emotions for Intelligent Agents

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

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

Abstract. Starting out from a qualitative formalization of a well-known psychological model of emotions that we have developed in previous work, we now turn to the quantitative aspects of emotions, and investigate how these can be incorporated into the qualitative model.

1 Introduction

A popular computational model of emotions is that of Ortony, Clore & Collins [1] (henceforth to be referred to as "the OCC model" or simply "OCC"). Steunebrink, Dastani & Meyer [2] have formalized, in dynamic / doxastic logic, the conditions that elicit each of the emotions of the OCC model for cognitive agents. This formalization amounts to a *qualitative* model of emotions, specifying the conditions for *when* an emotion is experienced. However, a *quantitative* model for these emotions is missing. For example, the qualitative model does not specify how strong an emotion should be, for how long an agent should be aware of experiencing an emotion, in what ways and for how long an emotion should influence the behavior of an agent, according to what function the strength of an emotion should change over time, etc. Such quantitative aspects of emotions should be taken into account in any serious attempt at specifying a formal model of emotions, in particular the OCC model.

In this paper, we will treat the questions of which quantitative aspects of emotions (of the OCC model) should be considered, how these quantitative aspects can be formally modeled, and how a quantitative model of emotions can be combined with the qualitative model of Steunebrink *et al.* [2]. To illustrate the presented ideas, some formulas will be shown for "joy" (which is the shortest emotion label), but any emotion can be substituted for it.

This paper is outlined as follows. First the qualitative and quantitative aspects of emotions in the OCC model will be discussed. Then the way in which the qualitative aspects have been formalized will be reviewed. Finally, a way of formalizing and integrating quantitative aspects will be proposed.

2 Qualitative and Quantitative Aspects of Emotions in the OCC Model

The OCC model describes a hierarchy classifying 22 emotions, of which half are positive and half are negative. They are: joy, distress, hope, fear, satisfaction, disappointment, relief, fears-confirmed, happy-for, resentment, gloating, pity, pride, shame, admiration, reproach, love, hate, gratification, remorse, gratitude, and anger. The hierarchy contains three branches, namely emotions concerning aspects of objects (e.g., love and hate), actions of agents (e.g., pride and admiration), and consequences of events (e.g., joy and pity). 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). Each pair of emotions in the listing above (i.e. joy–distress, hope–fear, etc.) are called *opposing*, but this does not rule out the possibility of such a pair being triggered simultaneously. Indeed, agents are allowed to experience 'mixed feelings' with respect to a certain situation. The conditions that *elicit* these 22 emotions are described by OCC very concisely in natural language.

Steunebrink *et al.* [2] have formalized these eliciting conditions in dynamic logic, creating a formal qualitative model of emotions. (Note that in [2] only a subset of the qualitative model is described due to space limitations; a complete formalization exists nonetheless.) This model specifies precisely *when* emotions are triggered, but quantitative aspects are missing.

In the OCC model, quantitative aspects of emotions are described in terms of *potentials*, *thresholds*, and *intensities*. For each of the 22 emotions, OCC provide a list of variables that affect the intensity of that emotion if its eliciting conditions hold. Perhaps somewhat misleadingly, the idea is that the weighted sum of these variables equals the emotion's potential. The intensity of an emotion is defined as its potential minus its threshold, or zero if the threshold is greater than the potential. The workings of thresholds of emotions are not specified by OCC, but they probably depend on global variables indicating an agent's 'mood.' For example, if an agent is in a good mood, the thresholds of the eleven negative emotions are increased, causing a lower (or zero) intensity to be associated with a negative emotion, if one is triggered. Emotions that are assigned a positive intensity may in turn influence the mood of an agent, entangling the dynamics of long-term moods and short-term emotions.

The main idea behind separating emotion potentials and intensities is to be able to reason about why some emotions have no effect on an agent (i.e. the agent is not aware of the emotion and it does not influence its behavior) even though their eliciting conditions hold; namely, when the eliciting conditions of an emotion hold (as defined by a formal qualitative model), but its potential is calculated to be below its threshold, an agent can still recognize that an emotion has been triggered and reason about why it does not affect him (e.g., maybe the mood of the agent was 'too good' for him to be affected by shame, even though the agent is aware it performed a blameworthy action).

3 From a Qualitative to a Quantitative Formalization of Emotions

In the construction of a quantitative model of emotions, we build upon the qualitative model of Steunebrink *et al.* [2]. This qualitative model is an extension of the KARO framework [3, 4], which is a mixture of multi-agent dynamic logic and epistemic / doxastic logic, additionally providing several (modal) BDI operators for dealing with the motivational aspects of artificial agents. In the qualitative model, emotions are represented as *emotional fluents* that can be reasoned with in the object language. There are 22 types of emotional fluents, one for each of OCC's emotion types. Each emotional fluent takes between one and three arguments, namely, the *objects* of the emotion in question. For example, the single object of joy is a goal formula (containing one or more subgoals which the agent experiencing joy has just accomplished), whereas the three objects of gratitude are an agent, the (praiseworthy) action it performed, and a conjunction of subgoals it accomplished for the agent experiencing gratitude. Moreover, for each of the 22 emotion types, there is a corresponding set, defined in the semantics of the language, which contains the appropriate objects for that emotion type. An emotional fluent then holds in the object language if and only if the tuple of its objects is an element of the semantically defined set corresponding to that emotional fluent.

The crux here is that the conditions under which these semantically defined sets contain certain elements (i.e. the objects of an emotion) correspond closely to the eliciting conditions of the emotion type in question as prescribed by the OCC model. For example, because the OCC model defines joy as *being pleased about a desirable event*, the semantic set $Joy(a, s)$ for an agent a in state s only contains an event κ if an action has just been performed that has led the agent to become aware of the accomplishment of one or more previously unachieved (sub)goals. Defining emotions by using syntactic emotional fluents backed by semantic sets (i.e. $s \models \mathbf{joy}_a(\kappa) \;\Leftrightarrow\; \kappa \in Joy(a, s)$) has as advantage that propositions in the object language containing emotional fluents can actually be proven semantically. In other words, there is no need for defining emotions as abbreviations of other formulas; in fact, none but two of the emotions (i.e. hope and fear) can be written as one side of a bi-implication in the object language of the qualitative model.

The quantitative model that we propose is built on top of the described qualitative model as follows. The satisfaction of an emotional fluent in a certain state (e.g., $s \models \mathbf{joy}_a(\kappa)$) is regarded as a *trigger* for associating a potential, threshold, and intensity with the emotional fluent and for calculating their quantities. The potential and threshold of each triggered emotion are only defined for the state in which the corresponding emotional fluent holds. Emotion intensities, on the other hand, endure for some time, usually decreasing over time. In order to model enduring quantities, the dynamic logic underlying the qualitative model will need to be extended with a state history. In a dynamic logic (such as KARO), a history of states visited by the multi-agent system and actions taken in those states can easily be kept in the Kripke structure of actions. Given such a history of the multi-agent system, a (collective) memory of emotions experienced by each agent in each state can be constructed; namely, as a set containing for each state in the history all emotional fluents that hold in that state. Emotion intensities are then defined for all emotional fluents that are contained in the emotional memory of the multi-agent system. So potentials and thresholds are only associated with emotional fluents that hold in the actual state, whereas intensities are associated with each emotional fluent in the emotional memory, making them persist over multiple states of the multi-agent system.

