Ulm University
Department of Psychosomatic Medicine and Psychotherapy
Director: Prof. Dr. Harald Gündel
Section Medical Psychology
Head: Prof. Dr. Harald C. Traue

EMOTION CLASSIFICATION IN HUMAN-COMPUTER INTERACTION ON THE BASIS OF PHYSIOLOGICAL DATA

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the Doctoral Degree of Human Biology (Dr. biol. hum.) at the Medical Faculty of Ulm University

David Hrabal
born in Prague, Czech Republic
2013

Incumbent Dean: Prof. Dr. Thomas Wirth
First Correspondent: Prof. Dr. Harald C. Traue
Second Correspondent: Prof. Dr. Dr.-Ing. Wolfgang Minker
Third Correspondent: PD Dr. phil. Dr. habil. Jonghwa Kim
Day of Dissertation: 10/18/2013

Dedication

To my parents who have given up life in their home country to make it possible for my sisters and me to live a free life and who have always selflessly supported us.

Contents

List of Tables

List of Figures

Nomenclature

ANN .... Artificial Neural Network
BAS .... Behavioral Approach System
BIS .... Behavioral Inhibition System
BVP .... Blood Volume Pulse
ECG .... Electrocardiography
EIM .... Emotion Identification Module
EMG .... Electromyography
ERP .... Event-Related Potential
ERQ .... Emotion Regulation Questionnaire
ES .... Experimental Sequence
GUI .... Graphical User Interface
HP .... High Pleasure
IAPS .... International Affective Picture Set
kNN .... k-Nearest Neighbors
LA .... Low Arousal
LP .... Low Pleasure
Neo FFI ..... Neo Five-Factor Inventory
PAD .... Pleasure Arousal Dominance
RSP .... Respiration
SAM .... Self-Assessment Manikin
SC .... Skin Conductance
SRT .... System Response Time
WOZ .... Wizard Of Oz

Chapter 1

Introduction

As the urban environment is increasingly driven by technology, the evolution of technical systems has accelerated dramatically during the past decades. In spite of this, a typical interaction with a technical system has not changed significantly. Developing methods to improve the level of communication between man and machine is one of the goals of affective computing. Affective computing constitutes the effort of performing robust and reliable classification of a user's emotional state. The emotional state can be determined through the analysis of various modalities, for example prosody, facial expression, gesture and physiology. Because emotional behavior is a complex multidimensional process, several modalities contain partially redundant, convergent or conflicting emotion information. Compared to other modalities of human emotion expression, physiological data can be measured continuously and may contain emotional information. This is particularly the case when expressive behavior in the form of language, facial expression or gesture is not available. Therefore, emotion recognition on the basis of physiological data is a topic of great interest.
Being able to determine a user's emotional state would enhance simple assistive technical systems into so-called digital companions. Digital companions are described in a framework by Traue et al. (3) as embodied conversational agents who communicate in natural spoken language and carry out advanced and natural man-machine interactions.

The goal of digital companions is to provide not only functionality but also empathetic responses to the user's needs. In the future, digital companions will appear in everyday life in many different forms. They will provide us with helpful information, support us in decision-making processes and communicate their intentions to the social and technical environment. There are many roles a digital companion could play in everyday life, but still the answer as to what kinds of relationships people will require from companion entities remains unclear. Wilks (4) suggests that a companion should have the role of the Victorian lady's Companion.

Companion attributes within cognitive technical systems will ensure that the systems are perceived, accepted and utilised by their users personal and empathetic assistants that are socially and emotionally intelligent (see (5)). This is only possible if the functionality of such companion technologies is consistently and fully automatically oriented towards the user's abilities, preferences, requirements and current needs. This technological development opens up new, unforeseen opportunities of technical support and digital assistance. Companion technology can make an important contribution in this area, especially with regards to the future of our aging society. Its application potential ranges from a new generation of technology which provides versatile assistance with the organisation of everyday tasks to innovative support systems for patients in rehabilitation or people with limited cognitive abilities.

A clinical example of a simple Companion System is C. Bischoff's and K. Limbacher's E-Coaching system eATROS. It is a program for a handheld device which assists users, suffering from a psychological disorder, such as depression, to cope with their problems, activate resources and encourage self-regulation / self-control by leading him through a series of pre-determined questions (see for example (6) and (7)). If such a simple system could be equipped with emotion recognition capabilities, it could adapt its behavior better to a specific user and thereby become more effective.

One of the primary aims of companion technologies should undoubtedly be to provide access to helpful technology to older people. Because, as Wilks (4) writes, older people have, until today, been effectively excluded from information technology, as well as the web, the internet and some mobile phones, just because they are not able to cope with the buttons.

The following physiological responses are typically assessed for the purpose of emotion recognition / classification in the terms of affective computing (see for example (8), (9)) and (10)):

- the cardiovascular system - by recording the electrocardiography (ECG) or the peripheral blood volume (BVP)
- the electrodermal system - by measuring the skin conductance (SC)
- the facial expressive system - by recording the electromyogram (EMG) of the *corrugator supercilii* and the EMG of the *zygomaticus major*
- the respiratory system - by measuring respiratory excursion

Chapter 2

Theoretical Background

2.1 Emotion Theory and Historical Overview

In his work ‘The Expressions of the Emotion in Man and Animals’ that was published in 1872, Charles Darwin has analysed questions and theories supporting the thesis that humans evolved from animals. In his book, he put special emphasis on the analysis of facial expressions in humans and mammals. In the end, he was sure that a facial expression had a specific meaning in every emotion. For example in the emotion ‘surprise’, it meant to inform the subject about what happened. Therefore the eyes are opened widely to enable a better optical perception of the environment.
Another very early explanation about what an emotion is supposed to be, gave William James in 1884. James postulated that an emotion constitutes a series of procedures after a stimulus onset. His thoughts were put into the famous phrase: ‘Do we run from a bear because we are afraid or are we afraid because we are running from the bear?’. His answer to this question is based on the fact that some physiological reactions that occur are perceivable and that these can lead to an emotion.
Jame’s Theorem was revised by Walter Cannon. He was interested in the ‘flight-or-fight syndrome’ and postulated that in an emergency reaction, the physiological changes prepare the body to the upcoming event by, for example, pumping blood into the muscles. The physiological changes are regulated by the autonomic nervous system and seem to be too slow to be responsible for the origin of emotions. Another

counter-argument was brought up by Cobos et al. (11) who examined patients with spinal cord injuries who didn't show different emotional sensation than healthy subjects, meaning that the information of the periphery is not needed for a sensation of emotion.

A few years later, the James-Cannon debate was revised and discussed by Schachter and Singer (12). Their oppinion was that physiological feedback is very important for emotions, but not specific enough to evoke an emotion. They postulated that emotions must evolve from the cognitive interpretation of a situation. Their hypothesis was, that through a change in the physiological feedback, one should be able to influence the experienced emotion. They verified the hypothesis by injecting adrenalin to subjects who underwent an emotional experience and could show that emotions were sensed more intensely with the emotional quality being the same.

Paul Ekman postulated the Ekman's neuro-cultural theory of emotion. Similar to Darwin, Ekman studied the facial expressions and found cross-cultural similarities of emotion expressions. According to Ekman's theory, six basic emotions can be stated (see (13)) that are expressed and understood in every culture. These are the emotions fear, sadness, disgust, joy, surprise and anger. When one of the basic emotions is induced, the specific facial emotion pattern is triggered and the relevant facial muscles are activated. After that, Plutchik come to the conclusion, that there are two other emotions, adding to the previous six basic emotions - trust and anticipation. From these basic emotions, secondary emotions can be formed, for example shame being fear and disgust. The problem is that there are different concepts of basic emotions. Ekman's theory claims that there is an automated expression for each emotion, whereas Plutchik claims that emotion results from the disposition. Therefore it can not be strictly defined, how many basic emotions exist.

In the present study, the dimensional concept of emotions is used as a working definition. This concept differs from the discrete emotions model in that way, that there is not only the fixed basic emotions. In the dimensional concept, the emotional space is located within a coordinate system, and in this three-dimensional space, each emotion can be represented. The emotions are located in a two- or three-dimensional emotion space (PAD space) associated with the dimensions 'pleasure' (positive-negative), 'arousal' (aroused-relaxed) and 'dominance' (gain of control-loss of control) describing the location of a particular emotion within this three-dimensional space. This form of emotion theory is based on the concept of Russell and Mehrabian (14). Both, James as well as Darwin, postulated that emotion is an expression of a motivational behavior and basic instincts. Humans and animals either tend to a stimulus or they turn away from it. This motivational approach is in many ways still used by different research groups. It is biphasic, so it can be interpreted as positive and negative (see (15)), pleasant and unpleasant (see (16)) or appetite and aversive (see (17)).

2.2 Emotion and Physiology

Somatic reactions are part of emotional responding. The involved parameters, like breathing, sweating and muscle activation are regulated automatically (unconsciously) or consciously. The regulation of these parameters helps to adapt to specific situations. Even when emotions are constituted mimically and consciously, changes in the physiology of the autonomic nervous system can be observed (see (18)). It is also possible to derive emotional categories (positive, negative, neutral) from EEG data (see (19)). Therefore it is of great interest to find emotion patterns, patterns of physiological reactions to emotional content on subject-dependent as well as subject-independent level.

The use of physiological data for emotion classification can be very important, especially when the subject is emotionally less expressive. In such situations, classification using mimics or gestics can fail due to missing expressive data. People who tend to hide their emotional state in their mimics, but express them strongly in the physiological data are called internalized, while mimically expressive people who tend to show less physiological reactions are referred to as Externalizer (see (20), (21), (22)). There are four types of mechanisms that are responsible for emotional inhibition: genetic, repressive, suppressive and deceptive (see (23)). The suppression of emotional arousal is called suppressive inhibition (see (24) and (23)).

Emotion regulation processes can be investigated by the use of questionnaires. The Emotion Regulation Questionnaire (ERQ) (see (25)) has been designed to analyse two emotion regulation strategies suppression and reappraisal.

2.3 Emotion Induction

In order to obtain emotionally relevant physiological data, the subject must be emotionally stimulated. The most common ways of emotion induction are video clips (see for example (26), (27), (28), (29) and (30)) and audio clips (see for example (31), (32), (33), (34) and (35)). Another, very frequent method constitutes the emotional stimulation with the International Affective Picture Set (IAPS) (see (36)). In this study, two different emotion induction setups were employed. The first setup was a standardized IAPS picture presentation where the subject's task was to passively watch pictures with emotional contents. The IAPS pictures were presented as blocks. The description of the experimental setup of the blocked IAPS presentation is described in section 3.1.2. The second setup was a wizard of oz scenario where the subject actively took part in a mental trainer scenario. This experiment is described in section 3.1.3.
The physiological data was recorded by the Nexus32, a physiological monitoring device.

2.4 Current State of Scientific Research

To date there are two different approaches on how to detect an emotion from physiological data:

- Selected features are extracted from the physiological data and used as inputs for an automatic classification system for emotion classification
- One physiological channel is analysed and its changes are linked to a certain emotion

The first approach could be referred to as 'machine learning approach'. After the data acquisition and a feature extraction, selected features are used as inputs for an automatic classification system and a classification rate, that is mostly based on a leave-one-out validation is calculated. An overview over the field of affective computing using this method is given in Picard's work (10). One example adopting this approach can be found in Zhai's and Barettos's work (37) in which a supervised classification of two affective states, 'stressed' and 'relaxed,' was conducted using a support vector machine based on physiological features. The classification rates were determined by a 20-fold cross-validation.

Katsis and his team (38) also created a platform that classifies 'high stress', 'low stress', 'disappointment', 'euphoria' and 'neutral face' using support vector machines on physiological data and optical information from a facial detection system. Kim and his team (39) used support vector machines for classification between 'sadness', 'anger', 'stress' and 'surprise'. In (40), they performed a classification of four emotional states: 'high arousal, positive', 'high arousal, negative', 'low arousal, positive' and 'high arousal, negative' using a pre-clustered linear discriminant analysis (pLDA). Another approach was used by Rosalind Picard who performed a k-nearest neighbor (k-NN) classification on physiological data in (10), also testing the results via cross-validation. In this study, Picard points out the problem of day-dependence, showing the need for robust features / procedures when using different data sets that were recorded at different points of time.

Sark et al. (41) have conducted an experiment inducing stress with the Stroop Color-World Interference Test. They managed to achieve arousal classification rates of around 90% with an individual (subject specific) feature-selection and classification setup using support vector machines. Wu et al. (42) have classified arousal in threat situations. They achieved individual (subject specific) classification rates of 84-100%. Valenza et al. (43) have conducted an experiment, presenting blocked pictures from the International Affective Picture System (IAPS) with five pleasure and arousal levels to the subjects and classifying the levels of pleasure and arousal. They also present a small overview over used signals, induction methods, emotion classes and classification results that were reported in the last decade. In (44) Agrafioti et al. have analysed emotion recognition by means of Electrocardiographic (ECG) patterns. Out of the ECG morphology, around 89% of pleasure states were correctly classified.

Nicolau et al. (45) introduced a system that fuses multiple cues and modalities in order to perform a continuous prediction of the emotional state in the two dimensions of pleasure and arousal.
The second approach can be referred to as 'single channel analysis'. Physiological data is analysed and characteristics for a distinct emotion are searched in a single physiological channel. This physiological channel's changes can then be analysed in various situations. The heart rate, for example, can be a measure for emotion. Etzel et al. (46) have analysed this physiological feature and measured acceleration and deceleration of the heart rate while the subjects were listening to sad and fearful music. Bernat and his team (47) have used the heart rate to measure effects of picture content. Stemmler et al. (48) and Friedman et al. (8) have shown that skin conductance changes have an affect on emotion and Paus et al. (49) and Bradley and his team (50) used EMG as measure for emotion.
Kreibig (1) illustrates a detailed overview of studies on peripheral responding in fear and sadness, showing a very detailed overview of peripheral physiological channels and their status changes during these two emotional situations. In Figure 2.1 can be seen an overview over studies that found the heart rate rise and fall during fear and sadness induction. Kreibig found the heart rate to be one of the most sensitive measures to distinguish between the states of fear and sadness.

2.5 Aim and Organization of the Thesis

Neither of the methods described above have yet led to successful subject- and situation-independent emotion recognition. Using a machine learning approach for emotion classification can result in correct classification rates of about 80-90 percent. But since the classification rate drops alarmingly when trying to classify subject- or especially situation-independent data, this method cannot be realised in a companion system that needs to be able to identify its user's emotional state in various situations. The single channel analysis can be applied subject-independently, but if only one physiological channel is monitored, the cause of its changes will always remain

Dependent variable	Fear	Study	Sadness	Study
Cardiovascular system				
Heart rate	↗	Borkovec & O'Brien, 1977	↗	Ekman, Levenson, & Friesen, 1983
	↗	Eisenberg, et al., 1988	↗	Frazier, Strauss, & Steinhauer, 2004
	↗	Ekman, Levenson, & Friesen, 1983	↗	Gross & Levenson, 1997
	↗	Etzel, Johnsen, Dickerson, Tranel, & Adolphs, 2006	↗	Kunzmann & Grühn, 2005
	↗	Frazier, Strauss, & Steinhauer, 2004	↗	Levenson, Ekman, & Friesen, 1990
	↗	Lerner, Gonzalez, Dahl, Hariri, & Taylor, 2005	↗	Levenson, Ekman, Heider, & Friesen, 1992
	↗	Levenson, Ekman, & Friesen, 1990	↗	Palomba & Stegagno, 1993
	↗	Levenson, Ekman, Heider, & Friesen, 1992	↗	Ritz, George, & Dahme, 2000
	↗	Palomba, Sarlo, Angrilli, Mini, & Stegagno, 2000	↗	Sinha, Lovallo, & Parsons, 1992
	↗	Palomba & Stegagno, 1993		
	↗	Pauls & Stemmler, 2003		
	↗	Sinha & Parsons, 1996		
	↗	Stemmler, Heldmann, Pauls, & Scherer, 2001		
	↗	Tourangeau & Ellsworth, 1979		
	↗	Vrana & Rollock, 2002		
	↘	Bernat, Patrick, Benning, & Tellegen, 2006	↘	Eisenberg, et al., 1988
	↘	Bradley, Codispoti, Cuthbert, & Lang, 2001	↘	Etzel, Johnsen, Dickerson, Tranel, & Adolphs, 2006
	↘	Dimberg, 1986	↘	Krumhansl, 1997
			↘	Tsai, Levenson, & Carstensen, 2000

Figure 2.1: Overview over studies analysing the heart rate during the induction of fear and sadness taken from (1).

ambiguous.

The aim of this thesis was to develop a robust method for emotion-change recognition in the two-dimensional space of pleasure and arousal which can be applied to all subjects, thus subject-independently in various situations, thus situation-independently. The core idea was to observe changes of physiological channel-pairs instead of examining single channel changes and thereby obtaining the possibility of coding four emotional states, covering all edges of the two-dimensional emotion space of pleasure and arousal.

