Recognition and Simulation of Emotions

Christian Kleine-Cosack — October 18, 2006

"Seminar: Human-Robot Interaction"
SoSe 2006 Fachbereich Informatik Universität Dortmund

Abstract

Emotions play an important role in the process of human communication establishing an information channel beyond the verbal domain. Accordingly the ability to recognize and understand human emotions as well as the simulation of emotion is a desirable task in the field of Human-Computer-Intelligent-Interaction (HCII). Furthermore the acceptance of autonomous and especially humanoid robots will be directly dependent on their ability to recognize and simulate emotions. This paper will give an introduction into the field of emotion recognition focusing on facial and vocal emotion expressions. In addition the necessity of emotion simulation in HCII will be pointed out on the basis of various communication schemata and an example of an intelligent human robot interaction will be presented.

Contents

Chapter 1

Introduction

A short excursion into modern science fiction may well show what is expected to be an intelligent human-like robot. What can be seen are autonomously acting humanoid robots whose appearance and behaviour can not be held apart from the human ones. It becomes evident that for a humanoid robot it is not sufficient to look like a human being but to be able to recognize and simulate emotions [30]. The goal is truly affective human-computer intelligent interaction (HCII). As stated by Reeves and Nass [31] for this to achieve taking out one human in a classical human-human interaction and replacing him by a computer, the basic human-human issues should maintain.

Today rapid advances in technology make it possible to not limit the human-computer interaction interface to traditional keyboard and mouse activity. Especially the field of computer vision along with speech analysis opens up the interface for HCII. Humans interact mainly through speech, body gestures and facial expressions, displaying emotions by visual, vocal and other physiological means. Those emotional skills are considered to be part of what is called "intelligence" [34] [16] and have to be addressed in the field of robotic research.

Furthermore today autonomous robots already take their way out of the laboratory facing new challenges in the presence of humans. Traditionally designed to operate independently from humans, a new range of application domains, i.e. domestic, entertainment, health care, demand autonomous robots that interact and cooperate with humans. One key to natural robot-human interaction is the ability of the robot to recognize the person's emotional and attentional state. Another key is the simulatation of emotion by the robot in order to justify and support its own behavior. Both of them is needed to adapt human-human interaction protocols for robotic behavior and support natural communication modalities of humans. And for the acceptance of sociable humanoid robots in people's daily lives the usage of those interaction protocols will be necessary [4] [5].

In the following I will first give a short introduction into the field of emotion research. Afterwards different approaches of emotion recognition will be presented. The focus will be on facial and vocal emotions as these domains are of special relevance. Finally the necessity for a sociable robot to be able to simulate emotions will be pointed out, based on an expressive anthropomorphic robot called Kismet [19].

Two further aspects in this field are the synthesis of emotion and the design of emotion models. In contrast to emotion simulation, which means that an emotional expression is shown to serve a specific purpose in a given situation, the topic of emotion synthesis discusses how this emotion can be expressed, e.g. how the different actuators of the robot can be used to simulate a specific emotion. To determine which emotion is appropriate in a given situation to support the objective of the robot a so called emotion model is used.

In this paper I will not discuss the aspect of emotion synthesis nor different emotion models as both are mostly application dependent. The interested reader will find a generic model for personality, mood and emotion simulation for conversational virtual humans in [11]. Further references about sociable robots can be found within the publication area of the Sociable Machines Project [19].

Chapter 2

Human Emotion Research

Human emotions have attracted scientists of different fields. As a result a vast body of literature about emotion theory exists [10] [6] [17] preventing a comprehensive review. In particular not even a precise definition of emotion can be given but in general emotions can be considered short-term in opposite to moods, temperaments and personalities [18].

To be able to discuss human emotions a classification scheme is needed. In general observed emotions are described and classified by either labelling discrete categories of emotions or using multiple dimensions or scales. What is to be considered is the collecting of emotional data within the studies of human emotion. Until bio signals were recently introduced, there was no tool available for reliable emotion measurement. Instead, emotional observation data was either manually classified or so called emotion induction was performed, i.e. a stimulus was given to the subject in order to induce a specific emotion.

The introduction of emotion to computer science was done by Pickard who created the field of affective computing [30]. Pickard pointed out the importance of emotion in computer science. Since emotion is fundamental to human experience influencing cognition, perception and everyday tasks such as learning, communication and even rational decision-making, this aspect must be considered in human-computer interaction.

