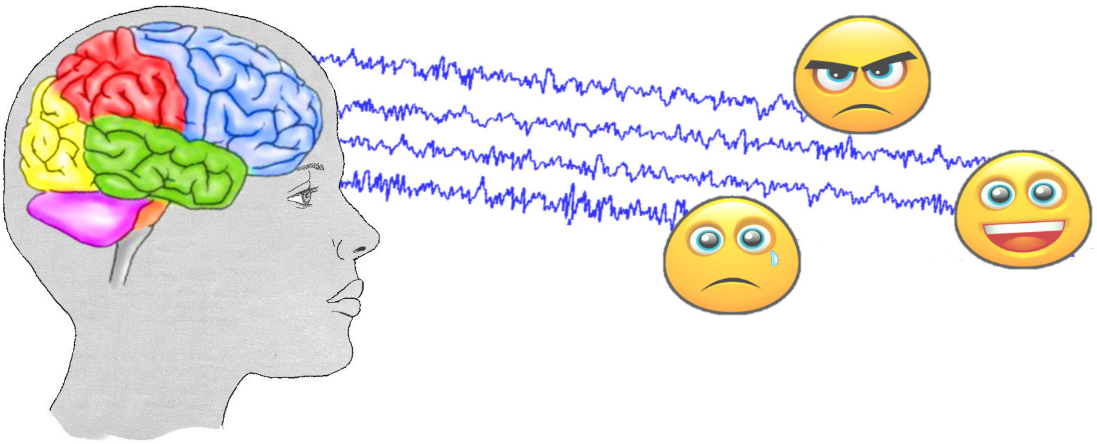
Emotion recognition using brain activity



Robert Horlings

Man-Machine Interaction Group
TU Delft

27 March 2008

Delft University of Technology

Emotion recognition using brain activity

Robert Horlings

Student No. 1099892

March 2008

Delft University of Technology
Faculty of Electrical Engineering, Mathematics, and Computer Science
Department of Mediamatics
Man-Machine Interaction Group

Graduation Committee
Dr. drs. L.J.M. Rothkrantz
Ir. P. Wiggers
Ir. H.J.A.M. Geers

Abstract

Emotion recognition by computers is becoming increasingly popular. Our project focused on recognizing emotion from human brain activity, measured by EEG signals. We have proposed a system to analyze EEG signals and classify them into 5 classes on two emotional dimensions, valence and arousal. This system was designed using prior knowledge from other research, and is meant to assess the quality of emotion recognition using EEG signals in practice.

In order to perform this assessment, we have gathered a dataset with EEG signals. This was done by measuring EEG signals from people that were emotionally stimulated by pictures. This method enabled us to teach our system the relationship between the characteristics of the brain activity and the emotion.

We found that the EEG signals contained enough information to separate five different classes on both the valence and arousal dimension. However, using a 3-fold cross validation method for training and testing, we reached classification rates of 32% for recognizing the valence dimension on from EEG signals and 37% for the arousal dimension when. Much better classification rates were achieved when using only the extreme values on both dimensions, the rates were 71% and 81%.

Contents

Acknowledgements

This thesis project has taken a lot of effort from a lot of people. I would like to give thanks to all the people who have helped me with working on this project. First of all, I would like to thank dr. Leon Rothkrantz for offering this research assignment, guiding me through the whole project and reading this thesis several times. I would also like to give thanks to professor Mirko Novak and Petr Bouchner from the *Department of Control and Telematics* group of the *faculty of transportation sciences* at the *Czech Technical University* for giving me the opportunity to gain experience with EEG measurements in Prague. Next, I would like to thank all participants who voluntarily participated in my experiment for data acquisition. I would like to thank Jabco for reading my thesis and giving useful comments along the way.

A special word of thanks for my girlfriend Elize. Thank you for listening to stories about technical details, motivating me when things did not go like they were planned and for your support during this project.

Chapter 1

Introduction

1.1 Emotion

Emotion is an important aspect in the interaction and communication between people. Even though emotions are intuitively known to everybody, it is hard to define emotion. The Greek philosopher Aristotle thought of emotion as a stimulus that evaluates experiences based on the potential for gain or pleasure. Years later, in the seventeenth century, Descartes considered emotion to mediate between stimulus and response [62]. Nowadays there is still little consensus about the definition of emotion. Kleinginna and Kleinginna gathered and analyzed 92 definitions of emotion from literature present that day [35]. They conclude that there is little consistency between different definitions and suggested the following comprehensive definition:

Figure 1.1: Different emotions

> Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can:
>
> 1. give rise to affective experiences such as feelings of arousal, pleasure/displeasure;
> 2. generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes;
> 3. activate widespread physiological adjustments to the arousing conditions; and
> 4. lead to behavior that is often, but not always, expressive, goal directed, and adaptive.

This definition shows the different sides of emotion. On the one hand emotion generates specific feelings and influences someones behavior. This part of emotion is well known and is in many cases visible to a person himself or to the outside world. On the other hand emotion also adjusts the state of the human brain, and directly or indirectly influences several processes.

In spite of the difficulty of precisely defining it, emotion is omnipresent and an important factor in human life. People's moods heavily influence their way of communicating, but also their acting and productivity. Imagine two car drivers, one being happy and the other being very mad. They will be driving totally different. Emotion also plays a crucial role in all-day communication. One can say a word like 'OK' in a happy way, but also with disappointment or sarcasm. In most communication this meaning is interpreted from the tone of the voice or from non-verbal communication. Other emotions are in general only expressed by body language, like boredom.

A large part of communication is done nowadays by computer or other electronic devices. But this interaction is a lot different from the way human beings interact. Most of the communication between human beings involves non-verbal signs, and the social aspect of this communication is important. Humans also tend to include this social aspect when communicating with computers [53].

1.2 Brain computer interfaces

The interaction between humans and computers is done in several ways. Most people use a simple keyboard and mouse to give instructions to the computer, and that works fine for most tasks. Newer forms of human computer interaction have been invented. The quality of speech recognition by computers is improving every day and becoming more and more common in consumer products. Arm movement is used when playing games on Nintendo's new game console, the Wii.

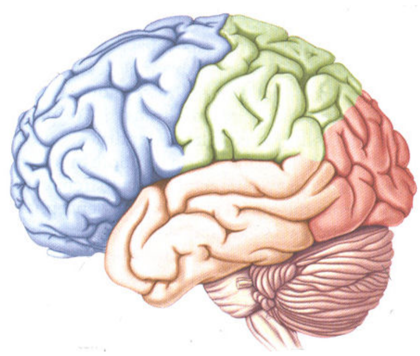


Figure 1.2: The human brain

In addition to these existing ways of interacting with computers, the use of brain activity is becoming increasingly popular. This method is called a brain computer interface (BCI). In theory, everything somebody does or says has its origin in his brain, and when using that activity for human computer interaction, many other information could be used, and the possibilities seem endless.

Brain computer interfaces have been constructed to be used as a spelling device [64, 22], for playing video games [18, 22] or even as a replacement for controlling a windows computer (combined with electromyography (EMG)) [20], but the list of possible applications that could be controlled by a BCI includes actually all programs known today.

1.2.1 TU Delft Project

In the past few years, a number of research projects have been done in the field of brain computer interfaces. Mark Wessel has started the BCI research by building a workbench for performing and analyzing EEG recordings of brain activity [67]. Zulfikar Dharmawan continued this research and investigated people's stress level while playing computer games [15]. Our project continues the work in the field of brain computer interfaces on TU Delft.

1.3 Emotion recognition using brain activity

Using a BCI to measure emotion could improve the quality and effectiveness of human-computer interaction. This communication can be done in a much more natural way [48], and open up new ways for computers to understand human behavior or meaning.

A lot of research is already done after recognizing emotions. For example, research has been done to make computers recognize emotion from speech [12, 11, 2], facial expressions [47, 11, 19] or a fusion of both methods [6].

Another method of measuring human emotion, is using physiological signals [24, 34]. These physiological signals are available at any moment, and it has been shown that emotional markers are present in electroencephalograms (EEG), and can be deduced from heartbeat rates and skin conductivity. Using these physiological signals also has the advantage that they can hardly be deceived by voluntary control and are available all the time, without needing any further action of the user. The disadvantage of using these signals is that the user has to wear some measurement equipment. This equipment could be very simple when measuring only heartbeat, but EEG measuring devices tend to be more demanding.

This method using EEG signals can however still use a lot of improvement. Earlier research has shown that these types of human computer interfaces do not perform very well. They are able to recognize emotion correctly from around 60-70% of the samples given when recognizing three emotions [10, 34]. Chopin [10] suggests there is a lot of improvement possible when looking at different features or using other methods or techniques. A fusion of different methods could be another solution [24].

1.4 Societal relevance and possible applications

Researching emotion recognition is done by many people. As explained before, many effort is put into emotion recognition from other modalities, such as speech, text or facial expressions. Emotion recognition from EEG signals is expected to outperform these modalities, since brain activity has direct information about emotion, where other modalities are a indirect reflection of the emotion. Moreover, EEG signals can be measured at any moment, and are not dependent on someone speaking or generating a facial expression.

Possible applications of this technique are many. First of all, the simple task of recognizing someones emotion automatically could assist therapists and psychologists in doing their job. Other applications can be found in the field of human machine interaction, human communication through computers and in assisting disabled people with communicating.

In the field of human machine interaction, the computer could adjust its behavior depending on the users emotion. When a game becomes too boring or too exciting, the game level can be changed. Another example might be that the computer changes the music or the windows background depending on the mood of the user sitting behind it. These types of human-computer interaction are already used in the commercial world [21, 18, 9].

Figure 1.3: Example of a commercial BCI system. Image from `www.gtec.at`

Communication between people through instant messaging is very common nowadays. This communication is unfortunately only capable of communicating through words, although speech and video are a great addition. The enormous use of smileys in many conversations shows the desire to share and express emotion while talking to others. A brain computer interface could be a useful tool to automatically add emotional information during the conversation.

Another application is emotion expression for people suffering from severe muscle diseases like amyotrophic lateral sclerosis (ALS). These patients may not be able to move their muscles or to speak. Some progress is made in creating a brain-computer interface for those patients to enable them to communicate again. A way to express emotion would be a great improvement to such a system. See for more information [10].

1.5 Research questions

The objective of this project is to *create a system for emotion recognition using brain activity and assess the quality of this emotion recognition in practice.* This goal can be broken down into several research questions:

1. *Is emotion recognition from EEG signals possible in practice?*
   Literature suggests that emotion recognition should be possible from EEG signals [36, 59]. However, the results in practice show very different results [10, 7]. We will design, build and test a system for the offline recognition of emotions from EEG signals based on literature from earlier research. After that we will create a dataset with EEG signals, which can be used to test the quality of emotion recognition. Using this dataset we will also investigate the quality of different parameters for our program.

   (a) *How to perform EEG measurements?*
   We will have to find out how to record EEG data. Using that knowledge, we

will setup an experiment to make EEG recordings with known emotion. These recordings will help in training and testing our program.

(b) *How to process* EEG *data?*
Measuring EEG data results in huge amounts of data, containing brain activity, but also a lot of noise and artifacts. We should find out how to handle the EEG data in our program in such a way, that we remove the noise and artifacts and keep the brain activity.

(c) *How to recognize emotion from the* EEG *data?*
Besides the removal of unnecessary part of the signal, we should decrease the size of the dataset by extracting only the useful information. From this information, the program should be able to detect the emotion of the user at that moment. We should find out how to do that.

1.6 Scope and limitation

The program we will create is specifically meant for research purposes, to see whether emotion recognition from brain activity is feasible at this moment. Researchers can use this program to test several methods or techniques easily. For that reason, there will focus on the techniques and the extendability, not so much on the user interface. However, while designing the software we keep in mind that the program should be easily extendable to include other methods or algorithms. While this program works on offline data, the ultimate goal is to build an online version of an emotion recognition system or even a hybrid system where vocal, facial and physiological signals are combined.

1.7 Methods

In order to accomplish the goal mentioned in the previous section, we have to take several steps:

1. *Literature research*
   The first step will be to gather knowledge about emotion, physiological measures such as EEG and brain computer interfaces (BCI). This information includes theoretical information about emotion and the way emotion works in the human brain, as well as information about the measurement of brain activity. Other literature will be consulted about signal processing and data mining, in order to use the latest techniques.

2. *Designing and implementing system to analyse emotional* EEG *signals*
   The next step of the process involves designing the software system that will be able to recognize emotion. The design will be based on earlier attempts to build an emotion recognition system. In this step, the structure of and the techniques we use in our system will be chosen and explained, as well as the way to represent emotion.

Based on the design of the system as chosen in the previous step, the system will be implemented for research use. The system will include a very simple user interface. After implementation, the different steps of the system will be tested.

3. *Data acquisition*
   We will have to teach the system how to recognize emotion from the brain activity. For this purpose, we need some measurements of brain activity, as well as the emotion someone has at that moment. Part of this data is gathered from the Enterface '06 project [55]. In order to increase the size of the dataset, we will collect some EEG data for which we know the emotion. This will be done by performing our own data measurements at the TU Delft.

4. *Training and testing*
   The gathered data will be used to teach our system the details about recognizing emotion from the EEG signals. We will use a part of the data for this training, and another part of the data for testing whether the learned relationship between the EEG signals and emotion is correct. We will change various parameters and choices we made, in order to find out the best working program.

1.8 Outline

Chapter 2 will provide some information on the theory behind emotion and ways to measure brain activity. In chapter 3 we will give an overview of the methods used when measuring EEG signals and recognizing emotion. Chapter 4 will explain the design of our system, where chapter 5 explains the details about the implementation of the emotion recognizer. The next chapters, chapter 6 and 7, present the ways our dataset of EEG signals is constructed, by scientists from Enterface '06 and by ourselves. In chapter 8 you can find information about the explorative analysis of the measured data. Chapter 9 presents the results about the performance of our system, using different parameters. Finally, the conclusions about recognizing emotion can be found in chapter 10.

CHAPTER 2

Literature

For recognizing emotion from brain activity, we need information about how the brain handles emotion, and information on how to measure the brain activity. Several methods on measuring brain activity are described in section 2.1. Section 2.2 will give information about literature on emotion, how emotion is handled by the human brain and emotional models. Section 2.3 will also provide an overview on earlier literature on emotion recognition using brain activity.

2.1 Measuring brain activity

The human brain is the part of the body that regulates almost all of the human activity. Therefore the activity of the brain contains a lot of information about someone. Several ways of measuring this activity exist, such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI) and electroencephalography (EEG).

2.1.1 Positron emission tomography

With positron emission tomography (PET), a radioactive isotope is injected into someones blood. Because this isotope emits positrons, and is taken with the blood flow, a machine can measure the blood flow. Since it is believed that the blood flow in the brain is highly correlated with brain activity, the machine can show brain activity. PET is able to measure activity with a high spatial resolution, but it has a low time resolution and a time delay due to the time it takes before the radioactive material has arrived in the brain. Another disadvantage of this method is the radiation a person is subjected to.

Figure 2.1: Example of a PET scanning system. Image from `www.radiologyinfo.org`

2.1.2 Functional magnetic resonance imaging

Figure 2.2: Example of a fMRI scanning system. Image from www.ftcm.org.uk

Functional magnetic resonance imaging (fMRI) is another method that depends on the blood flow. Active neurons are known to consume oxygen that is carried by hemoglobin. This consumption of oxygen changes the magnetic properties of hemoglobin, and these magnetic properties are measured by the fMRI system. So, indirectly the brain activity is measured. Just as PET, fMRI has a high spatial resolution, but a low temporal resolution. Moreover, the equipment needed is very expensive.

2.1.3 Electroencephalography

Electroencephalography (EEG) uses the electrical activity of the neurons inside the brain. When the neurons are active, they produce an electrical potential. The combination of this electrical potential of groups of neurons can be measured outside the skull, which is done by EEG. Because there is some tissue and even the skull itself between the neurons and the electrodes, it is not possible to measure the exact location of the activity.

Figure 2.3: Example of a cap for measuring EEG data

For EEG measurements an array of electrodes is placed on the scalp. Many caps available use 19 electrodes, although the number of caps using more electrodes is rising. The electrodes are placed according to the international 10-20 system, as is depicted in figure 2.4. This system makes results from different research easily comparable.

The brain activity can be coarsely divided into different classes. This division follows the way the activity is produced by the brain.

Rhythmic activity

The neurons of the brain produces together a rhythmic signal that is constantly present. This signal can be divided into several bands, based on the frequency.

Delta band The delta band is the frequency band up to 3Hz. Delta activity is mainly seen in deep sleep.

Theta band The theta band consists of frequencies between 4Hz and 7Hz. This activity can be observed with drowsiness or meditation

Alpha band The alpha band is the so-called 'basic rhythm' and contains the frequencies between 8Hz and 12Hz. It is seen when people are awake, and is known to be more apparent when eyes are opened.

Beta band The beta band contains frequencies between 13Hz and 30Hz. This band is apparent with active thinking or concentration.

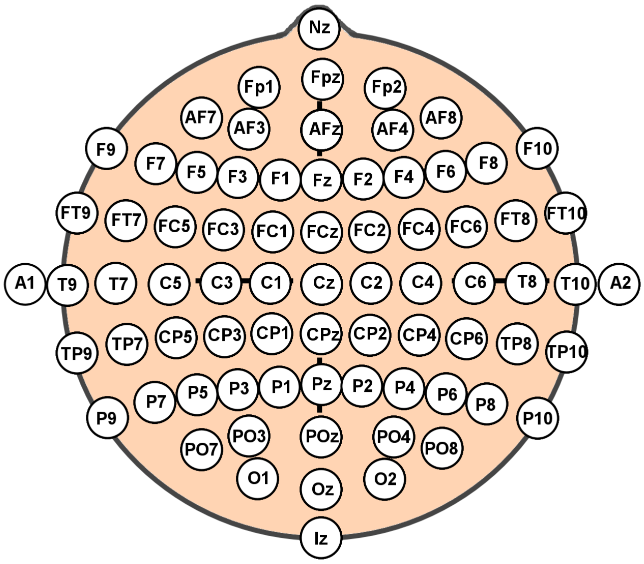


Figure 2.4: The 10-20 electrode placement system. Picture taken from `www.wikipedia.org`

Event-related potentials

Event-related potentials (ERP) are potentials that occur after some event, most apparently after unpredictable events. This potential occurs in the stream of ongoing brain activity, and most of the time it is difficult to extract some specific activity. If it is known when such an event occurs, it is possible to time-lock the EEG signal, and extract the ERP from the signal. One of the best known ERPs is the P300, which occurs around 300 milliseconds after the event.

Event-related (de-)synchronization

Besides the ERPs, the rhythmic activity of the brain can also be altered by external or internal events. These events can cause an increase or decrease in the power of one of the frequency bands. This change is said to be caused by synchronization or desynchronization of the activity of different neurons, resulting in band power changes.

2.1.4 Conclusions

fMRI and PET have a good spacial resolution, but are very slow, as they depend on changes in composition of blood. EEG on the other hand gives immediate responses, but is not able to reflect the exact location of the brain activity. Because of this high time-resolution, the simple procedure and its low cost, EEG is used in our project to measure brain activity.

2.2 Emotion

As explained in the introduction, emotion is a concept that is hard to define. Scientists agree on the fact that the human brain is the main source of emotion. However, the precise role of the brain is not clear. Many theories exist about the exact involvement of the brain.

2.2.1 The emotional brain

For years, scientists have reasoned about and researched emotion. One of the first theories proposed was the James-Lange theory, in the end of the 19th century. This theory states that emotion is the experience of some changes in the body. For example: when one sees a bear, he will run away, and the running away will provide him the feeling of fear.

In the 1920s Cannon and Bard challenged this theory, among others because bodily changes are too slow to be responsible for emotions and artificial body changes are insufficient to generate emotions. They state that the hypothalamus is the brain region that translates stimuli into emotions.

Some years later, Papez augmented this idea and made a nice scheme of how emotion would be processed in the brain. It consists of two streams, the stream of feeling and the stream of thinking. He claimed that emotional experiences are a function of activity of both streams, that is computed in the cingulate cortex. Many of this theory, especially the pathways he proposed, are proven to exist, although not all the regions he mentioned are important in emotion processing.

In 1949, Maclean introduced another, more accurate model, and invented the term limbic system. He claimed (as a neo-Jamesian) that perception of the world, combined with information on body changes generates emotional experience, and this integration of knowledge occurs in the limbic system.

This theory of the limbic system has been the dominant theory for a long time, although it is now thought that some structures of the limbic system play a less important role, and some other structures play a more important role in the 'emotional brain'. Some of these structures are:

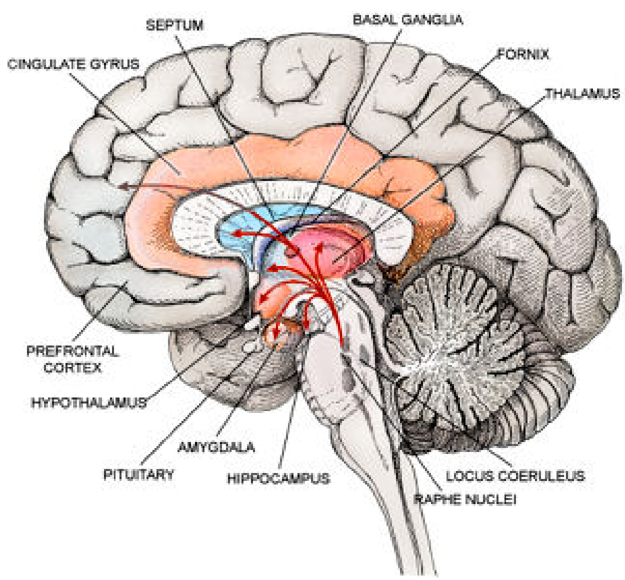


Figure 2.5: The human brain. The colored part is considered part of the limbic system

Amygdala The amygdalae are two groups of neurons deep inside the human brain. Together, they are one of the most important brain regions for emotion. Together they are called the amygdala. One of its functions is to interpret the emotional value of incoming signals. When a person receives some sort of signal with some emotional load, it is recognized by the amygdala. Another function of this organ is fear conditioning: it helps in learning a connection between some stimulus and a threatening event. For example, when a rat hears a sound and gets an electric shock, he will connect these two things after a number of combinations. Afterwards, he will become afraid of the sound. A third important function of the amygdala is the consolidation of long-term emotional memories.

Prefrontal cortex Just as the amygdala, the prefrontal cortex (PFC) plays a role in reward processing. Neurons in the PFC can detect changes in the reward value of learned stimuli [42]. Furthermore, the PFC is involved in planning, making decisions based on earlier experiences and working towards a goal. The combination of functions of the PFC is described as the 'executive function'.

Anterior cingulate cortex This part of the brain is generally subdivided into a 'cognitive' and a 'affective' part. The affective part is suggested to monitor disagreement between the functional state of the organism and any new information, that might have affective consequences [5].

Hypothalamus The hypothalamus is the part of the brain that controls many processes in the body, such as body temperature, hunger and thirst. It also handles the release of some hormones. As such, the hypothalamus is involved in processing emotions and sexual arousal.

Insular cortex The insular cortex is said to be associated with emotional experience and produces conscious feelings. It combines sensory stimuli to create a emotional context.

2.2.2 Models of emotion

Emotion is a phenomenon that is difficult to grasp, as described in section 2.2. In order to classify and represent emotions, some models have been proposed. There are two dominant models.

Basic emotions

One of the models uses the idea that all emotions can be composed of some basic emotions, just as colors can be composed of primary colors. Plutchik [50] relates eight basic emotions to evolutionary valuable properties, and based thereupon he claims these to be the basic emotions: anger, fear, sadness, disgust, surprise, curiosity, acceptance, and joy. Ekman [17] has chosen other emotions to be basic, and has found that these emotions and expressions of these emotions are universal. Ekman's list of basic emotions is: anger, fear, sadness, happiness, disgust and surprise.

Dimensional view

Another model is composed of multiple dimensions, and places every emotion on a multidimensional scale. The first dimension is emotional valence, with positive on the one and negative on the other side. The second axis contains the arousal, from calm to excited. The third dimension is not always used, and when used, it is not always the same. Sometimes dominance is used as the third dimension. Other times it is motor activation (approach or avoid) that is on the third axis.

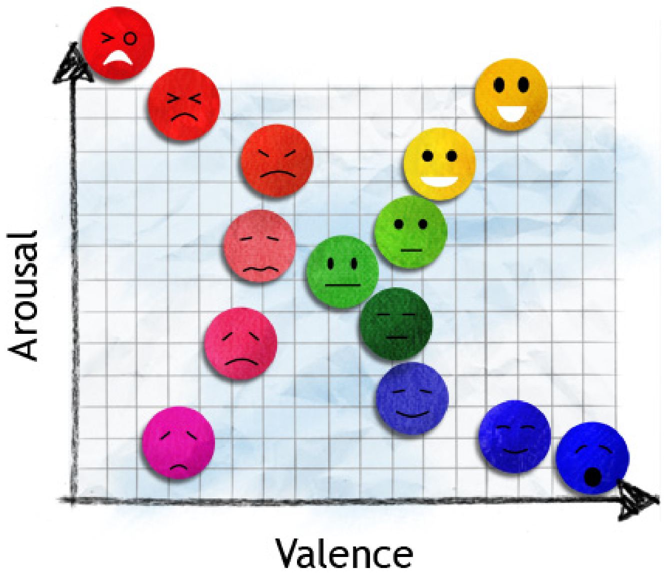


Figure 2.6: Different emotions on the valence and arousal scale. Image from `www.exmocare.com`

This model is used in most studies, because of two reasons

- Simplicity: it is easy to express an emotion in terms of arousal and valence, whereas it is much more difficult to decompose an emotion into basic emotions.
- Universality: there is little controversy about the first two dimensions of the model. Valence and arousal are natural terms to use, when speaking about emotion, and they are understood between all cultures.

Model used in our project

Most literature on emotion recognition uses the dimensional model of emotions because of its simplicity and universality. This makes comparison with and using methods from other literature easy. Basic emotions are more difficult to use, and there is agreement only about a small number of basic emotions. For that reason we will use the dimensional model in our project.

2.2.3 Measuring emotion using EEG

Although we know something about the way the brain handles emotion, we also want to be able to measure this activity. Unfortunately, EEG uses electrodes outside the skull. It is a known problem for EEG measurements that the received electrical signals are not only signals from one spot, but the skull spreads out the brain activity from the whole brain. Because most of the mentioned structures lie deep inside the brain, it is hard to get some information from it using EEG.