The organization of the thesis is as follows:

Both, a machine learning method (see chapter 4.1) as well as single channel analysis (see chapter 4.3.1) will be applied to the acquired data in order to analyse weather one of these methods is suited for a two-dimensional emotion recognition and to reveal their advantages and disadvantages. In chapter 3.1, the experimental setup including an IAPS picture presentation, a mental training scenario and questionnaires is described. In 3.2 the physiological Data that was measured is shown and in 3.4 the basic concept of Artificial Neural Networks is described to the reader.

In chapter 4.1 a common classification is performed on the EMOREC I data. Finally,

in chapter a hybrid method is developed, combining the machine learning approach and the single channel analysis approach in Emotion Identification Modules that can be used for subject- and situation-independent emotion change identification.
The thesis of this work is that by the linking of each two physiological channels and a weighted majority voting, it should become possible to determine an emotion change and its strength in both dimensions - pleasure and arousal.

Chapter 3

Materials

3.1 Experimental Setup

The physiological data was collected during the acquisition of the EmoRec II corpus, an emotion-induction corpus designed for data acquisition for the use of multimodal emotion recognition. Therefore not only physiological data was recorded but also high quality video and audio material.

3.1.1 The EmoRec II Corpus

The EmoRec II corpus contains multimodal data from 110 subjects (69 female and 41 male). The subject's average age was 35.4 years (SD = 18.8). All subjects signed an informed consent form. The corpus data was recorded at the Emotion Lab, Ulm. In the first part of the experiment blocked pictures with emotional content were presented to the subjects (see section 3.1.2). In the second part of the experiment the 'mental trainer' - a Wizard-of-Oz (WOZ) experiment was performed. This experiment is based on six experimental sequences in which different emotional states were induced employing praise, dispraise, time pressure etc. (see section 3.1.3).
A graph illustrating the recorded data sets as well as the data handling can be seen in Figure 3.2. The physiological data recorded during the IAPS picture presentation

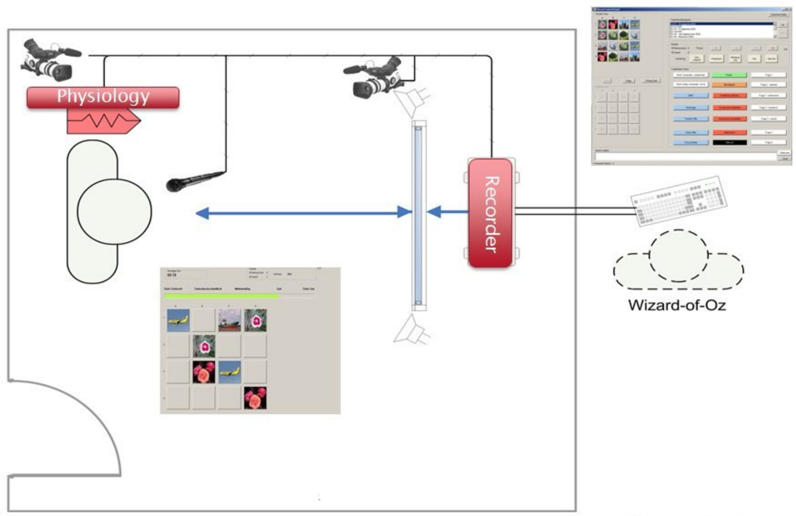

Figure 3.1: The setup of the mental training WOZ experiment

was divided into two parts (IAPSg1 and IAPSg2) containing physiological data of 55 subjects each. The first group's data (IAPSg1) was used for empirical detection of the most emotionally relevant feature-pairs and for construction of five Emotion Identification Modules (EIMs) for each of the two emotion dimensions pleasure and arousal. This procedure is described in section 4.4.1. These modules were then applied for emotion identification on the second group's IAPS picture presentation data (IAPSg2) as well as the WOZ interaction experiment data (WOZg1) to verify the subject- and situation-independent identification performance of the EIMs (see section 4.4.2).

3.1.2 Blocked IAPS Picture Set

As mentioned in section 1, the international picture set (IAPS) is a very common method of inducing emotions. Since these pictures have been presented to and rated by thousands of subjects, the validity of them is undoubted. For each picture, there exists a mean rating for each of the three dimensions of pleasure, arousal and dominance.

Previous tests that used IAPS pictures as stimulus material (see (51) have achieved

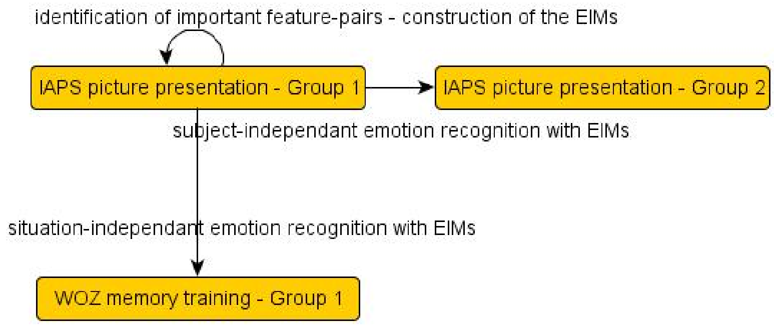

Figure 3.2: An illustration of the splitting of the acquired data set (EmoRec II) into the parts IAPSg1, IAPSg2 and WOZg1 for the purpose of EIM creation as well as subject- and situation-independent testing of the EIM's emotion change identification.

good results in the field of event-related potential (ERP) classification research that is applied in EEG analysis. For the purpose of ERP recognition, it is sufficient to present the induction material for only a few milliseconds, since the ERPs occur within milliseconds. The temporal resolution of peripheral physiological data is much greater. Therefore we blocked the pictures and presented ten pictures of approximately the same arousal and pleasure rating score subsequent without a pause to ensure a longer lasting emotion induction.

This procedure put the subject's emotion into one place on the two-dimensional space of pleasure and arousal and kept it there for 20 seconds. This method will allow a segmentation of the affective physiological data for classification purposes and also a better comparison with the affective physiological data that was gathered from the mental trainer wizard of oz experiment (see section 3.1.3) where the subjects also remain in one state of affect for even a few minutes.

Eight IAPS picture blocks, containing ten pictures from the International Affective Picture Set (36) each, were presented to the subjects in random order[1]. Each picture

[1]IAPS pictures: HPLA: 1610, 1620, 2360, 2388, 2540, 5200, 5551, 5760, 5779, 2530 (52), 11-20 (53), mean pleasure SAM rating = 7.58, mean arousal SAM rating = 3.26, SD(pleasure) = 0.39, SD(arousal) = 0.47 LPLA: 2210, 2280, 2715, 2722, 6010, 7046, 7700, 8010, 9045, 9101, 9360, 939 (52), 1, 2, 5-10 (53), mean pleasure SAM rating = 3.74, mean arousal SAM rating = 3.63, SD(pleasure) = 0.43, SD(arousal) = 0.51 HPHA: 4611, 4690, 4695, 5629, 8161, 8503, 5621, 8080, 8370, 8490, 1650, 5700, 5950, 8191, 8341, 5470, 8030, 8180, 8400 (52), 30 (53), mean pleasure SAM rating = 7.0, mean

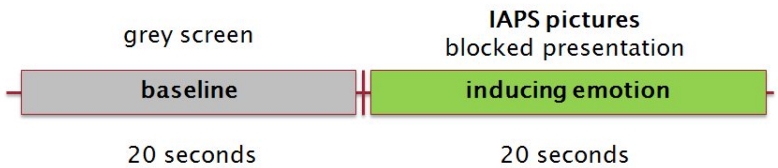

Figure 3.3: The subjects were shown a gray screen for 20 seconds, following 20 seconds of IAPS picture material with emotional content.

was presented for two seconds resulting in a block of 20 seconds duration. The blocks were separated by a 20 second pause. Each of the ten pictures in one block was very similar in pleasure and arousal rating, inducing one of the emotion categories low pleasure and low arousal (LPLA), low pleasure and high arousal (LPHA), high pleasure and low arousal (HPLA), high pleasure, high arousal (HPHA) (see Figure 3.4).

The blocked picture presentation was chosen due to the advantage of emotional stimulation with longer sequences. This approach causes deeper emotional involvement and physiological reactions. As Smith et al. (54) could show, corrugator activity and skin conductance responses increase with ongoing exposure to unpleasant IAPS pictures. A further advantage of this stimulation method is the comparability with the WOZ experiment data with experimental sequences of similar length.

- LPLA (low pleasure, low arousal)
- LPHA (low pleasure, high arousal)
- HPLA (high pleasure, low arousal)
- HPHA (high pleasure, high arousal)

arousal SAM rating = 6.5, SD(pleasure) = 0.57, SD(arousal) = 0.51 LPHA: 1050, 2800, 3266, 3400, 5971, 6230, 6350, 9250, 9410, 9424, 9810, 9920, 9921, 9925, 3000, 3053, 3120, 3170, 3500, 6300, mean pleasure SAM rating = 2.20, mean arousal SAM rating = 6.66, SD(pleasure) = 0.64, SD(arousal) = 0.57

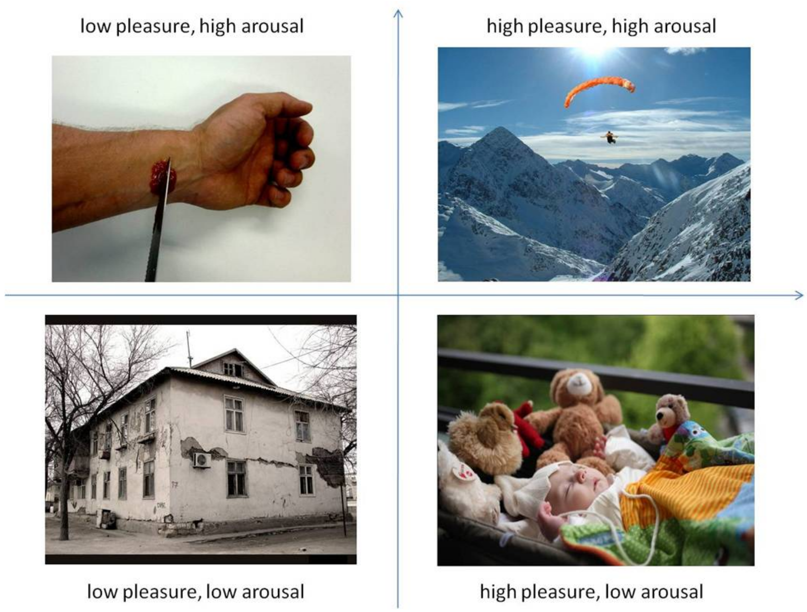

Figure 3.4: Eight picture series were presented to the subjects. Two series of pictures inducing low pleasure and low arousal (LPLA), two inducing low pleasure and high arousal (LPHA), two inducing high pleasure and low arousal (HPLA) and two inducing high pleasure and high arousal (HPHA). These pictures are similar to the IAPS collection (2).

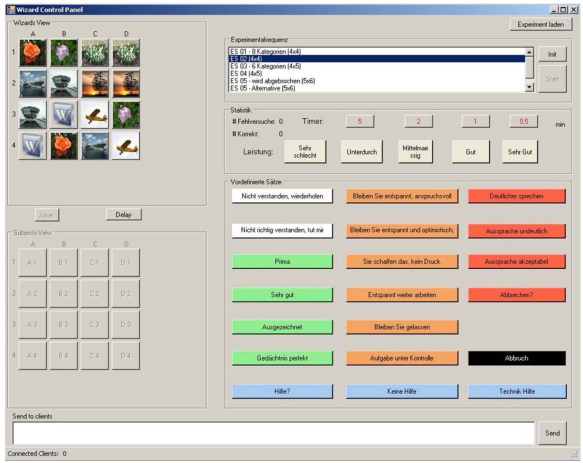

Figure 3.5: The wizard's operating screen for the WOZ experiment. On the top left the wizard sees the subject's screen. On the bottom left the cards can be uncovered. On the top right the sequences are started and on the bottom right positive, neutral or negative commands can be triggered.

3.1.3 The Mental Trainer WOZ Experiment[2]

To perform the mental trainer WOZ experiment, an application was programmed by Andreas Scheck and David Hrabal in C#. It consists of two parts: the subject's screen and the experimenter's user interface (see Figure 3.1).

The subject's screen shows the picture matrix with the hidden pictures, the elapsed and remaining times and a performance bar showing the user's performance in the current experimental sequence. Ranging from 'very bad' to 'very good' (see Figure 3.6).

The concept of WOZ experiments is widely used for software development and prototyping in the area of Human-Computer-Interaction and interface design (see for example (56), (57), (58), and (59)). The subject is told to be interacting with a computer system, being unaware of the fact that the experiment is controlled or manipulated by the experimenter, the so-called 'wizard'. In our WOZ experiment, the subject was told to be performing a mental training experiment with an autonomous

[2]Contents of this section can also be found in (55)

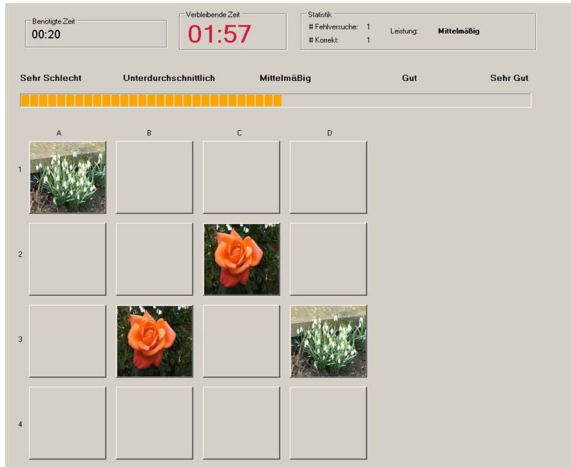

Figure 3.6: The subject's view onto the experimental screen. On the top the consumed time and the time that is left are shown. On the top right the subjects performance is displayed varying from 'very bad' to 'very good'. Underneath the performance bar and the card deck are shown.

computer system that is operated by voice. The WOZ scenario consisted of six experimental sequences (es01-es06) in each of which a certain emotional state was induced. This emotional state, represented as a point in the two-dimensional space of pleasure and arousal, can be seen in Figure 3.7.

The induced emotion depends on the following factors:

- size of the picture matrix
- alikeness of the pictured motives
- time pressure
- indicated rating of the subject's performance
- positive and negative comments of the system
- incorrect recognition of the subject's commands

- delays in execution of the subject's commands

The subject's screen during the experimental sequences four (es04) and six (es06) of the WOZ experiment can be seen in Figure 4.10. In es04 the pictures looked very much alike, there was time pressure and the performance bar showed 'under average'. The induced emotion was low pleasure and high arousal (LPHA). In es06, the pictures were diverse, there was no time pressure, positive comments were given to the subject and the performance bar showed 'very good'. The induced emotion was high pleasure and low arousal (HPLA). After the experiment, a manipulation check via a semi-structured interview which included questions about the subject's feelings in all six experimental sequences of the WOZ experiment as well as a rating of these sequences via Self Assessment Manikin (SAM) (36) were assessed. The mean pleasure rating for es04 was 5.8 (SD = 1.9) whereas the mean pleasure rating for es06 was 7.9 (SD = 1.5). The mean arousal rating for es04 was 4.8 (SD = 2.2) whereas the mean arousal rating for es06 was 2.9 (SD = 1.9).

Detailed Description of the Mental Trainer WOZ Experiment

The simulation of the natural verbal human-computer interaction was implemented as a Wizard-of-Oz (WOZ) experiment [41], [42]. The WOZ experiment allows the simulation of computer or system properties. The subjects get the impression that they have a completely natural verbal interaction with a computer-based memory trainer. The design of the memory trainer follows the principle of the popular mental trainer 'concentration'. The variation of the system behavior in response to the subjects was implemented via natural spoken language, with parts of the subject's reactions taken automatically into account. However, the system does not work with algorithmic speech recognition and automatic response control, but is controlled by an experimenter in an adjoining room. This method allows simulating human-computer interactions that technically are not yet possible (or only rudimentary so), such as a reliable recognition of natural spoken language. In the following we describe details of our experiment:

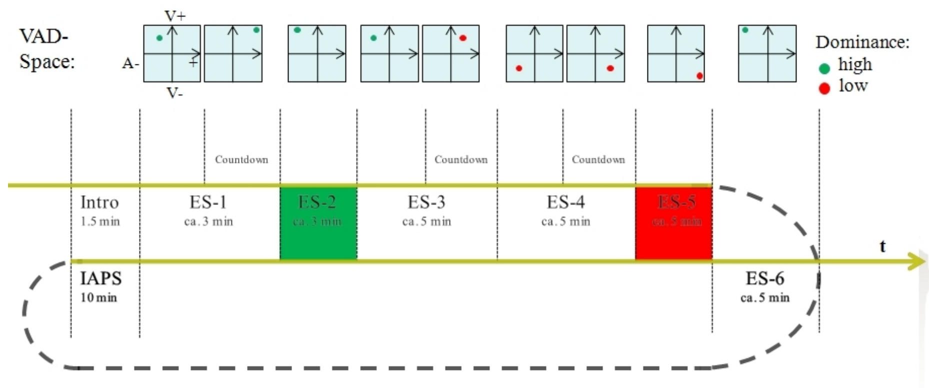

Figure 3.7: The two rounds of the WOZ experiment with their experimental sequences. The first round consists of five experimental sequences, the second round consists of six experimental sequences.