Human emotion is expressed by different modalities. In the following I will present the most important indicators for human emotion: Facial expression, speech, gestures and bio signals.

2.1 Facial expression

Concerning non-verbal communication the facial communication channel is the most important in human cognition. The assumption already made by Darwin [10] that the origin of nonverbal forms of communication are species-specific and not cultural-specific could be proved by the extensive studies of Eckman and Friesen [12].

By studying facial expression in different cultures Eckman and Friesen found evidence for "universal facial expressions". Despite the influence of different social settings and circumstances - so called display rules - those expressions are consistent across different cultures and called the Eckman six: Happiness, sadness, anger, fear, surprise and disgust.

Furthermore Eckman and colleagues developed the Facial Action Coding System (FACS). By defining 48 different action units (AUs) in relation to different muscle groups each emotion is described by a combination of AUs activity. The recognition or vice versa the synthesis of facial emotion based on FACS can be done by simply identifying resp. stimulating the corresponding AU. Based on FACS a subset called Emotion FACS (emFACS) is defined by Eckman and colleagues containing only relevant AU and activity combinations for emotional facial expression. This subset reduces the computational cost in emotion recognition applications.

FACS example: Happiness Expression *(one possible combination)*

- pulling lip corners (AU12 + AU13)
- and/ or mouth opening (AU25 + AU27)
- with upper lip raiser (AU10)
- bit of furrow deepening (AU11)

FACS / emFACS are still under development and further information is available on the corresponding website [14]. Many recent work in the area of facial emotion analysis and recognition has used FACS or a subset [2] [7] [8] [9] [13] [22] [23] [24] [25] [28] [29] [38]. Whereas FACS does not take a time component of description into account there are enhanced models of emotional description developed on basis of FACS. An interactive demonstration of emotion synthesis based on a model of action units, combinatoric activity and a time component is available in [1].

With advances in computer technology and computer vision the work of Eckman and Friesen [12] more and more gained impact on emotion recognition and emotion simulation in affective computing.

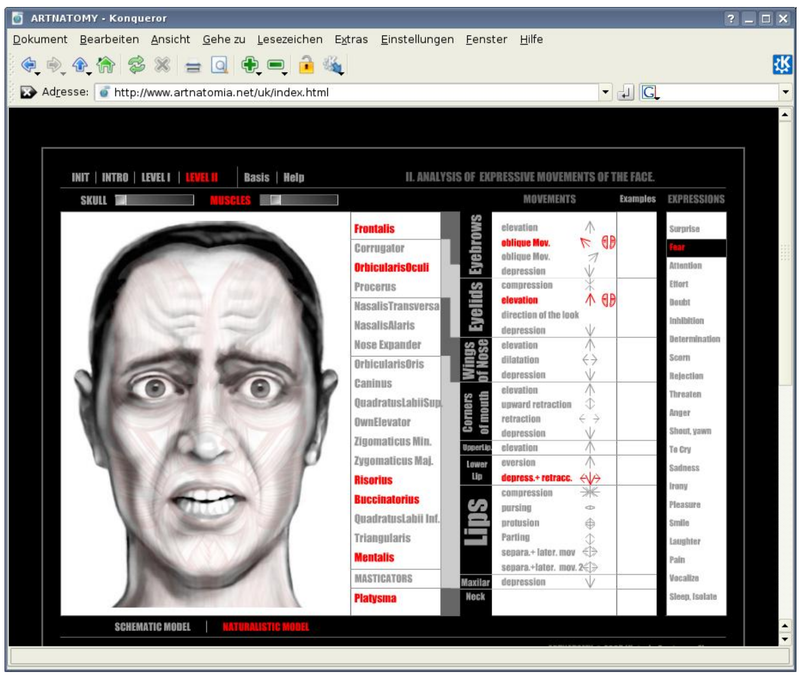

Figure 2.1: Facial expression learning tool [1]

2.2 Speech

In addition to facial emotional expressions, vocal aspects of communicative messages carry various non-verbal information. If only the verbal part of a spoken message is considered the meaning of the message might be completely misunderstood. For example by utilizing irony the meaning of a message can even be inverted. With respect to speech recognition and human-computer interaction the emotional aspect of speech must be studied and taken into account.