Fortunately, there have been a lot of people doing research after the measurements of emotion using EEG. One of the most apparent results is the important role of the alpha frequency band. Kostyunina et al. [36] have found that different emotional states show different peak frequencies in the alpha band. More recent work from Shemyakina et al. [59] shows that significant differences in local EEG power and spatial synchronization (EEG power at different spots to behave in the same way) can be observed with different emotions. This effect was largest in the temporal area of the brain.

Other researchers have concentrated on one of the two emotional dimensions, and determined how the value in both dimensions could be found from the EEG data.

Valence

One of the most apparent things about emotion in EEG measurements is the difference between the left and the right hemisphere. It has been suggested that information about the valence dimension can be found in this difference, which is called the hemispherical emotional valence (HEV). Basically, there are two theories about HEV:

- Right hemisphere hypothesis: this hypothesis states that the right hemisphere is mostly involved in processing emotion. This lateralization should be most apparent with negative emotions.

- Valence asymmetry hypothesis: this hypothesis poses that the involvement of both hemispheres depends on the valence of the emotion. The right hemisphere is dominant with negative emotions whereas the left hemisphere is more active with positive emotions.

Both hypotheses have been supported by a lot of research. However, some recent papers have compared and analyzed most of this research, and have not found empirical evidence for any of the theories. Murphy et al. have performed a meta-analysis on 106 studies on human emotion [43]. They found partial support for the valence asymmetry hypothesis, because greater activity was observed in the left hemisphere for positive emotions, where the activity was symmetrical for negative emotions. Schiffer et al. have collected a lot of literature on this subject, and have found no strong empirical support for any of the theories [56]. They hypothesize that the direction of the hemispherical emotional valence is a personal matter, just like gender or hand-writing. They have found support for this hypothesis by their own research.

Arousal

Where there is a lot of information about EEG properties of the valence dimension, less is known about how to measure arousal using EEG signals. Several studies have been done, but the results are less obvious than for the valence dimension. Heller et al. [27] gives an overview of many of these studies. The main conclusion from earlier research shows that the right posterior hemisphere has something to do with arousal. This has also been confirmed by their own research. Aftanas et al. [1] state that not only the right posterior hemisphere is involved with arousal, but also the left anterior hemisphere. They show that emotional states with a high or low arousal value, as compared with neutral signals, show asymmetrical increases in activity in the posterior areas of the right and the anterior areas of the left hemispheres. Krause et al. [37] take another approach, and studied the differences between EEG bands, instead of the brain areas. They showed that the 4-6Hz EEG band elicited a greater synchronization when viewing an aggressive film than when viewing a neutral or sad film.

2.3 Other literature on emotion recognition using physiological signals

There have been other researchers that have used physiological signals to detect emotion. For example, Choppin has built an emotion recognizer for patients suffering from amyotrophic lateral sclerosis (ALS) [10]. These patients lose, at some moment during their disease, their ability to use their muscles, and those patients lose their ability to speak or to communicate on any other way. For that reason Choppin uses only EEG signals to recognize emotion, in order to provide those people with the possibility to express their feelings. He uses neural networks to classify the EEG signals online, and achieved a correct classification rate for new unseen samples of about 64%, when using three emotion classes. Mainly because of the low number of training data, but also because the project was very preliminary, these results could be improved.

Musha et al. [44] also use EEG signals to read a persons emotion. They have extracted cross-correlation coefficients between the EEG activity from different locations, to find information about the correlation between the EEG signals from different locations. They have computed an 'emotion matrix' to transform these coefficients linearly into a four-element vector that corresponds with 4 basic emotions. The numbers in the vector denote how strong that particular emotion is found in the EEG signals. The authors do not mention how well their system performs.

Another experiment of Chanel et al. [7] tried to recognize only the arousal dimension of emotion from EEG and other physiological measures. The activity in specific EEG bands on specific locations on the scalp were used as features to describe the EEG data. The authors used different classifiers to classify the EEG signals, the physiological signals and a combination of both. Results showed that the EEG signals performed equally well as the physiological signals, but a combination of both improved the results. Classification rates were around 60% when using two classes and 50% when using three classes.

Kim et al. [34] have used other physiological measures, such as electrocardiogram (ECG) and skin temperature to build a user-independent system for recognizing emotion. The support vector machine they used to implement the system was able to correctly classify 78% and 62% of the samples, in case of three and four different emotions respectively.

Haag et al. [24] explore the use of bio-sensors, to find out whether they are suitable for complementing or replacing other modalities in measuring emotion. In particular, the authors use electromyography (EMG), ECG, skin conductivity, blood volume pulse and respiration rate. Two separate neural networks are trained to predict the arousal and valence values respectively, for one subject. Results that were within 20% of the correct answer, were counted as correct. This resulted in 96.6% correct classification for arousal and 89.93% for valence.

2.4 Summary

Electroencephalography (EEG) seems to be the most practical way of measuring brain activity, because it is cheap and easy to use. Several researchers have shown that it is possible to measure emotional cues using EEG measurements, which is an important condition to be able to find emotion from EEG activity. In literature, a distinction is made between two dimensions of emotion, the valence dimension ranging from negative to positive and the arousal dimension, ranging from calm to excited. Information about both dimensions has been found to be present in EEG signals, which shows that emotion recognition from EEG signals should be possible. We will also use this distinction in our project.

CHAPTER 3

Methods

Measured EEG signals are a valuable source of information about brain activity. However, since brain activity only produces very weak signals, the EEG signals contain a lot of background noise. Before using the signals for emotion recognition, they have to be preprocessed, in order to remove unwanted noise. Although EEG signals contain a lot of information, they also tend to result in very large amounts of data. As mentioned in chapter 2, the information in EEG signals includes information about emotions. Our program has to extract the valuable information from the large amount of data. For this task, we will first reduce the amount of data available. This process is known as feature extraction and extracts specified measures that is useful for our task from the signals. These features should contain enough information about the emotion. After having reduced the size of the data, the emotion has to be recognized from the features. For this purpose several classification methods exist to classify a given EEG signal into a number of emotional classes.

This chapter will provide information about methods to preprocess the measured EEG signals (section 3.1), present different EEG features (section 3.2), a way of selecting the best features (section 3.3) and introduce several methods for classification (section 3.4). Finally the methods for eliciting emotion (section 3.5) and recording EEG data (section 3.6) will be explained.

3.1 Preprocessing

Preprocessing is a step to process raw EEG signals in such a way that they are ready to be used. As depicted in figure 3.1, the measured EEG signal is a combination of brain activity, reference activity and noise. The goal of this preprocessing step is to reconstruct the original brain activity. All steps can be executed using different methods and algorithms. These steps will be explained in this section.

3.1.1 Referencing

EEG signals are a recording of the voltage at different electrodes. Ideally, this measurement should only represent the electrical activity on that spot. However, since voltage is a relative measure, the measurement is compared to the voltage at a reference site. Unfortunately, this results in measurements that reflect the local activity, but also the activity at the reference site (see figure 3.1). Because of this, the reference should be

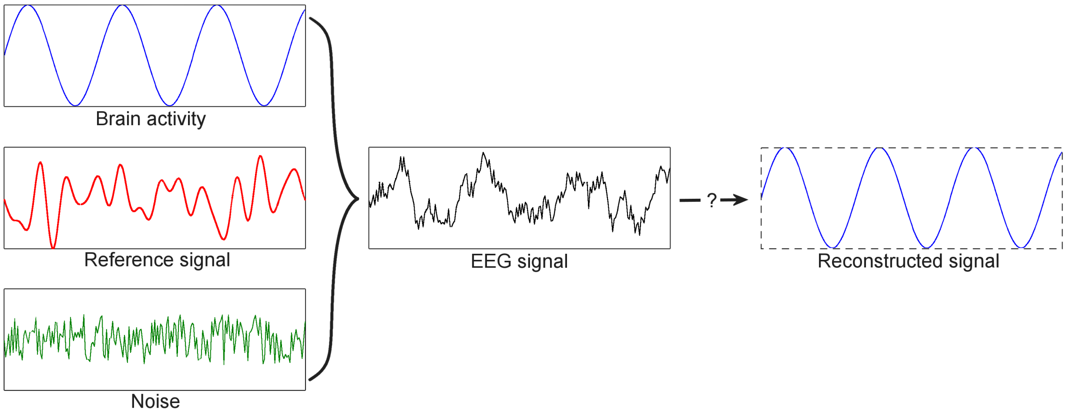


Figure 3.1: EEG signals as a combination of brain activity, reference activity and noise

chosen such that the activity at the reference site is almost zero. The nose, mastoids and earlobes are typically used [16].

Hagemann et al. have stated that the differences between results of different studies (as described in section 2.2.3) are partly due to the differences in referencing [25]. This underlines the importance of the subject.

Common reference

A widely used method of referencing is the common reference technique. This method uses one common reference for all electrodes, preferably a reference at a large distance from all electrodes. The activity at the reference site influences all measurements equally, and differences between electrode measurements still contain all information needed.

Average reference

Another method is the average reference. The average reference subtracts the average of the activity at all electrodes from the measurements. This method is based on the principle that the activity at the whole head at every moment sums up to zero. Therefore, the average of all activity represents an estimate of the activity at the reference site. Subtracting this average produces in principle a dereferenced solution. However, the relatively low density of the electrodes and the fact that the lower part of the head is not taken into account, bring some practical problems along [16].

Current source density

A technique that is frequently recommended for asymmetry research it the current source density (CSD) approach. The current source density is "the rate of change of current flowing into and through the scalp" [65]. This quantity can be derived from EEG data, and it may be interpreted as the potential difference between an electrode and a weighted average of their surrounding electrodes. The CSD can be estimated by computing the laplacian. This comes down to computing the sum of the differences between an electrode and its neighbors. A problem with this estimation is that it is actually only valid when the electrodes are in a 2d plane and equally distant. Despite of this problem, the CSD is used in many studies.

Conclusions

Even though there is a lot of literature available about referencing, it is still not very clear which method to use. All techniques have some problems in practice, and thus one can only choose the least bad one. Preliminary research from Hagemann et al. [25] points to the CSD method as the most promising method, when researching hemispherical asymmetry. We will use this method in our project.

3.1.2 Noise and artifacts

The brain produces electrical activity in the order of microvolts. Because these signals are very weak, it usually contains a lot of noise. Sources of noise are static electricity or electromagnetic fields produced by surrounding devices. In addition to this external noise, the EEG signal is most of the times heavily influenced by artifacts that originate from body movement, as depicted in figure 3.2. For example, eye blinks or other eye movements produce large spikes in the EEG signal. Other muscle movements also leave their mark in the brain signal. In many studies the participants are asked not to move or to blink as few times as possible. However, in many practical situations this is not feasible.

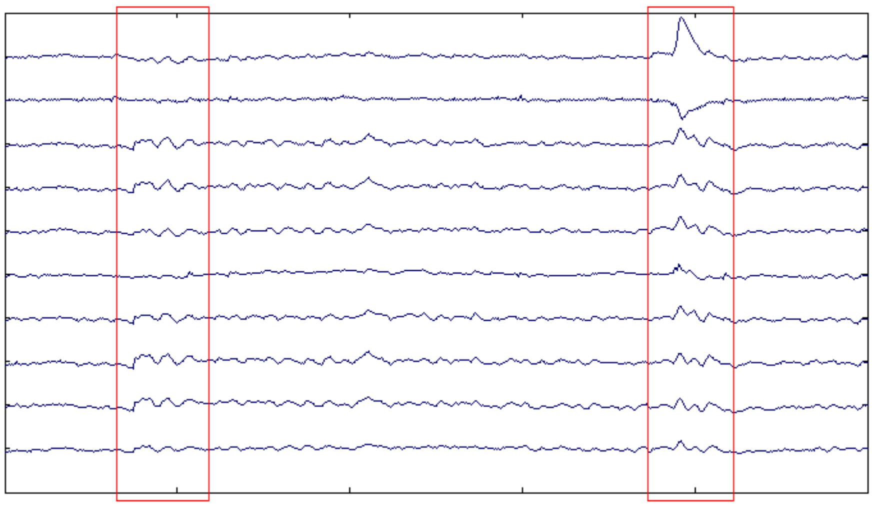


Figure 3.2: EEG signal contaminated with artifacts. The red boxes show the artifacts

Filtering

Many noise that is present in the EEG signals can be removed using simple filters. The relevant information in EEG, at least for emotion recognition, is found in the frequencies below 30Hz. Therefore, all noise with higher frequencies can be removed using a low pass filter. For example, noise from the electrical net has a fixed frequency of 50Hz. Bandpass filters can also divide the EEG signals into frequency bands (see section 2.1.3), which can be analyzed separately. Low pass and bandpass filters are implemented by EEGLab using FIR filters.

Artifacts

Artifacts are somewhat more difficult to remove, because they are not present all the time, and not always in all electrodes. Another disadvantage of artifacts over other noise is the relative high voltage as compared to the normal EEG signals. However, since artifacts contaminate the signal very much, they are a huge problem when using EEG signals. Fortunately, for the same reason many solutions for the problem have been suggested, ranging from very simple to mathematically complex.

Filtering A very simple method of removing artifacts is simply using a high pass filter to remove the frequencies below 1 or 2Hz. This method assumes there is not much brain activity with very low frequencies, and that artifacts occur at a lower frequency. A great advantage of this method is that its implementation is very simple and it can be combined with the low pass filtering to remove noise with high frequencies. Unfortunately, the method is not very precise, and might remove some useful information.

Artifact rejection One method of removing artifacts is the rejection of the data sample. Although this is a valid method of removing artifacts, the problem is that there are often not much samples available. If some of the samples are removed, there are even less samples left, which might decrease the quality of the system. Another problem of rejecting data in online systems, is the possibility of losing some important part of the data, which might not return again.

Artifact subtraction Another method involves the subtraction of electrooculographic (EOG) signals from the EEG data [31, 32]. EOG measures the eye movement. For each EEG channel, a scale factor is estimated for how much it is involved by eye movements. The EOG signal is scaled with this factor, and subtracted from the EEG signal. The main disadvantage of this method is the fact that the EOG signal contains traces of brain activity, especially from the frontal lobe. This activity is also subtracted from the EEG signals, and some useful information is removed.

Adaptive filtering The use of adaptive filtering in another method for the online removal of EOG artifacts. This algorithm estimates the clean EEG signals by subtracting the filtered EOG signals from the measured EEG signal. The filter is adjusted to the signal automatically, and therefore does not need any calibration or learning steps. He et al. [26] have investigated the use of adaptive filtering for EOG artifact removal. They report good artifact removal and a good and fast convergence of the algorithm, which makes it suitable for online filtering.

Blind source separation Another more difficult method is using Blind Source Separation (BSS). BSS decomposes signals from multiple sources, in the case of EEG the electrodes, into another set of signals, such that the signals' independency is maximized. This method has the advantage that is does not need separate EOG measurements. BSS relies on the assumption that these independent sources indeed exist. For EEG signals, this assumption holds. As the EEG signals on a electrode are said to be a mix of underlying signals of brain processes, the signals can be decomposed [41, 70]. An example of BSS is shown in figure 3.3, where all different samples can be coarsely separated into two groups, corresponding to the blue and the red line.

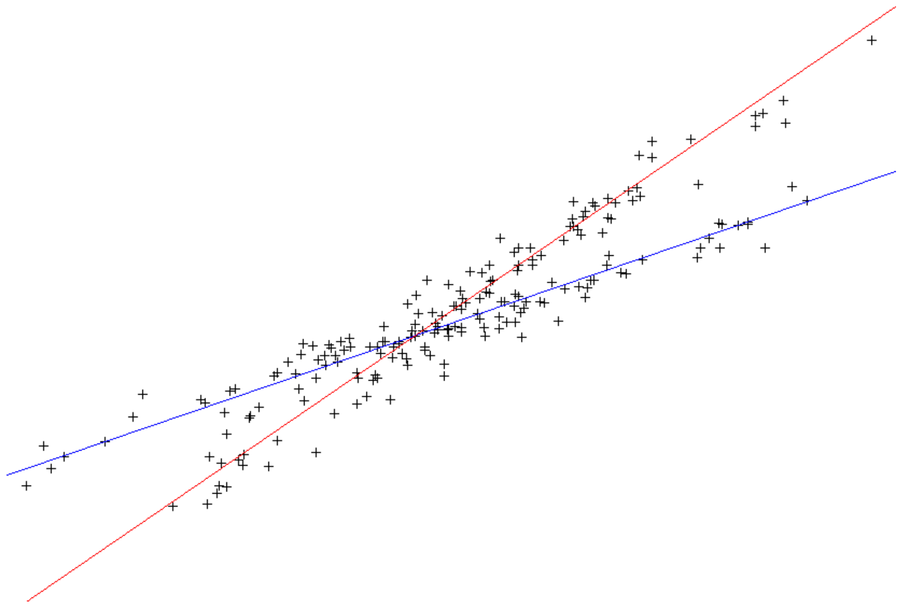


Figure 3.3: Example of BSS. Two components are found within the data

A classic example of BSS, that might clarify this method, is the cocktail party problem. A number of people are talking in a room, and this is recorded with some microphones. These microphones record all voices together, and the goal of BSS is to identify the underlying sources of the sound, in this case the different voices.

Mathematically, BSS comes down to the following formula

$$\begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \\ w_{n1} & w_{n2} & & w_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

where $x_{1..n}$ are the measured signals and $s_{1..n}$ are the independent sources of the signals.

The matrix W determines how to unmix the signals x to get the original signals. The inverse matrix W^{-1} is the referred to as the mixing matrix and represents the signal propagation from the sources to the different signals. The trick of BSS is that it is able to find W and s from x, without any other knowledge, but since only x is known, BSS algorithms have to make assumptions about the signals. Different algorithms make different assumptions

Independent Component Analysis A well-known and widely used method of BSS is Independent Component Analysis (ICA). ICA assumes the that the component are statistically independent at each time point, but does not consider the relationships across time. This method is very powerful and some fast implementations exist. However, when separating EEG signals, the user has to manually distinguish the real brain signals from the artifacts.

Second Order Blind Inference To use a particular method for removing ocular artifacts in an online system, it has to be able to automatically remove the artifacts, without manual intervention. Joyce et al. [33] have tried to devise such a method, using the Second Order Blind Inference (SOBI) method. This method makes other assumptions on the components. Where the ICA algorithm does not take the relationship across time into account, SOBI insists that the relationship between components at different time lags are as uncorrelated as possible. The remaining correlation can isolate highly correlated components. This property of SOBI makes it especially useful for automatic ocular artifact detection, because signals from both eyes are highly correlated. The correlation is used to automatically determine which components contain ocular signals, and which ones contain brain activity.

Automatic removal without EOG measurements Unfortunately we do not have the equipment to measure EOG signals separately. This is needed for most before mentioned methods. BSS is capable of removing artifacts without the separate EOG measurement. However, that method requires manual intervention, and the automatic method described earlier in this paragraph does need the EOG measurement again. One attempt was made using Kalman filters [54, 57]. In [54], an extended Kalman filter is used to detect dynamic changes in the signal. Using that information they were capable of detecting artifacts. The main disadvantage of this method is that is seems to work well on muscle artifacts, but not so well on ocular artifacts. Furthermore, it is a computationally complex method. Another solution proposed by Delorme et al. [14] is using higher-order statistics on ICA components, in order to select some pieces of the EEG signal for rejection. They compute the entropy and the kurtosis for the ICA components. However, the authors still check the selected pieces manually, in order to avoid mistakes. Nicolaou et al. [45] went in the same direction. They used Temporal Decorrelation Source Separation (TDSEP, [71]), an temporal extension to the standard ICA algorithm, to separate the EEG from artifacts. Using a support vector machine (SVM, see for more information section 4.3.4), the different components could be divided in artifact-related components and components containing useful information.

Table 3.1: Use of different features in earlier research for emotion recognition

	Musha 1997 [44]	Vourkas 2000 [63]	Choppin 2000 [10]	Chanel 2005 [7]	Savran 2006 [55]
EEG frequency band power			x	x	x
Cross-correlation EEG band power	x		x		x
Peak frequency in alpha band			x		
Hjorth parameters		x			

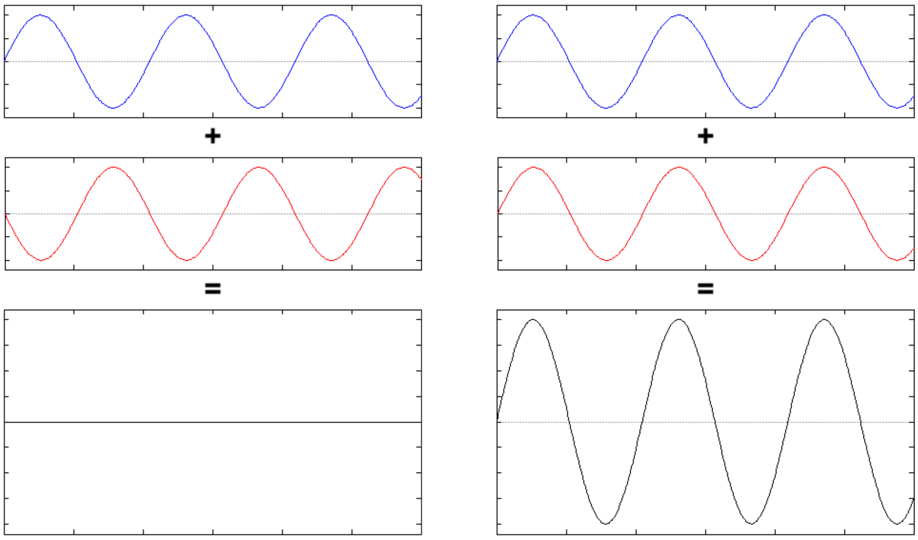
3.2 Feature extraction

Feature extraction is the process of extracting useful information from the signal. Features are characteristics of a signal that are able to distinguish between different emotions. For example, when trying to assess the health of a person, you can check his weight, heartbeat rate, blood pressure etc. Those measures contain a lot of information about the health of a person. However, the length of that same person (most of the times) does not have anything to do with his health. All features together should contain all information that is needed for the specified task. Since there are many features that could be extracted, the difficulty is which features to extract from the signal.

The features we will use are based on earlier research. Several studies on recognizing emotion have already been performed. The studies all use different features and claim good results using those features. An overview of the features that were used in those studies can be found in table 3.1 and are explained in the next sections.

3.2.1 EEG frequency band power

As explained in section 2.1.3, the brain produces an always present rhythm. The activity in different frequency bands in this rhythm reveals information about the brain activity. The activity within a frequency band is expressed by the frequency band power. Instead of using the absolute band power, most of the times a measure of event related (de)synchronization (ERD/ERS) is used. The term synchronization refers to groups of single neurons in the brain. When these neurons are desynchronized (i.e. generate brain activity in different phases), the different neurons extinguish each others activity. This is shown in picture 3.4(a). When they are synchronized, the activity that is measured will be very high, as seen in picture 3.4(b). ERD occurs when the band power on a specific location decreases, as compared to a baseline condition, and ERS occurs when the band power increases. Aftanas et al. [1] and Chanel et al. [7] have successfully used the synchronization and desynchronization features of specific bands on specific locations to recognize the arousal component of emotion.



(a) Event Related Desynchronization (b) Event Related Synchronization

Figure 3.4: Event Related Synchronization and Desynchronization

3.2.2 Cross-correlation between EEG band powers

Not only do the frequency bands contain information on their own, some information can also be obtained from comparing different locations and different frequency bands. Since the asymmetry between the hemispheres is said to tell a lot about emotion (see section 2.2.3), it is useful to use features that express this comparison. For this purpose the cross-correlation between EEG band powers is used.

3.2.3 Peak frequency in alpha band

Kostyunina et al. [36] have found that the peak frequency (i.e. the frequency of the maximal amplitude peak of the power spectrum) in the alpha band can differentiate between emotions. This peak frequency has a tendency to increase with joy and anger and to decrease for sorrow and fear, as compared to a baseline measurement. This measure will also be used in our system.

3.2.4 Hjorth parameters

Already in the '70s of the 20th century, Hjorth [28] has described three parameters that characterize the EEG signal in terms of amplitude, time scale and complexity. The parameters are measured in the time domain, as opposed to the other features, which are measured in the frequency domain. It has been shown that these parameters are capable of discriminating between different mental states [63]. The parameters are:

Activity The first parameter is activity. It is a measure of the mean power of the signal. It is measured as the standard deviation (Equation 5.6).

Mobility The second parameter, called mobility, represents the mean frequency in the signal. The mobility can be computed as the ratio of the standard deviation of the slope and the standard deviation of the amplitude (Equation 5.7).

Complexity The last parameter, complexity, tries to capture the deviation from the "softest" possible curve, the sine wave. It is expressed as the number of standard slopes actually seen in the signal during the average time required for one standard amplitude, as given by the mobility (Equation 5.8).

3.3 Feature selection

Because the feature extraction still results in a lot of data, the number of features should be decreased. Some of the best features should be selected. Fortunately, the extracted features contain a lot of redundancy, due to different features capturing the same information.

Several methods of selecting appropriate features exist. Guyon et al. [23] give an overview of available methods. One of the methods described is using mutual information between features as a measure for the quality of a subset of features. Peng et al. [49] have extended this method, resulting in the max relevance min redundancy (mRMR) algorithm.

The mRMR algorithm tries to maximize the mutual information between the selected features and the desired output (relevance), and to minimize the mutual information between the selected features (redundancy).

$$max D(S,c), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{3.1}$$

$$min R(S), D = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \tag{3.2}$$

where $I(x;y)$ is the mutual information between the two variables. These two factors are combined into one criterion, according to equation 3.3

$$max \Phi(D,R), Phi = D - R \tag{3.3}$$

In practice, incremental search methods can be used to search for a (sub)optimal selection of features. The algorithm seems to select good measures, as is shown by the researchers. They also provided a ready-to-use implementation for Matlab. For these reasons, we have used this feature selection method.

3.4 Classification

After extracting the desired features, we still have to find the emotion. This process will be done by a classifier. A classifier is a system that divides some data into different classes, and is able to learn the relationship between the features and the emotion that belongs to that part of the EEG signal. Several methods for classification have been proposed. We will explain some of the most well-known algorithms.

3.4.1 Artificial Neural Networks

The human brain consists of billions of neurons. These neurons are connected to thousands of other neurons, and together form a neural network. Through these connections, the neurons are able to communicate with each other, and as a whole the network is able to perform highly intelligent actions, such as memorizing things or controlling parts of the human body.

(Artificial) neural networks (ANN or NN) originate from the desire to mimic the human brain in a computer. ANN s can be seen as a simplified model of the human brain and also consists of several neurons, where some neurons are connected. With all neurons together, an ANN tries to learn a relationship between the input and the output, without having knowledge about the underlying model.