Because functions defining the values of intensities are obviously dependent on a time parameter, the dynamic logic of the qualitative model must be further extended to include temporal information so that an explicit representation of time is available to the intensity functions. The intensity of an emotion can be modeled using any function (this is application-dependent), but an inverse sigmoid function could be a natural default choice. However, as an inverse sigmoid function only reaches zero in the limit, an additional cut-off threshold must be set so that the intensity of an emotion (together with its effect on an agent's behavior) can be discarded within a finite amount of time. Note that in the qualitative model there exist several conditions under which a pair of opposing emotions are triggered simultaneously, but this implies by no means that the intensities associated with these opposing emotions should be the same, or be related at all (with the exception of hope and fear, who's intensities should be complementary and always sum to a constant [1]). For example, it is possible that gloating and pity with respect to the same agent and event are triggered simultaneously, but that the emotional fluent corresponding to gloating gets assigned zero intensity, whereas pity gets assigned some positive intensity, because the other agent is being admired. Here we see that even though the qualitative model prescribes that in a certain situation opposing emotions can be experienced together, the quantitative model does not have to lead to mixed feelings.

In order to actually define the emotional 'characters' of agents for some application, 66 functions must be specified for each agent, i.e. a potential, threshold, and intensity function for each of the 22 emotion types. (Although this may appear to be a lot, many of these functions may be very similar.) For the running example, consider for the emotion "joy" the potential function $\mathcal{P}_{\mathbf{joy}}$, the threshold function $\mathcal{T}_{\mathbf{joy}}$, and the intensity functions $\mathcal{I}_{\mathbf{joy}}$ (these are three of the 66 mentioned functions). In general, the parameters of each of these 66 functions are the current state of the multi-agent system, the time, and the objects of the emotion in question, because from these parameters all necessary information about the state of mind of an agent can be derived. The po*ten*tial functions must be defined as the weighted sum (denoted as $\overrightarrow{w} \cdot \_$) of the variables (denoted as $\overrightarrow{x}$) that according to OCC affect intensity, e.g., $\mathcal{P}_{\mathbf{joy}}(a, s) := \overrightarrow{w} \cdot \overrightarrow{x} = q_p$ (we assign this quantity to q_p for brevity). How these variables themselves (e.g., effort, desirability, likelihood) are calculated and what weights are used are again application-dependent and defining for the emotional 'character' of the agent. The OCC model does not prescribe any particular constraints for the threshold functions, but they are hinted to depend on the 'mood' of the agent, e.g., $\mathcal{T}_{\mathbf{joy}}(a, s) := -w \cdot \mathit{Mood}(a, s) = q_t$ (we assign this quantity to q_t for brevity).

Note that because joy is a positive emotion, the mood must negatively influence the emotion's threshold (denoted as $-w \cdot \_$), i.e., a positive mood must lower the threshold so that the intensity will be greater; indeed, there should be no minus sign in the threshold functions for the negative emotions. As mentioned above, a (collective) emotional memory *Emem* is defined over the state history of the multi-agent system, so if at some later time point t' the multi-agent system is in state s', one can retrieve the potential and threshold quantities (q_p and q_t in this example) of the triggered emotions, e.g., $(\mathbf{joy}_a(\kappa), q_p, q_t, t) \in \mathit{Emem}(s')$. (Note that the framework should be extended with temporal information so that the time of triggering t can also be recorded.) An intensity quantity is then associated with each entry in *Emem*. The intensity functions can by default be inverse sigmoid functions, with the difference between the potential and threshold as the numerator (if this difference is negative, the intensity should be set to zero altogether), and the difference between the current time and triggering time in the exponent, e.g., $\mathcal{I}_{\mathbf{joy}}(a, s', t') := \max(0, \frac{q_p - q_t}{1+e^{(t'-t-\mu)\delta}} - \theta)$. Some variables specifying the inverse sigmoid function also need to be set, i.e. the half-life μ, fall-off speed δ, and cut-off threshold θ, such that $\mu\delta \approx -\ln 0.01$ and $0 < \theta \ll q_p - q_t$. Note that other types of intensity functions are also possible (especially for hope and fear).

4 Related Work, Conclusions, and Future Work

Previous work on specifying and implementing emotions carried out by Meyer [3] and Dastani [5] follows Oatley & Jenkins' model of emotions [6] and comprises only four emotion types (happy, sad, angry, fearful), but quantitative aspects of emotions were not considered. Our work is also similar to other computational models of emotions, such as EMA [7], CogAff [8], and the work of Picard [9].

In this paper, we have provided an overview of how the conditions eliciting the 22 emotions of the OCC model have been formalized in a qualitative model. We have described how this qualitative model can be extended to a quantitative model of emotions. Specifically, we have described how potentials, thresholds, and intensities can be associated with the qualitative emotions, and how and when their values should be calculated. Issues such a computational complexity and the possible need to empirically determine parameter settings have been taken into account when developing this formalization, but we have not yet explicitly dealt with these issues and justified our choices.

In future work, the dynamics of a 'mood' and its influence on threshold functions has to be further taken into account. It also remains to be investigated how functions for calculating e.g. effort, desirability, and likelihood can be defined. Furthermore, it remains to be studied what the actual effects of emotions with positive intensities should be on the deliberation of an agent, i.e., in what way emotions can function as heuristics for focussing the attention of the deliberation process of an agent.