Main Instruction

The subject is being introduced to the system by the following instruction:

'You will be communicating with a computer and completing a memory test. The experiment will be carried out in this room. You will receive commands from the system. The system can react to you on the basis of language, gestures, and facial expressions. Therefore, there are a microphone and a camera. The system has many different functions. At the beginning of the test you will be asked to answer some questions. When you exceed certain thresholds with regard to the physiological parameters, the system will ask you whether to terminate the task. Poor language recognition may lead to technical delays. Try to communicate with the system like you would with a human being.'

Structure of the experimental sequences (ES)

The experiment involves a memory-training sequence consisting of two parts (see Figure 3.7). Both parts consist of identical experimental sequences. Part two also includes a debriefing sequence at the end of the scenario. Between round one and round two, 24 images (International Affective Picture System, IAPS) are shown to the subjects, so that they could distance themselves emotionally from the first round. The presentation time for each image is six seconds, with a break of four seconds to ten seconds.

Experimental sequence 1 (ES1)

System instruction:

'On the screen in front of you is a deck of covered cards with different images. Every image exists twice. Your task is to successively uncover the image pairs. Please uncover them as quickly as possible as there will be a time bonus. At the same time, you should make as few mistakes as possible, because every wrong pair will be deducted from your score. On the upper screen you see a bar showing your performance status. Only verbal communication is possible. I understand many different commands. So just speak the way you usually would. We will now begin the memory test. The test is divided into several rounds. I wish you success in solving the tasks! You begin the first test with the command: start test.'

Procedure

In ES1, a deck of a 4 x 4 matrix with highly discriminative images was presented. After three correctly uncovered pairs the subject received the feedback *'Your performance is improving'*. After half of the ES (four remaining pairs), a countdown of five minutes was started. *'Please finish the game quickly! You have five minutes!'* At the end of the ES the subject received the following comment: *'You have successfully solved the first task. Please describe how you feel at the moment.'*

Targeted Emotional State

In the first half (four pairs), ES1 was inducing the emotional state 'high pleasure, low arousal, high dominance' and after the countdown the emotional state 'high pleasure, high arousal, high dominance' (see Figure 3.8).

Experimental sequence 2 (ES2)

System instructions

'Now comes the second round. Try to be as good as before. Begin again with the command: start test.'

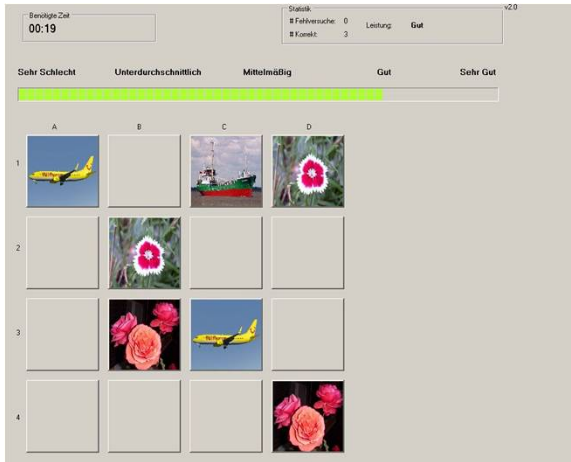

Figure 3.8: A screenshot of the experimental sequence 1 (ES1). There are 16 cards in the card deck, seven are already uncovered, the subject's performance is 'good'.

Procedure

A deck of a 4 x 4 matrix with highly discriminative images was presented. After three correctly uncovered pairs the subject received the feedback *'Your performance is improving'* with a visual bar feedback of very good. There was no countdown. At certain intervals the wizard sent positive feedback, such as:

- Very good!
- Keep it up!
- You are doing great!
- Great!
- Your memory works perfectly!

At the end of the ES, the subject received the following comment: *'You have also very successfully solved this round! Please describe your emotional state. If necessary: Could you describe this in more detail?'*

Targeted Emotional State

In ES2 the emotional state 'high pleasure, low arousal, high dominance' was induced (see Figure 3.9).

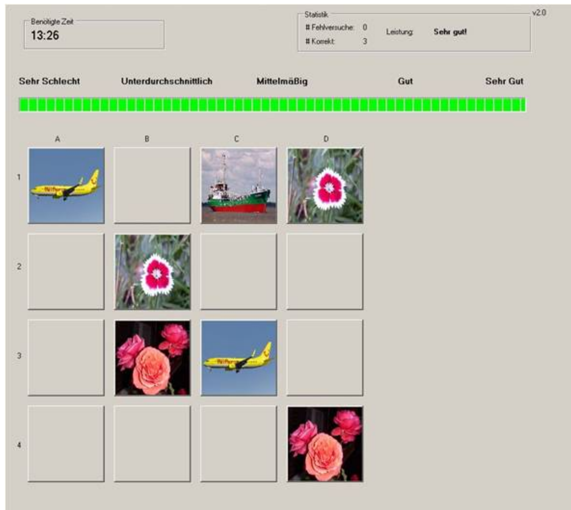

Figure 3.9: A screenshot of the experimental sequence 2 (ES2). There are 16 cards in the card deck, seven are already uncovered, the subject's performance is 'very good'.

Experimental sequence 3 (ES3)

System instructions

'The third deck is larger. The majority of subjects in your age group solve the task without problems. For this deck you can, however, ask for help. After you have uncovered a card, the second card will be uncovered. Again, begin with the command: Start test.'

Procedure

A deck of a 4 x 5 matrix with highly discriminative images was presented. After three incorrectly uncovered pairs the subject received the feedback, *'Your performance is declining'* with a visual feedback of mediocre. After half of the ES (5 remaining pairs), a countdown of two minutes was begun. *'Please finish the game quickly now! You have two minutes!'*. At the end of the ES the subject received the following comment: *'Task solved. This round was more difficult because of the number of cards. Nevertheless, you have again successfully solved the task. How do you feel now?'*

Targeted Emotional State

In the first half (five pairs correctly uncovered), ES3 was inducing the emotional state 'high pleasure, low arousal, high dominance' and after the countdown the emotional

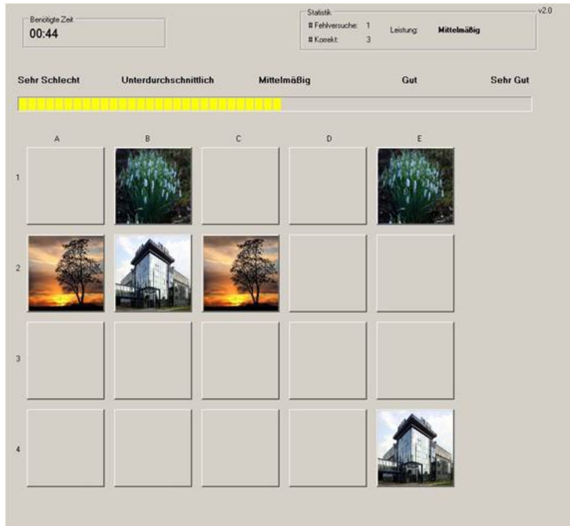

Figure 3.10: A screenshot of the experimental sequence 3 (ES3). There are 16 cards in the card deck, six are already uncovered, the subject's performance is 'fair'.

state 'high pleasure, high arousal, low dominance' (see Figure 3.10).

Experimental sequence 4 (ES4)

System instructions

In the fourth round the difficulty will be increased again. You can again ask for help with this deck.

Procedure

A deck of a 4 x 5 matrix with two discriminative images (ships and airplanes) was presented. After 3 incorrectly uncovered pairs the subject received the feed-back *'Your performance is declining'* along with a visual feed-back showing 'below average'. There are approximately six delays of six seconds when the cards are uncovered. In some cases the cards were uncovered incorrectly. After half of the ES (5 pairs), a countdown of one minute was begun. *'Please finish the game quickly now! You have one minute!'* At the end of the ES the subject received the following comment: *'The round was terminated. This deck was very stressful for you. How do you feel now?'*

Targeted Emotional State

In the first half, ES4 was inducing the emotional state 'low pleasure, low arousal, high

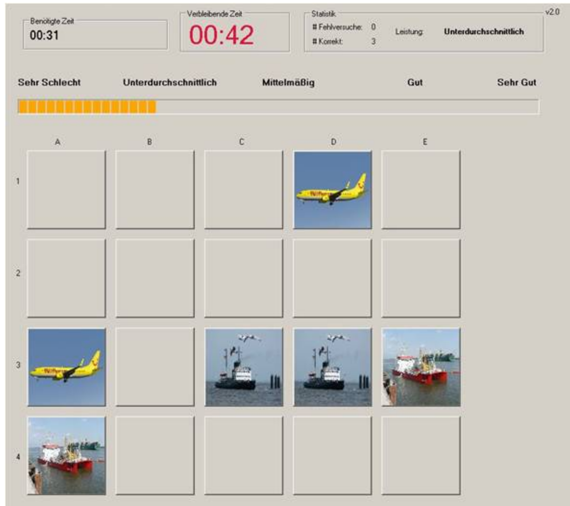

Figure 3.11: A screenshot of the experimental sequence 4 (ES4). There are 16 cards in the card deck, seven are already uncovered, the subject's performance is 'under average'.

dominance' and after the countdown the emotional state 'low pleasure, high arousal, low dominance' (see Figure 3.11).

Experimental sequence 5 (ES5)

System instructions

'Try to not ask for any help in Round 5. Begin with the command Start test!'

Procedure

A deck of a 5 x 6 matrix with very similar images (snowdrops) was presented. After three incorrectly uncovered pairs the subject received the feedback *'Your performance is declining'* with a visual bar feedback of very poor. There were ca. ten delays in uncovering the cards or cards were uncovered incorrectly. After the first third of the ES, the subject was given the comment *'Your articulation is unclear'!* After processing half of the matrix the subject received the following comment four times in a row: *'Would you like to terminate the task?'* At the end of the ES the subject received the comment: *'The task was very difficult for you, therefore I have terminated the task. Don't be disappointed. How do you feel now?'*

Targeted Emotional State

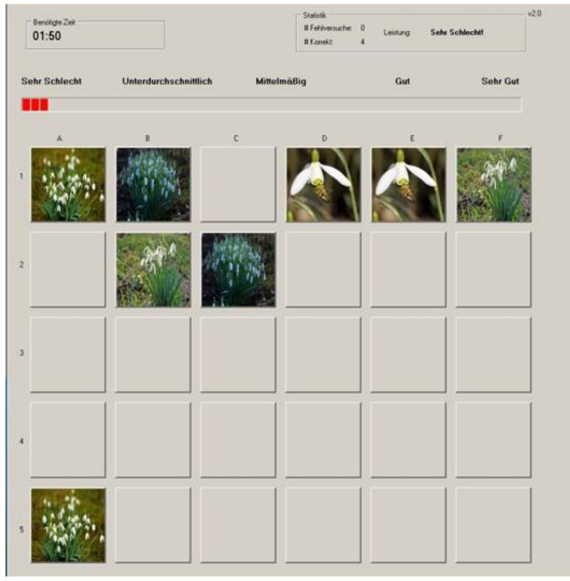

Figure 3.12: A screenshot of the experimental sequence 1 (ES1). There are 16 cards in the card deck, eight are already uncovered, the subject's performance is 'very bad'.

In the first half, ES5 was inducing the emotional state 'low pleasure, high arousal, low dominance' and after the question regarding termination the dimension 'low pleasure, high arousal, low dominance' was to be strengthened further; i.e. pleasure decreases, arousal increases, dominance decreases (see Figure 3.12).

3.1.4 Questionnaires

BIS/BAS

In his biophysiological emotion theory, Gray (60) postulates three different systems. He makes those responsible for subject-specific differences in behavior and experience. The respective reactivity of the behavioral inhibition system (BIS) and the behavioral approach system (BAS) are linked to the personality dimensions of anxiety and impulsivity on the basis of animal studies and pharmacological studies (61). The German questionnaire comprises a total of 20 items (7 BIS, 13 BAS).

Emotion Regulation Questionnaire

The Emotion Regulation Questionnaire (ERQ) (62) was developed to detect individual differences in emotion regulation. Two variables are identified: reappraisal and suppression. The questionnaire consists of 10 items.

Neo Five-Factor Inventory (Neo-FFI)

The Neo FFI is the standard personality test in research. It consists of 60 questions. The variables covered by the Neo FFI are neuroticism, extroversion, openness, agreeableness and consciousness.

Self-Assessment-Manikin

The Self-Assessment-Manikin or SAM rating was introduced by Bradley and Lang (36) and constitutes a pictogram rating technique for three the dimenstions of emotions (3.13).

3.2 Physiological Measures

3.2.1 Physiological recording

The physiological signals were acquired using a NEXUS-32 polygraph, a flexible 32 channel monitoring system. Recorded channels were: electromyogram of the *corrugator supercilii* and *zygomaticus major*, skin conductance (SC), peripheral blood volume (BVP), and respiration (RSP). The sampling rate was set to 512 Hz.

3.2.2 Electromyography - EMG

The EMG was measured via Nexus EXG Sensors, dual channel sensors that deliver two signals of EEG, EMG, ECG, EOG, or any combination. Carbon technology combined with active noise cancellation technology are used to deliver best signal quality at micro volt level. These electrodes were placed above the eyebrow of the

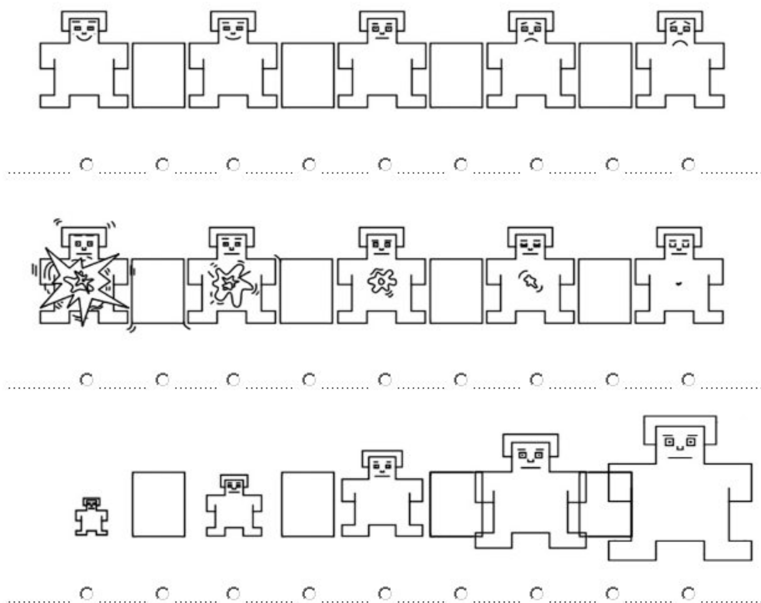

Figure 3.13: The Self-Assessment Manikin (SAM) rating. In the first row the pleasure dimension is rated, ranging from one (very high pleasure) to nine (very low pleasure). In the second row the arousal dimension is rated, ranging from one (very high arousal) to nine (very low arousal). In the third row the dominance is rated, ranging from one (very low dominance) to nine (very high dominance).

subject in order to record the *corrugator supercilii* muscle and between the mouth and the ear in order to record the *zygomaticus major* muscle.
For an introduction to EMG signals refer to (63).

3.2.3 Skin Conductance - SC

There are two different ways of measuring the skin conductance. The endosomatic and the exosomatic measurement. In the endosomatic measurement, potential differences that are produces in the skin, are passively measured. In the exosomatic measurement method, a current is applied to the skin. It is the most common method.
The electrodermal activity is divided into a tonic range (1-30 μ S) and a phasic range (0.05-5 μ S). The tonic signal (skin conductance level, SCL) changes slowly and is often used for measurements of arousal. The phasic signal (skin conductance response, SCR) overlays the signal and is called the stimulus-specific reaction. In this case, the skin conductance (SC) is measured with a set of electrodes that are placed on the finger tips of the subject (see (64)). Figure 3.14 shows a typical skin conductance curve.
The skin conductance was measured via the Nexus Skin Conductance Sensor. This sensor is designed to measure very small relative changes in skin conductance (1/1000 micro siemens) with the Ag-AgCl electrodes which are attached to two fingers.