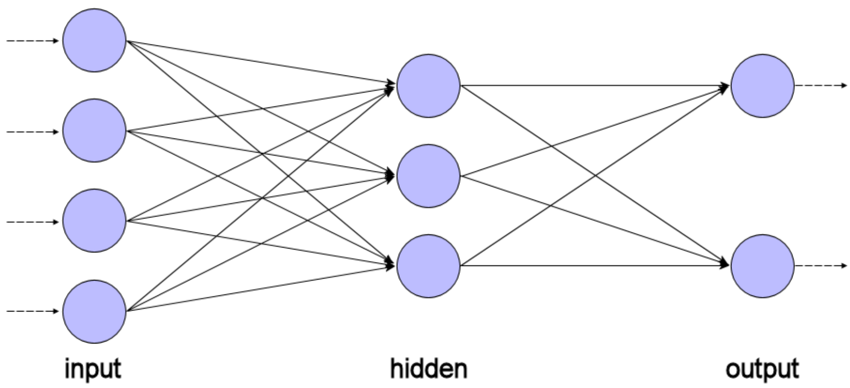


Figure 3.5: The structure of a neural network

The neurons are also referred to as processing elements (PE), as they process the inputs given to it, combine all values and produces one output value. ANN s are typically organized in several layers. One input layer and one output layer are always present, but one or more hidden layers can be added. The neurons from one layer are connected to the neurons in the next layer, giving its output as input to the next layer neurons. An example of a neural network can be seen in figure 3.5. The input to the input layer is the original data to be classified, in this case the feature values. The output from the output layer are the values on the arousal and valence scale.

The neurons in an ANN performs a function on the linear summation of the inputs. The weights of the summation are adapted during the training phase, and reflect the importance of the input to this neuron. Typically, the process in the neuron will look like $y(x) = f(s(x))$ where $s(x)$ is the linear combination of inputs x with weights w and bias b:

$$s(x) = \sum_i w_i x_i + b \tag{3.4}$$

The neural networks do not need to know the underlying model of the data. Instead, an ANN is a general structure that is capable of learning the relation between input and output itself. The most widely used training scheme is the back-propagation method.

This method uses the difference between the output of the ANN and the expected output to change the weights w for all neurons in output layer. This process is repeated for the other layers in the network. The total training consists of the repetition of this training step, where the weights are improved every step. The back-propagation training is performed in the following way:

$$w(n+1) = w(n) + \eta(d(n) - y(n))x(n) \tag{3.5}$$

where $w(n)$ is the weight vector for a certain neuron at training step n, d is the desired output of that neuron, y is the actual output of the neuron, x is the input vector and η is the learning vector. In every training step the difference between the desired output d and the actual output y will be smaller. Therefore, the training will converge to the optimal weights.

In an ANN there is always one input layer, with as many neurons as there are features. The number of neurons in the output layer is the same as the number of desired outputs or classes. The number of hidden layers can be chosen freely. Using one or more hidden layers gives the ANN the capability of learning more complex relations between the input and output. However, it also increases the number of samples needed to train the system. Within the hidden layers, the number of neurons can also be chosen freely. More neurons in a layer can result in a better overall result, but as with the number of layers, the number of samples needed for training also increases with a higher number of neurons in a layer.

3.4.2 Support Vector Machines

A support vector machine (SVM) is another type of classifier. It is capable of separating highly dimensional data. A SVM separates the data from two classes by constructing a hyperplane between the data points from both classes. It searches for the hyperplane that separates the two classes with the highest margin. The data points closest to the separating plane are called support vectors. This process is depicted in figure 3.6.

We will explain the mathematical details of support vector machines with a two-class example. If we have data points

$$(x_1, c_1), (x_2, c_2), \ldots, (x_n, c_n) \tag{3.6}$$

where x_i is the point to be classified and c_i is the class to which the point belongs, -1 or 1. The SVM searches for a hyperplane to divide the two classes with the formula

$$w \cdot x - b = 0 \tag{3.7}$$

where vector w points perpendicular to the separating hyperplane and b is an offset. If the two classes are linearly separable, the parameters w and b can be chosen such that

$$c_i = sign(w \cdot x_i - b) \tag{3.8}$$

for all samples. The intuitive trick is to find the hyperplane that separates the data points best, the hyperplane that keeps the largest margin to the closest datapoints.

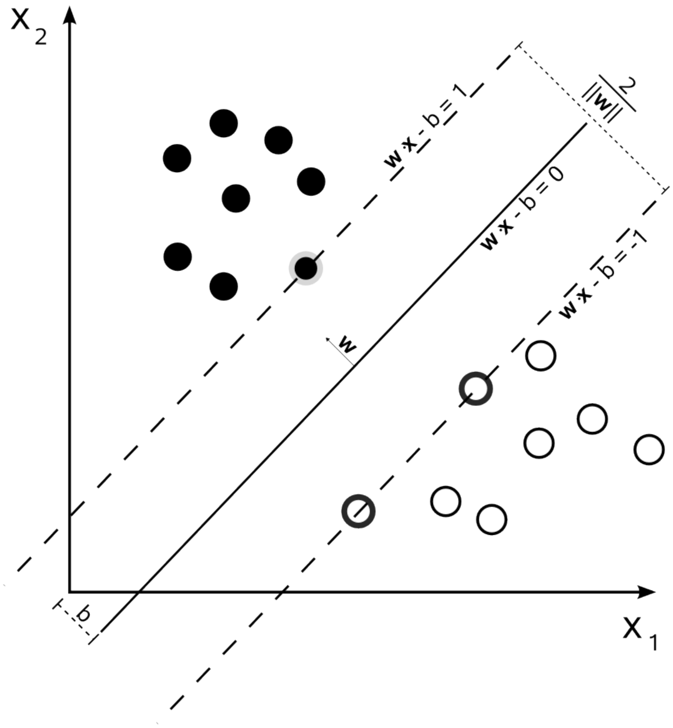


Figure 3.6: Separation using a support vector machine

Those closest datapoints are the support vectors, and two support planes are created through those data points:

$$w \cdot x_i - b = 1 \tag{3.9}$$
$$w \cdot x_i - b = -1 \tag{3.10}$$

The margin between the separating hyperplane and the two support planes is equal to $1/|w|$, so in order to maximize this distance we should minimize $|w|$. When we choose w and b also such that

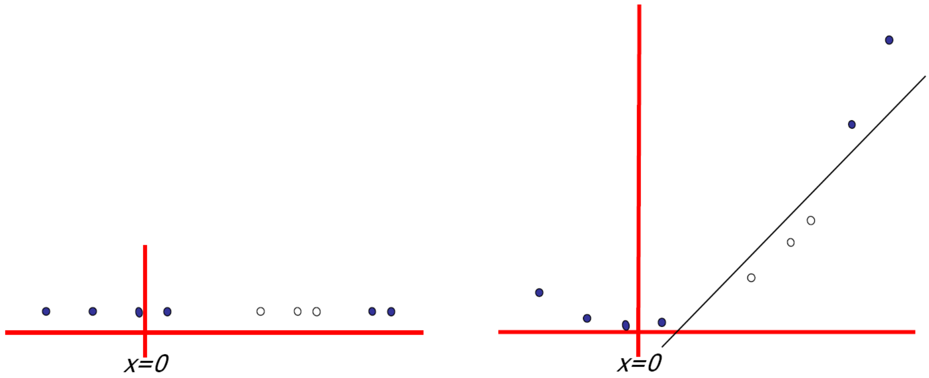
$$w \cdot x_i - b > 1 \text{ for } c_i = 1 \text{ and} \tag{3.11}$$
$$w \cdot x_i - b < -1 \text{ for } c_i = -1 \tag{3.12}$$

we can rewrite the problem as

$$\text{minimize} \quad |w| \tag{3.13}$$
$$\text{subject to} \quad c_i(w \cdot x_i - b) \geq 1, 1 \leq i \leq n \tag{3.14}$$

which can be solved using existing minimization algorithms.

SVM s are very powerful classifiers. Even though it only searches for a separating hyperplane, it is capable of finding very complex divisions between classes. The power of this method lies in the fact that the data is transformed into a high-dimensional space, using a so-called kernel function. Using a proper transformation, it will be easier to separate the points in this higher dimensional space. This process is depicted in figure 3.7. If the separating hyperplane is constructed in a N-dimensional space, where N is the number



(a) 1 dimension (b) 2 dimensions

Figure 3.7: Support vector machine kernel function. The 1-dimensional points in figure 3.7(a) can not be linearly separated. By a transformation to the points (x, x^2) in figure 3.7(b), a linear separation can be made

of training vectors, the data points can always be perfectly separated. However, the memory and computational demands increase dramatically with a larger training set. This is not really a problem in our case, as the training set is small.

3.4.3 Naive Bayes Classifier

A naive Bayes classifier is a simple probabilistic classifier. This classifier computes the probability that some data points belong to a specific class. To perform the classification, the algorithm chooses the class with the highest probability, as its result.

The goal of this classifier is to find the posteriori probability $p(C|F_1, \ldots, F_n)$, where C is the class variable and $F_1, \ldots, F_n$ are the data points. Unfortunately, this probability is hard to determine. Therefore we use the Bayes theorem, a statistical theorem that states that

$$p(C|F_1, \ldots, F_n) = \frac{p(C)p(F_1, \ldots, F_n|C)}{p(F_1, \ldots, F_n)} \tag{3.15}$$

For all classes, the denominator of the fraction is the same, so we only look at the numerator. This numerator can be written as

$$\begin{aligned} & p(C)p(F_1, \ldots, F_n|C) \\ &= p(C)p(F_1|C)p(F_2, \ldots, F_n|C, F_1) \\ &= p(C)p(F_1|C)p(F_2|C, F_1)p(F_3, \ldots, F_n|C, F_1, F_2) \end{aligned} \tag{3.16}$$

and so forth. Here the naive part of the classifier come into play. It assumes that all input variables (features) F_i are independent. That fact makes that

$$p(F_i|C, F_j) = P(F_i|C) \tag{3.17}$$

which can be substituted into equation 3.15, resulting in

$$\begin{aligned} p(C|F_1,\ldots,F_n) &= \frac{p(C)\prod_{i=1}^{n} p(F_i|C)}{p(F_1,\ldots,F_n)} \\ &= \frac{1}{Z} p(C) \prod_{i=1}^{n} p(F_i|C) \end{aligned} \tag{3.18}$$

where Z is a scaling factor, only dependent on $F_1,\ldots,F_n$. This formula is much easier to use and to compute than the first equation, equation 3.15.

Even though this classifier puts huge assumptions on the independence of all input features, it still performs surprisingly well, also when the assumptions are not really accurate. Recent research has shown that there are theoretical reasons for the quality of the classifier [69]. A huge advantage of this type of classifier is the low number of training samples needed to estimate the probabilities.

3.5 Emotion elicitation

When we want our system to learn the relationship between EEG signals and emotion, we need a dataset with EEG signals, for which we know what the emotion of the user at that moment was. Since we need samples with different emotions, we need a method to elicit different kinds of emotions.

3.5.1 International Affective Picture System

Several methods exist for this purpose. For example, participants could be looking at movies, listening to sounds or playing games. However, most experiments (e.g. [7, 34, 55]) that measure emotion from EEG signals use pictures from the International Affective Picture System (IAPS, [38]).

The IAPS is a database of images with an emotional content. The database contains 956 images with a wide range of subjects, like happy people, household objects, car accidents, war, happy couples etc. Each image has been rated for emotional value by a group of participants. Participants rated the images on the valence and arousal score. The scores for all participants rating a picture (approximately 100, half female) are averaged to result in an overall score for that image.

The use of IAPS allows better control of emotional stimuli and simplifies the experimental design. Because of the standardized picture set, experiments using IAPS can easily be verified by other scientists. Another advantage of using images is the fact that it minimizes muscle and eye movements, compared to playing games or looking at movies, which results in cleaner EEG measurements.

3.5.2 Self Assessment Manikin

When using the IAPS to elicit emotion, we assume that showing an image with some specific emotional content will trigger that emotion on the participants. However, it

Figure 3.8: Image similar to IAPS pictures

is possible that the emotion that a participant experiences differs from the expected value. For that reason, the participant is asked to rate his emotion on a Self-Assessment Manikin (SAM, [4]). This is a standardized system to assess emotion on the valence and arousal scales (and dominance, if wanted). Using this system makes it easy to compare the self assessments with the expected emotion, based on the IAPS ratings of the images. This procedure of filling in self assessments is also suggested by Chanel et al. [7], because they found that the self assessments had a stronger relationship with the EEG signals than the preset IAPS scores. That would support our expectations that the pictures do nog always raise similar emotions with different people.

3.6 Measuring EEG

We want to create a dataset with EEG signals, for which we know the emotion. To accomplish that goal we also have to do some measurements ourselves. The process of measuring EEG signals is not a very simple task. This section will provide some information on the methods and equipment we used for the EEG measurements.

3.6.1 Practical internship in Prague

In order to gain experience with measuring and analyzing of EEG signals, I have done an internship at the Faculty of Transportation Sciences of the Czech Technical University in Prague with professor Novak and Petr Bouchner. Their group is involved in experiments after micro-sleeps and workload in driving simulators [3]. They have done a large number of experiments where participants drove in a driving simulator and had to perform additional tasks. During these rides, EEG signals were measured and saved for later use. In order to gain the needed experiences, but also assist with the project in driving simulators, I have tried to measure workload levels from brain activity, during a ride in the driving simulator. More details on this project can be found in appendix D.

3.6.2 Equipment

To perform our measurements, we used the equipment that was available on the university. The equipment was the same as at the university in Prague, so the experience gained in Prague was useful for these measurements. The equipment consists of the following parts:

- EEG-*cap*
  To measure brain activity through EEG signals, we used the Truscan EEG-caps (see figure 3.9(a)) provided by Deymed. The cap uses 19 Silver Chloride (AgCl) electrodes, placed according to the international 10-20 system (see for more information section 2.1.3). Three EEG caps are available for different sizes of the head.

- *Earlobe electrodes*
  In order to measure a reference signal that is (as much as possible) free from brain activity, we have two electrodes to attach to the participants earlobes. See figure 3.9(b).

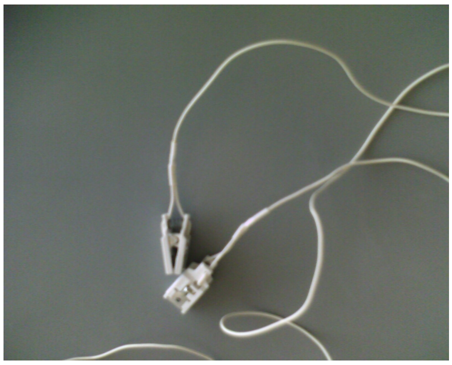
- EEG *headbox*
  The signals from the EEG-cap are fed into the EEG headbox (figure 3.9(c)). This device has the sole task of amplifying the signals from the EEG-cap and leading it to the computer. We used the amplifier that came with the Truscan system.

- EEG *adapter*
  The EEG adapter (figure 3.9(c)) converts the analog EEG signals to a digital version. We used the EEG adapter that came with the Truscan system.

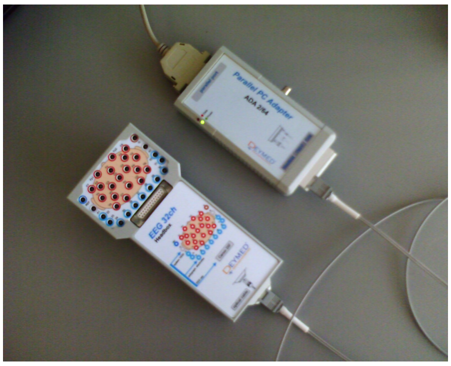
- EEG *gel:* ECI *electro-gel*
  In order for the EEG cap to function properly, we have to make sure that the electrodes are connected to the head. This is done by putting some conducting gel between the electrodes and the skin. See figure 3.9(e).

- EEG *conductive paste: Ten20*
  Ten20TM conductive EEG paste (figure 3.9(d)) is used to lower the impedance on the earlobe electrodes. This paste is more viscous than the EEG gel used on other electrodes, and is used on the earlobes because other gel would dribble from the earlobes.

(a) EEG cap



(b) Earlobe electrodes



(c) EEG headbox and adapter

(d) EEG paste: Ten20

(e) EEG gel: ECI electro-gel

Figure 3.9: Hardware used for EEG recordings

Usage

The EEG adapter is connected to a PC using the parallel port. An optical cable connects the EEG adapter and the EEG headbox, to prevent any electrical charge from reaching the participant. The EEG cap and the earlobe electrodes are connected to the EEG headbox with a flat cable, containing one wire for each electrode.

The EEG cap is placed on the participants head, with the two frontal electrodes on his forehead, just under the hairline. The cap is connected to a belt around the participants chest, to make sure it will stay in the same place.

After the EEG cap is placed, the most time-consuming part of the preparation starts: filling the electrodes with conductive gel. This gel will lower the impedance between the electrode and the participants skin. This process is necessary, since voltages in EEG signals are in the range of microvolts. Higher impedance will increase the noise available in the signals, and reduce the usability.

3.6.3 Preparation

When the participant is wearing the EEG cap, all electrodes are filled with ECI electro-gel. This gel is injected under the electrode through a hole on top of each electrode using a syringe. The earlobes are prepared with Ten20TM conductive EEG paste, to lower the impedance on those electrodes. When injecting the EEG gel or applying the conductive paste, we look at the TA software, which presents a screen with the impedance levels for each electrode. We keep injecting EEG gel until the impedance will reach a level below 10 kΩ. That level is used by Dharmawan [15] and Bouchner [3] and was also suggested by people who worked with the system before.

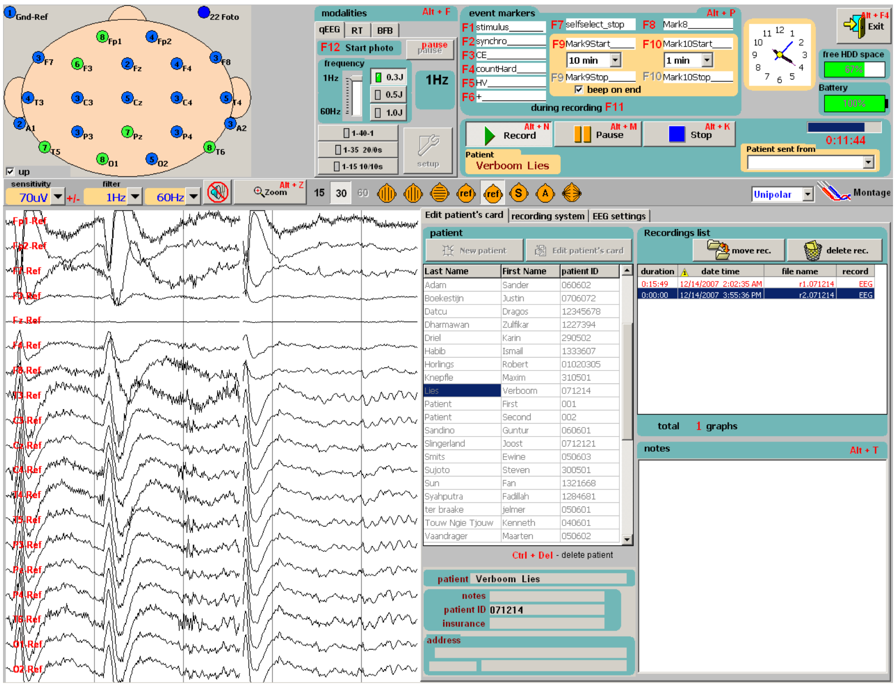
In order to avoid most of the artifacts, the participant is sitting in a comfortable chair where he has no need to move any muscles during the EEG measurements. The only moment when a participant has to move is when he is filling in the self-assessment survey.

3.6.4 Software

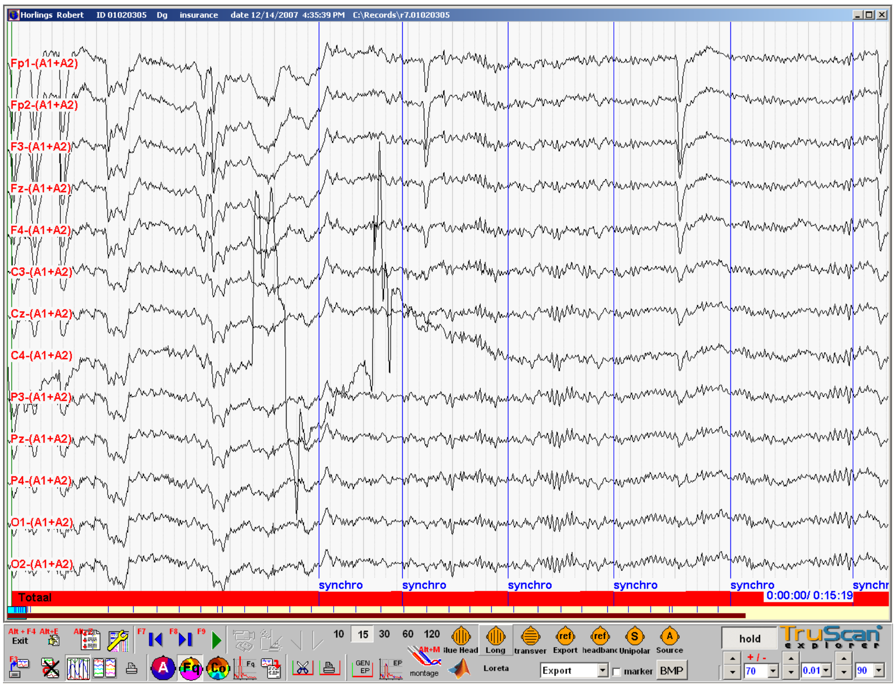
The Truscan hardware only worked well with the accompanying software. Two programs were available, Truscan Acquisition and Truscan explorer.

- *Truscan Acquisition*
  Truscan Acquisition (TA) is the program that assists in measuring the EEG data. TA handles the low-level communication with the hardware, and supplies an interface where the recordings can be made. See figure 3.10(a).

- *Truscan Explorer*
  Truscan Explorer (TE) can be used to analyze the recordings by visual inspection, as well as to perform basic frequency analysis. we have used TE only to export the data to a simple text file, as the program is not very intuitive. See figure 3.10(b).



(a) Truscan Acquisition user interface during a measurement



(b) Truscan Explorer user interface for analyzing the recordings.

Figure 3.10: Software used for EEG recordings

3.7 Summary

In order to remove most of the noise and artifacts from the EEG signals, we will use a combination of methods to retrieve a clean output. A bandpass filter is applied to remove frequencies under 2Hz and above 40Hz. This filter will remove all high frequency noise, and also removes most of the artifacts. Because of the simplicity and the good results of this filtering method, we prefer this method over other artifact removal methods. After this filtering step, the data is rereferenced using the CSD technique, because this seems to be the best method when looking at hemispherical asymmetries.

Several authors have shown good results in recognizing emotion with different sets of features. For that reason, we will extract all features that were found in those papers. Since this results in a rather large set of features, we will use a feature selection algorithm to select the best subset of these features to use for our classification process.

In most experiments for recognition of emotion, a neural network is used to classify the features, because of its capability to learn complex associations [10, 24, 63]. However, we have only few training samples, which makes the training of the neural network very hard. The same disadvantage applies for support vector machines. However, a naive Bayes classifier still seems to be too simple for our purposes. For that reason we will implement all types of classifiers and decide the specific type of classifier we use later on (see section 9.4).

For emotion elicitation we will use a subset of the IAPS images.

Chapter 4

System design

The main goal of this project is to build an emotion recognition system. This chapter describes the overall structure and design of the system. For designing the program, we have looked at literature about other research, in order to find out good practice and proven ideas in this field. We will implement several possibilities for the different processes in our system. Those methods all have their advantages and disadvantages, and have to be tested to see which one is the best. When we have a good dataset to test our system, we will perform some explorative tests to find out the best choices.

4.1 Goal of the system

The system is intended to recognize emotion from offline EEG signals. The input of the system consists of EEG signals and the output of the system will be some indicator about what emotion the subject experiences. The first step will be to return some values on the valence and arousal scale. With these values, it is possible to represent all emotions, such that the output is not restricted by predefined emotions. Afterwards, it might be possible to put a tag on the valence and arousal values, in order to return a specific emotion.

4.1.1 Intended use

The system will be designed for research purposes, to test and analyze different techniques and methods for emotion recognition from EEG signals. We focus on the techniques and the performance, instead of creating a fancy user interface. However, we keep in mind that the methods we choose must be extendable to online emotion recognition and use in commercial products.

4.2 Requirements

To design a software program, we need to know the requirements. This section describes the requirements that are put onto the system.

4.2.1 Functional requirements

The functional requirements of the system are the following:

- *Analyze biophysical measures*
In order to achieve the main goal, the system must be able to analyze biophysical measures. The raw signals that are measured should be preprocessed, in order to remove noise and artifacts. After this preprocessing, features will be extracted from the signals, that are used for classification.
- *Derive valence and arousal values*
The extracted features of the signal will be fed into some expert system, that will compute the values on the valence and arousal axes.
- *Adjustment to different persons*
Emotion is a very subjective matter. One person can experience fear or happiness in one way, where someone else will feel the same, but his body reacts differently. EEG signals for the same emotion might therefore be different for several people. The system should be able to handle this difference.

4.2.2 Nonfunctional requirements

In addition to the functional requirements, the system should also meet some nonfunctional requirements:

- *Performance*
Because the ultimate goal is to provide an online system, the system should respond fast. Although EEG measurements result in a lot of data, and sometimes a large amount of computation is needed, this should not result in a delay of more than a second.
- *Flexibility*
The system we will build, will not satisfy the ultimate goal yet. For that reason, it should be able to modify the system easily or add new functionality. One might want to add new input devices, other data analysis techniques or novel classification algorithms.
- *Hardware considerations*
Measuring EEG signals and other physiological modalities requires special hardware. For the EEG measures we have used the Deymed TruScan 32 device. See for more information section 3.6.2. However, the system should be capable to be extended to use other hardware. The program itself will be able to run on standard PC hardware.

4.3 Structure

The global structure of the system is depicted in figure 4.1. It shows the two main parts of the system. The part on the left is for training the system. We will have to teach the system the relationship between the EEG signals and emotion. This is done automatically, by the training part. This process does only have to be run once, since the system will then be able to recognize emotion from new EEG signals (in the ideal case). The other part shows the normal execution of the system. This part does exactly the same preprocessing and feature extraction steps as the training part. After the feature extraction, it will use the relationships extracted in the training phase to classify the EEG signals into emotions.