References

1. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, UK (1988)
2. Steunebrink, B.R., Dastani, M., Meyer, J.-J.Ch.: A logic of emotions for intelligent agents. In: Proceedings of AAAI'07, AAAI Press (2007)
3. Meyer, J.-J.Ch.: Reasoning about emotional agents. In de Mántaras, R.L., Saitta, L., eds.: Proceedings of ECAI'04, IOS Press (2004) 129–133
4. Meyer, J.-J.Ch., Hoek, W.v.d., Linder, B.v.: A logical approach to the dynamics of commitments. Artificial Intelligence **113** (1999) 1–40
5. Dastani, M., Meyer, J.-J.Ch.: Programming agents with emotions. In: Proceedings of ECAI'06. (2006) 215–219
6. Oatley, K., Jenkins, J.M.: Understanding Emotions. Blackwell Publishing, Oxford, UK (1996)
7. Gratch, J., Marsella, S.: A domain-independent framework for modeling emotions. Journal of Cognitive Systems Research **5**(4) (2004) 269–306
8. Sloman, A.: Beyond shallow models of emotion. Cognitive Processing **2**(1) (2001) 177–198
9. Picard, R.W.: Affective Computing. MIT Press (1997)

Using the Emotional Relation of Topics for Text Mining Based Recommendation Systems

Bernd Ludwig, Stefan Mandl, and Jan Schrader

Chair for Artificial Intelligence, University of Erlangen-Nürnberg
Am Weichselgarten 9, 91058 Erlangen (Germany)
<firstname>.<lastname>@informatik.uni-erlangen.de

Abstract. This paper presents a computional approach to classify (german) descriptions of TV-programmes based on emotional generalization of queries. Emotions are represented formally in a two dimensional coordinate system. Each emotion is positioned with respect to its valence and arousal – two fundamental categories in psychological research on emotions. Text mining is enabled by applying WHISSEL's and PLUTCHICK's results on the relatedness between emotions and words (see [1]). The approach is implemented in a TV recommender system that allows the user to set the emotional profile for desired TV programmes.

1 EPG Data, Recommender Systems, and the Real User

Via satellite the Electronic Programme Guide (EPG) provides an enormous amount of information about TV programmes with natural language descriptions about the content of programmes. Viewers are overwhelmed by the huge number of channels and programmes when they select a programme to watch.

For the design and implementation of TV recommendation systems sophisticated user models such as [2] are used. In order to allow for default reasoning, stereotypes for users are applied which are based on the analysis of the average user's lifestyle (see [3]). Much attention is paid on the issue on how to design an attractive, functional, and easy-to-use graphical interface between users and the recommendation system (see [4]). In order to increase the user's confidence in the system proposals, the generation of trust-worthy suggestions that take programmes watched ealier into account has been studied in detail (see [5]). For the implementation of the search, different approaches and theories of reasoning have been applied: (statistical) classifiers such as neural nets [6], fuzzy logic [7], similiarity based reasoning [8] – to name the most prominent ones. We use a text mining based approach as the necessary computations can be carried out quickly. Performance is an important factor as the system is dedicated to run in real-time on an embedded platform in TV sets.

In a user study [9] conducted as part of the research project EMBASSI (see [10]), candidates were situated in front of a computer display that suggested an automatic recommendation system to be at work. The users were allowed to ask arbitrary questions about available TV programmes. A Wizard-of-Oz provided

Engl. adj.	*German adj.*	*Dornseiffgroups*	*Activ*	*Eval*	*Angle*
Adventurous	Abenteuerlich	9.72,10.23,10.38	4.2	5.9	270.7
Affectionate	Herzlich,Liebevoll	4.50,10.51,15.33,10.49,10.52	4.7	5.4	52.3
Afraid	Besorgt,Bange	10.41,10.13,4.29,9.40	4.9	3.4	70.3
Aggressive	Aggressiv,Dynamisch, Energisch,Feindlich	10.30,10.55,18.32,9.35,5.35, 8.21,9.6,15.78,15.46	5.9	2.9	232.0
Agreeable	Angenehm,Annehmlich	9.54,10.10,10.52,12.13,11.46	4.3	5.2	5.0
Amazed	Erstaunt	10.29	5.9	5.5	152.0

Fig. 1. Some emotion words with V/A coordinates

responses. The experiments showed that users express emotional attitudes they desire the programme to have, or even their own emotions hoping the system would come up with proposals that match their mood:

Liebe, Romantik (Love, romance) | *Entspannen (relax)* | *Show, Witz (Show, fun)*

2 Our Approach

The examples above indicate that current recommender systems do not use the optimal feature for proposals. Ideally, the best user interface for a TV recommender system would support natural language and be operated with a microphone. However, current TV hardware does not provide the resources for real-time and speaker independent speech recognition. Additionally, processing the semantics of such kind of natural language input is infeasible in practice.

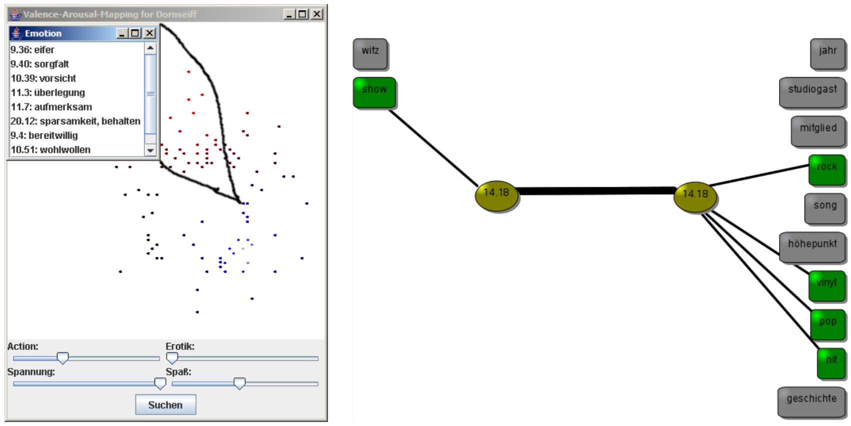
2.1 Psycholinguistic Knowledge for Emotion Detection in Texts

Still, we would like to keep the emotional aspect as a recommendation criterion. For this purpose, we use the valence/arousal model by Whissel (described in [1], pp. 39), which gives each emotion represented by an adjective a fixed position in a two dimensional coordiante system. Valence indicates emotion quality, hence whether it is considered a positive or negative emotion. Arousal stands for the activation level an emotion possesses "i.e. the strength of the person's disposition to take some action rather than none" ([1], p. 39). The table in Fig. 1 shows some English adjectives along with their German translation and their position in the two dimensional valence/arousal model (see colums *activ* and *eval*).

The valence/arousal model allows us to relate emotions and vocabulary. On the other hand, it allows to localize moods such as *fun*, *sadness*, *chill*, or *action* (see the sliders in Fig. 2(a)) in certain regions of the model. This relationship between moods and vocabulary can be exploited for finding recommendations: if the user wants to see something chilling then the adjectives assigned to the emotions in the region for chill are good features to search for.

2.2 Extending the vocabulary

We search for features in EPG free text descriptions of TV programmes (see section 1) that contain many words that are not related to a point in the va-



(a) User interface

(b) Relation between words and topics in a query and in a free text description

Fig. 2. From emotions to words

lence/arousal model. However, we can still extract more information out of free text descriptions as we use a lexicon that organizes German words around topics [11]. Each adjective in the valence/arousal model is also found in at least one topic group together with a number of other words (see Fig. 2(a)). All these words serve as features as well. Fig. 2(b) gives an example.