3.2.4 Blood Volume Pulse - BVP

The blood volume pulse sensor optically measures the blood flow in a peripheral body region - in this case the finger tip. The sensor measures the amount of light that is absorbed by the blood cells, allowing inference on the density of the blood volume (see (65) and (66)). From the peak-to-peak intervals that correspond to the R-R interval in a QRS complex of the ECG (see(67) and (68)) the heart rate can be calculated. Changes that occur within the heart rate reflect changes in the autonomic nervous system. Very often, it is used to identify stress and mental tension (40).
In Figure 3.15 a plot of a raw BVP data sequence can be observed. In Figure 3.16 the Q-peak of the QRS signal are marked with crosses. The algorithm for Q-peak

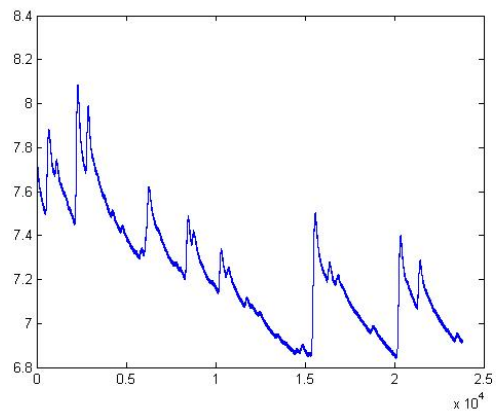

Figure 3.14: A typical skin conductance curve.

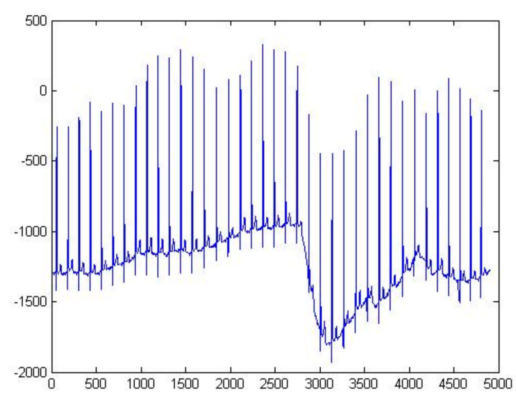

Figure 3.15: Plot of a raw BVP data sequence. The QRS sequences can be seen very clearly.

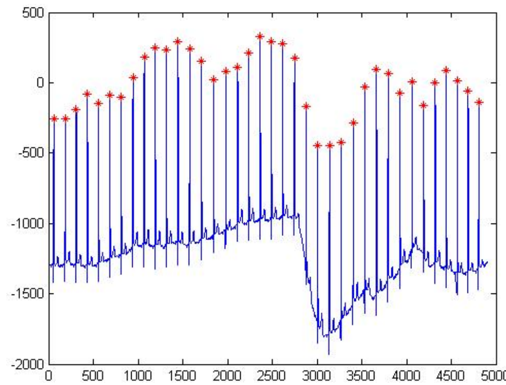

Figure 3.16: The Q-peaks in the BVP signal are marked with crosses.

identification has been programmed in Matlab using the find peak function. Figure 3.17 shows the heart rates that were calculated from the Q-peak distances in Figure 3.16.

3.2.5 Respiration - RSP

The respiration was recorded via Nexus Respiration Sensor. This sensor is a belt that is placed in the abdominal area, with the central part of the sensor just above the navel. It can be worn over clothing, although for best results it is advised that there are only one or two layers of clothing between the belt and the skin. Besides measuring breathing frequency, the relative depth of breathing can also be calculated from the Nexus Respiration Sensor data.

3.3 Machine Learning

Many methods of computer scientists have their origins in machine learning and pattern recognition. There are two different machine learning algorithms called supervised and unsupervised learning. Unsupervised learning methods are used for splitting the data into unknown clusters. Supervised learning methods, also called classification, are used when the categories (classes) are already known - in this case

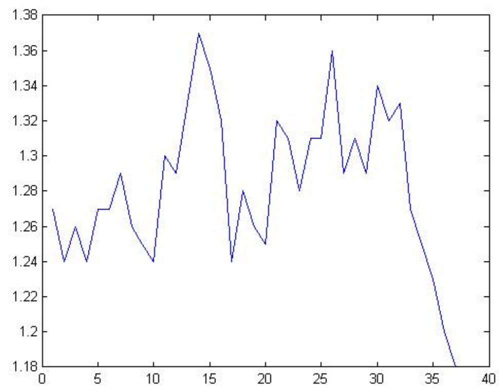

Figure 3.17: From the raw BVP signal, the heart rate can be calculated when the Q-peaks have been identified.

for example low pleasure vs high pleasure. The known data set is used for training the algorithm (training data set). Unsupervised methods cannot be used for prediction or separation of data into classes. Therefore they are not suitable for emotion classification of data with known labels. Known examples for supervised learning algorithms are Artificial Neural Networks (see section 3.4) and Support Vector Machines (SVM), ((69), (70), (71)), boosting (72), K-nearest neighbors (kNN), (73), (74), and Random Forests (75).

The advantage of methods like SVM or kNN are capable of fitting a model for high dimensional data. But when confronted with completely new data, their performance is usually unsatisfactory. This problem was already described by Bellman (76). Bellman called it the curse of dimensionality. In computer science, high dimensional data is used for training of the algorithms, therefore the training is a sampling problem (see (77) for details). It is very important for valid training data to cover the whole feature space, because the quality of the trained model depends on the training data set. In some cases, the data is sparse that means the sampling is bad and the underlying structure cannot be covered by the fitted model. Consequently, the predictive power of such a model is poor.

In modern algorithms, feature selection is used to overcome the curse of dimensionality

- meaning that the best predicting features are selected after a validation procedure. By removing unnecessary features the curse of the dimensionality is avoided during model fitting. Of course selecting features from the same data used for training as well as for testing bears the risk of overfitting. In this case the performance of the training data would be over-optimistic while the performance on unseen data is poor. The trained data would be classified correctly, but the classification rate of unseen data would drop dramatically. Feature selection can be a separate step or a part of the learning algorithm (see (78) for an overview on feature selection methods) and was used for example in section 4.1.2.

3.4 Artificial Neural Networks

An Artificial Neural Network is being characterized as a directed graph (N,V) with edge weights, which are described via a weighting function ω. N denotes the set of neurons, V the sets $\{(i,j)|i,j \epsilon N\}$ of the neuron connections. The function ω : V $\rightarrow$ $\mathbb{R}$ defines the edge weights, with ω_{ij} being the weight of the neuron connection from neuron i to neuron j.

The layers of an Artificial Neural Network are:

- one input layer
- n hidden layers
- one output layer

The input vector is being applied to the input layer and propagated to the hidden layers. The hidden layers are connected with the output layer which will output the result after the propagation of the input vector through the whole network.

3.4.1 Topology of Artificial Neural Networks

One possibility to group Neural Networks is:

- feed forward networks
- recurrent networks

Feed forward Networks

By convention, in Neural Networks, the input layer is the lowest layer and the output layer is the highest layer. In feed firward Neural Networks, there is only a connection between each neuron of the lower layer to each neuron of the higher layer. Thus the connections have only one direction - from the outputs of one layer to the inputs of the next layer. All neurons of one layer have the same activation function.

3.4.2 Recurrent Networks

In recurrent Neural Networks, there is direct connections[3], indirect connection[4] and lateral connection[5]. Recurrent Neural Networks can also be completely connected. These have connections between all neurons of the network.
The spetial feature of recurrent networks is the fact that they can simulate a memory, because of the possibility of integrating the time step t-1 into calculation. This fact allows an even better simulation of biological systems.
In the general case, the neurons of a neural network can be connected arbitrarily. For each neuron a unique activation function can be defined.

3.4.3 The Error Function

To minimize the error vector of a Neural Network, the global minimum of the error network needs to be found. This is often put into affect by application of a gradient descent algorithm as for example the backpropagation algorithm. If illustrated as a function of their weights $E(W)$ with the weight matrix W, the errors of a Neural Network form a hyperplane. The error vector of a Neural Network when applying an input vector p is is the average square deviation between the expected output t_{pj} and

[3] connection from a neuron to itself
[4] connection transcending layers
[5] connection within a neuron layer

the actual output o_{pj}:

$$E_p = \sum_j (t_{pj} - o_{pj})^2$$

Alternatively, the error E_p can be diplayed as the mean square deviation of the error rate:

$$E_p = \frac{\sum_j^n (t_{pj} - o_{pj})^2}{n}$$

This is then called MSE[6] and is used by MATLAB by default.

3.4.4 Learning Algorithms

Without a learning algorithm, an Artificial Neural Network allegories solely a transform function that maps input vectors to the outputs.
With the help of a learning algorithm and a set of training vectors a designated mapping can be learned by the Neural Network.
In the present work, only the supervised learning method was used. This means that the output vectors are made available to the network with the input vectors of the training set. The training vector set is propagated through the Neural Network and generates an output at the output layer. The actual output is compared to the anticipated output and generates an error vector. The error is then propagated backwards through the Neural Network from the output layer to the input layer and creates changes in the connection weights of the neurons. In the next step, the weights are changed, resulting in the minimization of the error.
The weights of the Neural Network are stepwise being adapted, in order for the Neural Network to deliver outputs that are as similar as possible to the anticipated output of the training output vector set.

[6] mean square error

3.4.5 The Backpropagation Algorithm

P. Werbos' ((79)) Backpropagation is one of the most used algorithms for supervised training of multilayer Neural Networks. It also is the most common learning algorithm used in Neural Networks.
The Backpropagation learning algorithm is a gradient descent algorithm where the weights of the network that are connecting the neurons are changed proportionally to the gradient of the error function, minimizing it:

$$\Delta\omega_{ij} = -\eta \frac{\partial}{\partial\omega_{ij}} E(W)$$

The proportionality factor η is called learning rate and represents the range of the increment with which the weights are adapted:
The groundwork for the backpropagation algorithm is the Widroww-Hoff learning method (see (80)), where the weights between the neuron i and j are adapted in the following way:

$$\Delta w_{ij} = \eta o_{pi} \delta_{pj}$$

mit

$$\delta_{pj}(t_{pj} - o_{pj})$$

Where η represents the learning rate, o_{pj} the output of the neuron j when pattern p is applied and t_{pj} represents the anticipated output.
The disadvantage of the Widrow-Hoff learning method is the fact that it can only be applied in networks that possess only one layer of adaptive weights. Another disadvantage is the constrain of the Widrow-Hoff learning method to linear activation functions.
The backpropagation algorithm is an expansion of the Widrow-Hoff algorithm and can be applied also to the more potent nonlinear multi-layered Neural Networks, as the feed-forward networks with nonlinear activation functions. Here the difference δ_j is calculated:

$$\delta_j = f_j'(net_j)(t_j - o_j)$$

if j is an output neuron and

$$\delta_{pj} = f_j'(net_j)\sum_k \delta_k w_{jk}$$

if j is a hidden neuron. Where net_j is the network input j with:

$$net_j = \sum_i o_i \omega_{ij}$$

The output o_i of the neuron i results from the activation function and the network input:

$$o_i = f_a ct(net_j)$$

Chapter 4

Methods of Analysis and Results

4.1 Emotion Classification using Artificial Neural Networks and 13 Statistical Features[1]

After having acquired the physiological data, the first step was to see whether the typical machine learning classification would suffice to satisfy our needs for a robust subject- and situation-independent classification. Therefore a classification using Artificial Neural Networks was conducted using individual and inter-individual classification on the experimental sequences 2 (es02) and 5 (es05).

4.1.1 Feature Extraction

As described in Section 3.2, we used four physiological channels (2 x EMG, BVP, and SCL) for emotion recognition. The signal duration is about three minutes for es02 and 5 minutes for es05, representing antithetic emotions. Using a window length of 10 seconds with an overlap of 5 seconds, the signals were segmented into about 20 samples of es02 and 20-30 samples of es05 depending on different experiment durations of each subject. For the classification of the two emotional states, a total of 13 statistic features were extracted based on mean, minimum, and maximum values from each segment. Please note that, since the goal was to analyse 'general' tendency of

[1]Contents of this section can also be found in (55)

physiological specificity between different situations and individuals, such a compact feature set was used, including only basic statistics, rather than using an extended one.

For the EMG signals obtained from corrugator supercilii ($x_c[n]$) and zygomaticus major ($x_z[n]$), which are by nature noisy, were smoothed with help of a symmetrical moving-average (MA) filter over 20 points. From the filtered signal $\hat{x}_c[n]$, three statistics (mean (f_1), max (f_2), min (f_3) values) with window size of ten points was calculated.
From the signal $x_z[n]$, f_4, f_5, and f_6 were calculated through the same process.

To obtain the heart rate from the continuous BVP signal, a simple peak searching method was developed (based on the QRS detection method (81)) for locating R-peak points indicating heart beats. Next, a time series of R-R distances was generated, similar to heart rate variability (HRV) analysis. From the time series, the mean (f_7), spectral ratio (f_8), max (f_9), and min (f_{10}) were calculated. The feature f_8 is the ratio of spectral power between the low-frequency (0.003-0.15 Hz) and high-frequency band (0.15-0.4 Hz). Such ratio is generally thought to distinguish sympathetic effects from parasympathetic effects, since the parasympathetic activity dominates at high frequency. The features f_{11}, f_{12}, and f_{13} are the mean, max, and min value of the first differences of the SCL signal, respectively.

4.1.2 Classification

As the classification instrument, the feed-forward Artificial Neural Network (ANN) in 13-40-20-1 architecture with two hidden layers (82) was employed.

Observation of different classification cases. Based on the various training sets the following five different cases for individual and inter-individual classification were considered:

- Case 1: leave-one-out and leave-one-subject-out classification of ES-2 (training set 1) vs. ES-5 (training set 2) of round one, without feature selection.

- Case 2: leave-one-out and leave-one-subject-out classification of ES-2 (training set 3) vs. ES-5 (training set 4) of round two, without feature selection.

- Case 3: training of the classifier on ES-2 (training set 1) and ES-5 (training set 2) of round one and testing on ES-2 (training set 3) and ES-5 (training set 4) of round two, without feature selection.

- Case 4: training of the classifier on ES-2 (training set 1) and ES-5 (training set 2) of round one and testing on ES-2 (training set 3) and ES-5 (training set 4) of round two, with automatic feature selection on round one only by using exhaustive optimization (brute-force search) with linear regression.

- Case 5: training of the classifier on ES-2 (training set 1) and ES-5 (training set 2) of round one and testing on ES-2 (training set 3) and ES-5 (training set 4) of round two, with a manual feature selection (like sequential backward search) on round one and round two.

Leave-one-subject-out classification. To test its performance for inter-individual classification, the ANN-classifier was trained with the preprocessed data of ES-2 (target = 0) and ES-5 (target = 1) of all but one subject. Then the classifier was tested with the data of the subject that was left out for training. The output was recorded. Here we also rated an output $x >= 0.5$ as 1 and $x < 0.5$ as 0.

Efficacy criterion. For the efficacy comparison of individual vs. inter-individual classifiers, paired t-tests and Cohen's effect sizes (d) were calculated (83),

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_1^2 + s_2^2)/2}},$$

where the $\bar{x}$ and s denote the mean value and standard deviation, respectively.

4.1.3 Results

Figure 4.1 shows the recognition results of the five cases:

- Case 1: The comparison between individual (N = 20; M = 92.6%, SD = 10.7%) vs. inter-individual classifiers (N = 20; M = 53.1%, SD = 13.4%) for round one was significant (p = .0001) with high effect size (d = 2.2).
- Case 2: The comparison between individual (N = 20; M = 92.3%, SD = 4.9%) vs. inter-individual classifiers (N = 20; M = 46.1%, SD = 7.4%) for round two was significant (p = .0001) with high effect size (d = 5.4).
- Case 3: The individual classifiers for all 20 subjects were trained on round one and tested on round two with a classification rate of 49.6% (SD = 19.6%).
- Case 4: We made individual specific feature selection for round one. We trained round one and recognized for round two with a classification rate of 53.0% (SD = 17.4%).
- Case 5: Subject-dependent feature selection for round one and two was performed. Round one was trained and round two was tested with a correct classification rate of 70.1% (SD = 17.8%).