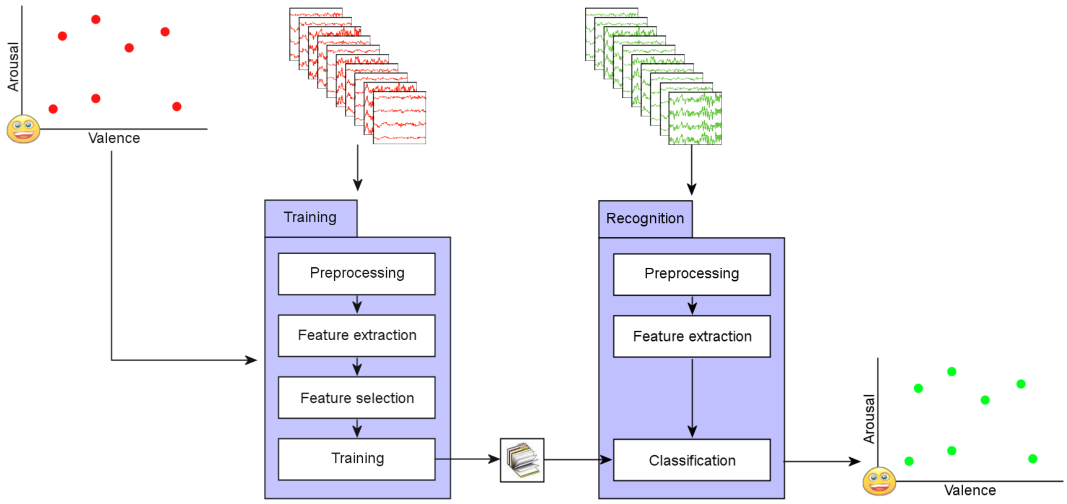


Figure 4.1: Global overview of the system with its input and output

4.3.1 Input

The input to the system consists of measured EEG signals. These signals can be measured by any device. It will be capable of importing several existing EEG file formats.

4.3.2 Data preprocessing

Raw EEG data is generally a mixture of several things: brain activity, eye blinks, muscle activity, environmental noise etc. After the data has been collected, the preprocessing step is meant to remove all unneeded noise and artifacts from the signal, and only keeping the interesting part of the signal, the brain activity.

Preprocessing requirements

In order to get satisfying results from the system, the preprocessing step should:

- remove all noise and artifacts from the signal, but preserve all the characteristics of the original signal,

- clean the signal from the influence of the reference electrode, if one is used,
- set the sampling rate to a fixed (low) number, in order to reduce memory requirements.

Baseline removal

Different EEG channels have different characteristics. This has to do with the electrode montage, the electrode itself but also with the location on the skull. These variations might introduce noise or artifacts (which is referenced in section 3.1.2), but can also result in a deviation of the mean of the signal from zero. Figure 4.2 shows the process of baseline removal.

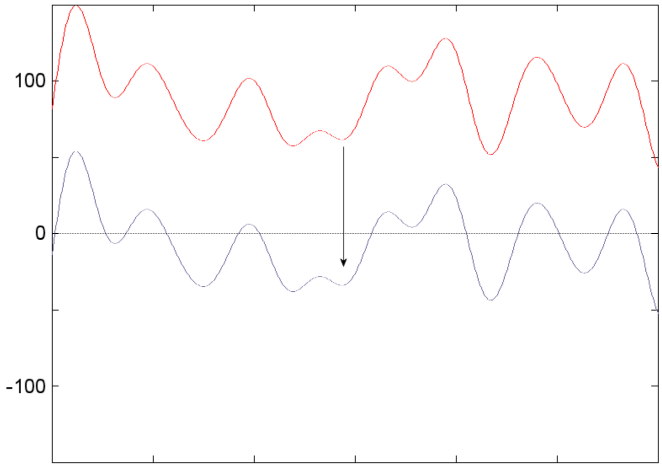


Figure 4.2: Example of baseline removal. The red signal is the original signal, which has a baseline of 100. The baseline is removed by subtracting 100 from all points

Downsampling

The EEG data is measured at a rate of 256Hz or sometimes 128Hz. The useful information in this data is only in the frequencies under 40Hz. Because of the Nyquist sampling rate [68], to measure this activity only 80Hz of measurements is needed. Therefore, reducing the amount of data while preserving all information in order to save memory space can be easily done by downsampling the data. We chose to downsample to a rate of 128Hz, because in that way the downsampling can be done easily without any inaccurate interpolation.

Referencing

As explained in section 3.1.1, it is important to choose the best reference for the EEG measurements. As suggested by research from Hagemann et al. [25], we will use the CSD method to retain the valuable information in the signals.

Noise and artifacts

In order to remove most of the noise from the measured signals, a band-pass filter is used that will remove frequencies below 2Hz and above 40Hz. This filter will also remove most of the artifacts, because their frequencies are far below 2Hz. We will not use any more sophisticated form of artifact removal.

4.3.3 Feature extraction and selection

After the data is filtered and artifacts are removed, there is still a huge amount of data. Because a lot of this data is unnecessary to recognize emotion, and it would take too much time to use all this data to perform the job, some features are extracted from the data. The goal of feature extraction is to find those features that contain all the information needed to recognize emotion.

Feature extraction requirements

The requirements for feature extraction are:

- reduce the size of the data by selecting appropriate features,
- the selected features should be minimally redundant and the expected results should maximally depend on these features,
- preserve all information from the signal that is needed for emotion recognition.

Features

In this system we will use all features that were found in literature about emotion recognition, as explained in section 3.2. All features seem to contain some information about different aspects of emotion. Therefore the combination of all features should be good enough to recognize emotion.

Time windows

As we want to recognize emotion on different moments, but also want to extract frequency information and derivatives, we cannot analyze the EEG samples on its own. We have to derive features from EEG sample windows of a specific length. Since emotion does not fluctuate too much, within a short period of time, this should not be a problem. We use 1 second windows feature extraction, as is also done by the Päivinen et al. [46] and Vourkas et al. [63], in order to have enough data to compute all the features.

Scaling

The features we have extracted may contain values in a wide range. Classifiers tend to be dominated by larger features (or features with larger spread). Since we do not want that to happen, we scale all features in such a way that they all lie between 0 and 1. This was also suggested by Hsu et al. [29].

This process is initialized when training the classifier. When actually classifying data, it is scaled in exactly the same way, and may thus lie outside the desired domain. However, we expect the trainingdata to be dense enough to cover all possible values for the features. That will make sure that the features never reach extremely deviating values.

Feature selection

The mRMR algorithm (see section 3.3) is used to select the best features to use for emotion recognition. The number of features to be extracted will be determined later on.

4.3.4 Analysis and classification

With the selected features, the system can try to recognize emotion. The dimensional model of emotion will be used in our program. As explained in section 2.2.2, this model provides a simple way to represent emotion.

We will implement different types of classifiers in our program, because we do not know the best one beforehand. When we have measured data to test our program, we will select the best classifier for our purpose.

The features are classified separately into a value on the valence and the arousal scale. We separate these two classifiers, because literature suggests that the arousal dimension is found in different parts or properties of the brain than the valence dimension.

Analysis and classification requirements

The requirements for the analysis and classification are:

- determine the right values on the arousal and valence scale, from the selected features.

4.3.5 Output

The output of the system will consist of two values on a 2-dimensional plane, arousal and valence. These values will be plotted real-time, so that a moving path through the arousal-valence plane will appear. The arousal and valence values will additionally be classified into different emotional classes. This process will be done based on the data from other research.

4.4 Graphical user interface

To provide the user an interface to control the program, we have created a simple graphical user interface. This interface only provides basic functionality, as it is not the focus of this project. The requirements for the user interface are:

- *Training*
  - select input files for training,
  - select classification method,
  - train the system.
- *Recognition*
  - select an input file with EEG data,
  - select an input file with marker data,
  - classify the input data and show the results.
- *Loading and saving*
  - save and load the configuration for the system,
  - save and load different classifiers.
- *Misc*
  - provide feedback to the user about the performance of the different parts of the system.

4.4.1 Main screen

The main screen of our program provides an easy way to train, test or execute the system. It is shown in figure 4.3. The screen is divided into three parts; the left part shows information about the training of the system, the middle part presents information about the tests that are run and the right part shows some things about the data that is loaded for execution. The traffic lights give an easy clue about the quality of the training data, the testresults or the execution data.

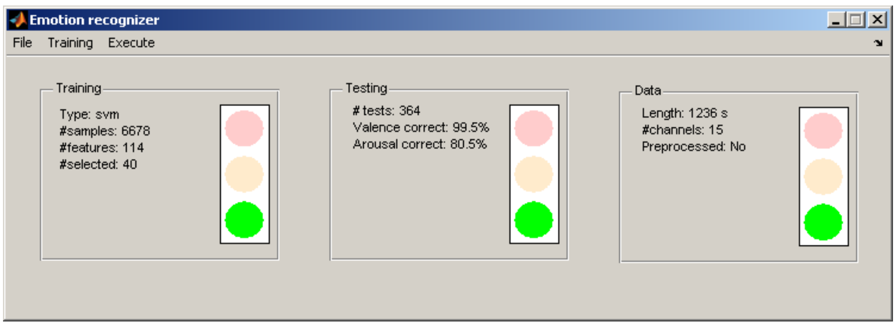


Figure 4.3: Main screen of the emotion recognition program

4.4.2 Results screen

The results screen (see figure 4.4 gives the results of the execution of the program. The top part shows information about the EEG, and shows a sample of the signal on one of the electrodes. The bottom part gives information about the emotion that was found by the system, as well as the real emotion someone felt, when it is available. The graph shows a graphical representation about the emotion through time.

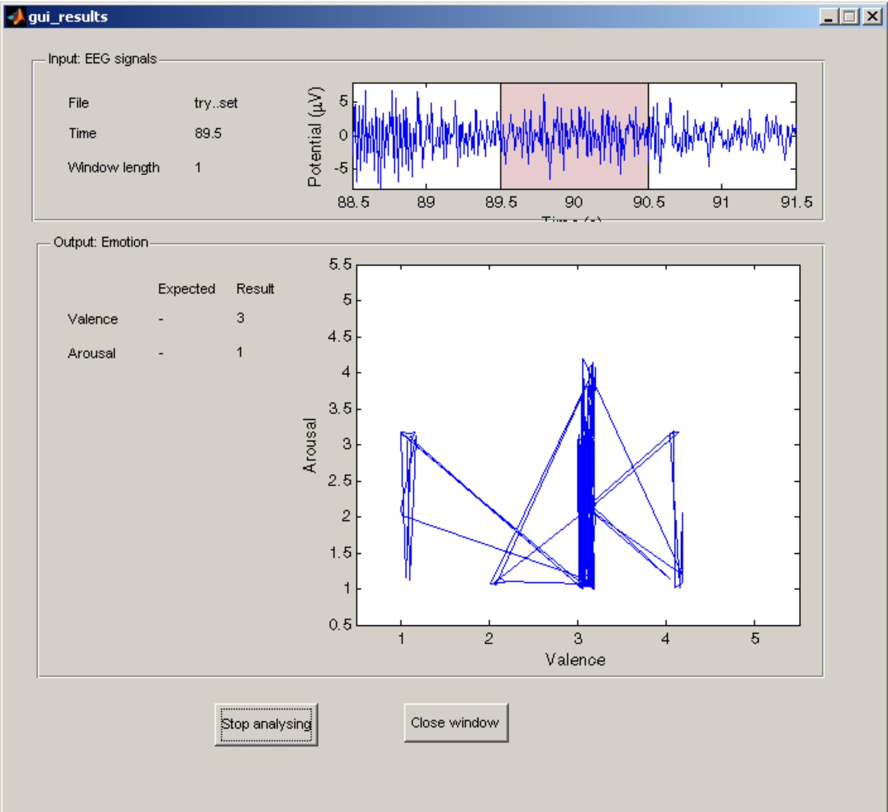


Figure 4.4: Results of execution of the emotion recognition system

Chapter 5

Implementation

Although all methods and techniques used are properly documented and described, there are still many non-trivial implementation issues. This chapter will describe the solutions we have chosen. Section 5.1 explains the software we used, section 5.2 describes the details about our implementation and in section 5.3 you can find the details about our testing.

5.1 Software

We have used some existing software as a base for our own system. Using this existing software resulted in a little bit less flexibility, but saved a lot of time and effort as several functions were already implemented.

5.1.1 Matlab

Emotion recognition consists of different parts, that require a lot of computational effort. As one of the requirements is the speed of the recognition, we have used Matlab to build the system. Matlab is a high level programming language in a numerical computing environment. It is capable of performing computationally complex tasks very fast and has a large userbase and many existing libraries and implementations of useful algorithms.

5.1.2 EEGLab

EEGLab [13] is a Matlab toolbox for processing and analyzing EEG and ECG data. It provides a large set of functions, that are available from the matlab command line. These functions range from simple averaging to ICA and advanced ERP analysis. The toolbox also provides a graphical user interface. In our system we will use many of the built-in functions of this toolbox.

5.2 Implementation details

This section handles the details of implementation. One of the issues we ran into when implementing this program is the lack of details about the implementation mentioned in most of the articles. This made the implementation a lot harder, and this section will help readers to reproduce the experiments, if needed.

5.2.1 Input

All data is imported into EEGLab structures. This structure provides flexibility and a consistent way to save information. Moreover, using these structures, many functions of EEGLab can be used, so they do not have to be implemented again.

5.2.2 Preprocessing

First the baseline is removed from each channel separately. After that, the signal is resampled to 128Hz, in order to reduce memory requirements. The signal is then rereferenced to CSD (see section 4.3.2), which can be computed as

$$csd_i(n) = -sc * l_i(n) \tag{5.1}$$

where $csd_i(n)$ is the CSD of the signal at electrode i at time n, sc is the skin conductivity at electrode i and $l_i(n)$ is the spatial laplacian of the signal at electrode i, which can be computed as:

$$l_i(n) = s_i(n) - \frac{1}{N} \sum_{m \in neighbours_i} s_m(n) \tag{5.2}$$

where $s_i(n)$ is the signal at electrode i at time n, and $neighbours_i$ contains the spatial neighbours of electrode i. Skin conductivity is unknown, so it is estimated at 33 Siemens per meter [65].

After that the signal is bandpass filtered between 4Hz and 40Hz, using a FIR filter (as implemented by EEGLab). This removes most noise due to electrical devices, as well as the lowest frequencies, which contain heartbeat and ocular artifacts and rarely contain useful information for EEG purposes. Another effect of this filtering is that most of the artifacts are removed as well, as they have frequencies of less than 2 Hertz. This fact, and the fact that the algorithm for online removal of artifacts was not very fast, made us decide not to use the artifact removal in our system.

5.2.3 Feature extraction

The data from the Enterface project contains EEG signals for 54 different channels. Our own EEG-cap only contains 19 channels. However, because both caps use a standardized placements of the channels, we can look at the channels that are available in both setups. Figure 5.1 shows the electrode placement according to the 10-20 system. Channels measured in the Enterface project are marked blue, channels only measured with our own cap are marked red. Common channels are colored green. Only the channels in both setups are taken into account when computing the features. This resulted in less features than originally designed and might decrease the performance.

Before extracting the features, the power spectral density (PSD) of the signals is computed using Welchs method [66]. The band power $bandpower_i(band)$ in a the alpha band on

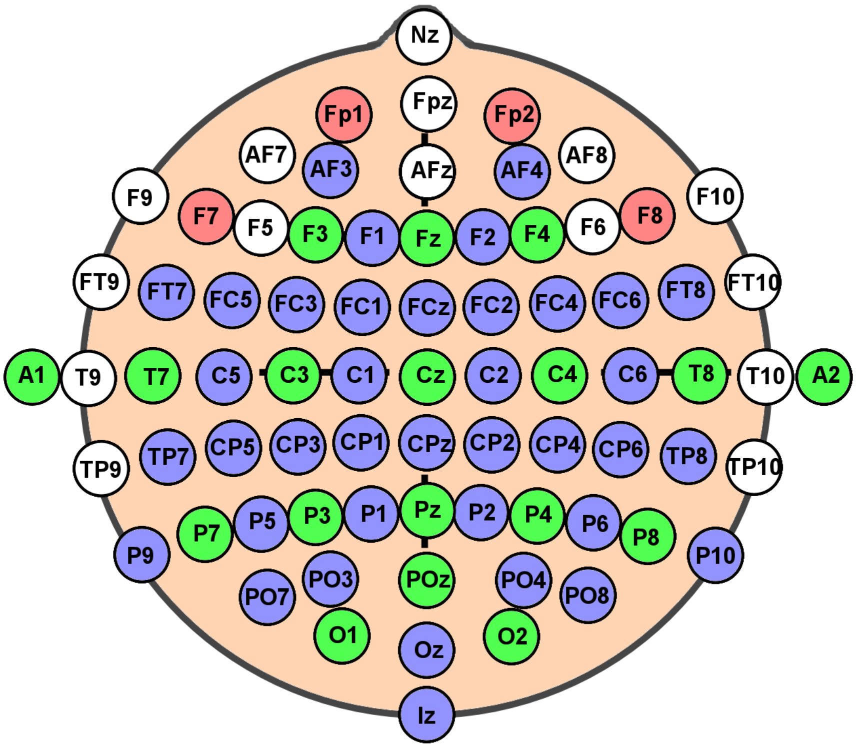


Figure 5.1: Used electrodes in Enterface project and own measurements. Red electrodes are only used in our own measurements, blue electrodes are only used in Enterface project and green electrodes are used in both measurements

electrode i is computed using

$$bandpower_i(\alpha) = \sum_{f \in \alpha} X_i(f) \tag{5.3}$$

where *band* is the frequency band and $X_i(f)$ is the Fourier transform of the EEG signal on electrode i in frequency bin f.

With this information, we compute the following features:

- *ERD/ERS*
  Divide the band power at a specific moment by the baseline band power.
- *Crosscorrelation*
  The crosscorrelation between electrodes at locations F3, F4, T7, T8, P3, P4 and O1 is measured in the θ, α and $\beta 1$ bands. The cross-correlation coefficient in the α band between electrodes j and k is given by

  $$c(\alpha; jk) = \frac{\sum_{f \in \alpha} X_j(f) X_k(f)}{\sqrt{\sum_{f \in \alpha} (X_j(f))^2} \sqrt{\sum_{f \in \alpha} (X_k(f))^2}} \tag{5.4}$$

  as explained by Musha et al. [44]. The researchers also took into account the channels T3, T4 and O2. Since O2 is not always available in our measurements, it is discarded. Channels T3 and T4 are also not available in our measurements (they even can not be found in descriptions of the standardized 10-20 system), but we have taken T7 and T8 as a substitute for them.
- *Peak frequency*
  The most dominant frequency in the alpha band, as computed by

  $$peak_i = argmax_{f \in \alpha} X_i(f) \tag{5.5}$$

  where $X_i(f)$ is the activity of frequency f on electrode i.
- *Hjorth parameters*
  The Hjorth parameters for activity, mobility and complexity. Mathematically these parameters can be computed as

  $$\begin{aligned} h_1 &= \sigma_x^2 && (5.6) \\ h_2 &= \sigma_d / \sigma_x && (5.7) \\ h_3 &= \frac{\sigma_{dd}}{\sigma_d} / \frac{\sigma_d}{\sigma_x} = \sigma_{dd} / \sigma_x && (5.8) \end{aligned}$$

  where σ_x is the standard deviation of the EEG signal, σ_d is the standard deviation of the first derivative of the EEG signal and σ_{dd} is the standard deviation of the second derivative of the EEG signal.

  In practice, the Hjorth activity has a too high variation, because they describe the activity of the signal, and that activity can be very different in different sessions. Since the activity depends largely on coincidence (sensor properties, impedance while measuring etc.) and the activity of the signal is also represented by the ERS features, we have discarded these features.

The resulting list of features can be found in appendix A.

Scaling

Before performing the classification, we scaled the features linearly in order to have all features in the same range. Although we initially used a range from 0 to 1, as suggested by Hsu et al. [29], we found out that using a larger range resulted in better classification. This probably has to do with the precision of the values that are used internally. To circumvent this problem we used a range from 0 to 100.

5.2.4 Feature selection

The mRMR algorithm as provided by Peng et al. [49] uses so-called Mutual Information toolbox, which is also provided by the same persons. However, this toolbox did not run on my machine, so we used another implementation of an algorithm to compute mutual information (found at `http://www.cs.rug.nl/~rudy/matlab/index.html`, used with permission from the author). The only parameter in the feature selection step is to select how many feature we have to keep. We have to keep enough information for the classification, but also decrease the computational demands of that task. The parameters will be chosen later on (see section 9.4), when we have enough data to analyze this matter.

5.2.5 Classification

For classification we used libsvm [8], a library for support vector machines and the Matlab Neural Network toolbox. The specific parameters for both classifiers are to be chosen in a later stage. The naive Bayes classifier was implemented manually, without the use of a toolbox. The linear vector quantizer was also implemented by the Neural Network toolbox.

5.3 Testing

After the implementation phase, we have to test whether the software works correctly. The system as a whole will be tested later on, when we have measured enough data to do that. Now we will focus on the different parts of the system, and see whether they do the jobs they are supposed to.

We have assessed the quality of the different parts within the system, by creating some artificial input to the part, for which we can predict the outcome. That way, we can exactly check whether the system meets our expectations.

5.3.1 Preprocessing

The preprocessing phase can be divided into 4 steps: baseline removal, downsampling, rereferencing and filtering (see section 4.3.2). We will construct 4 different datasets that all test one (and only one) of these steps. These tests consist of articially constructed EEG samples. The specific properties of the different tests can be found in table 5.1.

Table 5.1: Constructed testsamples to test the preprocessing phase. Bold lines denote the properties of the signal that should change during the test

Part under test	Properties	Test sample	Expected result
Baseline removal	Channels	1	1
	Sampling rate	128Hz	128Hz
	Constructed signal	10Hz	10Hz
	Baseline	**5μV**	**0μV**
Downsampling	Channels	1	1
	Sampling rate	**256Hz**	**128Hz**
	Constructed signal	10Hz	10Hz
	Baseline	0μV	0μV
Rereferencing	Channels	2	2
	Sampling rate	128Hz	128Hz
	Constructed signal	**both 10Hz**	**both 0**
	Baseline	0μV	0μV
Filtering	Channels	1	1
	Sampling rate	128Hz	128Hz
	Constructed signal	**1 + 10 + 60Hz**	**10Hz**
	Baseline	0μV	0μV

Results

For results, we compared the output visually with the expected output. That process showed very big similarities. As a numerical measure, we compared the mean difference between the output and the expected output at all moments. The results are shown in table 5.2, and look very good.

Table 5.2: Testresults for the preprocessing phase

Part under test	Properties	Expected result	Result
Baseline removal	Mean difference	0	0.02
Downsampling	Mean difference	0	0.02
Rereferencing	Mean difference	0	0
Filtering	Mean difference	0	0.02

5.3.2 Feature extraction

Just as with the preprocessing phase, we have 4 different parts to test. The parts consist of the 4 different types of features we extract. All types of features need a different approach to test:

- *ERS features*
  The ERS features measure the change in brain activity between the EEG signal and

a baseline signal. We will construct a signal that has some activity on the baseline, and has the same activity, but an amplitude twice as high in the signal under test. This should result in a feature value of 4.

- *Cross correlation*
  We will construct a signal for which two channels have a 5Hz signal, and a third channel has the same signal but inverted. That should produce a cross correlation in the theta band (4-8Hz) of 1 between all channels. Furthermore we will create a channel with a 2Hz + 5Hz signal, which should give a correlation of almost 1 with the other channels, since the 2Hz should be filtered out of this band. Yet another channel will have a signal of 7Hz, and should have a small correlation with the other channels.

- *Alpha peak frequency*
  The alpha peak frequency can be easily measured by creating a channel with low activity on all frequencies between 5 and 15, but a high activity on a specific frequency in the alpha band. We will test this for frequencies 8Hz, 9Hz, 10Hz, 11Hz and 12Hz, and expect the feature to take that frequency as value.

- *Hjorth parameters*
  For Hjorth parameters, the tests are somewhat more complex. Hjorth mobility is a measure for the mean frequency in the signal, so we construct a channel with 2Hz sine waves, a channel with 4Hz sine waves, one with 6Hz sine waves and one with 8Hz sine waves. The relation between the Hjorth mobility of different channels should be 1:2:3:4. The complexity should be approximately the same for these 4 channels. As an extra test, we add a channel with waves of 3Hz + 5Hz. This should give a significantly higher complexity.

Results

The feature extraction functions have all passed the tests, since the results are in line with the expected results. The results are shown in table 5.3.

5.3.3 Feature selection

Feature selection is actually a fairly simple process: it should return the features that have the highest mutual information with the output, and there should be a low mutual information between different features. To investigate that, we construct 4 features and some artificial output. The output is a list of ascending numbers, 1 up to 100. All features consist of the same numbers, but with some noise added, to decrease the mutual information between the output and the feature. Feature 3 has the most noise added, feature 1 the least. We add another feature with the same values as feature one. We expect the output to be that feature 1 is the best, followed by feature 2 and 3. Feature 4 could also be used instead of feature 1, but not together, because both features contain the same information about the output.

Table 5.3: Testresults for feature extraction

Part under test	Properties	Expected result	Result
ERS features	ERS Theta1	4	4
	ERS Theta2	4	4
Cross correlation	Channels 1 & 2	1	1
	Channels 1 & 3	1	1
	Channels 1 & 4	1	0.97
	Channels 1 & 5	$<< 1$	0.25
	Channels 2 & 3	1	1
	Channels 2 & 4	1	0.97
	Channels 2 & 5	$<< 1$	0.25
	Channels 3 & 4	1	0.97
	Channels 3 & 5	$<< 1$	0.25
	Channels 4 & 5	$<< 1$	0.43
Alpha peak frequency	Channel 1	8	8
	Channel 2	9	9
	Channel 3	10	10
	Channel 4	11	11
	Channel 5	12	12
Hjorth parameters	Mobility 2 / 1	2	2
	Mobility 3 / 1	3	2.99
	Mobility 4 / 1	4	3.98
	Complexity 2 / 1	1	1.02
	Complexity 3 / 1	1	1.05
	Complexity 4 / 1	1	1.10
	Complexity 5 / 1	$>> 1$	4.35

Table 5.4: Testresults for classification phase

Classifier	Properties	Expected result	Result
Support Vector Machine	Classification rate	1	1
	Mean squared error	0	0
Naive Bayes Classifier	Classification rate	1	1
	Mean squared error	0	0
Neural Network	Classification rate	1	1
	Mean squared error	0	0.0001

Results

The feature selection process resulted in the right sequence: 1 2 3 4, as expected. Other permutations of the same signals also resulted in the expected sequence.