2.3 Computing Recommendations

The most important advantage of this approach is that we do not need literal matches of words for a hit, but of topics only (as indicated by the black arc in Fig. 2(b). In order to rank the words which serve as features among each other we compute their semantic distance in GermaNet, the German version of WordNet. Along with the frequency of each topic, we use this semantic distance to compute the TF×IDF-value of each identified topic as in the following example:

Topic ID	... 15.1	15.16	15.61	15.70	15.9	16.9	17.20	19.1	2.14	...
TF × IDF	... 1.17	0.69	0.69	0.69	0.69	0.69	0.69	0	0	...

As most approaches to text mining do, we use the standard scalar product as similarity measure between a free text description d and a query q:

$$\mathrm{sim}(d,q) = \sum_i (d[i] \cdot q[i])$$

If d and q are normalized to length 1, the scalar product $\mathrm{sim}(d,q)$ is a real number that tells how big the angle between q and d is. For our task of finding texts that are as close to the query as possible, ideally $\mathrm{sim}(d,q) = 1$. This means that d and q are identical. In any case, if we have N texts with feature vectors d_k ($1 \leq k \leq N$) we can rank the texts so that $\mathrm{sim}(d_k, q)$ is descending when k is ascending. This means d_1 matches q best and therefore is the best recommendation the system can provide.

3 Results and Conclusion

The evaluation of a system such as ours is tricky as the correlation of programme descriptions with emotions is highly personal. Experiments with 1600 programme descriptions give good results for some slider bar categories like 'Erotik', but quite bad results for others, like 'Anspruch'. The different quality can be attributed to the fact that 'Anspruch' is hard to map properly to the V/A-diagram.

We conclude that the presented approach is feasible, still, large scale experiments with real users have to be conducted.

References

1. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.: Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine **18** (2001) 32 – 80
2. Ardissono, L., Gena, C., Torasso, P., Bellifemmine, F., Difino, A., Negro, B.: User modelling and recommendation techniques for personalized electronic program guides. In Ardissono, L., Kobsa, A., Maybury, M.T., eds.: Personalized Digital Television – Targeting Programs to Individual Viewers. Volume 6 of Human-Computer Interaction Series. Springer (2004) 3–26
3. Gena, C.: Designing tv viewer stereotypes for an electronic program guide. In: Proc. UM2001 Workshop on Personalization in Future TV (TV01), Sonthofen (2001)
4. van Barneveld, J., van Setten, M.: Designing usable interfaces for tv recommender systems. In Ardissono, L., Kobsa, A., Maybury, M.T., eds.: Personalized Digital Television – Targeting Programs to Individual Viewers. Volume 6 of Human-Computer Interaction Series. Springer (2004)
5. Buczak, A.L., Zimmerman, J., Kurapati, K.: Personalization: Improving ease-of-use, trust, and accuracy of a tv show recommender. In: Proceedings of the TV'02 workshop on Personalization in TV, Malaga (Spain) (2002)
6. Zhang, T., Iyengar, V.S.: Recommender systems using linear classifiers. Journal of Machine Learning Research (2002) 313–334
7. Yager, R.R.: Fuzzy logic methods in recommender systems. Fuzzy Sets Syst. **136** (2003) 133–149
8. McSherry, D.: A generalised approach to similarity-based retrieval in recommender systems. Artif. Intell. Rev. **18** (2002) 309–341
9. Nitschke, J., Hellenschmidt, M.: Design and evaluation of adaptive assistance for the selection of movies. In: Proceedings of IMC 2003 "Assistance, Mobility, Applications", Rostock (2003)
10. Herfet, T., Kirste, T., Schnaider, M.: Embassi – multimodal assistance for infotainment and service infrastructures. Computers and Graphics **25** (2001) 581–592
11. Dornseiff, F., Quasthoff, U.: Der deutsche Wortschatz nach Sachgruppen. De Gruyter, http://www.wortschatz.uni-leipzig.de (2004)

Application of emotion recognition methods in automotive research

Martin A. Tischler, Christian Peter, Matthias Wimmer, Jörg Voskamp

[1]Customer Research Center, DaimlerChrysler AG, Böblingen
[2]Fraunhofer Institut für Graphische Datenverarbeitung, Rostock
[3]Lehrstuhl für Informatik IX, Technische Universität München

Abstract. This paper reports on a pilot study applying emotion recognition technologies developed for Human-Machine-Interfaces in automobile research. The aim of the study was to evaluate technologies for quantifying driving pleasure in a close-to-reality scenario. Results show that car driving scenarios pose particular requirements on emotion recognition technologies which could be met by modifications of current systems.

Keywords: Driving Pleasure, Emotion Recognition, Affective Interaction, Human-Machine-Interaction, Affect Sensors, Facial Expression Interpretation

1 Introduction

A task in vehicle development is to configure vehicles not only according to security aspects, but also to ensure passengers feeling well and enjoying their drive. The pleasure of driving a car or, more general "the positive emotional interaction with the product car" is vitally important for the acceptance and the success of a vehicle model today [1]. Driving pleasure is determined particularly by the driving feeling, which arises from the complex reciprocal relationship driver-vehicle-environment.
Measuring emotional reactions of drivers in close-to-reality environments opens extended views of the interaction driver-vehicle and supplies important information to design vehicles to better comply to and satisfy constantly increasing customer demands.

This paper reports on a pilot study applying emotion recognition technologies developed for Human-Machine-Interfaces in automobile research. The aim of the study was to evaluate technologies to quantify driving pleasure in a close-to-reality scenario.

2 A Short Overview of Research Findings on Driving Pleasure

The topic “driving pleasure” was primarily explored by means of interviews or studies in driving simulators. Only very few studies have been performed to capture positive experiences of non-professional drivers in natural environments [2]. Beyond that no scientific investigation is known assessing driving fun of non-professional drivers with quantitative measures. While common studies rely solely on retrospective questionnaires and interviews, for product development purposes new methods to quantify the emotional state of the driver are needed. They would provide objective information in very small time intervals without interference of the driver. This accurate information is desirable, because in future research it is intended to combine objective conditions for instance driving dynamic measures with the emotional reaction of the driver. It is remarkable that the positive aspects of the activity "driving a car" were neglected so far by psychological research, although studies showed consistently that the majority of the people enjoy the activity driving [3].

3 Technologies Applied

Emotions are manifest in physiological changes controlled by the autonomous nervous system [4]. They can either be observed by humans, such as facial features, gestures, or changes of the voice, or they can be less obvious, like heart rate, blood pressure, skin temperature, electro-dermal activity, or pupil dilation. Sensors and modern computing technology today offer possibilities to access a variety of emotion-related parameters, including those not observable by humans, in a non- or minimally intrusive way. In the following we shortly describe the technologies used to infer the emotional state of drivers.

Emotion Recognition from Speech: Speech parameters have been examined for correlations with emotions for a long time. In connection with computers and other applications, results are increasingly acceptable [5].