Quantitative studies of vocal emotions started in the 1930s. Until today most of these studies used prosodic features in spoken language to recognize emotion. Those features include pitch, duration and intensity of the utterance [33]. By comparing the spectrograms of real emotional spoken language with acted data, William and Stevensen found similarities. Accordingly they reasoned the use of acted data in emotional speech analysis.

2.3 Analysis of Gestures

Another possibility to recognize human emotion in visual data is the analysis of gestures and body movement. Features of human emotion can be extracted from expressive gestures. These features mostly relate to arm movement. For example, tracing the spatial position of wrists over time, parameters of expressivity can be obtained from speed, acceleration and direction variation. In the context of interaction quantitative features derived from hand tracking indicate an observed emotion; as example, satisfaction turns to joy or even to exhilaration, as the speed and amplitude of the expressive gesture increase [20].

2.4 Bio Signals

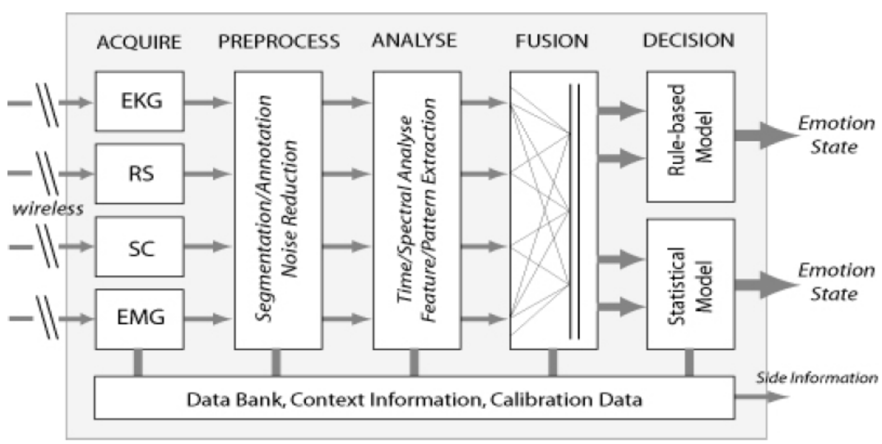

Figure 2.2: Systematic framework for emotion recognition using bio signals [20]

Today advances in technology make it possible to extract emotional cues and features from physiological signals called bio signals. Bio sensors, i.e. EKG, EMG, SC, RSP enable researchers to actually measure emotion continuously with high accuracy, comparable to the results of extreme facial expression or speech intonation (figure 2.2). Using musical emotion induction methods a bio-signal based database for emotion has been created and will be expanded to a multimodal emotional database combining audio/visual recordings as well as bio-features [20].

In contrast to the previously described techniques the usage of bio signals in general demands for a wired connetion to the observed person. As a result bio signals are not suitable for emotion recognition in daily situations. The importance of bio signals in the field of emotion recognition lies in the possibility to test and validate other emotion recognition techniques in laboratory conditions. It gives researchers a tool to measure human emotion in real time instead of using emotion induction techniques or manually annotated data.

Chapter 3

Emotion Recognition Approaches

As presented above human emotion can be recognized by analysing data gathered from different modalities. Accordingly different approaches have been taken to develop automatic ways for computers to recognize human emotions. In the past most of these approaches concentrated on emotion recognition by a single modality. Only very recent studies discuss the usage of multimodal data in this context.

In this chapter I will first present a vocal and a visual approach towards emotion recognition, afterwards one of the first studies on multimodal emotion recognition. In this context the different aspects of multimodal emotion recognition will be discussed.

3.1 Facial Emotion Recognition

Based on the work by Eckman and colleagues [12] automatic facial emotion recognition approaches were taken. In these studies the researchers use standard image processing techniques to measure the movement of facial regions. Afterwards these movements are mapped on the AUs defined by Eckman and colleagues and the FACS database is used for classification. The result is a defined category of emotion. A comprehensive introduction to the different approaches, the used features and classifiers is given by Sebe and colleagues [36].

Concerning the general process the methods are similar in the way that first some features are extracted which are then fed into classifiers. But the extraction of features can be divided into two categories: Feature based, where specific features/points like mouth corners are tracked and region based, where the amount of motion in distinct facial regions is measured (Figure 3.2).