5.3.4 Classification

To test our classification, we create a simple training set of 2 features and 1 output. The output is divided into 3 classes: 1, 2 and 3. The features describe points centered around (2,2) for class 1, around (5,4) for class 2 and around (2,5) for class 3. The distance between all points and the class center is less than 1, to make the problem linearly separable. If we train the system on this input, and feed all the same points into the system, it should be able to classify them correctly.

Results

All implemented classifiers perform equally well, and are able to classify all samples into the right categories. Details can be found in table 5.4.

Chapter 6

The Enterface '06 database

For our experiment, we needed as much EEG data as possible, in order to train a system that works for different people. For that reason we have used a database from Enterface '06 (referred to as the Enterface database) that was built by Savran et al. [55] as our initial dataset. This dataset was extended by additional data from our own experiment, which is explained in chapter 7. In this chapter we will present information on the details of the Enterface database, as well as an analysis of the usefulness of the data.

6.1 Experimental design

The Enterface experiment was meant to construct a database of EEG and other physiological signals for emotion recognition. Next to EEG, fNIRS signals and galvanic skin response (GSR), respiration and blood volume pressure are recorded. Because the fNIRS devices occupied the frontal part of the head, EEG signals are not measured for some frontal electrodes.

6.1.1 Experimental protocol

Savran et al. [55] showed each participant 3 sessions of 30 emotional stimuli. A stimulus consists of 5 images from the same class, as described above. Every image was shown for 2.5 seconds, for a total of 12.5 seconds. After this stimulus, there was a black screen for 10 seconds, and the participant was asked to give a self-assessment of his emotional state. These self-assessments are measured because emotions are known to be very subjective and dependent on previous experience. One can never be sure that the person feels the emotion that was intended by the pictures, and this self-assessment gives a good possibility to compare the results between persons. However, self-assessments are known to be very subjective, and hard to compare. For example, some persons tend to give extreme scores where others always choose the center.

Some problems with this experimental design are given by Savran et al. [55]:

- Participants reported a headache at the end of each session, due to the different caps (they also wore a fNIRS cap),
- Positive responses were not always felt, but the negative images were a bit too hard,

- Several participants said that the effects of the images decreases after viewing many images.

6.1.2 Instrumental setup

For recording all signals, three computers are used. One computer is used for stimulation presentation, the other two computers are used for recording fNIRS and EEG signals. For EEG measurements, the Biosemi Active 2 acquisition system with 64 EEG channels was used.

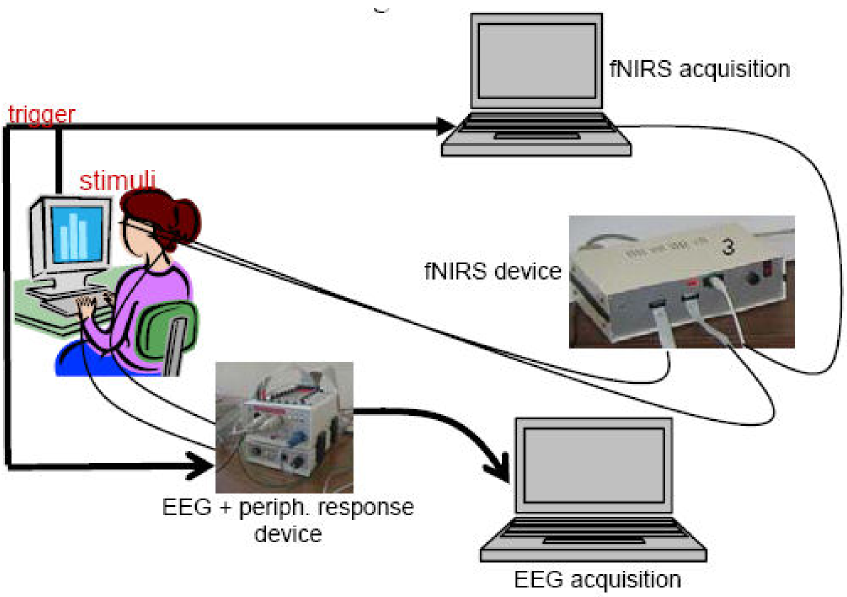


Figure 6.1: Instrumental setup in Enterface experiment

6.1.3 Emotional stimulation

To construct the EEG database, emotion had to be elicited with the participants. The researchers have chosen to use images for emotional stimulation, in particular images from the International Affective Picture System (IAPS). More information about IAPS can be found in section 7.1.2.

The subset of images used in this experiment could be divided into three different emotional classes: calm, exiting positive and exiting negative. The images in these classes were picked according to the rules in equation 6.1, where the arousal and valence scales were from 1 to 9. Particular images (for example erotic images) were removed from the selection.

$$
\begin{aligned}
calm &: \overline{arousal} < 4 \\
& \quad 4 < \overline{valence} < 6 \\
exitingpositive &: \overline{valence} > 6.8 \\
& \quad var(valence) < 2 \\
& \quad \overline{arousal} > 5 \\
exitingnegative &: \overline{valence} < 3 \\
& \quad \overline{arousal} > 5
\end{aligned}
\tag{6.1}
$$

This resulted in 106, 71 and 150 images for the calm, exiting positive and exiting negative class respectively. The distinction between the classes can be clearly seen from figure 6.2.

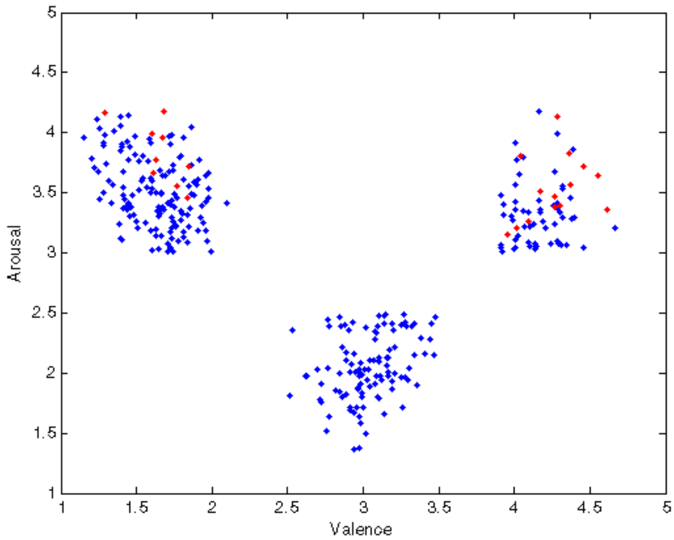


Figure 6.2: Spread of arousal and valence in selected images used by Savran et al. Red values denote stimuli with deviating self assessments

6.2 Participants

For constructing the database, recordings of EEG signals from five participants were made. All participants were male, right handed, with age ranging from 22 to 38.

6.3 Explorative analysis of the data

Before using the data, we performed some analysis on the data. Data from participant 2, session 1 was removed. This data file was not in line with expectations (no 30 markers and some consecutive markers within 1 second). Furthermore, as explained in the

Table 6.1: Samples removed from enterface data per session

Session 1	Session 2	Session 3
2, 8	2, 9, 13, 15	2, 4, 5, 25, 27, 28

documentation, the researchers have only used 54 electrodes to measure EEG data, but the database contained 73 channels with data. For that reason, the right channels were selected and the other data was discarded.

6.3.1 Bad stimuli

From a simple comparison between the self assessments and the expected values from the IAPS database, we found that some stimuli did not evoke the expected emotions. The arousal and valence from the images is shown in figure 6.2. For some of the stimuli, the participants noted that they felt really different, these stimuli are denoted in red. Apparently these stimuli were not clear enough to raise certain emotions on the participants. For that reason we did not use stimuli for which

$$\epsilon_{valence} = |\overline{selfassessment_{valence}} - E(valence)| > 1 \tag{6.2}$$

$$or$$

$$\epsilon_{arousal} = |\overline{selfassessment_{arousal}} - E(arousal)| > 1 \tag{6.3}$$

in the experiment for measuring additional data. This resulted in the removal of samples that are shown in table 6.1.

Chapter 7

Additional EEG recordings

The experiment mentioned in the previous chapter did not provide us with enough data. Since one of our requirements was that it should function properly on data from different persons, we have performed an additional experiment to gather data for testing purposes.

The goal of this experiment is to measure EEG data from participants that experience different emotions, in order to enlarge our data set of EEG measurements.

7.1 Experimental design

We want to use both our own data and the data from the Enterface project together. In order to measure comparable data, we have based our experiment on the experiment from the Enterface project. More information about the Enterface project can be found in chapter 6.

7.1.1 Experimental protocol

The participant is given information about the experiment beforehand. He has to watch the images that are shown, and remember his emotional state when he sees the images. He also has to minimize his muscle and eye movements. The exact instructions can be found in appendix C. He is asked to take a seat in the chair, and watch the screen. When the person is seated and wears the EEG cap, one test session is performed by showing one series of images and the SAM. This way, the participant will become familiar with the task he has to perform. The real experiment consists of 2 sessions of about 25 - 30 stimuli each, with a pause of about 5 minutes in between. One stimulus is a block of 5 pictures, shown on the screen for 2.5 seconds. One stimulus takes therefore 12.5 seconds to complete. Blocks of different emotions are shown in random order, to avoid the participant to become habituated. A dark screen with a white cross is shown for 5 seconds before each stimulus to attract user attention. After each block of 5 images, the self-assessment manikin is shown, and the participant can enter his emotional state. This is not time limited. The experimental protocol is shown in figure 7.1.

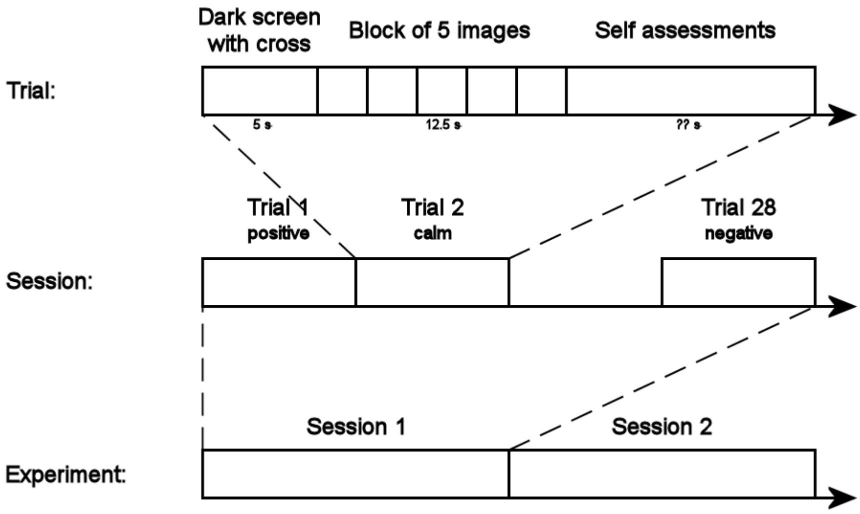


Figure 7.1: Protocol description

Instrumental setup

The EEG-cap was connected via the EEG headbox and the EEG adapter to the parallel port of a standard PC. On this PC, the Truscan Acquisition software recorded the measurements and saved it to the hard disk. The images to elicit emotion (see section 7.1.2) were shown on a second PC. The monitor of this PC is placed in front of the participant. On the same screen, a survey for self-assessment (see section 7.1.2) is shown after every stimulus. The instrumental setup is depicted in figure 7.2.

Both PCs are synchronized manually, because the TA software does not have any more sophisticated possibilities for synchronization. Every time a new series of images is shown on the screen, the supervisor hits a specific key on the PC with the TA software. When processing the data it is therefore known when the images were shown, and on what moment the emotions should be present.

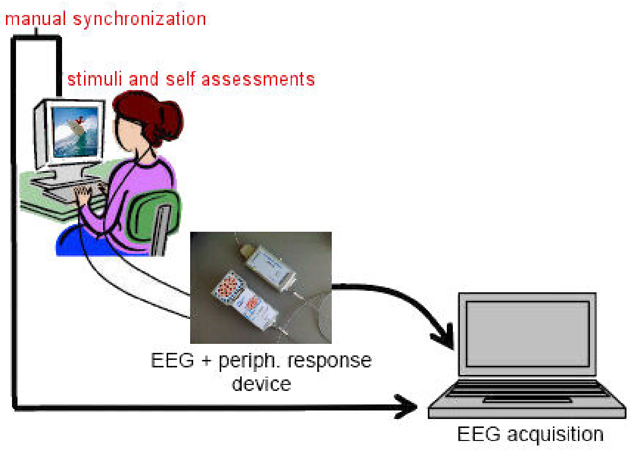


Figure 7.2: Instrumental setup

7.1.2 Emotional stimulation

For emotional stimulation we will use pictures from the IAPS (see section 3.5). The images used in this experiment are the same as in the Enterface '06 experiment [55]. However, some of the images were found not to evoke the right emotions, when comparing the IAPS annotations with the self-assessments (see section 6.3.1). These images were excluded from this experiment, resulting in 78 image series. Furthermore, the images from the last session were not used, because people would get tired and bored after viewing that much images. The full list of images used in this experiment is found in appendix B.

A digitized version of the Self Assessment Manikin (SAM, see section 3.5) was used to evaluate the emotion that was experienced by the participant during the time he saw the pictures. The version contained a 5-point scale for both the valence and the arousal dimension.

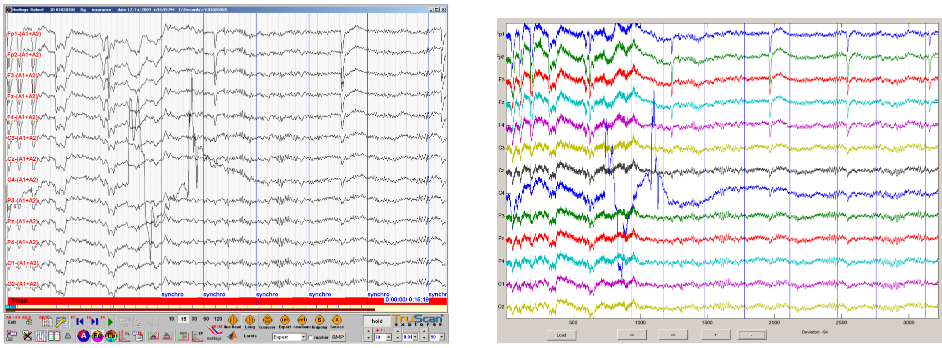
7.2 Participants

10 participants were found to voluntarily participate in the experiments, 8 male and 2 female. Most participants were students from Delft University of Technology. The participants were between 19 and 29 years old. Every participant was given information about the experiment, and was asked to fill in an informed consent, stating that they participated voluntarily and that they had enough knowledge about the experiment. This form can be found in appendix C.

Figure 7.3: Participant wearing the EEG cap

7.3 Data handling

Because the TA software stores its data in a closed format, the data is exported to ascii text files. These files can be read by matlab, and can be further processed. When exporting the data, it is bandpass filtered by the TE software in the range of 0.1 to 70Hz. When reading the exported data, the marks that were left for synchronization were not correct anymore. These marks seemed to be shifted in time for a fixed number of time points. Unfortunately, the TE software does not provide any information or documentation about this shift. For that reason, we added a number of extra marks in the beginning of each measurements. We visually inspected the signals in TE (figure 7.4(a)), where the marks are on the right spot, and the exported signals in a hand-made tool (figure 7.4(b)), which enabled us to move the marks forward and backward. Using this comparison, we were able to correct the shift. The shift appeared to be more or less constant. When using 256Hz data, the marks were shifted 222 samples forward (0.87s), and using 128Hz data, the shift was about 96 samples (0.75s).



(a) Signals in Truscan Explorer　　(b) Exported signals

Figure 7.4: Correcting the shift in markers caused by exporting the data. Figure 7.4(a) shows the original signals in Truscan Explorer. Figure 7.4(b) shows the exported signals in a hand made tool. The marks in figure 7.4(b) can be moved, and by comparison of the specific shape of the spikes in both signals, we found the shift period

7.4 Results

The results from these measurements are clear: recordings of EEG and knowledge of the emotion of the participant at the moment of the recordings. This data can be used to train and test our system.

7.4.1 User experiences

Volunteers who participated in the experiment reported their comments on the experiment. Some things that were mentioned by more than one person are:

- Negative pictures raise a more heavy emotion than positive pictures. Because of that reason, it is more likely to have a high arousal with negative valence, than with positive valence.
- Not all pictures within one stimulus evoked the same emotion. Although the pictures have the same valence and arousal values in the list from the IAPS, they are perceived differently by the participants. For example, some participants with medical background were not scared by clinical pictures with blood, but did feel negative emotions on pictures with crimes.
- Some pictures seem to be acted or are very old. These pictures do not raise real emotions with the participants, but sometimes even evoke a smile on someone's face, even though the picture is very shocking or sad.
- It is hard to fill in the self assessments consistently. Participants reported troubles with remaining consequent, when they feel approximately the same emotion. It is also difficult to compare the results between people, because some of the volunteers only used the scores 2 and 3 on the arousal scale, where others used the whole 1 to 5 scale.

Table 7.1: Samples that were removed due to mistakes

Participant	Session	Sample
2	2	13
3	2	1
5	1	21
5	2	7
9	1	23
10	2	26

These comments in fact do not compromise the results of our experiment. In our system the arousal and valence are determined independently, so it is not really a problem when some combinations occur more often. It is important that the different classes occur more or less equally often. Also, the subjectivity of the scores on the SAM reflect the subjectivity of emotion in different people. Most of the times people can not precisely describe their emotion, not even on a 1-5 scale. If we measure emotion, it is therefore not needed to have a very precise result from the system either. It is a problem though that people are inconsistent in their scoring of the same emotions, because that way, EEG samples with the same emotion are marked differently.

Another important thing to learn is that the differences between pictures in one stimulus and the old and acted pictures should be avoided. When repeating this experiment, one should therefore make a new selection of images. We would also suggest to take one image per stimulus (and showing it for 5 seconds e.g.). In the original experiment, the experimenters chose to use 5 images per stimulus to make sure the right emotion was raised, even if one or two images did not evoke that emotion. But when using the SAM to assess the participant's emotion, it does not matter if it is different between people.

7.4.2 Erroneous data

During the experiments, some mistakes were made with the marks in the TA software. With some samples, we forgot to press the right button, sometimes we were too late. Samples where these mistakes occurred were removed from the dataset. An overview of removed samples can be found in table 7.1.

Chapter 8

Explorative analysis

Now we have measured a lot of data, we will see whether the measured data meets our expectations. We will perform an explorative analysis of the data on itself, without looking at our emotion recognition system. This will be done by investigating the relationships between the self assessments, the IAPS scores as expected emotion and the EEG signals, as depicted in figure 8.1. In this analysis we only take the data into account that was measured in our own experiment (see chapter 7. The data from the Enterface experiment has already been analyzed before (see chapter 6 and the paper on the enterface project [55]).

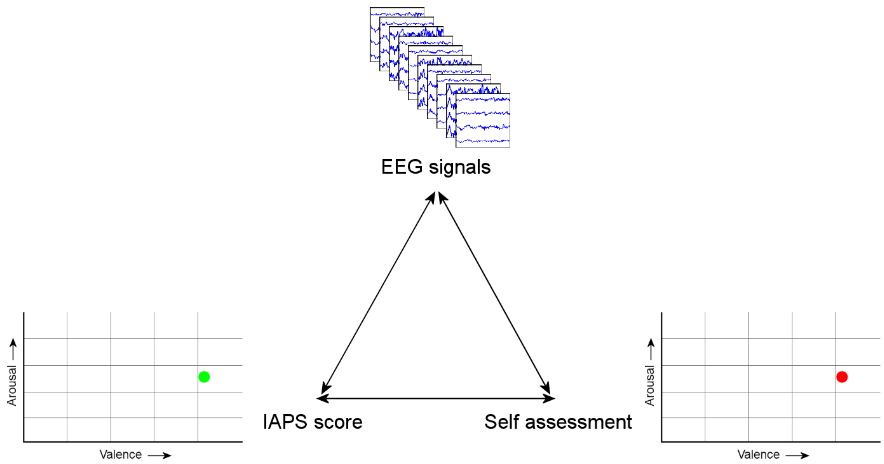


Figure 8.1: Explorative analysis of relationships between three modalities

8.1 Data preparation

The IAPS scores for the stimuli are computed as the mean value of the IAPS scores of the 5 images in that stimulus. The EEG signals are preprocessed as described in section 4.3.2. After that, features are extracted as explained in section 3.2. Features are extracted for 25 1-second windows, starting at multiples of 0.5 seconds after the stimulus (so starting at 0s, 0.5s, 1s etc.).

For the analysis of the relationship between EEG signals on the one hand and IAPS scores and self assessments on the other hand, we used a support vector machine (see section 3.4.2 for more information) using a polynomial kernel with default parameters for classification.

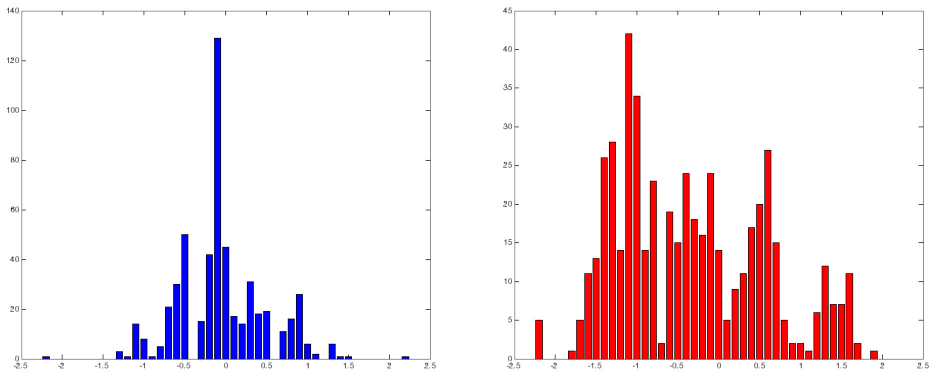
8.2 IAPS picture scores and self assessments

As mentioned in section 6.3.1, the images do not always elicit the emotion they are supposed to. This has to do with a number of reasons. The participant might have some trouble to assess his emotions when filling in the SAM. Another reason can be that the images, or series of images itself evoke other emotions than denoted in the IAPS list. To investigate this matter, we analyze the correspondence between the IAPS picture scores and the SAM scores as filled in by the participants.

8.2.1 Analysis for all participants

The Pearson correlation coefficient between the IAPS scores and SAM scores is .90 for the valence dimension and .67 for the arousal dimension. These values show that the correspondence between the expected emotion and the experienced emotion is very good, and that the images do evoke the right emotion, at least the right valence, most of the times.

This fact is also shown by the mean difference between the two scores, which is 0.40 and 0.91 for valence and arousal respectively. Figure 8.2 shows the distribution of the differences between the IAPS scores and the self assessments. The differences in the both dimensions are more or less normally distributed, as expected. However, the arousal arousal dimension is much less accurate. In the next sections we will investigate this correspondence in detail.



(a) Difference in valence dimension

(b) Difference in arousal dimension

Figure 8.2: Distribution of the differences between IAPS scores and self assessments for all participants. The difference is computed as SAM score - IAPS score

Table 8.1: Pearson correlation coefficient between IAPS scores and self assessments per participant

	Correlation Valence	Correlation Arousal
P1	0.9575	0.6306
P2	0.9199	0.8412
P3	0.9609	0.8567
P4	0.8865	0.7245
P5	0.9563	0.7918
P6	0.9315	0.7766
P7	0.9530	0.8926
P8	0.8553	0.5678
P9	0.9306	0.7189
P10	0.7246	0.5063

8.2.2 Analysis per participant

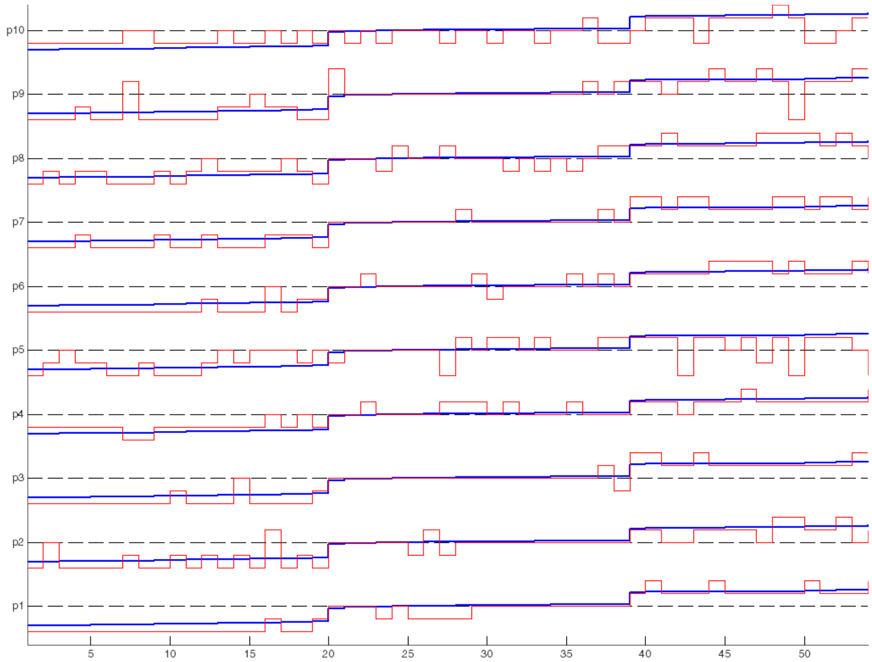
The self assessments are a very subjective measure. All SAM scores are depicted in figure 8.3. The images are on the horizontal axis, sorted by ascending valence or arousal score. The blue line shows the IAPS scores for the images, and the red lines show the self assessments per participant. The scores for the valence dimension show a very good correspondence with the IAPS values for all participants. The arousal self assessments on the other hand, show unexpected behavior with participants 8 and 10, who rated the vast majority of the pictures with low arousal. This can also be seen from the Pearson correlation coefficients in table 8.1. This table also shows a slightly lower correlation for the valence dimension for these participants.