At Fraunhofer IGD we work on tools to extract emotion information from a user's voice. While our research covers quality features of the speech signal, prosody, and semantics, in this study only emotion-related features of the speech signal have been analysed [6, 7]. For analysing the data, techniques from data mining and knowledge discovery have been applied. The extracted information then is used to build two classifiers based on the valence/arousal model [cf. 8, 9], one for detecting valence and one for arousal. They were trained on 10% of the collected data and tested using 10-fold cross-validation. They were then applied on the remaining 90%.

Emotion Recognition from Facial Features: During the last decade, a lot of research on facial expression recognition has been conducted [10]. Model-based approaches have proven to be highly appropriate for interpreting images in real-world scenarios. At the Technische Universität München, we developed algorithms that robustly localize facial features and seamlessly track them through image sequences in order to interpret facial expressions. They are able to determine the six universal emotions specified by Ekman [11], and to give evidence about the magnitude of the visible expression.

An Active Shape Model [12] represents the contour of the face and it is fit to the camera image via learned objective functions [13]. Successfully tracking a human face during some period of time, our system determines the motion of the individual parts such as eyes, chin, cheeks, and lips. A previously learned classifier uses this information to infer the currently visible facial expression.

Emotion Recognition from Physiological Data: Emotion-related changes of physiological parameters have been studied for a long time [14] and can be considered to be the emotional signs best understood today. A number of proof-of-possibility studies for emotion detection in non-lab settings have been performed, using either commercially available sensor equipment for physiology data or experimental devices. Among the latter are stationary, mobile, or wearable systems [15, 16], or furniture equipped with sensors [17].

At Fraunhofer IGD we developed a wireless and easy to use sensor system for collecting the emotion-related physiological parameters skin resistance, skin temperature, and heart rate [16]. As can be seen in figure 1, the EREC system (for *E*motion *REC*ognition) consists of a glove hosting sensing elements for skin temperature and skin resistance, a wrist pocket containing sensing electronics, and a base unit performing data enhancement steps and storing the data on an SD card. The system is particularly suitable for mobile applications since it provides for both, immediate data communication to a nearby computer, and storage of data on an exchangeable memory card. It has been used in several studies of different application fields, inside and outside a lab, and proved to be stable and reliable [18]. In this study, the wireless option has proved to be very useful for checking the quality of data in the setup phase. During the study, data have been recorded on memory card.

4 Study

The objective of this pilot study was to apply sophisticated emotion recognition techniques into a test car and to demonstrate their applicability for the described purpose. Three research institutes cooperated and provided two cars with sensory equipment to conduct a study in a close-to-reality environment. For safety reasons it was decided to do the study on a separated proving ground in Southern Germany.

A middle class saloon car ('new car') with a performance of 170 hp and a 25 year old previous model of this series ('forerunner') were used for a comparative study. Eight participants (age 33-53), all non-professional drivers, were invited. They drove on their own on three different courses: an autobahn-like course, a course similar to a rural road and a demanding handling course. Figure 1 shows the technical setup.

4.1 Setup

Technical setup: For recording the speech data, a capacitor lapel microphone has been connected to a Marantz PMD 660 audio recorder which stored the data on a removable flash card. The microphone was attached to the safety belt, at chest level. The recorder was placed on the backseat.

The EREC glove and chest belt were put on by the test persons at the beginning of the session. Sufficient time was allowed for the users to become familiar with the components and adjust them for most possible comfort. The base unit was placed in the center storage shelf of the car. For the facial data, a small video camera was fixed at the front column of the car. The video signal was recorded by a Sony Digital Video Walkman, which was placed on the backseat.

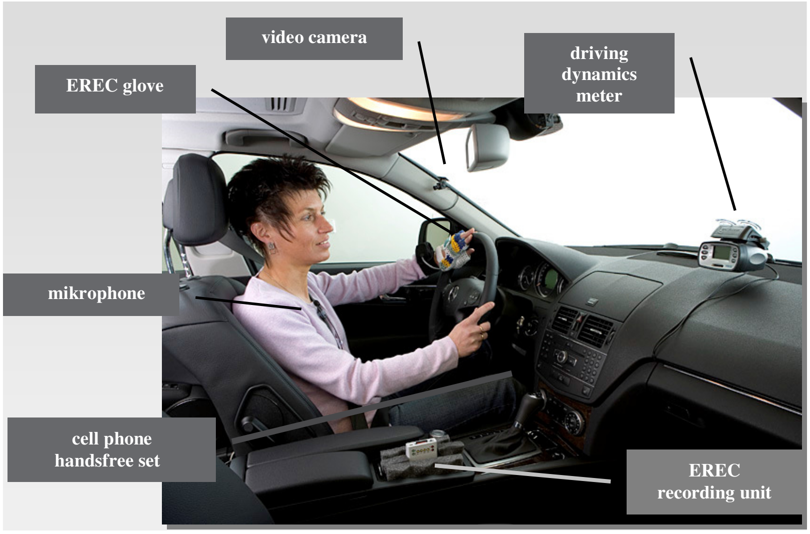


Fig 1. Technical setup of the study

Both cars were equipped with cell phones and hands free sets, in order to give instructions to the drivers. For logging the actual position and the basic driving dynamic data the GPS-based Racelogic VBOX Mini was used.

Additional data acquisition: Additional to the quantitative measurements of the emotional state, the participants were asked via questionnaires and in interviews before, during, and after driving about their subjective feelings. These results and the detailed methods are not presented here.

4.2 Results

Speech analysis: Parameters found to be most related with emotional speech were changes of pitch, intensity, and energy changes over frequency bands. We achieved an average confidence of 1.1-1.2 for valence and 0.9-1.0 for arousal.

Valence is found to be generally higher for the new car (6.11 vs. 5.45), particularly so for the handling course (6.62 vs. 5.58). Arousal is just slightly higher for the new car (6.0. vs. 5.95) but more significantly at the handling course again (6.13 vs. 5.47).

Facial features: Our system recognizes well-posted facial expressions that incorporate a lot of facial motion best, such as happiness or surprise. In these cases, facial motion is well-distinguished from head motion and background motion. In the car-scenario, we are interested in determining happiness, however, the results of our investigation show that drivers don't express this emotion strongly while being alone in the car. The presence of non emotional-related head motion while observing the traffic makes facial expression recognition very difficult.

The results of our investigation illustrate that the drivers express happiness with a magnitude of 5.5% on average and with a peak value of 19.8% in the new car. In contrast, happiness is expressed with a magnitude of 4.2% in the forerunner car (15.0% peak value). The underlying scale is derived from the image sequences of the Cohn-Kanade-Facial-Expression-Database [19], where the first image denotes a neutral face (0%) and the last image denotes the maximum expression (100%).