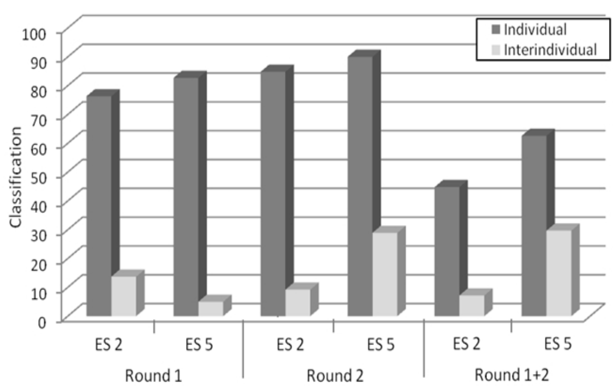

Figure 4.1: Recognition results of case 1-5 in percentage of correct classification

As shown in Figure 4.1, during the training of round one and the test of round two (Case 3-5) the classification performance decreased individually in comparison to the leave-one-out cross-validation of rounds one and two (Case 1 & 2). It is unlikely that habituation during round two caused recognition rates to drop, as the separate individual-specific classification of round two showed recognition rates of comparable accuracy to round one. In the Case 4, the individual-specific feature selection did not lead to a significant improvement on the recognition rates but just marginally of 3.4% in comparison to the Case 3, while the combined (situation-independent) feature selection on round one and two (Case 5) improved 20.5% of the recognition rate. Table 4.1 shows the automatically selected features for the classification of round one (Case 4) and the manually selected situation-independent features for round one and two (Case 5). In Table 4.2 the selection frequency of each feature after the situation-independent feature selection (Case 5) is listed. It turned out that both EMG signals from corrugator and zygomaticus muscles were highly relevant and effective as complementary variables for the differentiation of the emotional states 'high pleasure' and 'low pleasure' in Case 4 and Case 5 too. On the other side, the mean value of HRV time series (f_7) was effective in Case 4 but it was only selected for one subject in Case 5, while the spectral ratio (f_8) showed its consistent efficacy for both Cases.

To verify which descriptive emotional states are actually perceived by the subjects during the experiment, Table 4.3 shows the rating result from subject questionnaire. All participants stated that they experienced the same emotion in both the first and second round.

4.2 Emotion Classification using Artificial Neural Networks and 30 Features

The data was also analysed with 30 out of the standard 36 Picard features that were used in (10) to test the classification rates with a different feature set. Six features were calculated from each of the five physiological channels.

Subject #	Selected features ($f_{\#}$) for Case 4	situation-independent features ($f_{\#}$) for Case 5
112	[1, 2, 3, 5, 6, 7, 9, 12]	[3, 10, 12]
114	[1, 2, 3, 5, 6, 7, 8, 9]	[2, 5, 13]
118	[1, 3, 4, 5, 7, 9, 11]	[2, 11, 13]
125	[2, 3, 4, 5, 6, 7, 9, 10, 11, 13]	[4, 5, 6]
127	[3, 4, 5, 6, 7, 12]	[3, 6, 8]
129	[2, 4, 5, 6, 9, 11, 13]	[1, 3, 6]
208	[1, 2, 3, 4, 5, 6, 7, 10, 12]	[4, 7, 9]
212	[1, 3, 5, 6, 7, 8, 9]	[1, 2, 11]
215	[1, 2, 3, 4, 6, 13]	[8, 10, 11]
219	[2, 5, 7, 9, 13]	[1, 4, 5]
225	[1, 2, 4, 5, 6, 7, 9, 12]	[1, 3, 4]
226	[1, 2, 3, 4, 5, 6, 7, 9, 12]	[6, 11, 13]
308	[3, 4, 5, 6, 7, 11, 13]	[1, 6, 8]
423	[1, 3, 4, 5, 7, 8, 9]	[5, 10, 11]
427	[1, 2, 3, 6, 7, 9, 12, 13]	[2, 12, 13]
506	[1, 4, 5, 7, 10]	[4, 5, 6]
510	[4, 5, 6]	[2, 10]
511	[1, 2, 3, 4, 7, 9]	[2, 3]
518	[4, 5, 6, 7, 11, 13]	[5, 8, 9]
602	[3, 4, 5, 6, 7]	[1, 5]

Table 4.1: The list of selected features for Case 4 and Case 5

Feature ($f_{\#}$)	1	2	3	4	5	6	7	8	9	10	11	12	13
Frequency	6	6	5	4	6	5	1	4	2	4	5	2	4

Table 4.2: The selection frequency of each feature within the situation-specific feature selection (Case 5)

4.2.1 Feature Extraction

1. the mean of the raw signal

$$y_1 = \frac{1}{N}\sum_{n=1}^{N} X_n$$

2. the standard deviation of the raw signal

$$y_2 = (\frac{1}{N-1}\sum_{n=1}^{N}(X_n - y_1)^2)^{1/2}$$

Subject #	pleasure ES-2	arousal ES-2	dominance ES-2	pleasure ES-5	arousal ES-5	dominance ES-5
112	7	3	7	4	7	4
114	8	3	8	3	7	3
118	7	5	7	4	7	6
125	7	7	8	1	1	9
127	8	3	8	2	7	3
129	8	4	7	4	7	4
208	7	3	7	1	9	1
212	9	9	4	3	6	4
215	9	2	9	7	3	9
219	9	1	9	2	7	2
225	8	5	9	8	5	9
226	7	6	7	2	7	4
308	7	4	7	3	6	4
423	7	3	7	7	1	5
427	8	5	2	3	3	2
506	7	7	6	3	7	7
510	7	4	8	5	6	5
511	7	3	7	3	6	2
518	9	5	7	1	7	2
602	7	3	7	2	8	2
Average	7.65	4.25	7.05	3.4	5.9	4.4

Table 4.3: SAM Rating of ES-2 vs. ES-5

3. the mean of the absolute value of the first difference of the raw signal

$$y_3 = \frac{1}{N-1}\sum_{n=1}^{N-1}|X_{n+1} - X_n|$$

4. the mean of the absolute value of the first difference of the normalized signal

$$y_4 = \frac{1}{N-1}\sum_{n=1}^{N-1}|\tilde{X}_{n+1} - \tilde{X}_n| = \frac{y_3}{y_2}$$

5. the mean of the absolute value of the second difference of the raw signal

$$y_5 = \frac{1}{N-2}\sum_{n=1}^{N-2}|X_{n+2} - X_n|$$

6. the mean of the absolute value of the second difference of the normalized signal

$$y_6 = \frac{1}{N-2}\sum_{n=1}^{N-2}|\tilde{X}_{n+2} - \tilde{X}_n| = \frac{y_4}{y_3}$$

4.2.2 Classification

As the classification instrument, again a feed-forward Artificial Neural Network (ANN) in 13-40-20-1 architecture with two hidden layers (82) was employed.

4.2.3 Results

Very good classification results could be achieved in individual classification of pleasure (89%), arousal (88%), and es02 vs es05 (81%). When applying subject-independant classification, only about 50% classification rate could be achieved in all three cases (see Figure 4.2). This indicates the massive differences between results of individual classification, ranging up to almost 90% and subject-dependent classification residing around chance level.

4.3 Single Channel Analysis for Emotion Recognition[2]

4.3.1 Skin Conductance Response Behavior to Delayed Response Times

Introduction

System response time (SRT) research dates back to the late 1960s. But it is still an important issue in computer science research. As today network-based computing gains importance, software engineers have to be aware of network-related delays to be

[2]Contents of this section can also be found in (84)

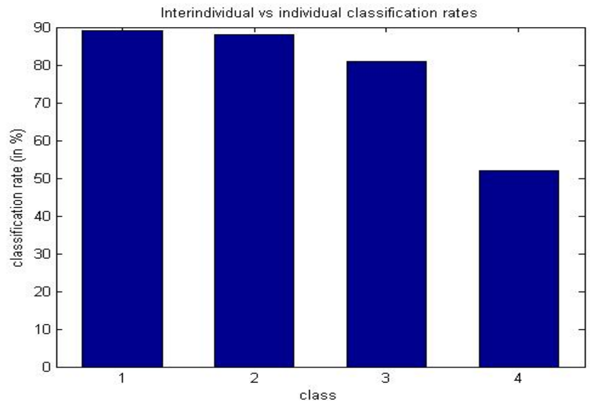

Figure 4.2: Subject-dependent vs subject-independent classification rates: subject-dependent classification of the pleasure-dimension (class 1), subject dependent classification of the arousal-dimension (class 2), subject dependent classification of experimental sequences es02 and es05 (class 3) and subject-independant classification (class 4).

able to improve user performance and satisfaction (85). Numerous studies concerning the best system response time for a user have been conducted, recommending specific response-time guidelines. These guidelines show that SRT durations of more than a few seconds are accepted when interaction tasks get more complex (86). But is a different situation with very simple, repetitive tasks. These, also called control tasks (85), should behave like physical devices and respond immediately or at least in a few tenths of a second. With more experience in usage, expectations about the SRT begin to establish. Therefore, novice users may wait much longer than experienced users are willing to (86). While a user is able to adopt to constant delays, variable, unexpected delays often disturb the process of interaction. Especially variations of twice the anticipated SRT might decrease the user's performance and cause frustration (87).

Numerous studies support that an increase in SRT leads to frustration, annoyance and irritation (see (85)). Rating the quality of the system reveals a decrease in perceived quality with increasing SRT. Furthermore, acceptance of such a system decreases. Especially under time pressure users get frustrated, annoyed or even angry by long system response times (87).

This study investigates whether a small delay of only 500ms is sufficient to elicit a physiological reaction and whether this response increases with longer delays lasting one or even two seconds. The results are compared to the results of the second scenario, where delays of six seconds are triggered in a stressful situation. Here we also have the possibility to compare the SC patterns during immediate display of a card compared to a delayed display of a card.
Whether delays in SRT could induce an emotional reaction and which kind of emotions participants could observe themselves during such an interruption was collected in a questionnaire at the end of the experimental session.
Depending on the situation at the time of interruption and on the expectancy of the organism, interruptions may produce a variety of different behavioral responses. According to (88) the interruption of many organized sequences produces repetition at first. Therefore we assume that after an unexpected delay of 500ms, the subject would tend to repeat his action and that this tendency increases with the duration of the SRT. In the second experiment, the delays are extended to six seconds. Here an even greater response in SC is expected.

Task Description - Experiment

Physiological data from 119 subjects (78 female, 41 male) who took part in a Wizard-of-Oz experiment was analysed in this study. Mean age was 47 years (SD = 23.27). Altogether 442 cases of delayed display of a card and 442 cases of immediate display of a card were analysed.
In the WOZ experiment, the participants had to solve a mental training task familiar to the game ‘concentration’. In each of the six experimental sequences (es01-es06) a number of hidden pictures was presented. The subject’s task was to uncover all matching card pairs. The whole interaction was controlled via voice. In es04, among other tools, delays were used in order to induce negative emotions.

Data Analysis

There are two paradigms in the second experiment:

- The card is displayed approximately one second after the user's request (immediately)
- The card is displayed approximately seven seconds after the user's request (delayed)

An overview over the time course of the experiment and in the paradigms 'no delay' (d = 0) and 'delay' (d = 6) can be seen in Fig. 4.3. The raw data was imported into MATLAB and for each case a baseline was calculated taking the mean SC value of one second before the user's request for the display of a card. This baseline was then subtracted from the next five seconds, ranging from the user's request for display of a card to five seconds after the request (see figure 4.3). In the case of immediate display, the card was displayed approximately one second after the request. In the case of a delayed display (in altogether 442 cases), the card wasn't displayed until two seconds after the analysed range.
After the experiment a SAM (Self Assessment Manikin) rating (36) was surveyed to gain emotional ratings of all six experimental sequences of the WOZ experiment.

Results

The course of the SC curves shown in Fig. 4.6 illustrates the mean of all 442 SC curves from all 119 subjects during a delay (red) and 442 SC curves from all 119 subjects during immediate display of a card (green) preceding the delayed displays.
As can be seen in Fig. 4.6 the divergence of the SC curves begins approximately 2.3 seconds after the display of a card / expected display of the card. It reaches its maximum (0.0125 μSiemens) at the end of the analysed range of 5 seconds.
The mean SAM ratings for the dimension 'pleasure' were 5.5 for es04 and 4.0 for es05. These were the experimental sequences where the delay was used. For the sequences es01, es02, es03 and es06 the mean SAM ratings were 7.6, 7.4, 7.2 and 7.8.

The mean difference in SC behavior between the trials immediate and the delayed trials over the range of five seconds is 0.007 μSiemens. Assuming a significance level of 0.05 the two conditions differ significantly (paired t-test, p-value <0.00001).

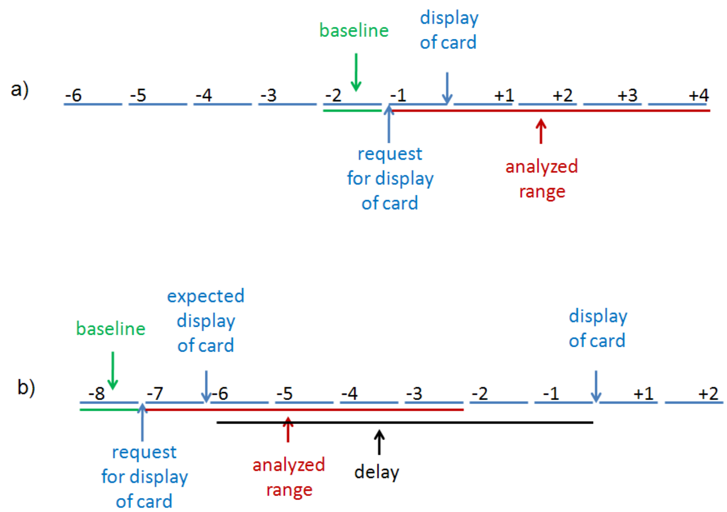

Figure 4.3: The time course in the two cases of immediate (a) and delayed (b) display. In (a), the card is displayed approximately one second after the subject's request. In (b), there is a delay of six seconds. The SC range plotted in Fig. 4.6 ranges from the user's request for the display of a card to five seconds after the request.

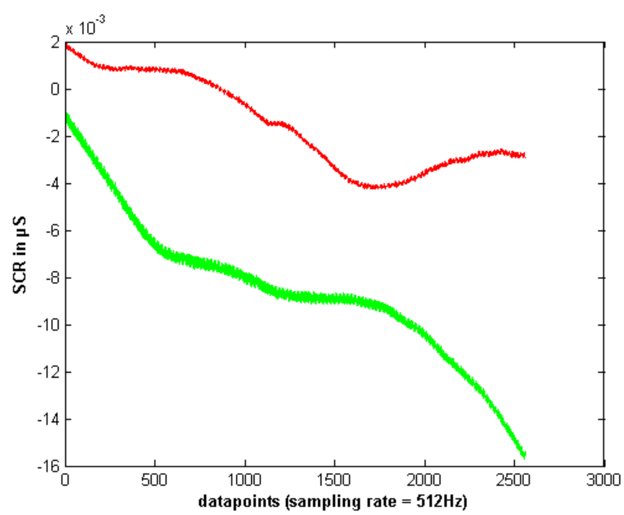

Figure 4.4: Plot of the mean SC reaction of 119 subjects to all 442 delayed card openings (red) and 442 immediate card openings (green) during the memory training experiment.

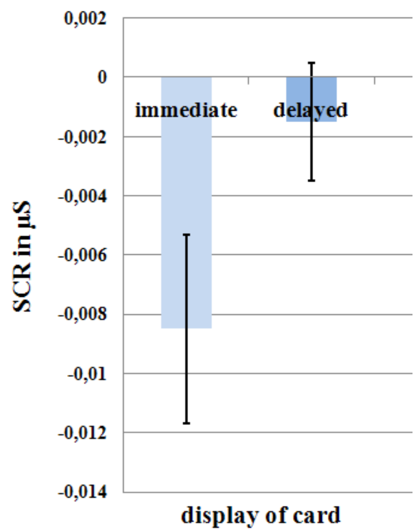

Figure 4.5: Mean and standard error of the SCR for the conditions of immediate card display and delayed card display (t = 6s). The SC response to immediate display is significantly lower than the response to delayed display.

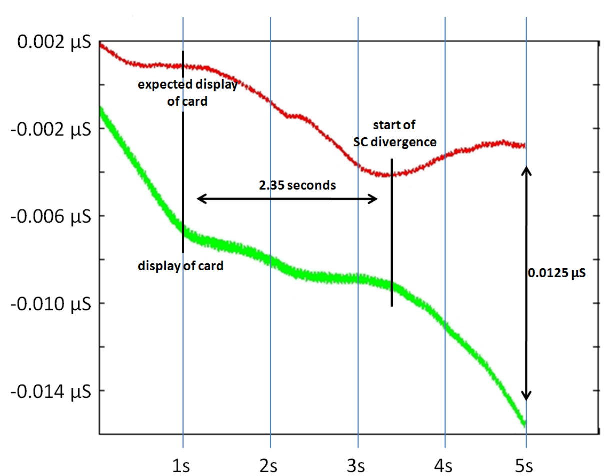

Figure 4.6: Analysis of the plot of the mean SC reaction of 119 subjects to all 442 delayed card openings (red) and 442 immediate card openings (green) during the memory training experiment. The divergence of the SC curve begins around 2.35 seconds after the display of a card / expected display of a card.