8.2.3 Analysis per stimulus

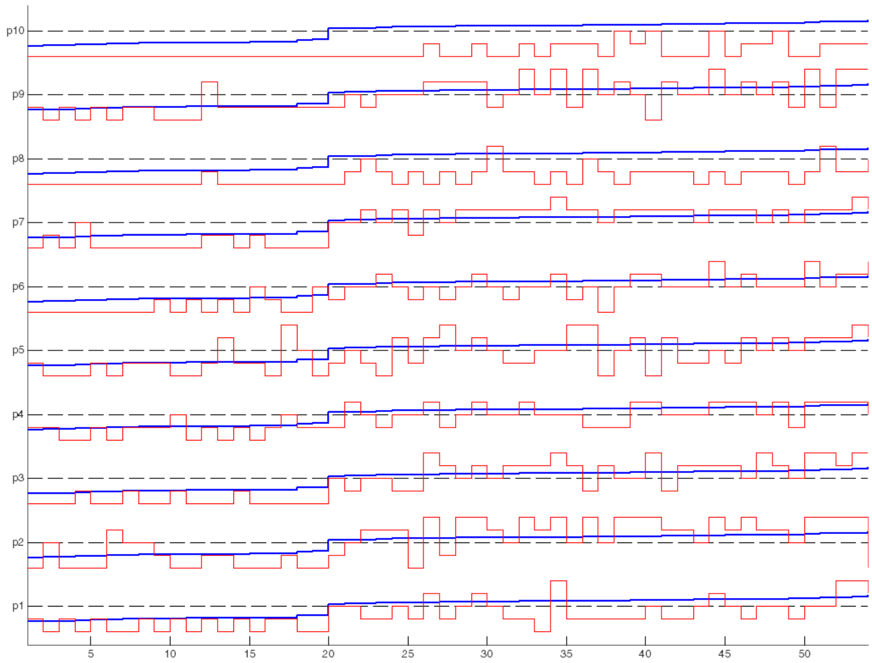
We have looked at the differences between the self assessments and the IAPS scores for the different stimuli. These differences are more or less stable between images, as can be seen in figure 8.4. There are no stimuli that obviously evoke the wrong emotions. The differences in arousal are higher than the differences in valence, but both are quite stable (standard deviation 0.14 and 0.18 respectively).

8.2.4 Analysis through time

Because of the setup of the experiment, it could be expected that participants become less concentrated during the experiment. This could introduce more errors in the self assessments scores. For that reason we investigated the difference between the errors in the first 10 stimuli and the last 10 stimuli. We found no significant difference for both dimensions ($\alpha < 0.01$).



(a) Valence



(b) Arousal

Figure 8.3: Self assessments per participant and IAPS scores. Self assessments are shown in red, the IAPS score is depicted in blue. Pictures are sorted in ascending order of IAPS score

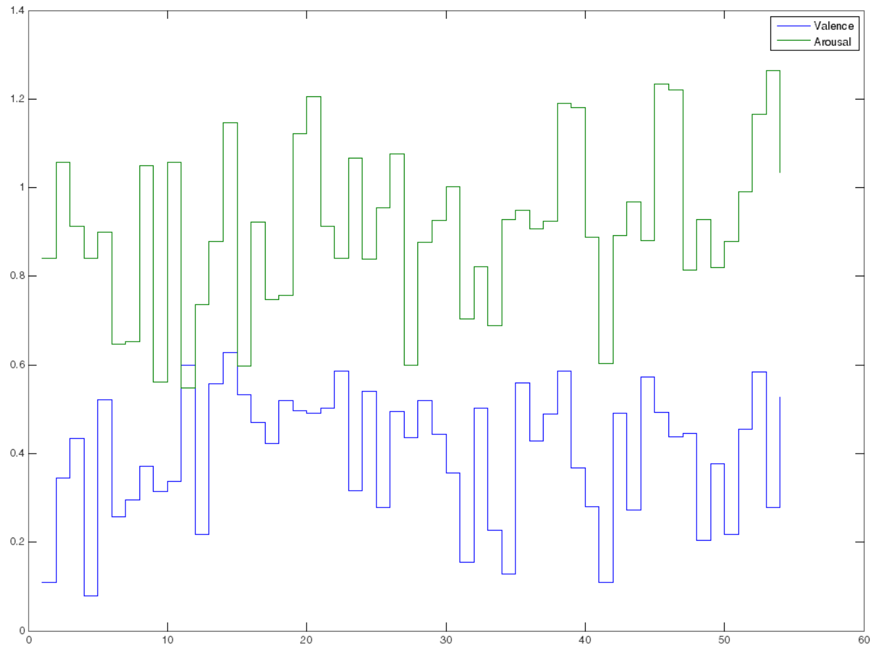


Figure 8.4: Mean differences between IAPS scores and self assessments for all stimuli used in the experiment.

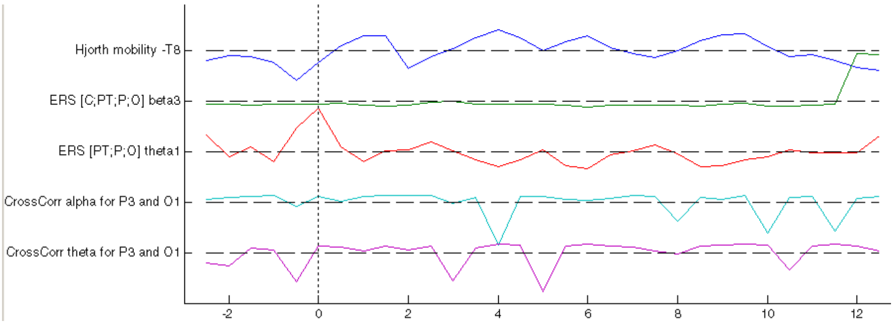
8.2.5 Results

At first glance the results look fine. The images elicited the right emotion in most cases, as seen from the self assessments. When looking at differences between participants, we found that two participants showed unexpected behavior, when looking at the self assessments on the arousal dimension.

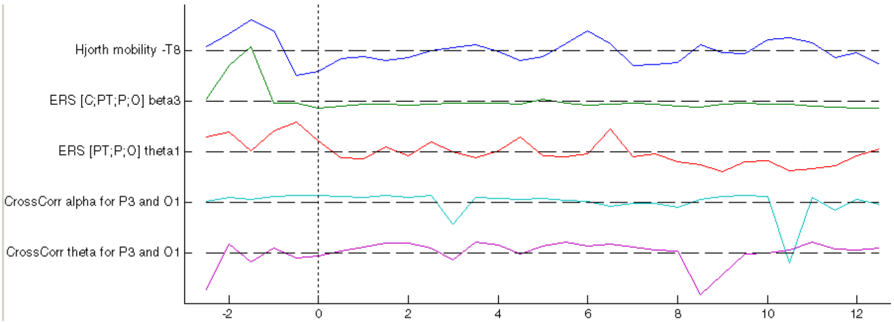
8.3 EEG data through time

We also hope too find a relationship between the EEG data and the self assessments. We have measured 12.5 seconds of EEG data per stimulus, since the images were shown on the screen for 12.5 seconds. But the emotion will not be very strong for the whole period, so we should choose the best part of the signal to train our classifier.

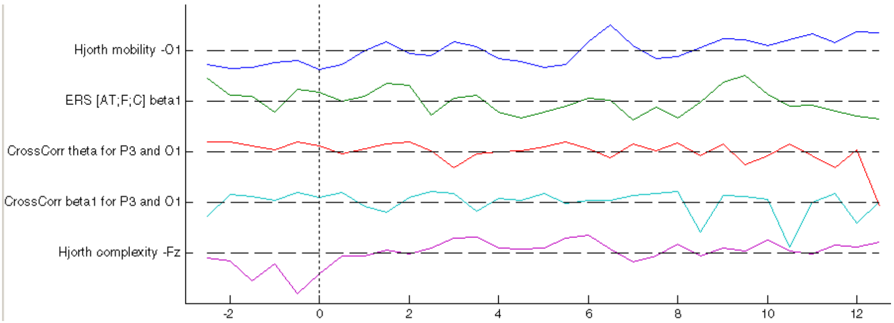
For first investigation, we looked at the values of the best features, to see whether they produce a specific pattern through time, for example low rising at the beginning, peaking after a few seconds and then declining again. Figure 8.5 shows the feature values of several features on different moments in time for some samples of participant 2. Because we expect to find at least some distinction between some extreme emotions, we compared feature values for the 5 samples with the highest and 5 samples with the lowest self assessment scores. Those feature values were averaged, to look at the overall trend. Unfortunately, we do not see any general trend in these features from visual inspection.



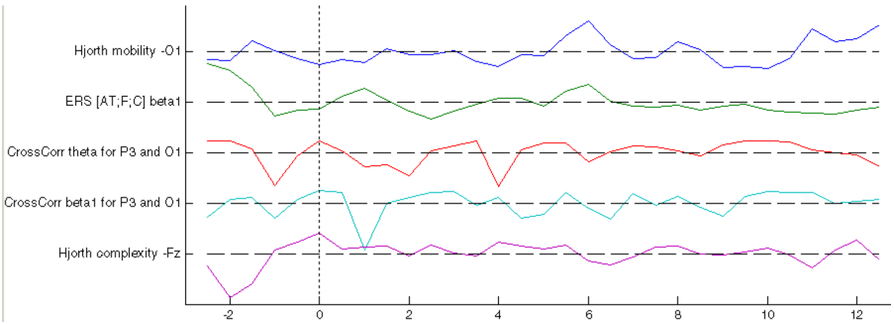
(a) Best features for high valence



(b) Best features for low valence



(c) Best features for high arousal

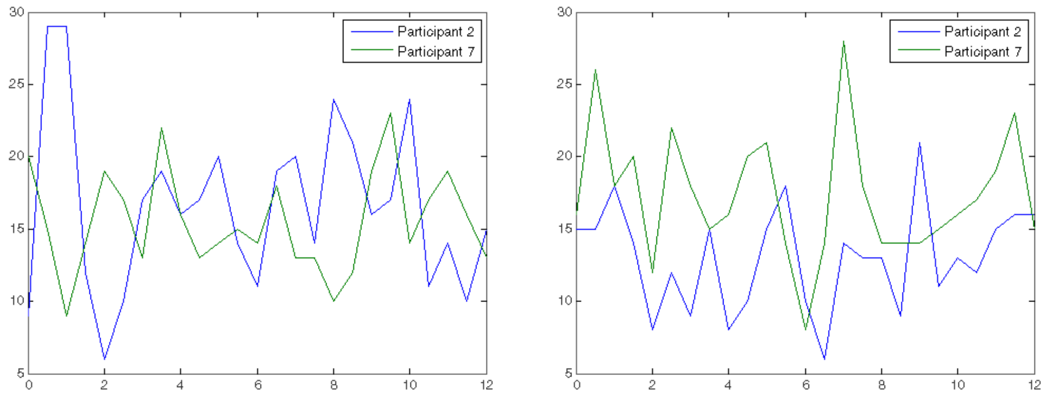


(d) Best features for low arousal

Figure 8.5: Feature values during and before the presentation of the stimulus. Best features are selected for both dimensions, and the 5 samples with the highest and lowest self assessment for valence and arousal were chosen and averaged to look at the general trend

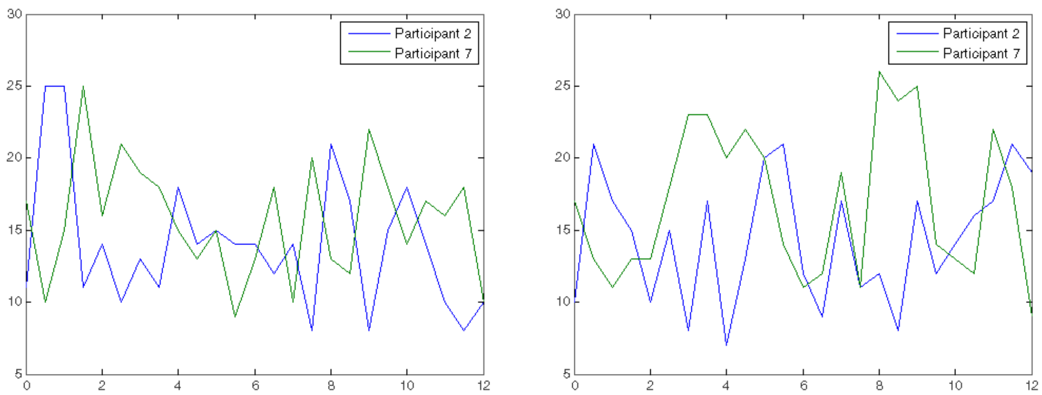
To further analyze this matter, we took the EEG data for two participants, participants 2 and 7. These participants are chosen based on the correspondence between self assessments and IAPS scores, since this correspondence shows at least the validity of the self assessments. We took for both participants and both valence and arousal dimensions the 5 stimuli with the highest and 5 with the lowest self assessments. Then we computed the correlation between the features of the EEG data for those stimuli and the self assessments, in order to find out whether the extreme values on a dimension are well separable. This procedure was repeated with the IAPS scores instead of the self assessments.

We looked at the number of features that have a high correlation (> 0.5) with the output (self assessments and IAPS scores). These are depicted in figures 8.6 and 8.7. This analysis does not show a very clear distinction between different time windows.



(a) Valence (b) Arousal

Figure 8.6: Number of features with high correlation between EEG features and self assessments, at different moments during the stimulus



(a) Valence (b) Arousal

Figure 8.7: Number of features with high correlation between EEG features and IAPS scores, at different moments during the stimulus

However, the data from the first seconds seems to perform slightly better than the rest of the sample. This could be related to the fact that participants experience the emotion heavily when the stimulus starts, but get accustomed to the feeling along the way. In further analysis we will use the the samples from 0-4 seconds, in order to train the system well.

8.4 EEG data/features and expected emotion

In order to look at the relationship between the EEG data and the self assessments, we used the same measure as in section 8.3, and now only for participant 2. Using this correlation measure, we selected the best features to correlate with arousal and valence respectively. Figure 8.8 shows the 5 samples with the highest value and the 5 samples with the lowest value for the valence and arousal dimension respectively. The samples are plotted in a graph with the first and second principal components, in order to find out whether the groups can actually be separated easily.

As seen from these pictures, these two groups of samples can not be very well separated using only two dimensions. When we train a SVM classifier on those features, we get classification rates of 65% and 58% respectively using those two best features. When using the 30 best features for training the classifier, the classification rates become 100% for both dimensions, so the data can be perfectly separated.

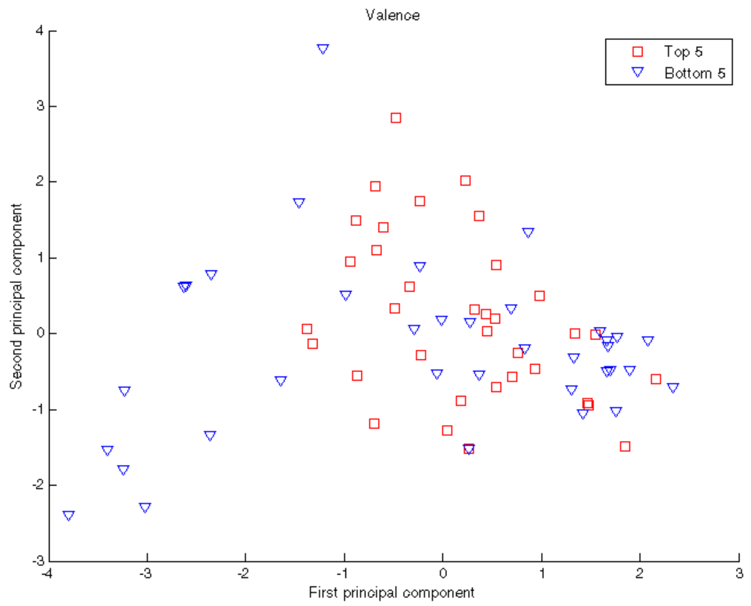
8.4.1 EEG data/features and IAPS scores

When performing the same analysis on the relationship between the EEG data and the IAPS annotations, we find that the relationship is weaker. When plotting a similar picture as in figure 8.8, we see very similar results. Classification based on the 2 best features gives classification rates of 63% and 54% respectively. However, increasing the number of used features to 30, the classification rates go up again to 100% for valence and 96% for arousal.

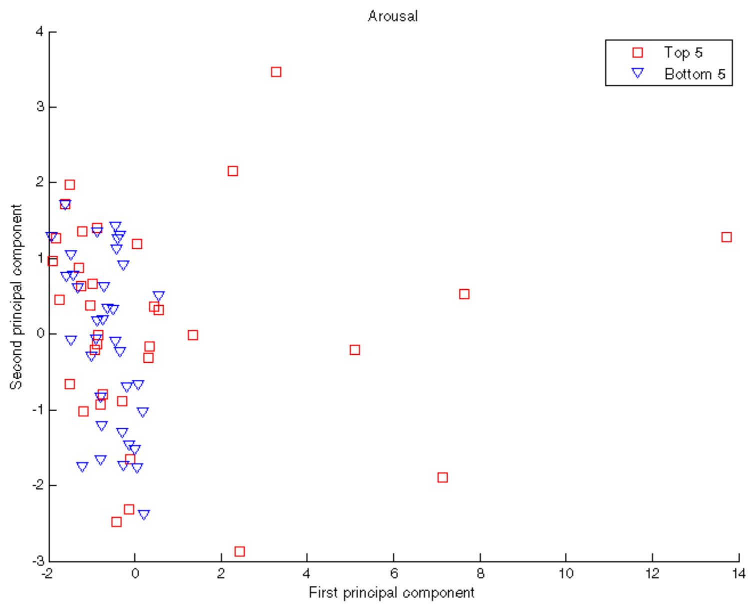
8.5 Conclusions

At first glance, the measured data seems to be quite useful. The emotions that were expected based on the IAPS scores of the image stimuli, were also found in the self assessments. This effect was found more clearly in the valence dimension than in the arousal dimension. Further analysis shows that the EEG signals contain enough information to recognize emotion from. There is a high correlation between some of the features and the expected emotion. We have taken some of the most apparent emotions (most positive or most negative) and were able to recognize those emotions from the EEG signals.

These results are promising for the rest of this project. The method of emotion elicitation meets our expectations and raises the wanted emotions. Those emotions can also be derived from the EEG signals, which is the basis of our system.



(a) Valence dimension plotted against the two best principal components



(b) Arousal dimension plotted against the two best principal components

Figure 8.8: Classification of two groups of samples with extreme self assessment values based on feature values

8.6 Training data

The data from both sources, the enterface database and our own experiment, was converted to EEGLab format, and preprocessed in the same way the system will preprocess the real data. The data was split up into parts, based on the moments when the stimulus was presented. The images to elicit some emotion were shown for a total of 12.5 seconds. Based on the analysis described in section 8.3, we assumed that the partici-

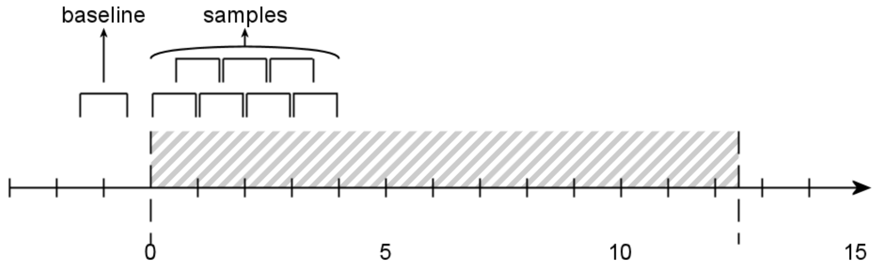


Figure 8.9: Division of measured data into training samples

pant experienced the specific emotion most clear during this first four seconds after the beginning of the stimulus. To train the system, we computed the features on data from some time windows of 1 second each, starting at 0, 0.5, 1, 1.5, 2, 2.5 and 3 seconds after the beginning of the stimulus. The time windows are chosen such that they cover all of the useful EEG data. These values were taken separately as 7 different samples, and fed into the system. Baseline values to compute the ERD and ERS were taken from the time window 1.5 to 0.5 seconds before the start of the image, assuming the participant was calm and relaxed at that moment. This scheme is shown in figure 8.9.

Chapter 9

Results

Now we have measured a lot of data, and know that the data contains information for emotion recognition, we will use specific parts of the measured data to see if the system performs well, and whether if could be improved by choosing other techniques or parameters.

When we use all our data to train our system, and use the same system afterwards to test the system, all samples are correctly classified on the valence dimension, as are 93% of the samples for the arousal dimension. This simple test shows that the different emotions can be very well separated automatically, based on EEG signals. However, this system is too much specialized in recognizing emotion from the given samples, and performs well on those samples. We are more interested in the quality of the system when it encounters samples that he has not seen before.

9.1 Setup

We want to assess the quality of the system in a structured way. We have chosen a basic setup of our system. This is our baseline measure. After that, we will change various parameters, one at a time, to see whether changing that parameters influences the results of the system.

9.1.1 System parameters

The specifications of the basic setup can be found in table 9.1. These specifications are chosen based upon simple preliminary tests. We have varied these parameters in various tests, and assessed the quality of the system with different parameters.

9.1.2 Quality measurement

The system we created will be tested using the 3-fold cross-validation method. This method divides the trainingdata into three parts. One of the parts is used for testing the classifier, where two of the parts are used for training. This process is repeated three times, every time with another part of the data. This method reduces the possibility of deviations in the results due to some special distribution of trainingdata and testdata, and ensures that the system is tested with different samples than it has seen for training.

Table 9.1: Specifications of the basic setup

Classifier type	Support Vector Machine
SVM type	C-SVC
Kernel type	Polynomial
# selected features	40
Number of classes	5 for both dimensions
Dataset	All available data
Selected trainingdata	0 - 4 seconds after the beginning of the stimulus

In order to assess the quality of the system, we need a measure for the quality of the system. The quality will be measured as the correct classification rate. That is the percentage of samples where the emotion is correctly recognized. This rate will be measured for both dimensions separately. However, the classification rate does not give all information. It is also important how bad the wrong guesses are. If a system always guesses the emotion almost right, it is better than when a system always guesses the exact opposite. For that reason, we will also use a second measure, the mean squared error (MSE). This measure squares the error made in every individual case and takes the mean of all errors.

9.2 Basic results

When the system is trained with the basic parameters on all available data, the classification rates for the valence and arousal are 31% and 32% respectively. Although the quality for both dimensions looks to be the same, the MSE is twice as high for arousal than for valence (1.5 for valence and 3.3 for arousal). These percentages are actually quite low. Statistically, when the system should randomly choose an emotional class, it would have a classification rate of 20%.

9.3 Variation in data

In our basic approach we used all available data for the training of our system. However, the data is actually not perfectly homogeneous. For that reason it might be that a subset of the data results in much better results. This is investigated in this section.

9.3.1 Using data from 1 participant

Because we have measured a lot of different participants, we expect the system to perform better or worse on data from specific persons. We have tested our basic system, trained on data from all participants, on test data from specific participants. We also trained different classifiers for different participants, which were also tested on other data from those participants. Results can be found in table 9.2.

Table 9.2: Classification rate for individual participants. The second and third column show the results with the basic classifier. The fourth and fifth column contain results with a classifier trained on the data from individual participants

	Valence	Arousal	Valence	Arousal
Participant 1	25%	20%	32%	48%
Participant 2	37%	24%	35%	27%
Participant 3	32%	28%	41%	29%
Participant 4	29%	12%	38%	34%
Participant 5	32%	22%	27%	28%
Participant 6	27%	25%	24%	37%
Participant 7	31%	25%	28%	45%
Participant 8	25%	50%	24%	56%
Participant 9	32%	13%	36%	36%
Participant 10	32%	68%	44%	69%
Participant 11	32%	33%	31%	46%
Participant 12	46%	52%	42%	49%
Participant 13	18%	26%	28%	34%
Participant 14	28%	24%	28%	28%
Participant 15	40%	70%	40%	73%

As was expected, the emotion recognition works generally better when the system is trained on data from one participant, and is also tested on data from the same participant. This is probably because the data from one participant is more homogeneous than the whole dataset with data from different participants.

Some participants (nr 8, 10 and 15) score high on the recognition of arousal in their EEG signals. This probably has to do with the fact that those participants showed unexpected behavior when filling in the self assessments (see section 8.2.2).

9.3.2 Removing noisy data

In order to improve the quality of the system, we removed the data from a number of participants from the training data, because their self assessment scores were abnormally distributed (see section 8.2). This resulted in 31% classification rate for the valence scale and 27% classification rate for the arousal scale. However, the MSE for arousal, went down from 3.3 to 1.6, so that's a large improvement.

9.3.3 Using data from one source

The data we collected was the result of two different experiments. This might have influenced the quality of the data or have introduced other details that are different. For that reason we have assessed the quality of the system for the data of both experiments separately. While the classification rate on the valence dimension was 31% for both experiments, the arousal dimension introduced a difference: 36% good classification on data from the enterface experiment, and only 29% good classification on data from our own experiment.

9.3.4 Conclusions

Based on this explorative analysis, we do not find large improvements when using only a subset of the data. The data from some participants can be classified with a better accuracy than the data from others, but these differences are not very large.

9.4 Different parameters

In this section we will look at the algorithms and parameters of the system. We will try various classifiers and parameters and compare the results with the results of our basic system.

9.4.1 Using another type of classifier

Other types of classifier are tried to improve the classification rates. A neural network and a naive bayes classifier could also give good results. When using a neural network, the classification rates were 31% and 28% respectively for valence and arousal. However, a neural network has the goal to optimize the mean squared error. However, the mean squared error is approximately 1.5 for both dimensions, which is only an improvement for thearousal dimension. A naive Bayes classifier produced slightly lower results of 29% for valence and slightly better results for arousal (35%).

9.4.2 Using other parameters for support vector machine

Support vector machines can be altered by setting several parameters. One of the most important parameters is the choice of the kernel. The kernel function takes care of the transformation of the data into higher dimensions, and the better the transformation is chosen, the better the classification will work.

Using another types of support vector machine

The LibSVM software package we use has also implemented different formulations of the SVM algorithm. Some types are used for classification, some are better for regression. We expect our data to perform better with regression, since our classes have a linear relation with eachother. In order to find the best way to classify our data, we have assessed the quality of different types and kernels, which is shown in table 9.3. Details about these parameters can be found in the LibSVM documentation [8].

Table 9.3: Classification rates for different types of support vector machines, and different kernels. Both valence and arousal results are noted.

	Polynomial	Radial Basis Function	Sigmoid
C-SVC	26.3% / 31.6%	31.1% / 25.8%	29.8% / 24.5%
nu-SVC	26.3% / 31.5%	31.6% / 26.1%	14.2% / 7.4%
epsilon-SVR	23.2% / 24.8%	32.1% / 24.1%	31.6% / 24.2%
nu-SVR		31.1% / 24.2%	31.1% / 25.4%

Table 9.4: SVM parameters that perform best

	C	γ
Valence	$2^{10.5}$	$2^{-12.5}$
Arousal	$2^{0.5}$	$2^{-14.5}$

Variation in specific parameters

There are a lot of parameters to set when running a support vector machine. As suggested by Hsu et al. [29], we have performed a search for the best parameters for a RBF kernel, by just trying all possibilities. The results are shown in figure 9.1.

It appears that the results for valence are more stable, in terms of varying parameters. The best parameters are shown in table 9.4. Using these parameters we can reach classification rates of 31,7% and 37% for both dimensions respectively.