Physiology: Physiological data have been gathered with different reliability. While skin resistance and skin temperature proved to be reliable data sources, heart rate data from the chest belt were less secure. As elaborated in consecutive tests, the chest belt data link was most likely affected by electro-magnetic interferences from servo and wiper drives. Further we found that the amount of training data for classifiers was too small to achieve results of sufficient confidence. Due to these circumstances we decided to not take physiological data into account for the final results.

5 Discussion

Speech data reveals that in the forerunner car arousal rises significantly on the speedway, compared to the other courses. In combination with results from interviews this leads to the conclusion that drivers of the forerunner car feel more fearful at this part of the course. With the new car arousal is higher on the speedway as well. Interview results lead to the conclusion that this increase in arousal is due to feelings of happiness, fun and joy.

Recognizing facial expressions in real world scenarios poses a higher challenge to both correctly fitting the face model and classifying the facial expression. In lab environments colour, texture, and motion of the background as well as the brightness and the position of the light source are controllable. While showing facial expressions in the laboratory, people are often asked to stop talking and reduce head movements as well as facial muscle activity that are not related to facial expressions. However, these activities occur often in real world scenarios and therefore, they have to be taken into account.

For the physiological data it can be said that other means for detecting heart rate needs to be found for future studies in cars. The main reason for the physiological data not being usable in this study was the lack of sufficient training data. This was mainly due to the tight schedule which didn't allow for an extra data collection session with each participant. Our plans to retrieve sufficient training data at the beginning of the main session proved to be unsuitable.

Concluding it can be said that all modalities give evidence that valence, i.e. the feeling well of the test persons was generally much higher in the new car. This study can also be seen as a prove that affect sensors can be applied to measure positive emotions in real-world settings, and that only a combined analysis of all modalities' results leads to useful results.

References

1. Automobilwoche, 23.4. 2007, p. 12
2. Toda, Y., & Kageyama, I.: Measuring a Driver's Pleasure during Driving. Rev. of Automotive Engineering 28, 89--93 (2007)
3. Ellinghaus, D., Steinbrecher, J.: Fahren auf Landstraßen. Traum oder Albtraum? Untersuchungen zum Fahrverhalten und Fahrvergnügen von Pkw-Fahrern auf Landstraßen. Uniroyal Verkehrsuntersuchung 28. Köln/Hannover (2003)
4. Lane, R.D., Nadel L. (eds.): Cognitive Neuroscience of Emotion. Oxford University Press (2000)
5. Douglas-Cowie, E., Cowie, R., Campbell, N.: Speech and Emotion. Speech Communication 40, 1--3 (2003)
6. Fernandez, R.: A Computational Model for the Automatic Recognition of Affect in Speech. Ph.D. thesis, Massachusetts Institute of Technology (2004)
7. Lee, C.M., Narayanan, S.S.: Toward detecting emotions in spoken dialogs. IEEE Transactions on Speech and Audio Processing, 13(2) (2005)
8. Picard R.W., Vyzas E., Healey J.: Toward Machine Emotional Intelligence - Analysis of Affective Physiological State. IEEE Trans. on Pattern Analysis and Machine Intelligence, 23(10) (2001)
9. Russell, J. A.: A Circumplex Model of Affect. Journal of Personality and Social Psychology 39, 161--1178 (1980)
10. Cohen, I., Sebe, N., Chen, L., Garg, A., Huang, T.: Facial expression recognition from video sequences: Temporal and static modeling. CVIU, 91(1--2), 160—187 (2003)
11. Ekman, P.: Universals and Cultural Differences in Facial Expressions of Emotion. In: Cole, J. K. (ed.), Nebraska Symposium on Motivation, vol. 19. Lincoln: University of Nebraska Press (1971)
12. Cootes, T.F., Taylor, C.J.: Active shape models - smart snakes. In: Proc. of the 3rd British Machine Vision Conference 1992, pp. 266--275. Springer (1992)
13. Wimmer, M., Pietzsch, S., Stulp, F., Radig, B.: Learning robust objective functions with application to face model fitting. 29th DAGM Symposium, Heidelberg, (2007)
14. Ax, A.: The physiological differentiation between fear and anger in humans. Psychosomatic Medicine 55(5), 433--442 (1953)
15. Haag A., Goronzy S., Schaich P., Williams, J.: Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System. In André et al. (eds.): Affective Dialogue Systems, pp 36--48, Springer (2004)
16. Peter, C., Ebert E., Beikirch, H.: A Wearable Multi-Sensor System for Mobile Acquisition of Emotion-Related Physiological Data. In: Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction, Beijing 2005. Berlin, Heidelberg, New York: Springer (2005)
17. Tan H.Z., Slivovsky L.A., Pentland, A.: A Sensing Chair Using Pressure Distribution Sensors. IEEE/ASME Transactions On Mechatronics, 6(3) (2001)
18. Peter C., Schultz R., Voskamp J., Urban B., Nowack N., Janik H., Kraft K., Göcke, R.: EREC-II in Use – Studies on Usability and Suitability of a Sensor System for Affect Detection and Human Performance Monitoring. In: Jacko, J. (ed.) Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, pp. 465--474. Berlin Heidelberg: Springer (2007)
19. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. International Conference on Automatic Face and Gesture Recognition, pp. 46--53 (2000)

True Feelings: Functionalist and Descriptionalist Modeling of Emotion

Joscha Bach

Universität Osnabrück, Institute for Cognitive Science, 49069 Osnabrück, Germany
joscha.bach@gmail.com

Introduction

While the individual sciences of cognition continue to make valuable contributions to understanding the mind, they do so within different and sometimes incompatible methodological and terminological frameworks. This is especially true with respect to the study of emotion. Yet the question what it is that makes a system emotional (i.e. “what does it mean to have an emotion?”), and how this can be captured in a computational model is rarely discussed in depth, nor is it differentiated from the question of how emotional behaviour can be described (i.e. “what does emotional behaviour look like?”). I will briefly look at some of the dominating paradigms in artificial emotion research, such as appraisal models and their conceptual relatives, at architectural models and at state space models, and compare them with respect to their explanative approaches. I would like to make the argument that most of the available models focus on descriptions and do not address the functional definition of emotion that necessarily underlies that description. Without such an understanding, however, it will be difficult to capture the role of emotion within cognitive processes, such as perception, memory, action control, planning and anticipation.

What does define the difference between *functionalist* and *descriptionalist* models?– Without a recourse into an entirely non-trivial domain of philosophy, let us just look at a simple example: a car-engine. Such an engine offers a lot to describe and measure: the dimensions and densities of different parts, the fluctuation of temperature patterns during the movement of the parts and so on. Yet to an engineer that wants to understand how the engine works, and how it could be recreated or replaces by an equivalent mechanism, these descriptions need to be applied to a functional frame of reference, which captures the underlying necessary and sufficient conditions for the observable performance. Thus, we will need to apply a conceptual structure which separates the engine into functional parts and relations. Without such a structure, we will not only be unable to distinguish car-engines from different mechanisms or to explain failures of the system, but we may also find us confronted with mysterious “explanatory gaps” between the description and the system we want to describe.