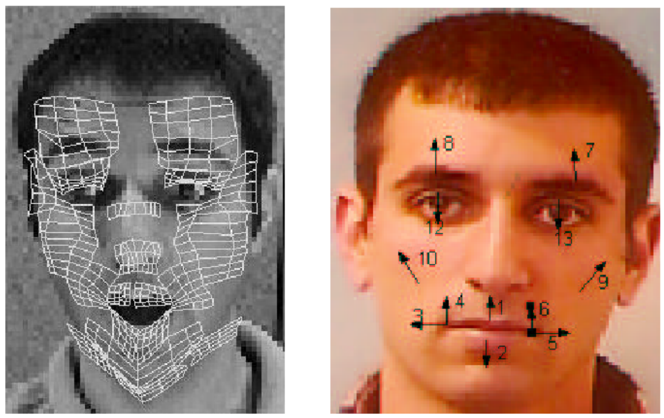

Figure 3.1: Feature extraction for facial emotion recognition [35]

3.2 Vocal Emotion Recognition

Derived from facial expressions as stated above, the Eckman Six are also widely used in recent vocal emotion recognition studies, although in the past many other categories for emotion classification were defined. As there is no evidence for universal emotional expressions in human voice - in opposite to facial expressions - the use of these categories is often not justified. Table 3.1 shows the relation between some vocal features and 5 emotions of the Eckman Six as reported by Murray and Arnott [26].

Even though not all human emotions can be recognized by analyzing speech good results in recent studies show the importance and possibilities of vocal emotion recognition. E.g. Kwon and colleagues [21] used pitch, log energy, formant, mel-band energies, and mel frequency cepstral coefficients (MFCCs) as base features. Velocity/acceleration components were computed and added in each frame to take the rate of speaking into account and model the dynamics within the signal. The resulting 15 feature streams were used as input vectors for various classifiers: support vector machine (SVM), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and HMM. From the many derived features given to the classifiers the ones that contributed most to the classification were identified (figure 3.1). For evaluation the SUSAS and AIBO databases were used. The results showed pitch and energy playing a major role in

	Anger	Happiness	Sadness	Fear	Disgust
Speech Rate	slightly faster	faster or slower	slightly slower	much faster	very much slower
Pitch Average	very much higher	much higher	slightly lower	very much higher	very much lower
Pitch Range	much wider	much wider	slightly narrower	much wider	slightly wider
Intensity	higher	higher	lower	normal	lower
Voice Quality	breathy	blaring	resonant	irregular	grumbled

Table 3.1: Summary of human vocal affects described relative to neutral speech [26]

recognizing emotions. With the text independent SUSAS database an accuracy of 96.3% was achieved for stressed/neutral style classification and 70.2% for 4-class speaking style classification. With the speaker independent AIBO database, an accuracy of 42,3% for 5-class emotion recognition was achieved.

Looking forward to multimodal approaches and considering the possibility of different techniques - based on different modalities - to complement one another, the importance of vocal emotion recognition becomes visible even more.

3.3 Multimodal Emotion Recognition

As shown in the field of speech recognition, using multimodal data can improve the accuracy of the detection. From the perspective of emotion recognition the question is, how can multimodal data especially be used in this application. De Silva and colleagues [37] propose a rule based method for joining vocal and visual information. In this setup visual and vocal data are recorded and evaluated independently, the separate results from visual and vocal emotion recognition are then combined by a weighting matrix (figure 3.3).

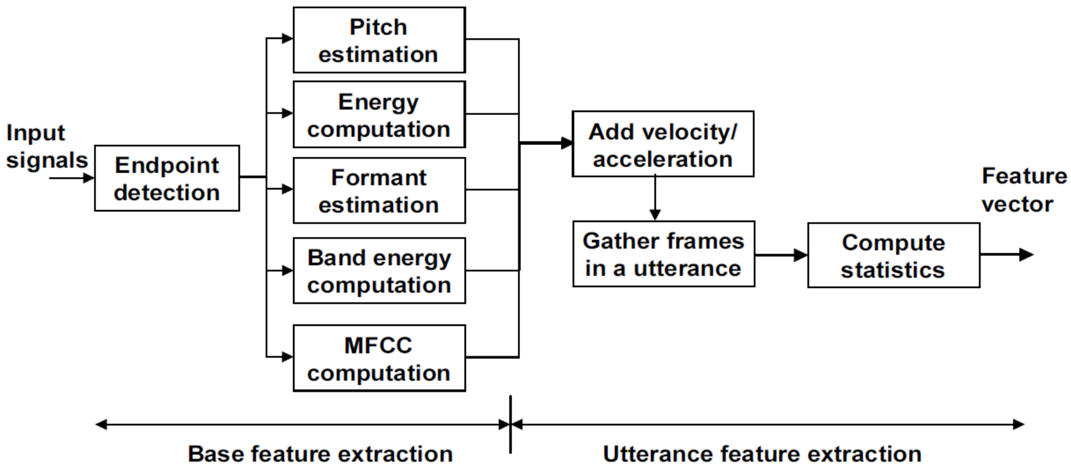

Figure 3.2: Block diagram of the feature extraction module [21]