9.4.3 Variation in the number of selected features

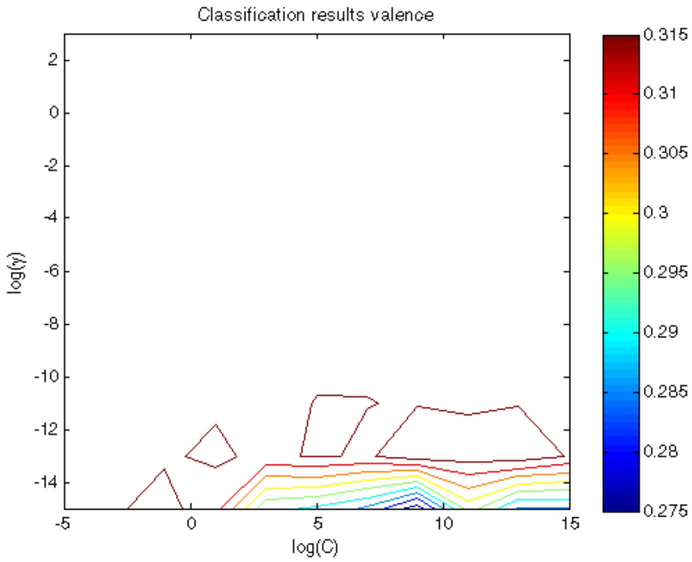
To determine the number of features to select, we ran the algorithm several times, with different numbers of selected features and looked at the classification rate with that number of features. The results are shown in figure 9.2.

The classification rate for valence seems to be slightly higher than the rates for arousal. More important, the classification rate for the valence dimension seems to be quite stable when changing the number of features. The rate for arousal seems to rise when increasing the number of features up to 30, but is decreasing again afterwards. The mean squared error is very stable at 1.5 for valence and 3.5 for arousal, for all numbers of features. The best number of features to use seems to be 30.

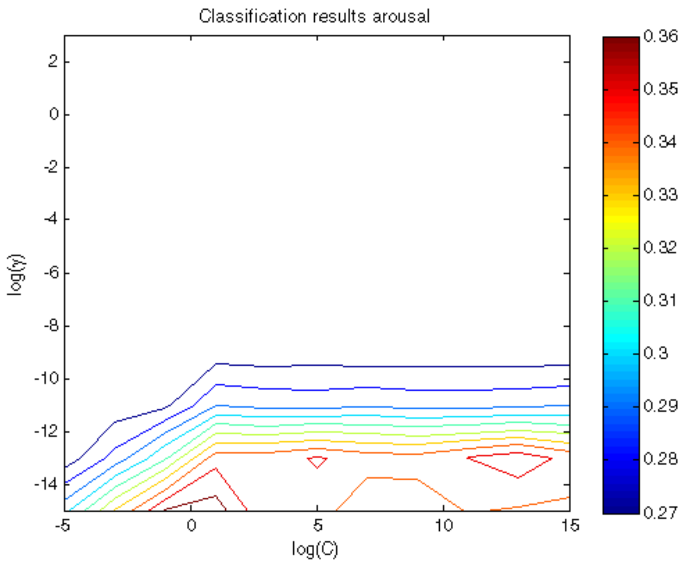
9.4.4 Using other time windows

From our preliminary tests, we took the first 4 seconds of data for each trial, to use as trainingdata. Now, we will look into this matter in detail.

We also have to find out whether the number of samples we take influences the results. It was expected that the more samples we take, the better the results will be, since this will provide the system with more training data. Figure 9.3 shows the results of this



(a) Valence



(b) Arousal

Figure 9.1: Classification results for SVM with different parameters

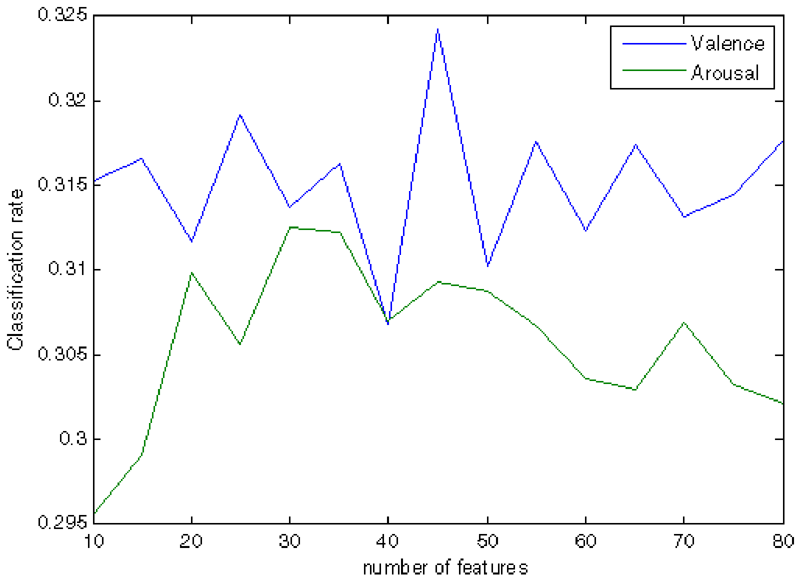


Figure 9.2: Classification rates with different numbers of features

analysis. Our expectations seem to be correct, since the higher the number of samples, the better the classification rate (overall).

This classification rate could even be improved if we only take samples from the part of the trial that contains valuable information. We have looked at the differences in classification rates when we use different parts of the trials for our training data. Figure 9.4 shows the classification rates for the cases that we use seconds 0-4 for classification, but also when we use seconds 3-7, 6-10 and 9-13. It shows that the first part (0-4s) and the middle part (6-10) seconds, give the best results. However, the differences are very small.

9.4.5 Combining data from different time windows

Until now we have always treated different parts of the trials as separate samples. This increased the number of training samples we could use, but also discarded any temporal coherence in the EEG signal. To see whether the temporal coherence could aid the system in recognizing emotion, we took the features of several samples from each trial and concatenated them. That way we got feature vectors that were twice or three times the length of the original feature vector. This procedure was done for several combinations of samples, as is shown in table 9.5. Unfortunately, all combinations resulted in classification rates of about 31% - 32% for valence and 24% - 27% for arousal, so this does not seem to be an improvement.

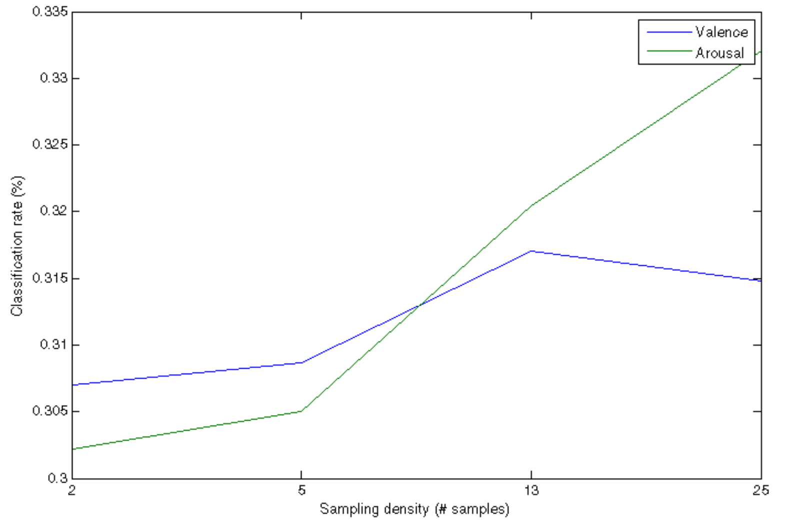


Figure 9.3: Classification rates on different sampling densities

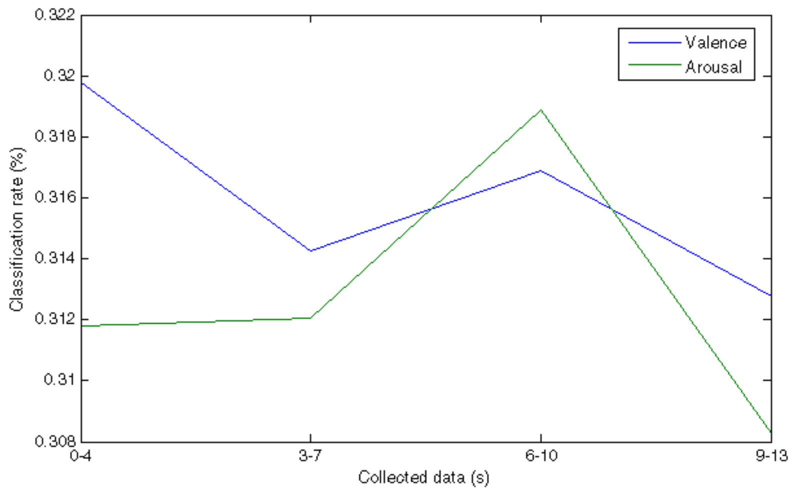


Figure 9.4: Classification rates on different parts of the data

Table 9.5: Classification rates for multiple samples combined.

	Valence	Arousal
Whole trial (12 samples)	0.3098	0.2755
Whole trial (5 samples)	0.3133	0.2566
Begin and end (2 samples)	0.2993	0.2406
First 4 seconds	0.3168	0.2336

9.5 Using different classes

In our experiment, we used a 5-point scale for the assessment of the emotions, because this scale was also used in the enterface experiment. However, when filling in the SAM, participants noted that it is difficult to choose between e.g. 1 and 2. It is however much easier to say if you feel positive, neutral or negative. This might introduce some difficulties in recognizing emotion from the EEG signals.

9.5.1 Classification into 3 classes

To see whether using three classes works better than using five classes, we brought down the self assessments to three classes. We chose two different distributions: the first took self assessments 1 & 2 as a class, 3 as a class and 4 & 5 as a class. This corresponds to the notion of positive, neutral, negative. After that we put self assessments 2, 3 and 4 together as one class, and 1 and 5 as another class, to see whether the extremes give better results.

The first class resulted in classification rates of 37% and 49% for valence and arousal respectively. That is not very much better than the results for 5 classes. The second distribution resulted in high classification rates of 72% for valence and 68% for arousal. However, this distribution resulted in too much samples in the middle class, which distorts the results.

9.5.2 Classification into 2 classes

To investigate this matter even more, we removed for both dimensions all samples with a score 2, 3 or 4. This resulted in much less samples, approximately 70% of the samples were removed, and this resulted in only the extreme values of 1 and 5 on both scales. When classifying these samples, we found classification rates of 71% for valence and even 81% for arousal. Using only two classes works apparently much better than using the whole range of data.

9.5.3 Classification using only extreme samples

Earlier results with data from participant 1 showed promising results with only the samples that have the highest and the lowest scores for both dimensions. We took for all participants the 5 samples with the highest score on valence, and the 5 samples with lowest score on valence. The output was converted to a 2-point scale. When feeding this data into the classifier and testing it, the results were 56% for both dimensions. This might look very good at first, but when you realize that there were only two classes, it is kind of disappointing.

Table 9.6: List of the 25 best selected features for both dimensions.

Valence	Arousal
Hjorth mobility -Pz	Hjorth mobility -Fz
ERS [C;PT;P;O] beta3	ERS [C;PT;P;O] beta3
ERS [AT;F] alpha2	ERS [AT;F] alpha2
ERS [PT;P;O] theta2	Alpha peak frequency F3
ERS [PT;P;O] gamma	ERS [PT;P;O] theta1
ERS [PT;P;O] theta1	ERS [PT;P;O] gamma
CrossCorr beta1 for F4 and T8	CrossCorr beta1 for P3 and O1
ERS [AT;F;C] beta1	ERS [AT;F;C] beta1
CrossCorr theta for F4 and P4	Hjorth mobility -O1
Alpha peak frequency F4	ERS [PT;P;O] theta2
CrossCorr beta1 for T7 and O1	Alpha peak frequency P8
Alpha peak frequency P7	CrossCorr theta for F3 and T8
CrossCorr theta for T7 and T8	CrossCorr beta1 for F3 and T7
CrossCorr theta for F3 and O1	CrossCorr theta for P3 and O1
Alpha peak frequency O2	Alpha peak frequency P4
CrossCorr beta1 for F3 and P4	CrossCorr beta1 for F4 and T8
CrossCorr alpha for P3 and O1	CrossCorr theta for F4 and T7
Alpha peak frequency Cz	Hjorth complexity -O2
CrossCorr alpha for F3 and T8	Alpha peak frequency P7
CrossCorr beta1 for T7 and T8	Alpha peak frequency T8
Alpha peak frequency T7	CrossCorr beta1 for F3 and F4
CrossCorr beta1 for F3 and F4	CrossCorr theta for T8 and P4
Hjorth mobility -P8	Hjorth mobility -F3
CrossCorr theta for P4 and O1	CrossCorr beta1 for T7 and T8
CrossCorr beta1 for P4 and O1	CrossCorr alpha for P4 and O1

9.6 Quality of features

One of the most important steps in recognizing emotion from EEG signals is the step of feature extraction. The extracted features determine what information is retrieved from the EEG signals, and what information is discarded.

We have extracted four different types of features: ERS, cross correlation, alpha peak frequency and Hjorth parameters (see section 3.2). A total list of all features that were extracted can be found in appendix A. Since we use a feature selection algorithm to determine the best features to use, we can assess the quality of the different types of features we used.

Table 9.6 shows the top 25 features for both dimensions. The ERS features are very useful for both dimensions, since they appear on top of the list. The other features all seem to be important. There is no clear 'best' feature type to extract.

9.7 Discussion

In this chapter we have investigated the quality of our system. Our program was intended to recognize emotion from EEG signals. As explained before, the emotion could very well be recognized from samples that the system was trained on. The system had a hard time though, generalizing its knowledge to new samples. The results for newly seen samples are much lower. About 30% - 40% correct classification was reached on both valence and arousal dimension for a 5-point scale.

We have tried several methods to improve these scores. The first thing we tried was taking only a part of the data. We saw very good scores for the data from some participants, while other participants scores were very low. This problem might be caused by the diversity in participants. Some of the diversity could be overcome, by selecting e.g. only right-handed males. Another part of the differences are caused by personal differences, like character and mood. We have done only little investigation into personal differences, for example between male and female, right- or left-handed, or age or character properties. Investigating these differences requires a much larger dataset, but could be a very useful addition to this research.

Using other types of classifiers or different parameters helped only a little bit in improving classification rates. Where different classifiers showed very similar scores on both dimensions, the time it takes to train them is different. The naive Bayes classifier is very fast, where the support vector machine and the neural networks are much slower. However, when tweaking those last two types of classifiers, the scores did improve somewhat, and that made them a better choice than a naive bayes classifier. Since the neural networks did not bring a significant improvement over a support vector machine, the latter is preferred because of its lower training time.

We have used a 5-point scale for measuring the participants emotion on both dimensions. This scale gave us the possibility to also use the data from the Enterface experiment, and gave the participants enough flexibility to rate their emotion. However, classification into 5 classes is much harder than using only three or two classes, as done by many other researchers. We have tried to convert our data into a three-point scale, but the classification rates were not very good. It is difficult to convert the data in such a way that all classes keep approximately the same number of samples. Using only the samples that were classified with a 1 or 5 on one of the dimensions, proved to be very successful. That could have been expected, because it is intuitively much easier to separate two 'extreme' classes, than a 5-point scale.

Chapter 10

Conclusions

10.1 Summary and results

During this thesis project we have focused on the construction of a program to recognize emotion from brain activity, using EEG signals. We have gathered theoretical information about the human brain and about emotion, which is described in chapter 2. Electroencephalography (EEG) seems to be the most practical way of measuring brain activity, because it is cheap and easy to use. Several researchers have shown that it is possible to measure emotional cues using EEG measurements. In this emotion research, a distinction is made between two dimensions of emotion, the valence dimension ranging from negative to positive and the arousal dimension, ranging from calm to excited. Information about both dimensions has been found to be present in EEG signals.

After gathering knowledge about the subject, we have looked at the methods to use. Several methods for signal processing, feature extraction and classification were discussed in chapter 3. We have introduced a method of emotion elicitation using images and practical information about measuring EEG signals. These methods were used in our own experiment to measure EEG signals.

A program to analyze EEG signals and recognize emotion from those signals was built using these methods. The program implements the theory that was found in literature. The system design and implementation have been discussed in chapters 4 and 5.

The program that was built was used to assess the quality of the chosen methods. In order to do that, we gathered EEG data for which we knew the emotion. We used an existing database that was created in the Enterface '06 project, which is described in chapter 6. To increase the size of the database, we have measured our own recordings (chapter 7). We have analyzed the measured data, to see whether the data met our expectations. We also did some simple tests to see whether we could separate different emotions using the EEG signals. This analysis is presented in chapter 8.

Using the data we have gathered, we tested our system. The different emotional classes could be perfectly separated, based on the EEG signals. However, the program should also work well on new data, and these results are a lot worse. The results showed 31% and 32% correct classification for the valence and arousal dimension respectively. Using other parameters, this rate could be improved to 32% and 37%. Significant improvements of these numbers were made when using the data from one participant at a time (up to 70% correct classification) or using only the extreme values for all participants on both dimensions (72% and 81%, but only for two classes).

We have found that EEG signals can be perfectly divided into emotional classes. When we group the EEG signals on their emotion, we can make a division between those groups based on the EEG signal. Unfortunately, it is hard to generalize this division to new EEG recordings. Apparently the division between different emotions is not (only) based on the emotion, but also on other properties of the signal. These properties are common for all samples used in training, but are not present in new EEG samples. When only using two extreme emotions (e.g. only positive and negative or aroused and calm), it is much easier to find a distinction, that can also be used for new EEG samples.

10.2 Contributions

In chapter 1 we formulated a number of research questions to answer during this thesis project.

1. *Is emotion recognition from* EEG *signals possible in practice?*
   We have created a program to recognize emotion from EEG signals using techniques and methods that were shown to work good on this subject. The samples for different emotions could very well be separated by our program, when using the same data for training and testing. However, on new samples the system performed much worse, around 35% correct classification rate. These results show that EEG data contains enough information to recognize emotion from it, but there is still a lot of diversity among different people and circumstances.

   In our experiment, we have used a 5-point scale for the analysis of emotion. The program appeared to work much better on recognizing the extreme values on both dimensions, but it was more difficult to recognize the intermediate states correctly. This is a promising way to go, since the distinction between the extreme emotions is more important than separating intermediate emotions. Specific details about these results can be found in chapter 9.

   (a) *How to perform* EEG *measurements?*
   During the internship in Prague, many details about measuring EEG data have become clear. This process is explained in section 3.6. It is still a demanding task to prepare EEG measurements, because the conductivity between the electrodes and the skull should be very high. To get such a conductivity, conductive gel must be applied to the participants skull, which takes a lot of time. The experiment itself should not last too long, because participants tend to get bored and emotions are not realistic anymore.

   (b) *How to process* EEG *data?*
   Raw EEG recordings have to be preprocessed before they can be used. The exact preprocessing steps depend on the specific task. Our EEG recordings had to be filtered and rereferenced, and split into slices of 1 second. From these slices, several features were extracted that described particular properties of the signal of interest for emotion recognition. The features to use were inspired by theoretical knowledge about emotion.

   (c) *How to recognize emotion from the data?*
   The features that were extracted from the EEG data provided us with informa-

tion about the emotion present in those signals. Using different classifiers, we thought the computer the relationship between the features and the emotion of a participant at that moment.

10.3 Future work

The field of emotion recognition with EEG signals is still very young and remains challenging. The results from this project show several possibilities for further work.

One of the main problems of using EEG signals is the great diversity in signals between different persons. This was shown by the fact that the system worked almost perfectly on classifying already seen samples, but had a very hard time on classifying new samples. This diversity has two main causes:

- The circumstances for the measurements change between participants. The lighting, background radiation or impedance levels may be different. Participants also get tired along the day, which may influence the recordings.
- Different participants may have different things on their mind. Even though people are triggered to experience some emotion, their brain activity may be still very much influenced by other thoughts and brain processes.

The first cause can be avoided by choosing the circumstances right. Participants can be measured all at the same moments in a shielded room. However, this solution is not always feasible. The second cause is inherent to EEG signals, and will always disturb the signals.

To overcome this diversity problem, several things can be done:

- *Measuring more data*
  The first and most straightforward thing to do is to measure more data. We assume that the one common brain activity when measuring EEG signals specific emotions is the emotion. The other brain activity or noise is actually some random activity, different for every person. The law of big numbers [61, chapter 2] tells us that the more samples we measure, the clearer we can separate the common data from the (random) noise. So when we measure enough samples, we should be able to cope with all differences and noise, given that there is some part of the EEG signal that is common for a specific emotion.
- *Other techniques for preprocessing and feature extraction*
  Another method of coping with differences and noise is using good techniques for preprocessing. Preprocessing is meant to remove most of the noise and other parts of the signal you do not need. Other techniques might be investigated to perform better on noise removal or artifact removal.

  In order to end up with only the needed information from the EEG signals, better features could also be investigated. The features used in our experiment were found in other literature, but more or better features could be investigated. From the features we used the ERS features proved to work very good, so that could be a starting point.

From our project we found that the recognition of two extreme emotions worked much better than also including intermediate states. Further research or measurements should focus simpler cases than we did in our experiment, for example only measuring two states on both dimensions or only some basic emotions. It seems to be too difficult to recognize a larger number of states on one dimension at this moment.

Another option for future research lies in using different methods of evoking emotion. We have used images to evoke some specified emotion, but it is not known whether these emotions are the same when evoked by movies or sounds. These methods could raise different brain activity with the same emotion, or result in better experiences of the expected emotion.

Overall, we conclude that emotion recognition from EEG signals is possible in practice. However, since there is only a very small dataset available, a lot of work still has to be done, by investigating better methods for EEG processing or recording more EEG data.

Appendix A

List of extracted features

Table A.1: List of extracted features

#	Description	#	Description
1	ERS [PT;P;O] theta1	4	ERS [AT;F] alpha2
2	ERS [PT;P;O] theta2	5	ERS [AT;F;C] beta1
3	ERS [PT;P;O] gamma	6	ERS [C;PT;P;O] beta3
7	CrossCorr theta - F3, F4	39	CrossCorr alpha - T7, T8
8	CrossCorr theta - F3, T7	40	CrossCorr alpha - T7, P3
9	CrossCorr theta - F3, T8	41	CrossCorr alpha - T7, P4
10	CrossCorr theta - F3, P3	42	CrossCorr alpha - T7, O1
11	CrossCorr theta - F3, P4	43	CrossCorr alpha - T8, P3
12	CrossCorr theta - F3, O1	44	CrossCorr alpha - T8, P4
13	CrossCorr theta - F4, T7	45	CrossCorr alpha - T8, O1
14	CrossCorr theta - F4, T8	46	CrossCorr alpha - P3, P4
15	CrossCorr theta - F4, P3	47	CrossCorr alpha - P3, O1
16	CrossCorr theta - F4, P4	48	CrossCorr alpha - P4, O1
17	CrossCorr theta - F4, O1	49	CrossCorr beta1 - F3, F4
18	CrossCorr theta - T7, T8	50	CrossCorr beta1 - F3, T7
19	CrossCorr theta - T7, P3	51	CrossCorr beta1 - F3, T8
20	CrossCorr theta - T7, P4	52	CrossCorr beta1 - F3, P3
21	CrossCorr theta - T7, O1	53	CrossCorr beta1 - F3, P4
22	CrossCorr theta - T8, P3	54	CrossCorr beta1 - F3, O1
23	CrossCorr theta - T8, P4	55	CrossCorr beta1 - F4, T7
24	CrossCorr theta - T8, O1	56	CrossCorr beta1 - F4, T8
25	CrossCorr theta - P3, P4	57	CrossCorr beta1 - F4, P3
26	CrossCorr theta - P3, O1	58	CrossCorr beta1 - F4, P4
27	CrossCorr theta - P4, O1	59	CrossCorr beta1 - F4, O1
28	CrossCorr alpha - F3, F4	60	CrossCorr beta1 - T7, T8
29	CrossCorr alpha - F3, T7	61	CrossCorr beta1 - T7, P3
30	CrossCorr alpha - F3, T8	62	CrossCorr beta1 - T7, P4
31	CrossCorr alpha - F3, P3	63	CrossCorr beta1 - T7, O1
32	CrossCorr alpha - F3, P4	64	CrossCorr beta1 - T8, P3
33	CrossCorr alpha - F3, O1	65	CrossCorr beta1 - T8, P4
34	CrossCorr alpha - F4, T7	66	CrossCorr beta1 - T8, O1
35	CrossCorr alpha - F4, T8	67	CrossCorr beta1 - P3, P4
36	CrossCorr alpha - F4, P3	68	CrossCorr beta1 - P3, O1

Continued on next page

Table A.1 – continued from previous page

#	Description	#	Description
37	CrossCorr alpha - F4, P4	69	CrossCorr beta1 - P4, O1
38	CrossCorr alpha - F4, O1		
70	Alpha peak frequency F3	78	Alpha peak frequency F4
71	Alpha peak frequency C3	79	Alpha peak frequency Cz
72	Alpha peak frequency T7	80	Alpha peak frequency C4
73	Alpha peak frequency P3	81	Alpha peak frequency T8
74	Alpha peak frequency P7	82	Alpha peak frequency P4
75	Alpha peak frequency O1	83	Alpha peak frequency P6
76	Alpha peak frequency Pz	84	Alpha peak frequency O2
77	Alpha peak frequency Fz		
85	Hjorth mobility - F3	100	Hjorth complexity - F3
86	Hjorth mobility - C3	101	Hjorth complexity - C3
87	Hjorth mobility - T7	102	Hjorth complexity - T7
88	Hjorth mobility - P3	103	Hjorth complexity - P3
89	Hjorth mobility - P7	104	Hjorth complexity - P7
90	Hjorth mobility - O1	105	Hjorth complexity - O1
91	Hjorth mobility - Pz	106	Hjorth complexity - Pz
92	Hjorth mobility - Fz	107	Hjorth complexity - Fz
93	Hjorth mobility - F4	108	Hjorth complexity - F4
94	Hjorth mobility - Cz	109	Hjorth complexity - Cz
95	Hjorth mobility - C4	110	Hjorth complexity - C4
96	Hjorth mobility - T8	111	Hjorth complexity - T8
97	Hjorth mobility - P4	112	Hjorth complexity - P4
98	Hjorth mobility - P6	113	Hjorth complexity - P6
99	Hjorth mobility - O2	114	Hjorth complexity - O2

Appendix B

List of IAPS images used in experiment

Table B.1: List of IAPS images used in experiment and their valence and arousal scores. Images are placed in groups of 5 similar images. Grey lines denote images that were used in the enterface project, but not in our own experiment. See section 6.3.1 for more information.