Most models of emotion account for the descriptional aspect ([1], [2], [3], [4]):

- What emotional states can a system have?
- Which events (in the environment) trigger emotions?
- What behavior is in turn triggered by emotions?
- How can emotional states be expressed?
- How can emotional states of humans be recognized?

While this is useful, the functional aspect of emotion–requiring the dissection of the emotional system itself in its *constituents*, and the identification of the *relationships* between them–is surprisingly often ignored. The functional modeling of emotion will require thorough conceptual analysis: how does emotion configure cognition? What is the relationship between emotion and motivation, and how does emotion interact with physiological and cognitive drives? What constitutes the phenomenal dimension of affects, and how and when are affects bound to objects?

The Example of Appraisal Models

Have a look at the well-known conceptual analysis of emotion by Ortony, Clore and Collins [2]: Here, emotions are distinguished into three main classes with respect to their *object*. Objects of emotion are either the consequence of some event, an aspect of some thing, or the action of some agent. (This marks a difference to 'emotion space' approaches–see below–which are relatively indifferent to the object of an affect and focus on the nature of the affect itself). From this perspective, the difference between social emotions (the appraisal of the actions of oneself or other agents) and event-based emotions (hope, relief) becomes visible.

At the first stage, the OCC model distinguishes for each group of emotions whether they are positively or negatively *valenced*; for events this is their degree of pleasurableness, resentment etc., for agents it is approval or disapproval, and for objects it is their desirability, and thus the degree of attraction or repulsion.

Event-related emotions are further separated depending on whether the consequences apply to others or oneself, and whether the event is already present or not. Present events may lead to joy or distress, anticipated events to hope and fear. If the anticipated events materialize, the reactions are either satisfaction or the confirmation of fear, if they do not occur, then the consequence is either disappointment or relief.

Emotions with respect to events happening to others depend on the stance taken towards these others–if they are seen positively, reactions may be happiness for them, or pity (if the event has negative consequences). If the others are resented, then a positive outcome may lead to envy and resentment, a negative to gloating.

Agent-oriented emotions (attributions) depend on whether the agent is someone else (who may be admired or reproached), or one self (in which case the emotion could be pride or shame).

Of course, appraisals may also relate to the consequences of events that are caused by the actions of agents. The OCC taxonomy calls the resulting emotions 'attribution/well-being compounds': Here, if oneself is responsible, the reaction may be gratification or remorse, and if the culprit is someone else, it could be gratitude or anger.

The OCC model is an engineering approach that constructs the emotion categories based on systematizing our commonsense understanding of emotion. Every emotion can be specified in a *formal language*, by using threshold parameters to specify intervals of real-valued variables in a weight-matrix. On the other hand, it does not say much about how the cognitive appraisals are realized–this is left to the designer of an architecture for a believable agent.

Another appraisal model, this time rooted in psychological methodology, has been suggested by Scherer [5], [6]. He proposes that emotional states are the result of *stimulus-evaluation-checks* (SECs), which are the equivalent to appraisals and performed by the human cognitive system.

According to Scherer, there are five major SECs, along with a number of sub-checks:

- novelty (with sub-checks for suddenness, familiarity and predictability)
- intrinsic pleasantness
- goal significance (with sub-checks for goal relevance, probability of result, expectation, supportive character, urgency)
- coping potential (with sub-checks for agent, motive, power and adaptability)
- compatibility (asks for conformance to social norms and standards, with sub-checks for externality and internality)

Scherer maintains that every emotion is uniquely defined by a combination of checks and sub-checks [3]. For a model with 14 emotions, which were organized with weights in a space of 15 dimensions (corresponding to the checks and sub-checks), Scherer achieved a level of agreement between his subjects and the model of 77.9%. Scherer also proposes a cognitive architecture with five different sub-systems to go with his emotion model, but his attempt at classifying emotions has a behaviorist core–emotions are only relevant as a link between external stimuli and externalizable responses.

Frequently, however, the onset and degree of emotional episodes does not stand in a strong relationship to a triggering external situation, especially with respect to undirected, basic emotions, like angst and extasy, which may simply result from the neurochemical setup of the brain at the time, or a stimulation of brain areas (see, for instance [7]). This is also true if they are bound to a cognitive content and angst becomes fear, extasy becomes joy. Sometimes, we just have the 'wrong' emotion, and a model that binds all emotional states to appraisals of external stimuli instead of examining the nature of the emotional state itself is going to meet a boundary, beyond which it can not explain its object any more.

Appraisal models [8], [9] like the ones mentioned above treat emotions as pre-defined categories and explain what events trigger them, and how they change behavior and expression. In this view, emotions are triggered by a causal interpretation of the environment [10], [4] with respect to the current goals, beliefs, intentions and relations of the agent. By evaluating these, a *frame* of the appraisal and a corresponding affective state of the agent are set, which in turn enable it to *cope* with the situation. These models are very useful for engineering applications, for instance in social simulation with multi-agent systems [11] and for the creation of believable behavior in animated characters for computer games and movies [12], [13]. From a functionalist point of view, however, appraisal models fall short when it comes to explaining what exactly happens between stimulus and response. To explicate how emotions *emerge* from cognitive functioning, they will have to be identified as features of a broader architecture of cognition [14], [15], and the decomposition of emotions into proto-emotional dimensions (*modulators* [16]) and/or primary or atomic emotions [1], [17].

Explaining the *How* of Emotion

When we treat emotions as aspects of a broader functionalist model of cognition, they become an emergent phenomenon: they appear on a different descriptional level, as the particular way cognitive processing is carried out. For example, *anger* is not just the cooccurrence of negatively valenced events in the environment of an agent, correlated with sanctioning behavior. Anger is a name for a family of complex cognitive configurations, initiated by the failure to attain a goal in the face of an obstacle and is characterized by a low resolution level, which leads to a limited problem solving capacity and neglect for details. Also, the failure increases the sense of urgency, which in turn amplifies the activation level, leading to more impulsive action, and a narrower examination of the environment.

Obviously, such a description rests on the premises of a *cognitive architecture* that supplies us not only with the ability to conceptualize goals and sense failure, but also with activation, urgency, resolution level and a motivational system in general. Here, let us just look at the proto-emotional dimensions (modulators) of such an architecture.