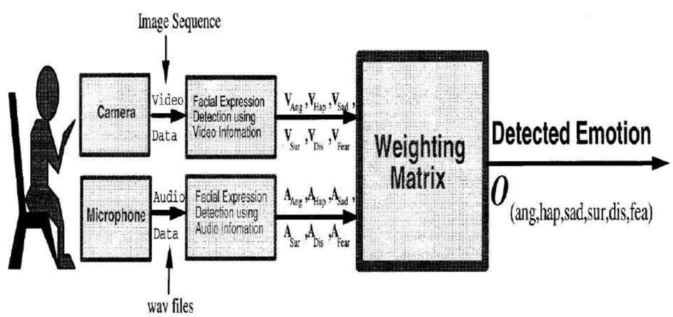

Figure 3.3: Rule based emotion classification [37]

The weighting matrix is defined as follows:

$$\begin{aligned}
O_{Anger} &= W_{1,Ang}V_{Ang} + W_{2,Ang}A_{Ang} \\
O_{Happiness} &= W_{1,Hap}V_{Hap} + W_{2,Hap}A_{Hap} \\
O_{Sadness} &= W_{1,Sad}V_{Sad} + W_{2,Sad}A_{Sad} \\
O_{Surprise} &= W_{1,Sur}V_{Sur} + W_{2,Sur}A_{Sur} \\
O_{Dislike} &= W_{1,Dis}V_{Dis} + W_{2,Dis}A_{Dis} \\
O_{Fear} &= W_{1,Fea}V_{Fea} + W_{2,Fea}A_{Fea}
\end{aligned}$$

The weights in the matrix are preliminary defined:

$$
\begin{array}{ll}
W_{1,Ang}=22.59 & W_{2,Ang}=0 \\
W_{1,Hap}=41.88 & W_{2,Hap}=0 \\
W_{1,Sad}=0 & W_{2,Sad}=20.65 \\
W_{1,Sur}=11.64 & W_{2,Sur}=0 \\
W_{1,Dis}=23.30 & W_{2,Dis}=0 \\
W_{1,Fea}=0 & W_{2,Fea}=6.54
\end{array}
$$

The Detected Emotion is then given by:

$$Max\{O_{Anger}, O_{Happiness}, O_{Sadness}, O_{Surprise}, O_{Dislike}, O_{Fear}\}$$

The idea behind the method proposed by De Silva and colleagues is to differ between vocal and visual dominated emotions and to set weights W to appropriate values. As can be seen in figure 5 the emotions sadness and fear which are difficult to hold apart by visual features can easily be distinguished from vocal features. This way De Silva and colleagues achieved a significant improvement in recognition accuracy compared to separate recognition by facial and vocal features.

Concerning multimodal approaches in the field of emotion recognition it must be discussed if the separate handling of features in different channels rule out further possibilities. Moreover using multidimensional features spaces better relates to human cognition and can improve emotion detection in the future. The recent Humaine project [20] combines features from visual and vocal data, bio signals and gesture recognition to develop a complex system for automatic emotion analysis. I.e. modern bio-sensors are used to deploy a so called "baseline channel" as it supplies a continuous flow of information. Moreover bio-signal features were identified to supplement visual and vocal information.

Chapter 4

Emotion Simulation in Affective Computing

The importance of emotion recognition has been pointed out previously. In order to "understand" human behaviour as well as to correctly interpret recognized speech, utilizing information channels beyond the verbal domain are of special importance.

The missing link to HCII is the ability for a computer/robot to simulate emotions. By achieving this an autonomous robot will be able to make use of a social protocol naturally understood by its human counterpart. This ability will be necessary not only for the factor of acceptance in socialized environment but for navigation and self-protecting reasons as well as to give reasoning for it's own behaving and feedback to the interacting user.