4.4 A Novel Approach - Emotion Identification Modules Based on Feature-Pair Transitions[3]

Assuming that in order to be able to identify an emotion change in a two-dimensional emotion space (pleasure and arousal), it is not sufficient to observe changes of single physiological features. By regarding **feature-pair changes** instead of single feature changes, four different states can be coded which are all possible states that can occur in the two-dimensional space of pleasure and arousal, mainly 'LA' (low arousal or the arousal is decreasing), 'HA' (high arousal or the arousal is increasing), 'LP' (low pleasure or the pleasure is decreasing) and 'HP' (high pleasure or the pleasure is increasing). As an example the feature-pair 'EMG_corr, EMG_zyg' - the EMG chanel of the corrugator supercilii and the EMG channel of the zygomaticus major can be regarded (see Figure 4.7). In the present study, the feature-pair data analysis showed that in 32 out of 55 subjects (58%) the activity of the corrugator supersilii (the frown muscle) has increased while the activity of the zygomaticus major (the laugh musle) has decreased when the pleasure dimension was manipulated from positive to negative thus decreasing. In 27 out of 55 subjects (44%) both, the activity of the corrugator supercilii and the zygomaticus major has decreased when changing the arousal dimension from positive to negative thus decreasing (chance level: 25%). Assuming that '1' represents an increase and '-1' represents a decrease, in the case of the feature-pair 'EMG_corr, EMG_zyg' '-1-1' indicates a decrease of arousal, '-11' indicates a decrease of pleasure, '1-1' indicates an increase of pleasure and '11' indicates an increase of arousal.

4.4.1 Methods

The aim of the present study was to empirically detect the most emotionally relevant feature-pairs and their specific transition directions within the physiological data of IAPSg1 and to combine those feature-pairs into Emotion Identification Modules with

[3]Contents of this section can also be found in (89), (90) and (91)

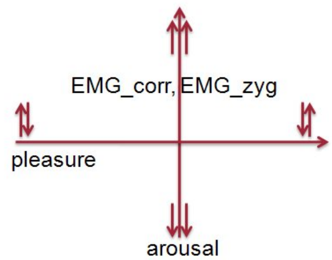

Figure 4.7: The four possible states (-1-1, -11, 1-1, 11) of the feature-pair 'EMG_corr, EMG_zyg' have the capability of representing all four possible emotion changes (decrease of arousal, decrease of pleasure, increase of pleasure, increase of arousal) in the two-dimensional space of arousal and pleasure.

the functionality of identifying and rating an emotion change when provided with a feature transition vector.

Identification of the Most Emotionally Relevant Feature-Pairs

In order to identify the most emotionally relevant feature-pairs within the IAPSg1 data the physiological data of all 55 subjects was partitioned into

- LA - physiological data that was recorded during the induction of low arousal
- HA - physiological data that was recorded during the induction of high arousal
- LP - physiological data that was recorded during the induction of low pleasure
- HP - physiological data that was recorded during the induction of high pleasure

For this purpose the blocked IAPS data inducing LPLA and HPLA was merged into LA data, LPHA and HPHA data was merged into HA data (see Figure 4.9), LPLA and LPHA data was merged into LP data and HPLA and HPHA data was merged into HP data (see Figure 4.8).
Afterwards, the six features described in chapter 3.3 were extracted from the physiological data. In order to obtain the emotionally relevant feature-pair changes for the transition from the state of 'low arousal' to the state of 'high arousal' (LA $\rightarrow$

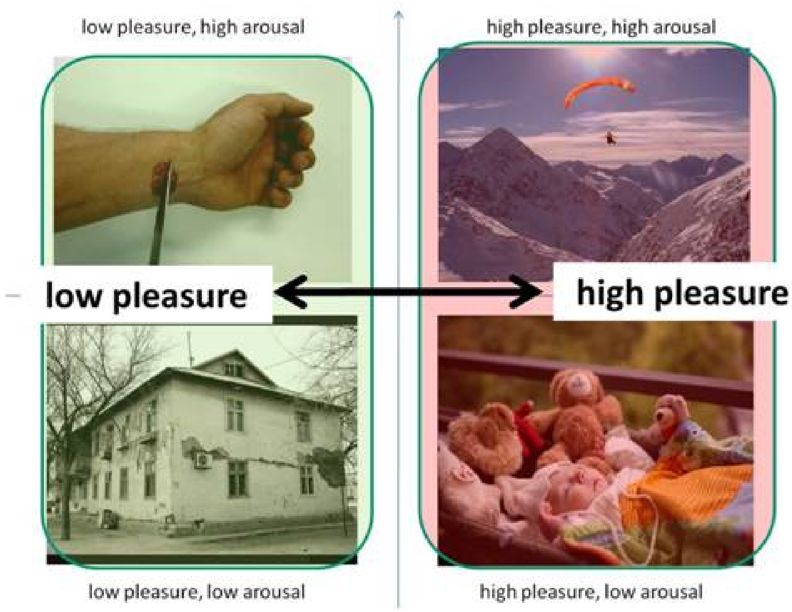

Figure 4.8: Blocking of the data acquired during the blocked IAPS presentation into the groups 'low pleasure' and 'high pleasure'. These pictures are similar to the IAPS picture set (2).

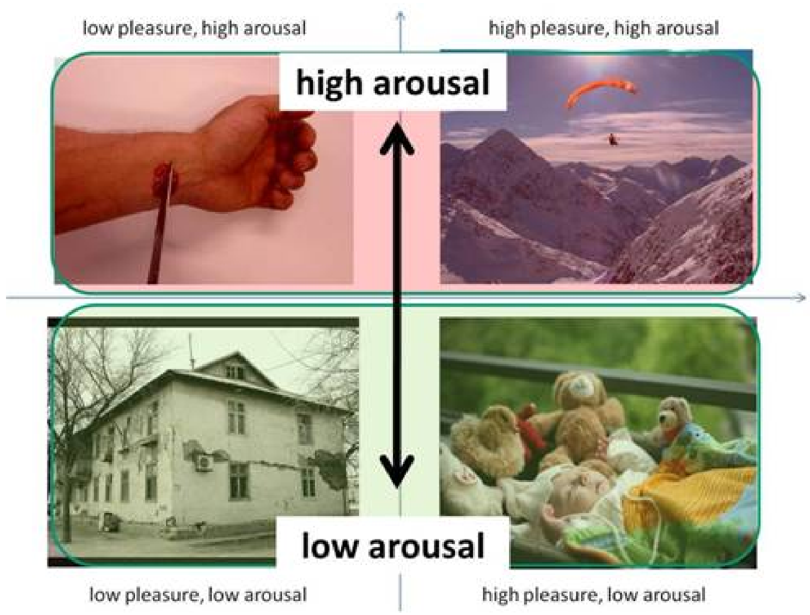

Figure 4.9: Blocking of the data acquired during the blocked IAPS presentation into the groups 'low arousal' and 'high arousal'. These pictures are similar to the IAPS picture set (2).

Table 4.4: Single feature transitions of the six features when the emotion-dimension 'arousal' increases (condition LA $\rightarrow$ HA) for 10 of the 55 subjects when watching blocked IAPS images. '-1': decreasing, '1': increasing.

	subjects									
features	1	2	3	4	5	6	7	8	9	10
f1 (EMG_corr)	1	1	1	1	1	1	1	1	1	1
f2 (EMG_zyg)	1	1	1	-1	1	-1	1	1	1	1
f3 (SC_fluc)	-1	1	1	-1	-1	-1	-1	-1	-1	1
f4 (SC_slope)	1	1	-1	1	1	-1	1	1	1	-1
f5 (SCL_mean)	1	1	1	1	1	1	1	1	1	1
f6 (HR)	-1	1	1	1	1	1	1	1	-1	-1

Table 4.5: Single feature transitions of the six features when the emotion dimension 'pleasure' increases (condition LP $\rightarrow$ HP) for 15 of the 55 subjects when viewing blocked IAPS images. '-1': decreasing, '1': increasing.

	subjects									
features	1	2	3	4	5	6	7	8	9	10
f1 (EMG_corr)	-1	1	-1	-1	1	-1	1	-1	-1	-1
f2 (EMG_zyg)	1	1	-1	1	1	-1	-1	1	-1	1
f3 (SC_fluc)	-1	-1	-1	-1	-1	1	-1	1	-1	1
f4 (SC_slope)	1	1	-1	1	1	-1	1	-1	1	-1
f5 (SCL_mean)	-1	1	1	1	-1	-1	1	1	1	-1
f6 (HR)	1	-1	1	1	-1	-1	1	1	-1	1

HA) representing an increase of arousal, the mean values of the LA features were subtracted from the mean values of the HA features. If the result was positive, the feature transition was set to 1, if it was negative, the feature transition was set to -1. In Table 4.4 the feature transition vectors for 10 subjects for the emotional condition LA $\rightarrow$ HA can be seen. The same procedure was applied to the pleasure dimension data (see Table 4.5) to obtain two 55x6 sized matrices of feature transitions containing a 1x6 feature transition vector for each subject.

From all of the possible $\binom{6}{2} = 15$ feature-pair combinations the occurrence frequency of all four possible states (-1-1, -11, 1-1, 11) was calculated. Table 4.6 lists the five most frequent feature-pair transitions for the condition LP $\rightarrow$ HP.

Table 4.6: The five most frequent feature-pair transitions for the dimension of pleasure (LP → HP) identified within the blocked IAPS data of 55 subjects. A weight was assigned to each of the feature-pairs according to their occurrence frequencies.

feature-pair	transition directions	freq	freq in %	weight
f1,f2 (EMG_corr, EMG_zyg)	-1,+1	32	58	5
f1,f5 (EMG_corr, SCL_mean)	-1,+1	30	55	4
f1,f4 (EMG_corr, SC_slope)	-1,+1	29	53	3
f1,f3 (EMG_corr, SC_fluc)	-1,-1	26	47	2
f2,f4 (EMG_zyg, SC_slope)	+1,+1	25	45	1

Table 4.7: The five most frequent feature-pair transitions for the dimension of arousal (LA → HA) identified within the blocked IAPS data of 55 subjects. A weight was assigned to each of the feature-pairs according to their occurrence frequencies.

feature-pair	transition directions	freq	freq in %	weight
f1,f6 (EMG_corr, HR)	+1,+1	27	49	5
f1,f2 (EMG_corr, EMG_zyg)	+1,+1	24	44	4
f1,f4 (EMG_corr, SC_slope)	+1,+1	23	42	2
f1,f5 (EMG_corr, SCL_mean)	+1,+1	23	42	2
f4,f6 (SC_slope, HR)	+1+1	23	42	2

Construction of the Emotion Identification Modules

Having determined the five most frequent feature-pairs with their specific transition combinations that occur within the analysed data, the feature-pairs were weighted with a weight ranging from one to five according to their occurrence frequencies (see the last column in Table 4.6 and 4.7). These five feature-pairs with their specific transition directions were centralized into EIM_p - the Emotion Identification Module for the pleasure dimension and EIM_a - the Emotion Identification Module for arousal. Hence the Emotion Identification Modules consist of five weighted feature-pair transitions with specific transition directions directions each.

4.4.2 Emotion Change Identification with the EIMs

The application of the EIMs is fast and requires low computation time. First a transition vector needs to be computed by subtraction of two feature vectors. Given a transition vector containing the transition states of all six features, the transitions of

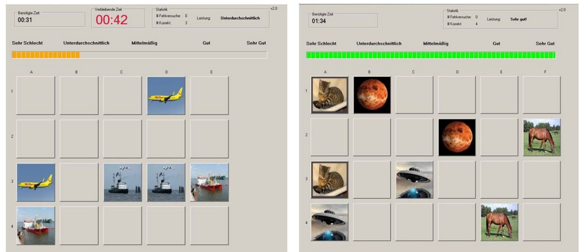

Figure 4.10: The subject's screen in the experimental sequences 4 (es04) and 6 (es06). The induced emotions were 'LPHA' in es04 (left) and 'HPLA' in es06 (right).

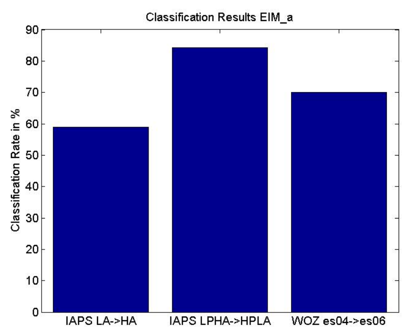

Figure 4.11: EIM_a's classification rates when comparing the IAPSg2's LA with HA, the IAPSg2's LPHA with HPLA and the WOZg1's es04 with es06.

the feature-pairs that are stored in the EIM_p and EIM_a must be compared to the transition states of the transition vector. In the case of a match the given weight is added to the EIM's output. For example if the transition state of f1 equals -1 and the transition state of f2 equals +1, the output of EIM_p is increased by five (see Table 4.6). If the transition states of the transition vector equal the exact opposite of the stored states, the weight is subtracted from the EIM's output. For example if the transition state of f1 equals +1 and the transition state of f5 equals -1, the output of EIM_p is decreased by four.

An example of the application of the Emotion Identification Module for arousal (EIM_a) can be seen in Table 4.8 where the EIM_a weights are calculated for four subjects.

Table 4.8: Feature transition vectors of four subjects and the votes of the EIM_a, based on a comparison of states of transition pairs.

subjects	1	2	3	4
f1 (EMG_corr)	-1	1	1	-1
f2 (EMG_zyg)	1	-1	-1	-1
f3 (SC_fluc)	-1	-1	-1	-1
f4 (SC_slope)	1	1	-1	1
f5 (SCL_mean)	-1	-1	1	-1
f6 (HR)	-1	1	1	1
f1,f6 weight	-5	5	5	0
f1,f2 weight	0	0	0	-4
f1,f4 weight	0	2	0	0
f1,f5 weight	-2	0	2	-2
f4,f6 weight	0	2	0	2
sum (overall weight of the EIM_a)	-7	9	7	-4

Table 4.9: Mean classification rates of the Emotion Identification Module for pleasure (EIM_p) and for arousal (EIM_a)

	EIM_p	EIM_a
IAPS (LP → HP / LA → HA)	56%	75%
IAPS (LPHA → HPLA)	84%	79%
WOZ (es04 → es06)	72%	70%

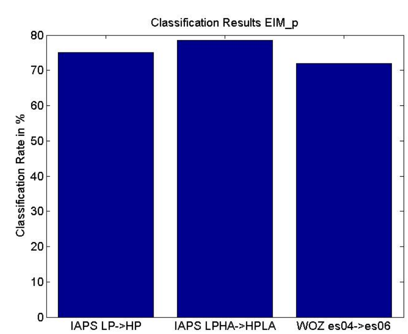

Figure 4.12: EIM_p's classification rates when comparing the IAPSg2's LP with HP, the IAPSg2's LPHA with HPLA and the WOZg1's es04 with es06.

The most common feature transition pairs, found in the IAPSg1 data set of 55 subjects when analysing changes in the pleasure dimension:

- EMG_corr, EMG_zyg
- EMG_corr, SCL_mean
- EMG_corr, SC_slope
- EMG_corr, SC_fluc
- EMG_zyg, SC_slope

The most common feature transition pairs, found in the IAPSg1 data set of 55 subjects when analysing changes in the arousal dimension:

- HR, EMG_corr
- EMG_corr, EMG_zyg
- EMG_corr, SC_slope
- EMG_corr, SCL_mean
- SC_slope, HR

4.4.3 Results

The results of the EIM emotion identification are shown in Table 4.9, Figure 4.11 and Figure 4.12. The mean EIM rating for the emotion transition LP $\rightarrow$ HP was +6.2 and +4.5 for the transition LA $\rightarrow$ HA. Overall in 56% (pleasure) and 75% (arousal) the induced emotion change was identified correctly by the modules. In the second case, the transition HPLA $\rightarrow$ LPHA was tested. The correct classification rate was 84% for the pleasure dimension and 79% for the arousal dimension. In the case of WOZg1 data, the correct classification rate was 72% (pleasure) and 70% (arousal).

To compare the EIM's identification rates to classification rates of commonly applied procedures, a classification was performed on the same data applying a leave-one-out classification via Artificial Neural Networks (ANN). For each subject, a feed-forward Artificial Neural Network (ANN) in 6-40-20-1 architecture with two hidden layers (82) was trained and tested via leave-one-out cross-validation. The same six features were used for each subject, no feature selection was performed. Applying this method, classification rates of 88% were achieved in classification of the IAPSg2 data set and classification rates of 81% were achieved in classification of the WOZg1 data set. But when the IAPSg2 data set was used for training the ANN and WOZg1 es04 and es06 were used for testing, the classification rates did not exceed 55% (almost chance level).

Chapter 5

Discussion

In this work, both a machine learning approach as well as a single channel analysis were tested for their applicability in the field of emotion classification from physiological data.