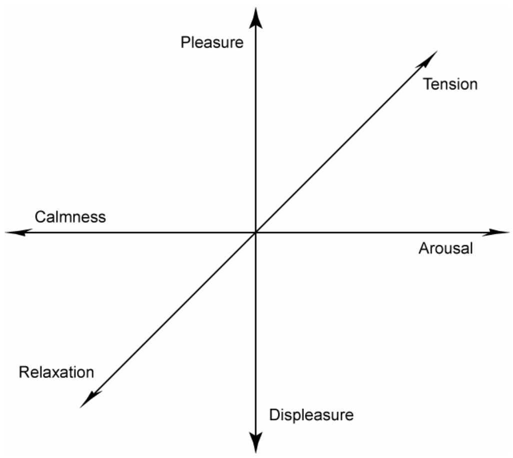


Fig. 1. Dimensions of Wundt's emotional space (see Wundt 1910).

One of the first attempts to treat emotion as a continuous space was made by Wilhelm Wundt in 1910 [18]. According to Wundt, every emotional state is characterized by three components that can be organized into orthogonal dimensions. The first dimension ranges from pleasure to displeasure, the second from arousal to calmness, and the last one from tension to relaxion (figure 1), that is, every emotional state can be evaluated with respect to its positive or negative content, its stressfulness, and the strength it exhibits. Thus, an emotion may be pleasurable, intense and calm at the same time, but not pleasurable and displeasurable at once. Wundt's model has been re-invented by Charles Osgood in 1957, with an *evaluation* dimension (for pleasure/displeasure), *arousal*, and *potency* (for the strength of the emotion) [19] and re-discovered by Ertel [20] as *valence*, *arousal*, and *potency*.

Wundt's model does not capture the social aspects of emotion, so it has been sometimes amended to include extraversion/introversion, apprehension/disgust and so on, for instance by Traxel and Heide, who added *submission/dominance* as the third dimension to a *valence/arousal* model [21].

Note that *arousal*, *valence* and *introversion* are themselves not emotions, but mental configuration parameters that are much closer to the physiological level than actual emotions – we could call them *proto-emotions*. Emotions are areas within the space spanned by the proto-emotional dimensions.

A much more detailed model is supplied by Dörner's Psi theory [16], [15], where the modulators span at least a six-dimensional continous space: Katrin Hille [22] describes it with the following proto-emotional dimensions: *arousal* (which corresponds to the physiological *unspecific sympathicus syndrome* and subsumes Wundt's *tension* and *arousal* dimensions), *resolution level*, *dominance* of the leading motive (usually called *selection threshold*), the level of *background checks* (the rate of the securing behavior), and the level of *goal-directed behavior*. (The sixth dimension is the *valence*, i.e. the signals supplied by the pleasure/displeasure system.) The dimensions are not completely orthogonal to each other (resolution is mainly inversely related to arousal, and goal-orientedness is partially dependent on arousal as well). This way, the emotional dimensions are not just classified, but also explained as result of particular demands of the individual. The states of the modulators (the proto-emotional parameters) are a function of the urgency and importance of motives, and of the ability to cope with the environment and the tasks that have to be fulfilled to satisfy the motives.

Thus, emotions can be *functionally divided* into a lower emotional level (affects):

- Not an independent sub-system, but *aspect of cognition*
- Emotions are emergent property of the modulation of perception, behavior and cognitive processing
- *Phenomenal qualities* of emotion are due to
  - effect of modulatory settings on perception on cognitive functioning
  - experience of accompanying physical sensations

and a higher level (directed emotions):

- *Directed* affects
- Objects of affects are given by motivational system
- Behavioral propensities are given by motivational system or acquired by learning

The available space does not permit to give more than a shallow introduction into the models mentioned above. Yet I think it becomes clear that functional models of emotion do not contradict the descriptional approaches. Instead, they supply them with an architecture, a framework that is eventually necessary to clearly capture the object of our work: to understand the nature of emotion.

References

1. Plutchik, R. (1994): The Psychology and Biology of Emotion. New York: Harper Collins
2. Ortony, A., Clore, G. L., Collins, A. (1988): The Cognitive Structure of Emotions. Cambridge, England: Cambridge University Press
3. Scherer, K. (1993): Studying the Emotion-Antecedent Appraisal Process: An Expert System Approach. In: Cognition and Emotion, 7 (3/4), 325-355
4. Gratch, J., Marsella, S. (2004): A framework for modeling emotion. Journal of Cognitive Systems Research, Volume 5, Issue 4, 2004, 269-306
5. Scherer, K. (1984): On the nature and function of emotion: a component process approach. In K.R. Scherer, and P. Ekman (eds.). Approaches to emotion. Hillsdale, N.J., Erlbaum
6. Scherer, K. (1988): Criteria for emotion-antecedent appraisal: A review. In: V. Hamilton, G.H. Bower, N.H. Frijda (eds.): Cognitive perspectives on emotion and motivation. Dordrecht, Kluwer
7. LeDoux, J. (1992): Emotion and the Amygdala. In The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, 339–351. Wiley-Liss
8. Roseman, I. J. (1991): Appraisal determinants of discrete emotions. In: Cognition and Emotion, 3, 161-200
9. Roseman, I. J., Antoniou, A. A., Jose, P. A. (1996): Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. In: Cognition and Emotion, 10 (3): 241-277
10. Smith, C. A., Lazarus, R. (1990): Emotion and Adaptation. In Pervin (ed.), Handbook of Personality: theory & research, 609-637. NY: Guilford Press
11. Schmidt, B. (2000): PECS. Die Modellierung menschlichen Verhaltens. SCS-Europe Publishing House, Ghent
12. Norman, D. A. (2004): Emotional Design: Why We Love (or Hate) Everyday Things. New York: Basic Books
13. Norman, D. A., Ortony, A., Russell, D. M. (2003): Affect and machine design: Lessons for the development of autonomous machines. IBM Systems Journal, 42 (1), 38-44
14. Minsky, M. (2006): The Emotion Machine. Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. New York, Elsevier
15. Bach, J. (to appear): Principles of Synthetic Intelligence. An Architecture of Motivated Cognition (in preparation)
16. Dörner, D. (1999): Bauplan für eine Seele. Reinbeck: Rowohlt
17. Sloman, A. (2001): Beyond shallow models of emotion. Cognitive Processing: International Quarterly of Cognitive Science, 2(1), 177–198
18. Wundt, W. (1910): Gefühlselemente des Seelenlebens. In: Grundzüge der physiologischen Psychologie II. Leipzig: Engelmann
19. Osgood, C. E., Suci, G. J., Tannenbaum, P. H. (1957): The measurement of meaning. Urbana: University of Illinois Press
20. Ertel, S. (1965). Standardisierung eines Eindrucksdifferentials. Zeitschrift für experimentelle und angewandte Psychologie, 12, 22-58
21. Traxel, W. (1960): Die Möglichkeit einer objektiven Messung der Stärke von Gefühlen. Psychol. Forschung, 75-90
22. Hille, K. (1998): A theory of emotion. Memorandum Universität Bamberg, Lehrstuhl Psychologie II. available online at www.uni-bamberg.de/ppp/insttheopsy/dokumente/Hille_A_theory_of_emotion.pdf