Another aspects of emotion simulation in the context of an autonomous acting robot is the possibility to reduce the complexity of what is received. In complex environments attention can be directed to selected objects which is of special importance in learning tasks.

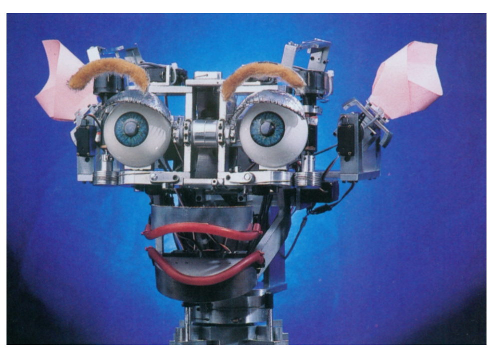

Figure 4.1: Kismet, an autonomous, talking, emotional robot head [19]

An excellent example for the utilization of emotions in context of robotics and human-computer intelligent interaction can be found in the laboratory of the MIT under the name of Kismet (figure 4.1). In this project the researchers developed a robotic head with human-like abilities to express emotion through facial expression and head movement. Moreover, not only the ability to simulate emotions relates to human physiognomy but also the sensing is oriented on human reception. Accordingly Kismet's sensing is mainly based on a stereo vision and a stereo sound system. The cameras are placed in the eyes and the microphones in the ears to achieve a human-like perception. In addition to sensor equipment for audio and vision, Kismet's hardware consists of sensors for distance measurement as well as actuators for speech synthesis, visual expression simulation and head movement.

4.1 A Socialized Robot

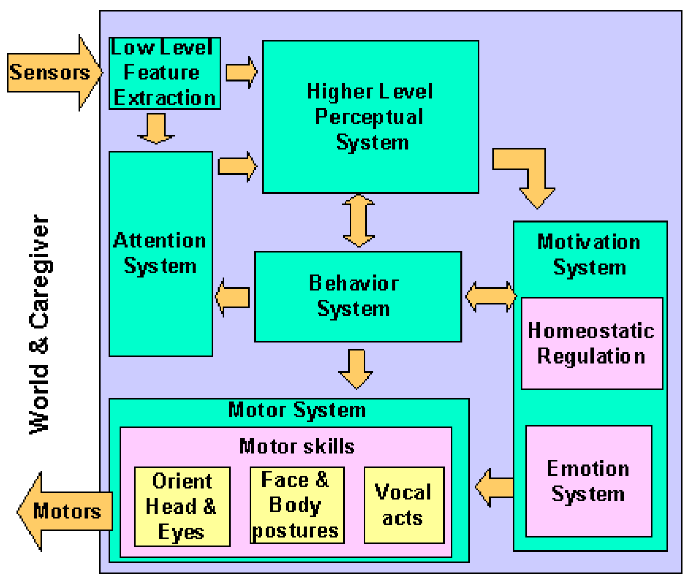

Figure 4.2: The robot's framework [19].

The robot's system architecture consists of six subsystems:

- Low level feature extraction: Extraction of sensor based features.
- Higher level perceptual system: Encapsulation of detected sensor based features into percepts.

- Attention system: Determination of the most salient and relevant stimulus at any time.
- Behavior system: Implementation and arbitration of competing behaviours.
- Motivation system: Regulation of the robot's state in form of homeostatic regulation processes and emotive responses.
- Motor system: Synchronizing and controlling of output modalities.

A simplified view of the architecture is shown in figure 4.2 and detailed information can be found in [3].

4.2 Utilization of Social Protocols

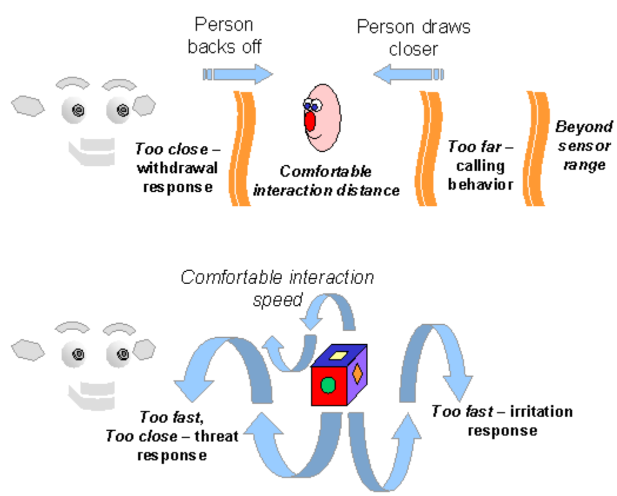

Figure 4.3: Kismet utilizing social protocols [19]