Single channel analysis (see chapter 4.3) was performed on data from an experiment in which skin conductance changes during different system response times were recorded. The findings confirm the frequent observation that users get annoyed and frustrated when unexpected delays in the system responses occur. In such a situation the user is faced with two possible questions: 'Was my action registered?' and 'Do I have to repeat it again?' (see (92)). The physiological data reveals an increase in skin conductance during unexpected delays compared with immediate feedback presentation. Even delays of only 500ms were sufficient to trigger this physiological change, but longer delays elicited a greater response. The increase of skin conductance is in line with findings of Kuhlmann (93). Kuhlmann found a greater number of spontaneous skin conductance responses and a higher skin conductance level during blocks of longer system response times (8s) compared to blocks of shorter system response times (2s).

In chapter 4.3 the effects of delays in a stressful, realistic scenario were observed and the differences in the average skin conductance level were found to be even greater. This effect can be explained by the setup of the experiment - the deliberate failure to display a card requested by the subject being a negative event. The pattern of the skin conductance curve changed 2.35 seconds after the display of a card. With the latency of a skin conductance reaction being 1.0 - 3.0 seconds ((94)), the introduced delay perfectly explains the change of the skin conductance pattern after 2.35 seconds. Since nearly all subjects reported negative emotions elicited by the delays in the SAM ratings, the rise of skin conductance level can, in this case, be interpreted as negative arousal and should be avoided in a human-computer-interaction.

When machine learning analysis was applied in chapter 4.1, classification rates of 89% for the pleasure dimension, 88% for the arousal dimension, 81% for es02-es05-classification and approximately 50% for subject-independent classification (see chapter 4.2) were achieved. These results are in accordance with other groups' findings in subject-dependent emotion classification using equivalent techniques.

By way of comparison, Haag et. al. (95) achieved classification rates of 96.6% for the arousal dimension and 89.9% for the pleasure dimension having collected 1000 data segments, using 700 for training and 300 for testing via cross-validation. The shortcoming of Haag's study was the limited sample size - only one subject's physiological data was used for the analysis.

Kim et. al. (40) achieved mean classification rates of 87% in subject-dependent emotion recognition applying linear discriminant analysis. In subject-independent emotion recognition the classification rate only came to 65%. As with Haag's study the sample size was once again a significant shortcoming - only three subjects' data was used.

Bailenson et. al. (96) achieved mean classification rates of 82% when classifying amusement versus sadness using physiological data and individual subjects, but only 46% in subject-independent classification. In contrast to the two aforementioned studies, Bailenson's results were considerably more robust, given that 41 subjects' data was used.

The massive discrepancy between the results of subject-dependent classification (ranging up to almost 90%) and subject-independent classification (residing around chance level) illustrates the major drawback of the machine learning approach for emotion classification given the requirement of recognising an emotion in a new and unseen user.

Walter and Kim demonstrated in (97) that situation-independent classification - being referred to as 'the hard problem' - is also an unsolved challenge. This can be seen in chapter 4.1 where results of only 57% were achieved in situation-independent classification. One explanation for the lack of satisfactory results in situation-independent classification could be the social judgment theory put forward by Sherif (98). Another explanation is the fact that physiological signals are never exactly the same on two days and therefore patterns that have been trained on day one can never exactly match patterns on day two.
This indicates that the machine learning approach to emotion classification is very limited since it cannot be applied whenever subject-independent or situation-independent classification abilities are required.
On the other hand, single channel analysis may be able to indicate certain changes, but it is not possible to make a statement about emotional changes in a two-dimensional space using this method. A change in a single parameter cannot represent more than one state, but for an emotion representation in a two-dimensional space, four states are required - which in turn necessitates a minimum of two variables.
Therefore a novel approach was introduced in this work through the use of Emotion Identification Modules. These Modules were created on the basis of physiological feature-pair changes and identify an emotional change by comparing the feature-pair transitions from physiological data of two distinct situations. With this newly developed method, correct recognition rates of 79% and 84% were achieved during subject-independent emotion recognition and correct recognition rates of 70% and 72% were achieved in situation-independent emotion recognition (see chapter 4.4). This fact constitutes a major progress and represents a considerable and promising result in emotion recognition from physiological data. Since the EIMs do not perform an actual classification, it is hard to compare the recognition results with classification results found in the literature. But it can be said that the aforementioned problems of subject- and especially situation-independent emotion recognition have now been resolved utilising the EIMs.
Recently, Kolodyazhniy and his team (99) were able to achieve classification rates of around 75% in subject- and situation-independent classification using nonlinear

classifiers and a feature-selection procedure, but these results are based on a cross-validation procedure which is controversial, as described in 3.3.

Since the creation of the EIMs is based on empirical analysis of data from as many as 55 subjects, very robust subject-independent recognition rates could be achieved. The superiority in situation-independent testing was achieved because the EIMs compare two consecutive sequences of physiological data instead of training inflexible patterns during one particular situation. Therefore the day-to-day changes in the human physiological signals do not effect the EIMs' recognition ability.

Additionally, the EIMs have low computational complexity and therefore are predestined to be used for real-time emotion recognition.

Because of the EmoRec II corpus' size - consisting of data from 110 subjects (see chapter 3) - which was used for this study, the results are representative and therefore the EIMs can legitimately be referred to as universal.

As far as the relevance of feature-pairs to emotion recognition is concerned, it is not surprising that the feature-pair that was extracted from the facial EMGs of the corrugator supercilii and the zygomaticus major (EMG_cor, EMG_zyg) was found to be the most predictive feature-pair in terms of emotion change identification within the pleasure dimension. Decreasing activity of the corrugator major and at the same time increasing activity of the zygomaticus supercilii (feature-pair change 'EMG_corr, EMG_zyg: -1+1') indicate an increase of pleasure and vice versa. It is however an interesting fact that this feature-pair also contributes to the arousal dimension of emotions. Simultaneous decrease of activity of both facial muscles ('EMG_corr, EMG_zyg: -1-1') indicates a decrease in the overall arousal whereas simultaneous increase of facial muscle activity ('EMG_corr, EMG_zyg: +1+1') indicates an increase in arousal.

Chapter 6

Outlook

The extensive EmoRec II data set comprising physiological data of 110 subjects undergoing diverse emotional paradigms enables us to continue empirical analysis on physiological feature-transitions in emotional situations. In the future, we will increase the number of extracted features in order to obtain a larger feature-pair pool. The subjects will also be arranged into groups of similar age and gender for subject- and group-specific feature-pair detection.

We see the EIM's field of application in critical situations where sudden emotion changes need to be identified. These might occur during human-computer interaction or when a human is operating a machine. Cues from different modalities could then trigger the EIMs, receive the calculated emotion change of the subject and enable the system to react accordingly.

The EIMs could also be used to compare consecutive sequences of physiological data resulting in a local emotion curve that allows continuous emotion state tracking. For this purpose, the EIMs need to be calibrated individually - with different modules being weighted according to the individual's specific physiological reactions.

For the purpose of continuous emotion state recognition, a Graphical User Interface (GUI) was programmed in Matlab that made it possible to load a subject's data and to view the experiment videos. Together with the videos, the relevant triggers and a plot showing the corresponding 2D representation of the subject's emotion was displayed. When only the activity of the corrugator's EMG channel is used to determine

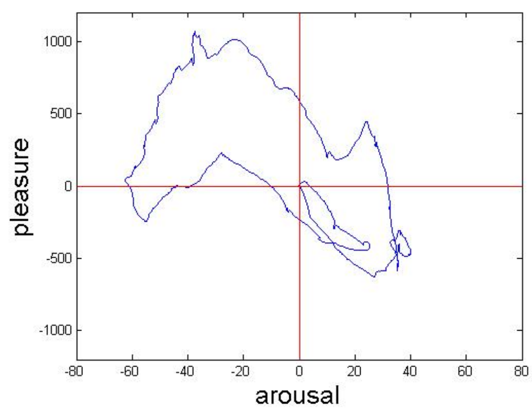

Figure 6.1: Skin conductance (x-axis) vs Corrugator EMG (y-achsis) during the WOZ experiment.

the pleasure dimension and the mean value of the skin conductance is used to determine the arousal level, in some cases, a curve can be drawn showing the subject's emotional course during the WOZ experiment (see Figure 6.1).
A precise calibration of the EIM's will be crucial for this kind of application in order to avoid drifts in the EIM's output. Since the EIMs have low computational complexity, such a tool could perform online emotion state recognition which would be very useful for future companion systems. Such a system could always know the user's emotional state within the two-dimensional space of arousal and pleasure, enabling it to react to it appropriately.

Chapter 7

Summary

A lot of effort has been put into emotion classification from physiological data in the recent years. The reported classification rates are fairly high but only when a cross-validation is performed, given that the test data and the training data derive from the same single data set. The classification rates may be improved still further when a good feature selection is used as there will always be some data set specific features that will give solid results using a leave-one-out cross-validation. However using too many features and applying a feature selection may result in the problem of overfitting (see chapter 3.3). Emotion classification of data sets that are completely new, unseen and which may be different depending on the emotional context is still an unsolved problem when using machine learning methods (see for example (97) or (99)).

Single channel analysis, as can be seen in 4.3 is another way of extracting information about the emotional state of a user. But there is one big drawback - the fact that a change in the value of a physiological channel can often be interpreted in several ways. Kreibig et al. (1) give an overview of studies on peripheral physiological responses to fear and sadness. It shows that previous studies have asserted that the heart rate, the skin conductance level, and the EMG of the *corrugator supercilii* rise during fear, but also during sadness. One can therefore conclude, that by examining only a single parameter, determining the cause of an observed effect, such as an acceleration of a subject's heart rate, becomes problematic. In this case, either sadness or fear could be the source of change in the particular physiological parameter; by examining only a single parameter one can not determine whether fear or sadness is the emotion that was responsible for the observed physiological change. The primary advantage of this method is that these particular physiological changes have been identified by many groups independently and therefore seem to be largely universal.

Neither of the methods described above have yet led to successful subject- and situation-independent emotion recognition. For this reason a new method for emotion recognition is needed. It is submitted herein that the best approach, and solution to this problem, is to extract only meaningful features in order to avoid the problem of data set specific feature training and overfitting. Furthermore it is necessary to explore the differences between selected features when comparing two emotional states at two points in time, and combine these feature tendencies into feature-pair-modules to be able to distinguish between changes in the arousal dimension and changes in the pleasure dimension of the emotion space.

These so-called Emotion Identification Modules were constructed using the IAPSg1 data set and tested on the IAPSg2 data set (see chapter 3.1.1) - in which the subjects were watching pictures without interacting with the system in order to identify subject-independent emotion-change. The WOZg1 data set was also used for testing the Emotion Identification Modules - in this instance the subjects were interacting with the WOZ mental trainer which was controlled by voice whilst a cognitive task was solved in order to identify situation-independent emotion-change. With each data set containing physiological data of 55 subjects, the size of this corpus exceeds most of the current tested data sets in this area of research.

The concept behind the Emotion Identification Modules is to empirically detect feature-pairs that are consistent and robust in their behaviour throughout many subjects. It is essential that the feature-pairs remain so when a subject's emotion changes in order to guarantee universal subject- and situation-independent results. Emotion change recognition rates of 70-84% in subject- and situation-independent environment show the potential of this newly developed method. In future it is anticipated that this method will be improved further by individual calibration for even more accurate and robust results in terms of user-specific companion abilities.

Acknowledgment

I would like to thank all my friends and colleagues Stefanie Rukavina, Jun-Wen Tan, Andreas Scheck, Holger Hoffmann, Steffen Walter, Hang Li, Sascha Gruß, Kerstin Limbrecht, Timo Schuster, Lucia Jerg-Bretzke and Doreen Poeppel for the extraordinary teamwork as well as fun that was only possible due to their efforts within our group. Also I would like to thank my friends and colleagues in Magdeburg, Andreas Meineke, Ingo Siegert, Ronald Böck, Susann Wolff and Christin Kohrs.

I owe thanks and respect especially to Professor Harald C. Traue who is the best boss, team-mate and friend one can wish for.

This research was supported by grants from the Transregional Collaborative Research Centre SFB/TRR 62 'A Companion-Technology for Cognitive Technical Systems' founded by the German Research Foundation (DFG).

Attachments - Publications

D. Hrabal, S. Rukavina, J. W. Tan, A. Scheck, S. Walter, and H. C. Traue, “Feature-pair-based emotion change identification from physiological data in pleasure-arousal space,” *IEEE Transactions on Affective Computing, submitted*, vol. PP, pp. 1–9, 2013.

D. Hrabal, S. Rukavina, K. Limbrecht, S. Gruss, S. Walter, V. Hrabal, and H. C. Traue, “Emotion identifiaction and modelling on the basis of paired physiological data features for companion systems,” in *22nd International Conference on Automated Planning and Scheduling*, 2012.

D. Hrabal, C. Kohrs, A. Brechmann, J. W. Tan, S. Rukavina, and H. C. Traue, “Physiological effects of delayed system response time on skin conductance,” in *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction*, Springer, 2013.

S. Walter, J. Kim, D. Hrabal, S. Crawcour, H. Kessler, and H. C. Traue, “Transsituational individual-specific biopsychological classification of emotions,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. PP, pp. 1–8, 2013.

S. Walter, S. Scherer, M. Schels, M. Glodek, D. Hrabal, M. Schmidt, R. Böck, K. Limbrecht, H. C. Traue, and F. Schwenker, “Multimodal emotion classification in naturalistic user behavior,” in *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments*, Springer, 2011.

R. Böck, S. Gluge, A. Wendemuth, K. Limbrecht, S. Walter, D. Hrabal, and H. C. Traue, “Intraindividual and interindividual multimodal emotion analyses in human-machine-interaction,” in *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*, 2012.

M. Schels, M. Kächele, D. Hrabal, S. Walter, H. C. Traue, and F. Schwenker, "Classification of emotional states in a WOZ scenario exploiting labeled and unlabeled bio-physiological data," in *Lecture Notes in Computer Science*, Springer, 2012.

J. W. Tan, S. Walter, A. Scheck, D. Hrabal, H. Hoffmann, H. Kessler, and H. C. Traue, "Repeatability of facial electromyography (EMG) activity over corrugator supercilii and zygomaticus major on differentiating various emotions," *Journal of Ambient Intelligence and Humanized Computing*, vol. 3, pp. 3–10, 2012.

J. W. Tan, S. Walter, A. Scheck, D. Hrabal, H. Hoffmann, H. Kessler, and H. C. Traue, "Facial electromyography (fEMG) activities in response to affective visual stimulation," in *IEEE Workshop on Affective Computational Intelligence*, 2011.

J. Katicic, A. Bachvarov, S. Walter, and D. Hrabal, "Emotional customer feedback on virtual products in immersive environment," in *20th Scientific and Technical Conference with International Participation*, 2011.

S. Walter, D. Hrabal, A. Scheck, H. Kessler, G. Bertrand, F. Notdurft, W. Minker, and H. C. Traue, "Individual emotional profiles in Wizard-of-Oz-experiments," in *3rd International Conference on Pervasive Technologies Related to Assistive Environmants*, 2010.

K. Limbrecht, D. Hrabal, H. Hoffmann, A. Lubeck, and H. C. Traue, "The FEEL-R - facial emotion recognition with stimulus material from five perspectives," in *13th European Conference on Facial Expression*, 2010.

Bibliography

[1] S. D. Kreibig, F. H. Wilhelm, W. T. Roth, and J. J. Gross, Cardiovascular, electrodermal and respiratory response patterns to fear- and sadness-inducing films *Psychophysiology*, vol. 44, pp. 787–806, 2007.

[2] P. Lang, M. Bradley, and B. Cuthbert, International affective picture system (IAPS): instruction manual and affective ratings tech. rep., University of Florida, 1999.

[3] H. C. Traue, F. Ohl, K. Limbrecht, S. Scherer, F. Schwenker, M. Kotzyba, A. Brechmann, H. Hoffmann, A. Scheck, and S. Walter, *Framework for Emotions and Dispositions in Man-Companion-Interaction.* Science Publishers, in press, 2013.

[4] Y. Wilks, *Close Engagements with Artificial Companion.* John Benjamins Publishing Company, 2010.

[5] D. M. Romano, Look, emotion, language and behavior in a believable virtual Companion in *Close Engagements with Artificial Companions*, John Benjamins Publishing Company, 2010.

[6] C. Bischoff, S. Schmädeke, C. Dreher, M. Adam, D. Bencetic, and K. Limbacher, Akzeptanz von elektronischem Coaching in der psychosomatischen Rehabilitation *Verhaltenstherapie und Verhaltensmedizin*, vol. 31, pp. 274–287, 2010.

[7] C. Bischoff and S. Schmädeke, eATROS - eine Smartphone-gestützte Intervention für Patienten mit affektiven Störungen in der ambulanten Reha-Nachsorge-Phase *Gesundheit im sozialen Wandel*, vol. 38, p. 65:73, 2011.

[8] I. Christie and B. Friedmann, Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach *International Journal of Psychophysiology*, vol. 51, pp. 143–153, 2004.