The basis of Kismet's behaviour in human-computer interaction is the utilization of social protocols (figure 4.3) [19]. In the shown examples the robot is able to control the distance to a presented object to make it "comfortable" for it's sensors. In the first example the distance to the person is influenced with respect to further recognition purposes in the context of conversation. The second example shows Kismet's behaviour to avoid damage to sensors caused by fast moving objects close in distance. If demanded the head movement can be supported by facial emotional expressions. What makes this behaviour socialized is the way a reaction is implied to the interacting human being. I.e. a withdrawal reaction of Kismet will cause the user to adjust the distance resp. the distance of the object to a comfortable value for the robot. The user intuitively understands the robot's behavioural purpose.

Concerning more complex social interactions, the ability to control the flow of conversation is of special importance. It can be observed that people tend to glance aside when they begin their turn, establish eye contact when preparing to relinquish their turn and await a response. In conversation people tend to raise their eyebrows when listening. Kismet is able to use such regulatory mechanisms in a way which is naturally understood by it's human partner in conversation.

Besides emotion recognition, emotion simulation and usage of social protocols, researchers in the Artificial Labs of the M.I.T. implemented learning schemata into the robot. Kismet is able to recognize objects and learn the structures of presented objects for later recognition.

Chapter 5

Discussion

Due to technological development autonomously acting robots emerge from laboratories opening up new application domains. Thus interaction between robots and human beings is shifted more and more into the focus of robotic research. As pointed out in the previous chapters when talking about Human-Computer Intelligent Interaction emotional skills are of eminent importance. Emotions do not have a slight but a great influence to human communication modalities. One simple but impressive example comes from speech recognition where even acceptance cannot be distinguished from negation concerning not the word itself but the meaning of the message.

In the past emotion recognition was mainly done by extracting low level features from different channels. It has been shown that by combining these features in multimodal feature spaces good recognition results can be achieved using probabilistic methods. To be able to recognize the affect and intention of a person, combining the extracted multimodal low-level features with information about the person's context, situation, goal and preference will be needed [30].

Concerning emotion simulation the findings in the field of Affective Computing and HCII underline the challenge robotic research encounters. Thinking about humanoid robots in socialized environments the acceptance of such machines will be dependent on emotional skills. Simulating emotions the robot can give reasoning for its behavior to the environment, communicating its intention or attentional state. It is believed that multimodal context-sensitive human-computer interaction will become the single most widespread topic of the artificial intelligence research community [36].

Bibliography

[1] Anatomical Basis of Facial Expression Learning Tool: `http://www.artnatomia.net`

[2] M.J. Black, Y.Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *Proc. International Conf. on Computer Vision*, pages 374-381, 1995.

[3] C. Breazeal. Learning by Scaffolding. Ph.D. Thesis Proposal, M.I.T. Department of Electrical Engineering and Computer Science, 1998.

[4] C. Breazeal, B. Scassellati. A context-dependent attention system for a social robot. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99)*, pages 1146-1151, Stockholm, Sweden, 1999.

[5] C. Breazeal. Sociable Machines: Expressive Social Exchange Between Humans and Robots. Sc.D. dissertation, Department of Electrical Engineering and Computer Science, M.I.T., 2000.

[6] W.B. Canon. The James-Lange theory of emotion: A critical examination and an alternative theorie. *American Journal of Psychology*, 93, pages 196-124, 1927.

[7] L.S. Chen. Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction. Ph.D. Thesis, University of Illinois at Urbana-Champaign, Dept. of Electrical Engineering, 2000.

[8] I. Cohen, N. Sebe, F. Cozman, M. Cirelo, T.S. Huang. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 595-601, 2003.

[9] I. Cohen, N. Sebe, A. Garg, L. Chen, T.S. Huang. Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding*, 91(1-2):160-187, 2003.

[10] C. Darwin. The Expression of the Emotions in Man and Animals, John Murray, London, 2nd edition, 1890.