[9] G. Stemmler, M. Heldmann, C. Pauls, and T. Scherer, Constrains for emotion specificity in fear and anger: The context counts *Psychophysiology*, vol. 38, pp. 275–291, 2001.

[10] R. Picard, E. Vyzas, and J. Healey, Toward machine emotional intelligence: Analysis of acffective physiological state *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1175–1191, 2001.

[11] P. Cobos, M. Sánchez, N. Pérez, and J. Vila, Effects of spinal cord injuries of the subjective component of emotions *Cognitive Emotion*, vol. 18, pp. 281–287, 2004.

[12] S. Schachter and J. E. Singer, Cognitive, social and physiological determinants of emotional state *Psychological Review*, vol. 69, pp. 379–399, 1962.

[13] P. Ekman, Are there basic emotions? *Psychological Review*, vol. 99, pp. 550–553, 1992.

[14] J. A. Russel and A. Mehrabian, Evidence for a Three-Factor Theory of Emotions *Journal of Research in Personality*, vol. 11, pp. 273–294, 1977.

[15] J. T. Cacioppo and G. G. Berntson, Relationships Between Attitudes and Evaluative Space: A Critical Review with Emphasis on the Separability of Positive and Negative Substrates *Psychological Bulletin*, vol. 115, pp. 401–423, 1994.

[16] P. J. Lang, M. Bradley, and B. Cuthbert, Emotion, Attention, and the Startle Reflex *Psychological Review*, vol. 97, pp. 377–395, 1990.

[17] A. Dickinson and M. F. Dearing, *Mechanisms of learning and motivation*. Routledge, 1992.

[18] R. W. Levenson, P. Ekman, and W. V. Friesen, Voluntary Facial Action Generates Emotion-Specific Autonomic Nervous System Activity *Psychophysiology*, vol. 27, pp. 363–384, 1990.

[19] H. T. Shupp, B. N. Cuthbert, M. M. Bradley, J. T. Cacioppo, T. Ito, and P. J. Lang, Affective picture processing: The late positive potential is modulated by motivational relevance *Psychophysiology*, vol. 37, pp. 257–261, 2000.

[20] H. E. Jones, *Feelings and Emotion*. McGraw-Hill, 1948.

[21] R. Buck, Measuring individual differences in the nonverbal communication of affect: The slide-Viewing paradigm *Human Communication Research*, vol. 6, pp. 47–57, 1979.

[22] R. Buck, R. E. Miller, and W. F. Caul, Sex, Personality and Phsyiological Variables in the Communication of Affect via Facial Expression *Journal of Personality and Social Psychology*, vol. 30, pp. 587–596, 1974.

[23] H. C. Traue, Emotional Inhibition and Health in *The International Encyclopedia of the Social & Behavioral Sciences*, Elsevier, 2001.

[24] R. M. Deighton and H. C. Traue, Emotional inhibition and somatization across cultures *International Review of Social Psychology*, vol. 18, pp. 109–140, 2005.

[25] B. Abler and H. Kessler, Emotion Regulation Questionnaire - eine deutschsprachige Fassung des ERQ von Gross und John *Diagnostica*, vol. 55, pp. 144–152, 2009.

[26] D. Hagemann, E. Naumann, S. Maier, G. Becker, A. Luerken, and D. Bartussek, The Assessment of Affective Reactivity Using Films: Validity, Reliability and Sex Differences *Personality and Individual Differences*, vol. 26, pp. 627–639, 1999.

[27] J. Hewig, D. Hagemann, J. Seifert, M. Gollwitzer, E. Naumann, and D. Bartussek, A Revised Filmset for the Induction of Basic Emotions *Cognition and Emotion*, vol. 19, pp. 1095–1109, 2005.

[28] A. J. Tomarken, R. J. Davidson, and J. B. Henriques, Resting Frontal Brain Asymmetry Predicts Affective Responses to Films *Journal of Personality and Social Psychology*, vol. 59, pp. 791–801, 1990.

[29] J. J. Gross and R. W. Levenson, Emotion Elicitation Using Films *Cognition and Emotion*, vol. 9, pp. 87–108, 1995.

[30] P. Philippot, Inducing and Assessing Differentiated Emotion-Feeling States in the Laboratory *Cognition and Emotion*, vol. 7, pp. 171–193, 1993.

[31] T. Baumgartner, M. Esslen, and L. Jaencke, From Emotion Perception to Emotion Experience: Emotions Evoked by Pictures and Classical Music *International Journal of Psychophysiology*, vol. 60, pp. 34–43, 2006.

[32] E. Heene, R. De Raedt, A. Buysse, and P. Van Oost, Does Negative Mood Influence Self-Report Assessment of Individual and Relational Measures? An Experimental Analysis *Assessment*, vol. 14, pp. 86–93, 2007.

[33] P. N. Juslin and J. A. Sloboda, *Music and Emotion*. Oxford University Press, 2001.

[34] P. M. Kenealy, Mood-State-Dependant Retrieval: The Effects of Induced Mood on Memory Reconsidered *The Quaterly Journal of Experimental Psychology*, vol. 50, pp. 290–317, 1997.

[35] E. Schubert, Measuring Emotion Continously: Validity and Reliability of the Two-Dimensional Emotion-Space *Australian Journal of Psychology*, vol. 51, pp. 154–165, 1999.

[36] M. Bradley and P. Lang, Measuring Emotion: The Self-Assessment Minikin and the Semantic Differential *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25 (1), pp. 49–59, 1994.

[37] J. Zhai and A. Barretto, Stress Detection in Computer Users Based on Digital Signal Processing of Noninvasive Physiological Variables *Engineering in Medicine and Biology Society*, vol. 1, pp. 1355–1358, 2006.

[38] C. D. Katsis, G. Ganiatsas, and D. I. Fotiadis, An Integrated Telemedicine Platform for the Assessment of Affective Physiological States *Diagnostic Pathology*, vol. 1, pp. 16–25, 2006.

[39] K. H. Kim, S. W. Bang, and S. R. Kim, Emotion recognition system using short-term monitoring of physiological signals *Medical and Biological Engineering*, vol. 42, pp. 419–427, 2004.

[40] J. Kim and E. Andre, Emotion Recognition Based on Physiological Changes in Music Listening *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 2067–2083, 2008.

[41] G. E. Sakr, I. H. Elhajj, and H. A.-S. Huijer, Support Vector Machines to Define and Detect Agitation Transition *IEEE Transactions on Affective Computing*, vol. 1, pp. 98–108, 2010.

[42] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie, and T. D. Parsons, Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task *IEEE Transactions on Affective Computing*, vol. 1, pp. 109 – 118, 2010.

[43] G. Valenza, A. Lanata, and E. Scilingo, The Role of Nonlinear Dynamics in Affective Valence and Arousal Recognition *IEEE Transactions on Affective Computing*, vol. 3, pp. 237–249, 2011.

[44] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, ECG Pattern Analysis for Emotion Detection *IEEE Transactions on Affective Computing*, vol. 3, pp. 102–115, 2012.

[45] M. A. Nicolaou, H. Gunes, and M. Pantic, Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence-Arousal Space *IEEE Transactions on Affective Computing*, vol. 2, pp. 92–105, 2011.

[46] M. Etzel, E. L. Johnsen, J. Dickerson, D. Tranel, and R. Adolphs, Cardiovascular and respiratory responses during musical mood induction *International Journal of Psychophysiology*, vol. 61, pp. 57–69, 2006.

[47] E. Bernat, C. J. Patrick, S. D. Benning, and A. Tellegen, Effects of picture content and intensity on affective physiological response *Psychophysiology*, vol. 43, pp. 93–103, 2006.

[48] G. Stemmler, Physiological processes during emotion in *The Regulation of Emotion*, Erlbaum, 2004.

[49] C. A. Pauls and G. Stemmler, Repressive and defensive coping during fear and anger *Emotion*, vol. 3, pp. 284–302, 2003.

[50] M. Bradley, M. Codispoti, B. Cuthbert, and P. Lang, Emotion and motivation 1: Defensive and appetitive reactions in piture processing *Emotion*, vol. 1, pp. 276–298, 2001.

[51] T. Schuster, S. Gruss, H. Kessler, A. Scheck, H. Hoffmann, and H. C. Traue, EEG pattern classification during emotional picture processing in *7th International Conference on Cognitive Science*, 2010.

[52] P. Lang, M. Bradley, and B. Cuthbert, International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual tech. rep., University of Florida, 2005.

[53] S. Walter, H. Kessler, S. Gruss, A. Scheck, and H. C. Traue, The influence of neuroticism and psychological symptoms on the assessment of images in three-dimensional emotion space *GMS Psycho-Social-Medicine*, vol. 8, 2011.

[54] J. Smith, M. Bradley, and P. Lang, State anxiety and affective physiology: effects of sustained exposure to affective pictures *Biological Psychology*, vol. 69, pp. 247–260, 2004.

[55] S. Walter, S. Scherer, M. Schels, M. Glodek, D. Hrabal, M. Schmidt, R. Böck, K. Limbrecht, H. C. Traue, and F. Schwenker, Multimodal Emotion Classification in Naturalistic User Behavior in *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments*, Springer, 2011.

[56] J. Höysniemi, P. Hämäläinen, and L. Turkki, Wizard of Oz prototyping of computer vision based action games for children in *IDC '04: Proceedings of the 2004 Conference on Interaction Design and Children*, 2004.

[57] L. Molin, Wizard-of-Oz prototyping for co-operative interaction design of graphical user interfaces in *NordiCHI '04: The third Nordic conference on Human-computer interaction*, 2004.

[58] N. O. Bernsen, H. Dybkjaer, and L. Dybkjaer, Wizard of Oz Prototyping: How and When? in *CCI Working Papers in Cognitive Science and HCI*, 1994.

[59] D. Akers, Wizard of Oz for participatory design: inventing a gestural interface for 3D selection of neural pathway estimates in *CHI '12: Extended abstracts on human factors in computing systems*, 2006.

[60] J. A. Gray, *The neuropsychology of anxiety: An enquiry into functions of septo-hippocampal sytems.* Oxford University Press, 1982.

[61] J. A. Gray, Three fundamental emotion systems in *The nature of emotions: Fundamental questions*, Oxford University Press, 1994.

[62] D. M. Sloan, Emotion Regulation in Action: Emotional Reactivity in Experiential Avoidance *Behavior Research and Therapy*, vol. 4, pp. 1257–1270, 2004.

[63] A. L. Hof, EMG and muscle force: an introduction *Human Movement Science 3*, vol. 1-2, pp. 119–153, 1984.

[64] D. T. Lykken and P. H. Venables, Direct measurement of skin conductance: A proposal for standardization *Psychophysiology*, vol. 8, pp. 656–672, 1971.

[65] M. W. Wukitsch, M. T. Petterson, D. R. Tobler, and J. A. Pologe, Pulse oximetry: Analysis of theory, technology and practice *Journal of Clinical Monitoring*, vol. 4, pp. 290–301, 1988.

[66] G. A. Millikan, The Oximeter, an Instrument for Measuring Continuously the Oxygen Saturation of Arterial Blood in Man *Review of Scientific Instruments*, vol. 13, pp. 434–445, 1942.

[67] N. Selvaraj, A. Jarayal, J. Santhosh, K. K. Deepak, and S. Anand, Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography *Journal of Medical Engineering and Technology*, vol. 32, pp. 479–484, 2008.

[68] C. D. Katsis, N. S. Katertsidis, and D. I. Fotiadis, An integrated system based on physiological signals for the assessment of acffective states in patients with anxiety disorders *Biomedical Signal Processing and Control*, vol. 6, pp. 261–267, 2011.

[69] B. E. Boser, I. M. Guyon, and V. Vapnik, A training algorithm for optimal margin classifiers in *5th Annual Workshop on Computational Learning Theory*, 1992.

[70] V. Vapnik, *The nature of statistical learning theory.* Springer, 1999.

[71] B. Schoelkopf and A. J. Smola, *Learning with Kernels - Support Vector Machines, Regularization, Optimization and Beyond.* The MIT Press, 2002.

[72] Y. Freund, Boosting a weak learning algorithm by majority *Information and Computation*, vol. 121, pp. 256–285, 1995.

[73] T. M. Cover and P. E. Hart, Nearest Neighbor Pattern Classification *IEEE Transactions on Information Theory*, vol. 13, pp. 21–27, 1967.

[74] E. Fix and J. L. Hodges, Discriminatory analysis - nonparametric discrimination: Consistency properties tech. rep., U.S. Air Force, School of Aviation Medicine, 1951.

[75] L. Breiman, Random Forests *Machine Learning*, vol. 45, pp. 5–32, 2001.

[76] R. E. Bellman, *Adaptive control processes: A guided tour.* Princeton University Press, 1961.

[77] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2009.

[78] I. Guyon and A. Elisseeff, An Introduction to Variable and Feature Selection *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[79] P. Webos, *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting.* John Wiley & Sons, 1994.

[80] B. Widrow, *Self-Organizing Systems.* Spartan Books, 1962.

[81] R. M. Rangayyan, *Biomedical Signal Analysis - A Case-Study Approach.* John Wiley & Sons, 2002.

[82] J. C. Principe, Artificial Neural Networks in *The Electrical Engineering Handbook*, 1997.

[83] J. Cohen, A Power Primer *Psychological Bulletin*, vol. 112, pp. 155–159, 1992.

[84] S. Walter, J. Kim, D. Hrabal, S. Crawcour, H. Kessler, and H. C. Traue, Transsituational Individual-Specific Biopsychological Classification of Emotions *IEEE Transactions on Systems, Man and Cybernetics*, vol. PP, pp. 1–8, 2013.

[85] J. Dabrowski and E. V. Munson, 40 years of searching for the best computer system response time *Interacting with Computers*, vol. 23, pp. 555–564, 2011.

[86] B. Shneiderman and C. Plaisant, Quality of services in *Designing the user interface - Strategies for effective human-computer interaction*, Pearson Addison Wesley, 2005.

[87] B. Shneiderman, Response time and display rate in human performance with computers *ACM Computing Surveys*, vol. 16, pp. 265–285, 1984.

[88] G. Mandler, The interruption of behavior in *Nebraska Symposium on Motivation*, Levine, D., 1964.

[89] D. Hrabal, S. Rukavina, J. W. Tan, A. Scheck, S. Walter, and H. C. Traue, Feature-Pair-Based Emotion Change Identification from Physiological Data in Pleasure-Arousal Space *IEEE Transactions on Affective Computing, submitted*, vol. PP, pp. 1–9, 2013.

[90] D. Hrabal, S. Rukavina, K. Limbrecht, S. Gruss, S. Walter, V. Hrabal, and H. C. Traue, Emotion Identification and Modelling on the Basis of Paired Physiological Data Features for Companion Systems in *7th Vienna International Conference on Mathematical Modelling*, 2012.

[91] D. Hrabal, C. Kohrs, A. Brechmann, J. W. Tan, S. Rukavina, and H. C. Traue, Physiological Effects of Delayed System Response Time on Skin Conductance in *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction*, Springer, 2013.

[92] M. A. Pérez-Quiñones and J. L. Sibert, A collaborative model of feedback in human-computer interaction in *SIGCHI Conference on Human Factors in Computing Systems: Ccommon Ground*, 1996.

[93] W. Kuhmann, W. Boucsein, F. Schaefer, and J. Alexander, Experimental investigation of psychophysiological stress-reactions induced by different system response times in human-computer interaction *Ergonomics*, vol. 30, pp. 933–943, 1987.

[94] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, *Handbook of Psychophysiology.* Cambridge University Press, 2007.

[95] A. Haag, S. Goronzy, P. Schaich, and J. Wiliams, Emotion Recognition Using Bio-Sensors: First Steps toward an Automatic System *Affective Dialogue Systems*, vol. 3068, pp. 36–48, 2004.

[96] J. N. Bailenson, E. D. Pontikakis, I. B. Mauss, J. J. Gross, M. E. Jabon, C. A. Hutcherson, C. Nass, and O. John, Real-Time Classification of Evoked Emotions using Facial Feature Tracking and Physiological Responses *International Journal of Human-Computer Studies*, vol. 66, pp. 303–317, 2008.

[97] S. Walter, J. Kim, D. Hrabal, S. Crawcour, H. Kessler, and H. C. Traue, Transsituational Individual-Specific Biopsychological Classification of Emotions *IEEE Transactions on Systems, Man and Cybernetics*, vol. PP, pp. 1–8, 2013.

[98] M. Sherif and C. I. Hovland, *Social Judgement: Assimilation and Contrast Effects in Communication and Attitude Change*. Oxford University Press, 1961.

[99] V. Kolodyazhniy, S. D. Kreibig, J. J. Gross, W. T. Roth, and F. H. Wilhelm, An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of filminduced emotions *Psychophysiology*, vol. 48, pp. 908–922, 2011.