[11] A. Egges, S. Kshirsagar, N. Magnenat-Thalmann: Generic personality and emotion simulation for conversational agents. *Computer Animation and Virtual Worlds*. Volume 15, Issue 1 , pages 1-13, 2004.

[12] P. Eckman, W.V. Friesen. Facial Action Coding System: Investigator's Guide, Consulting Psychologists Press, 1978.

[13] I.A, Essa, A.P. Pentland, Coding analysis, interpretation, and recognition of facial expressions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):757-763, 1997.

[14] The Facial Action Coding System (FACS Manual). `http://face-and-emotion.com/dataface/facs/description.jsp`

[15] B. Fasel, J. Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36:259-275, 2003.

[16] D. Goleman. Emotional Intelligence, Bantam Books, New York, NY, 1995.

[17] W. James. The Principles of Psychology. Henry Holt, New York, NY, 1890.

[18] J.M. Jenkins, K. Oatley, N.L. Stein. Human Emotions: A Reader, Blackwell Publishers, Malden, MA, 1998.

[19] Kismet: An expressive anthropomorphic robot, M.I.T. `http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet`

[20] S. Kollias, N. Amir. D4C: Description of potential exemplars: Signs, 2004. `www.emotion-research.net`

[21] O. Kwon, K. Chan, J. Hao, T. Lee. Emotion Recognition by Speech Signals, Eurospeech 2003, Geneva, 2003.

[22] A. Lanitis, C.J, Taylor, T.F. Cootes. A unified approach to coding and interpreting face images. In *Proc. International Conf. on Computer Vision*, pages 368-373, 1995.

[23] J. Lien. Automatic recognition of facial expressions using hidden Markov models and estimation of expression intensity. Ph.D. Thesis, Carnegie Mellon University, 1998.

[24] A. Martinez. Face image retrieval using HMMs. In *IEEE Workshop on Content-based Access of Images an Video Libraries*, pages 35-39, 1999.

[25] K. Mase. Recognition of facial expression from optical flow. IEICE Trans., E74(10):3474-3483, 1991.

[26] I.R. Murray, J.L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature of human vocal emotion. *Journal of the Acoustic Society of America*, 93(2), pages 1097-1108, 1993.

[27] A. Nefian, M. Hayes. Face recognition using an embedded HMM. In *IEEE Conf. on Audio and Video-based Biometric Person Authentication*, pages 19-24, 1999.

[28] N. Oliver, A. Pentland, F. Berard. LAFTER: A real-time face and lips tracker with facial expression recognition. *Pattern Recognition*, 33:1369-1382, 2000.

[29] T. Otsuka, J. Ohya. Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences. In *Porc. International Conf. on Image Processing*, pages 546-539, 1997.

[30] R.W. Picard. Affective Computing, M.I.T. Media Lab Perceptual Computing Section Tech. Rep. No. 321, Cambridge, MA, 1995.

[31] B. Reeves, C. Nass. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, Cambridge Univ. Press, Cambridge, 1996.

[32] M. Rosenblum, Y. Yacoob, L.S. Davis. Human expression recognition from motion using a radial basis function network architecture. *IEEE Trans. on Neural Network*, 7(5):1121-1138, 1996.

[33] Y. Sagisaka, N. Campbell, N. Higuchi. Computing Prosody. Springer-Verlag, New York, NY, 1997.

[34] P. Salovey, J.D. Mayer. Emotional intelligence. *Imagination, Cognition and Personality*, 9(3):185-211, 1990.

[35] N. Sebe, I. Cohen, A. Garg, M. Lew, and T. Huang. Emotion recognition using a Cauchy Naive Bayes classifier. In ICPR, 2002.

[36] N. Sebe, I. Cohen, T.S. Huang. Multimodal Emotion Recognition. In *Handbook of Pattern Recognition and Computer Vision*, World Scientific, London, 2005.

[37] De Silva (1997):L.C. De Silva, T. Miyasato, R. Nakatsu, Facial emotion recognition using multimodal information. In *Proc. ICSP*,pages 397-401, 1997.

[38] N. Ueki, S. Morishima, H. Yamada, H. Haishima. Expression analysis/synthesis system based on emotion space constructed by multilayered neural network. *Systems and Computers in Japan*, 25(13):95-103, 1994.

[39] C.E. Williams, K.N. Stevens. Emotions and speech: Some acoustical correlates. *Journal of the Acoustic Society of America*, 52(4), pages 1238-1250, 1972.