File	Val	Ar	File	Val	Ar	File	Val	Ar
7090.jpg	5.19	2.61	6570.jpg	2.19	6.24	2058.jpg	7.91	5.09
7006.jpg	4.88	2.33	3266.jpg	1.56	6.79	8090.jpg	7.02	5.71
7030.jpg	4.69	2.99	6540.jpg	2.19	6.83	8300.jpg	7.02	6.14
5530.jpg	5.38	2.87	6825.jpg	2.81	5.36	7570.jpg	6.97	5.54
9700.jpg	4.77	3.21	7359.jpg	2.92	5.36	4610.jpg	7.29	5.1
1710.jpg	8.34	5.41	6570.1.jpg	2.54	6.12	3150.jpg	2.26	6.55
7270.jpg	7.53	5.76	9425.jpg	2.67	5.92	3220.jpg	2.49	5.52
8210.jpg	7.53	5.94	9140.jpg	2.19	5.38	2053.jpg	2.47	5.25
8350.jpg	7.18	5.18	6230.jpg	2.37	7.35	6370.jpg	2.7	6.44
2160.jpg	7.58	5.16	3400.jpg	2.35	6.91	6360.jpg	2.23	6.33
7053.jpg	5.22	2.95	1616.jpg	5.21	3.95	5740.jpg	5.21	2.59
2580.jpg	5.71	2.79	7058.jpg	5.29	3.98	7050.jpg	4.93	2.75
2221.jpg	4.39	3.07	7002.jpg	4.97	3.16	7150.jpg	4.72	2.61
7010.jpg	4.94	1.76	7004.jpg	5.04	2	7180.jpg	4.73	3.43
7235.jpg	4.96	2.83	5510.jpg	5.15	2.82	7546.jpg	5.4	3.72
3191.jpg	1.95	5.95	4599.jpg	7.12	5.69	2208.jpg	7.35	5.68
6243.jpg	2.33	5.99	8540.jpg	7.48	5.16	5270.jpg	7.26	5.49
9252.jpg	1.98	6.64	2346.jpg	7.05	5.28	5470.jpg	7.35	6.02
6550.jpg	2.73	7.09	5700.jpg	7.61	5.68	8496.jpg	7.58	5.79
3063.jpg	1.49	6.35	5623.jpg	7.19	5.67	8371.jpg	6.82	5.12
8501.jpg	7.91	6.44	8185.jpg	7.57	7.27	8470.jpg	7.74	6.14
5629.jpg	7.03	6.55	7502.jpg	7.75	5.91	8080.jpg	7.73	6.65
5260.jpg	7.34	5.71	4601.jpg	6.82	5.08	5600.jpg	7.57	5.19
8300.jpg	7.02	6.14	8503.jpg	7.02	5.22	7330.jpg	7.69	5.14
7220.jpg	6.91	5.3	8180.jpg	7.12	6.59	8190.jpg	8.1	6.28
9800.jpg	2.04	6.05	8311.jpg	5.88	3.57	6315.jpg	2.31	6.38
6571.jpg	2.85	5.59	7056.jpg	5.07	3.07	6821.jpg	2.38	6.29
9571.jpg	1.96	5.64	7020.jpg	4.97	2.17	3130.jpg	1.58	6.97

Continued on next page

Table B.1 – continued from previous page

File	Val	Ar	File	Val	Ar	File	Val	Ar
6250.jpg	2.83	6.54	7140.jpg	5.5	2.92	6260.jpg	2.44	6.93
3530.jpg	1.8	6.82	7025.jpg	4.63	2.71	6021.jpg	2.21	6.06
7052.jpg	5.33	3.01	9250.jpg	2.57	6.6	7035.jpg	4.98	2.66
6150.jpg	5.08	3.22	9429.jpg	2.68	5.63	7161.jpg	4.98	2.98
7710.jpg	5.42	3.44	9530.jpg	2.93	5.2	7234.jpg	4.23	2.96
7179.jpg	5.06	2.88	9253.jpg	2	5.53	7495.jpg	5.9	3.82
5500.jpg	5.42	3	2688.jpg	2.73	5.98	2206.jpg	4.06	3.71
8400.jpg	7.09	6.61	8485.jpg	2.73	6.46	3015.jpg	1.52	5.9
8170.jpg	7.63	6.12	9180.jpg	2.99	5.02	3181.jpg	2.3	5.06
8185.jpg	7.57	7.27	3102.jpg	1.4	6.58	3061.jpg	2.32	5.28
4641.jpg	7.2	5.43	3068.jpg	1.8	6.77	2683.jpg	2.62	6.21
8503.jpg	7.02	5.22	3100.jpg	1.6	6.49	3060.jpg	1.79	7.12
3016.jpg	1.9	5.82	8420.jpg	7.76	5.56	7547.jpg	5.21	3.18
6830.jpg	2.82	6.21	8502.jpg	7.51	5.78	5471.jpg	5.21	3.26
2095.jpg	1.79	5.25	8501.jpg	7.91	6.44	7100.jpg	5.24	2.89
3301.jpg	1.8	5.21	5450.jpg	7.01	5.84	8465.jpg	5.96	3.93
3101.jpg	1.91	5.6	5480.jpg	7.53	5.48	5130.jpg	4.45	2.51
9401.jpg	4.53	3.88	8380.jpg	7.56	5.74	7096.jpg	5.54	3.98
7025.jpg	4.63	2.71	8200.jpg	7.54	6.35	9360.jpg	4.03	2.63
7050.jpg	4.93	2.75	4640.jpg	7.18	5.52	7705.jpg	4.77	2.65
9360.jpg	4.03	2.63	2216.jpg	7.57	5.83	7590.jpg	4.75	3.8
2880.jpg	5.18	2.96	8030.jpg	7.33	7.35	9210.jpg	4.53	3.08
6250.1.jpg	2.63	6.92	6312.jpg	2.48	6.37	3030.jpg	1.91	6.76
3080.jpg	1.48	7.22	2981.jpg	2.76	5.97	9900.jpg	2.46	5.58
7380.jpg	2.46	5.88	6834.jpg	2.91	6.28	9254.jpg	2.03	6.04
3064.jpg	1.45	6.41	2800.jpg	1.78	5.49	9902.jpg	2.33	6
3051.jpg	2.3	5.62	9419.jpg	2.55	5.19	9630.jpg	2.96	6.06
2346.jpg	7.05	5.28	8496.jpg	7.58	5.79	9424.jpg	2.87	5.78
7508.jpg	7.02	5.09	8370.jpg	7.77	6.73	9611.jpg	2.71	5.75
5833.jpg	8.22	5.71	8340.jpg	6.85	5.8	2717.jpg	2.58	5.7
8531.jpg	7.03	5.41	8034.jpg	7.06	6.3	9007.jpg	2.49	5.03
4601.jpg	6.82	5.08	4623.jpg	7.13	5.44	6022.jpg	2.14	6.09
2209.jpg	7.64	5.59	8090.jpg	7.02	5.71	7059.jpg	4.93	2.73
8116.jpg	6.82	5.97	5910.jpg	7.8	5.59	7038.jpg	4.82	3.01
8470.jpg	7.74	6.14	1710.jpg	8.34	5.41	5731.jpg	5.39	2.74
1811.jpg	7.62	5.12	7260.jpg	7.21	5.11	2221.jpg	4.39	3.07
4624.jpg	6.84	5.02	5470.jpg	7.35	6.02	7233.jpg	5.09	2.77
7547.jpg	5.21	3.18	7055.jpg	4.9	3.02	1710.jpg	8.34	5.41
6570.2.jpg	4.86	3.85	7010.jpg	4.94	1.76	8170.jpg	7.63	6.12
7140.jpg	5.5	2.92	7040.jpg	4.69	2.69	2160.jpg	7.58	5.16
7035.jpg	4.98	2.66	7060.jpg	4.43	2.55	2352.1.jpg	7.27	5.16
7100.jpg	5.24	2.89	7057.jpg	5.35	3.39	2345.jpg	7.41	5.42
9421.jpg	2.21	5.04	7710.jpg	5.42	3.44	7270.jpg	7.53	5.76
9810.jpg	2.09	6.62	2445.jpg	5.39	3.83	8540.jpg	7.48	5.16
2730.jpg	2.45	6.8	2446.jpg	4.7	3.79	7501.jpg	6.85	5.63

Continued on next page

Table B.1 – continued from previous page

File	Val	Ar	File	Val	Ar	File	Val	Ar
3230.jpg	2.02	5.41	7190.jpg	5.55	3.84	8420.jpg	7.76	5.56
9428.jpg	2.31	5.66	7034.jpg	4.95	3.06	7502.jpg	7.75	5.91
7057.jpg	5.35	3.39	8400.jpg	7.09	6.61	2880.jpg	5.18	2.96
9210.jpg	4.53	3.08	2058.jpg	7.91	5.09	2890.jpg	4.95	2.95
7037.jpg	4.81	3.71	7330.jpg	7.69	5.14	7235.jpg	4.96	2.83
7031.jpg	4.52	2.03	1811.jpg	7.62	5.12	7285.jpg	5.67	3.83
5471.jpg	5.21	3.26	5660.jpg	7.27	5.07	6150.jpg	5.08	3.22
8500.jpg	6.96	5.6	7179.jpg	5.06	2.88	9430.jpg	2.63	5.26
5623.jpg	7.19	5.67	7036.jpg	4.88	3.32	2703.jpg	1.91	5.78
7400.jpg	7	5.06	5390.jpg	5.59	2.88	9006.jpg	2.34	5.76
5480.jpg	7.53	5.48	7052.jpg	5.33	3.01	3550.1.jpg	2.35	6.29
2058.jpg	7.91	5.09	7044.jpg	4.69	3.94	6510.jpg	2.46	6.96
7236.jpg	5.64	3.79	7700.jpg	4.25	2.95	8340.jpg	6.85	5.8
7180.jpg	4.73	3.43	1670.jpg	5.82	3.33	5621.jpg	7.57	6.99
7192.jpg	5.77	3.58	7009.jpg	4.93	3.01	8370.jpg	7.77	6.73
7130.jpg	4.77	3.35	7207.jpg	5.15	3.57	8186.jpg	7.01	6.84
7058.jpg	5.29	3.98	7175.jpg	4.87	1.72	5460.jpg	7.33	5.87
7020.jpg	4.97	2.17	9925.jpg	2.84	5.59	3069.jpg	1.7	7.03
7500.jpg	5.33	3.26	3350.jpg	1.88	5.72	9433.jpg	1.84	5.89
7207.jpg	5.15	3.57	3160.jpg	2.63	5.35	6213.jpg	2.91	5.86
7224.jpg	4.45	2.81	9500.jpg	2.42	5.82	6300.jpg	2.59	6.61
7590.jpg	4.75	3.8	9400.jpg	2.5	5.99	9050.jpg	2.43	6.36
2216.jpg	7.57	5.83	9570.jpg	1.68	6.14	7205.jpg	5.56	2.93
7260.jpg	7.21	5.11	6350.jpg	1.9	7.29	6570.2.jpg	4.86	3.85
7570.jpg	6.97	5.54	9635.1.jpg	1.9	6.54	7041.jpg	4.99	2.6
7501.jpg	6.85	5.63	3000.jpg	1.59	7.34	5520.jpg	5.33	2.95
8186.jpg	7.01	6.84	4664.2.jpg	2.79	6.13	7031.jpg	4.52	2.03
5533.jpg	5.31	3.12	7950.jpg	4.94	2.28	5260.jpg	7.34	5.71
7183.jpg	5.58	3.78	7550.jpg	5.27	3.95	8200.jpg	7.54	6.35
5740.jpg	5.21	2.59	2840.jpg	4.91	2.43	8030.jpg	7.33	7.35
7110.jpg	4.55	2.27	2749.jpg	5.04	3.76	8501.jpg	7.91	6.44
7285.jpg	5.67	3.83	7043.jpg	5.17	3.68	5629.jpg	7.03	6.55
3140.jpg	1.83	6.36	9560.jpg	2.12	5.5	2580.jpg	5.71	2.79
2751.jpg	2.67	5.18	3500.jpg	2.21	6.99	5534.jpg	4.84	3.14
9301.jpg	2.26	5.28	9340.jpg	2.41	5.16	2980.jpg	5.61	3.09
9910.jpg	2.06	6.2	9901.jpg	2.27	5.7	7110.jpg	4.55	2.27
9423.jpg	2.61	5.66	6200.jpg	3.2	5.82	7130.jpg	4.77	3.35
8080.jpg	7.73	6.65	7508.jpg	7.02	5.09	7283.jpg	5.5	3.81
5600.jpg	7.57	5.19	2208.jpg	7.35	5.68	7224.jpg	4.45	2.81
5910.jpg	7.8	5.59	8116.jpg	6.82	5.97	7491.jpg	4.82	2.39
8420.jpg	7.76	5.56	4641.jpg	7.2	5.43	9700.jpg	4.77	3.21
8499.jpg	7.63	6.07	5833.jpg	8.22	5.71	7500.jpg	5.33	3.26
2811.jpg	2.17	6.9	7160.jpg	5.02	3.07	6530.jpg	2.76	6.18
9300.jpg	2.26	6	7053.jpg	5.22	2.95	9181.jpg	2.26	5.39
3010.jpg	1.79	7.26	5531.jpg	5.15	3.69	9427.jpg	2.89	5.5

Continued on next page

Table B.1 – continued from previous page

File	Val	Ar	File	Val	Ar	File	Val	Ar
2900.jpg	2.45	5.09	7039.jpg	5.93	3.29	9921.jpg	2.04	6.52
3215.jpg	2.51	5.44	7217.jpg	4.82	2.43	3000.jpg	1.59	7.34
6415.jpg	2.21	6.2	7400.jpg	7	5.06	8380.jpg	7.56	5.74
9903.jpg	2.36	5.71	8499.jpg	7.63	6.07	8180.jpg	7.12	6.59
2710.jpg	2.52	5.46	8210.jpg	7.53	5.94	4599.jpg	7.12	5.69
8230.jpg	2.95	5.91	2345.jpg	7.41	5.42	5910.jpg	7.8	5.59
2352.2.jpg	2.09	6.25	8190.jpg	8.1	6.28	2209.jpg	7.64	5.59
7283.jpg	5.5	3.81	7242.jpg	5.28	3.83	2745.1.jpg	5.31	3.26
2745.1.jpg	5.31	3.26	2518.jpg	5.67	3.31	5532.jpg	5.19	3.79
7002.jpg	4.97	3.16	5120.jpg	4.39	3.07	9401.jpg	4.53	3.88
2980.jpg	5.61	3.09	7192.jpg	5.77	3.58	7236.jpg	5.64	3.79
7950.jpg	4.94	2.28	7595.jpg	4.55	3.77	7490.jpg	5.52	2.42
8190.jpg	8.1	6.28	6210.jpg	2.95	6.34	4640.jpg	7.18	5.52
5450.jpg	7.01	5.84	6242.jpg	2.69	5.43	4626.jpg	7.6	5.78
8371.jpg	6.82	5.12	9410.jpg	1.51	7.07	8531.jpg	7.03	5.41
4626.jpg	7.6	5.78	3180.jpg	1.92	5.77	7220.jpg	6.91	5.3
5700.jpg	7.61	5.68	3550.jpg	2.54	5.92	5833.jpg	8.22	5.71
4610.jpg	7.29	5.1	3062.jpg	1.87	5.78	4623.jpg	7.13	5.44
5460.jpg	7.33	5.87	3005.1.jpg	1.63	6.2	8500.jpg	6.96	5.6
8034.jpg	7.06	6.3	3017.jpg	2.45	5.34	8502.jpg	7.51	5.78
2352.1.jpg	7.27	5.16	3225.jpg	1.82	5.95	7230.jpg	7.38	5.52
8370.jpg	7.77	6.73	9920.jpg	2.5	5.76	5660.jpg	7.27	5.07
3110.jpg	1.79	6.7	4624.jpg	6.84	5.02	9911.jpg	2.3	5.76
9420.jpg	2.31	5.69	7230.jpg	7.38	5.52	3261.jpg	1.82	5.75
9405.jpg	1.83	6.08	8350.jpg	7.18	5.18	9620.jpg	2.7	6.11
3071.jpg	1.88	6.86	5270.jpg	7.26	5.49	9600.jpg	2.48	6.46
6313.jpg	1.98	6.94	5621.jpg	7.57	6.99	6838.jpg	2.45	5.8
6212.jpg	2.19	6.01	7080.jpg	5.27	2.32	3168.jpg	1.56	6
2799.jpg	2.42	5.02	7170.jpg	5.14	3.21	6560.jpg	2.16	6.53
3120.jpg	1.56	6.84	7006.jpg	4.88	2.33	9040.jpg	1.67	5.82
9520.jpg	2.46	5.41	7090.jpg	5.19	2.61	6831.jpg	2.59	5.55
3053.jpg	1.31	6.91	7000.jpg	5	2.42	3010.jpg	1.79	7.26

Appendix C

Experiment documents

C.1 Instructions

Thank you for coming today to participate in this experiment. This experiment is designed for gathering measurements of brain activity with different emotions. The brain activity will be measured using a EEG-cap. This cap measures the activity at 19 different positions on your scalp. This activity reflects the brain activity of that part of the brain. In figure C.1 you see an example of EEG measurements. Even though the cap might look scary, EEG measurements are harmless.

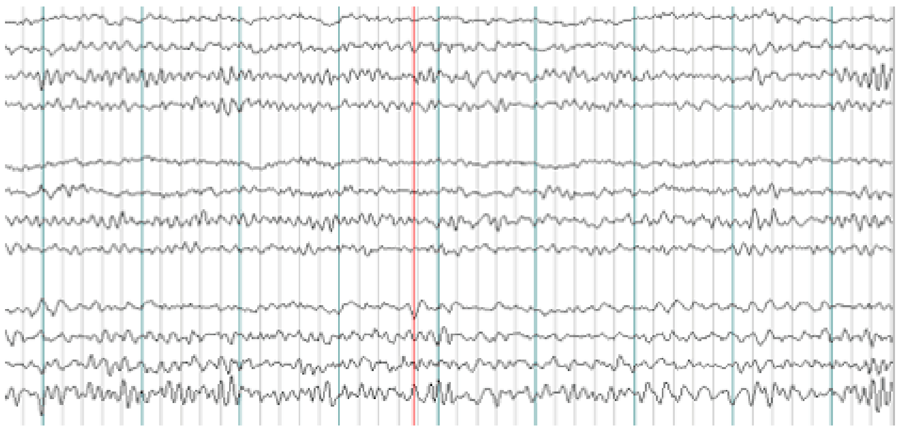


Figure C.1: Example of EEG measurements

You will be looking at several pictures, with specific emotional content, in order to evoke different emotions. These pictures are divided in groups of 5 pictures, with all pictures in one group having more or less the same emotional value. The pictures that will be shown could raise strong emotions, as they contain heavy content.

After you have seen one group of pictures, you are asked to rate your feeling while viewing the picture. These scores will not be graded as being right or wrong, but will serve as a measure for your emotion while viewing the pictures. Simply answer as honestly as you can.

C.1.1 Detailed outline

Every trial is divided in three parts. First you will see a white cross, meaning that the next sample will start soon. After three seconds, the pictures will be shown, without any further notice. Every picture will be shown for 2.5 seconds, for a total of 12.5 seconds.

Try to take a comfortable position in the chair. You are asked not to move during the viewing of the pictures. Also try to minimize eye movements or blinking. Every movement, even very small movements, influence the EEG data heavily, and it is a difficult task to clean the EEG measurements from these artifacts.

After the 5 pictures are shown, you will see a so-called self assessment manikin (SAM, see figure C.2). On this form you can fill in how you felt during the showing of the pictures.

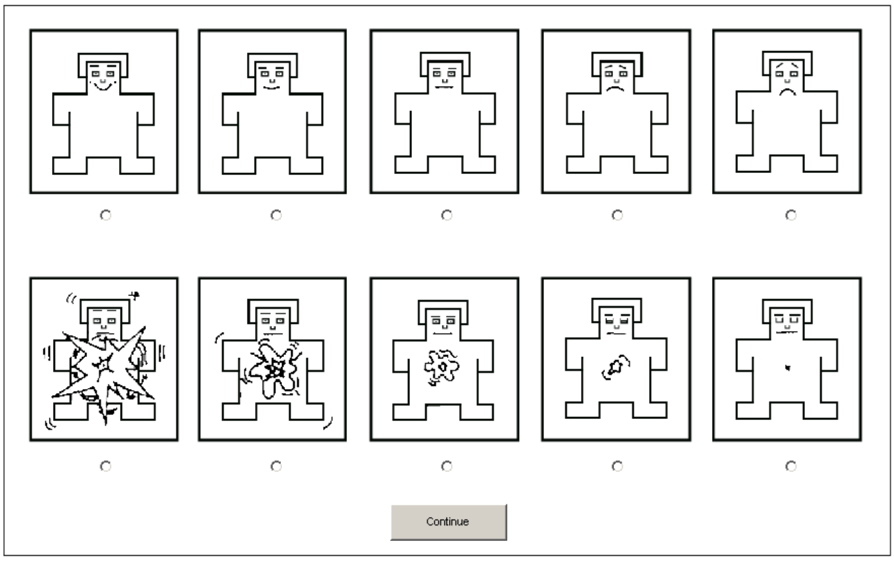


Figure C.2: Self assessment manikin

You see two sets of 5 pictures, both showing one kind of feelings.

The first set shows happy vs. unhappy, which ranges from a smile to a frown. At one extreme of the happy vs. unhappy scale, you felt happy, pleased, satisfied, contented, hopeful. If you felt completely happy while viewing the pictures, you can indicate this by checking the button under the figure at the left. The other end of the scale is when you felt completely, unhappy, annoyed, unsatisfied, melancholic, despaired, bored. You can indicate feeling completely unhappy by checking the box on the figure at the right. The figures also allow you to describe intermediate feelings of pleasure, checking the box over any of the other pictures. This permits you to make more finely graded ratings of how you feel in reaction to the pictures.

The second set shows the excited vs. calm dimension. At one extreme of the scale you felt stimulated, excited, frenzied, jittery, wide-awake, aroused. If you felt completely aroused while viewing the pictures, check the box under the figure at the left of the row. On the other hand, at the other end of the scale, you felt completely relaxed, calm, sluggish, dull, sleepy, unaroused. You can indicate you felt completely calm by checking the box at the right of the row. As with the happy-unhappy scale, you can represent intermediate levels by checking the box under any of the other figures.

Some pictures may evoke heavy emotions, others may seem to be relatively neutral. There is no right or wrong, so please rate each set of pictures as you actually felt while you watched the pictures.

When you are ready filling in the form, you can click ‘Continue’. The next set of pictures will start.

C.2 Informed consent

Title	Measuring brain activity on emotionally stimulated people
Purpose	Gathering EEG data on people that feel a specific emotion. This data is used to create a computer program to recognize emotion from EEG signals.
Description	This experiment will take about 1.5 hours. First the EEG cap will be prepared. This involves puttig the EEG cap on your head and decreasing the impedance on the electrodes using special conductive gel. After that, you will be shown 50 trials of pictures with emotional content. After each trial you will fill in a self assessment form, to assess your emotion. This will take 15.5 seconds for each trial plus the time needed to fill in the self assessments.
Discomforts	For decreasing the impedance on the electrodes, we will have to use some conductive gel. This gel won't do you any harm, and can be easily removed using water and soap afterwards. The pictures that will be shown may contain heavy content and raise strong emotions.
Data use	The data gathered with this experiment will be used for creating the emotion recognition program. It might also be used for further research on this topic.
	You know you can refuse to participate or withdraw at any time for any reason.
	For further questions you can reach me at r.horlings@student.tudelft.nl.

Name	....
Date	....
Signature	....

Appendix D

Measuring workload in a driving simulator using EEG

D.1 Introduction

There are a lot of devices available in cars, that are actually not needed to drive a car, like navigation systems, mobile phones, etc. Most of the times the goal of these devices is to aid the driver, or provide him with extra services or comfort. These devices have an influence on the drivers attention to the road. The more difficult it is to handle or control a device, the more attention a driver has to pay to the device. And that means, the driver can pay less attention to the road, which is generally bad for the driving skills, and might even be dangerous.

The driving simulator is often used for measuring the influence of a secondary task on the driving skills. Because it is not very easy to measure the attention to the road, or the driving skills, mostly the lane deviation and speed deviation are used as a measure for the level of driving skills. When these measures are compared, for the case the driver has no secondary task, and the case that the driver controls or pays attention to a device, you can deduce a measure for the level of distraction of such a device. On of the main goals in designing in-car devices should be the minimization of this distraction.

Using the driving performance as a measure for the workload of a device, might not be the best way to go. The driving performance is a highly subjective measure, and can is influenced by many factors, such as fatigue, emotion, things in mind, etc.

In order to find another measure, the use of electroencephalograms (EEG) is suggested, in particular event related potentials (ERPs). An ERP can be described as a response from your brain on an external event, for example a traffic light turning red, or a drop of rain falling on your arm. EEG signals come directly from the brain, and one can not easily manipulate these signals. For that reason, the EEG signals and ERPs could provide a more stable measure for the workload on a person.

Several papers describe a correlation between ERPs and mental task workload [39, 52, 51]. As described in [39], the amplitude of n2 and p3 ERPs can be used as a predictor for the workload of a task. [51] shows that the P300 belonging to the primary task and the one belonging to the secondary task, show a reciprocal relationship, with respect to their amplitudes. It also states that the latency of the P300 is longer, with higher workload.

D.2 Experiments

I have done some analysis on the existing EEG data from earlier experiments. After finishing this analysis, we have done two specially designed experiments, to see whether the ERPs could be used for workload prediction. Because we have no direct measure of the workload on a subject, we have taken the time it takes the subject to respond to a stimulus (reaction time) as an indirect measure.

D.2.1 Finding ERPs

In this first experiments, I used data from another experiment, where drivers had to drive some track, and enter a city name into the navigation system when told to. The given command was taken as the stimuli, so the ERPs in this experiment are the response to this command. I examined 24 subjects, all doing this same experiment.

The first thing I looked at is the appearance of the ERPs. Visual inspection supports the hypothesis that the ERP response can be observed in the EEG data. Figure D.1 shows an example of the measurements. All the samples are aligned on the time of the stimulus (t = 0) and averaged per channel. Every line in the graph represents one channel. Because we use about 1000 samples, and the samples are independent, the average before the stimulus is close to zero. After the stimulus, we see a clear response. The first response is at about 80ms post-stimulus, but is actually quite weak. At about 300 and 450ms post-stimulus, and later at about 650ms, distinct peaks are seen.

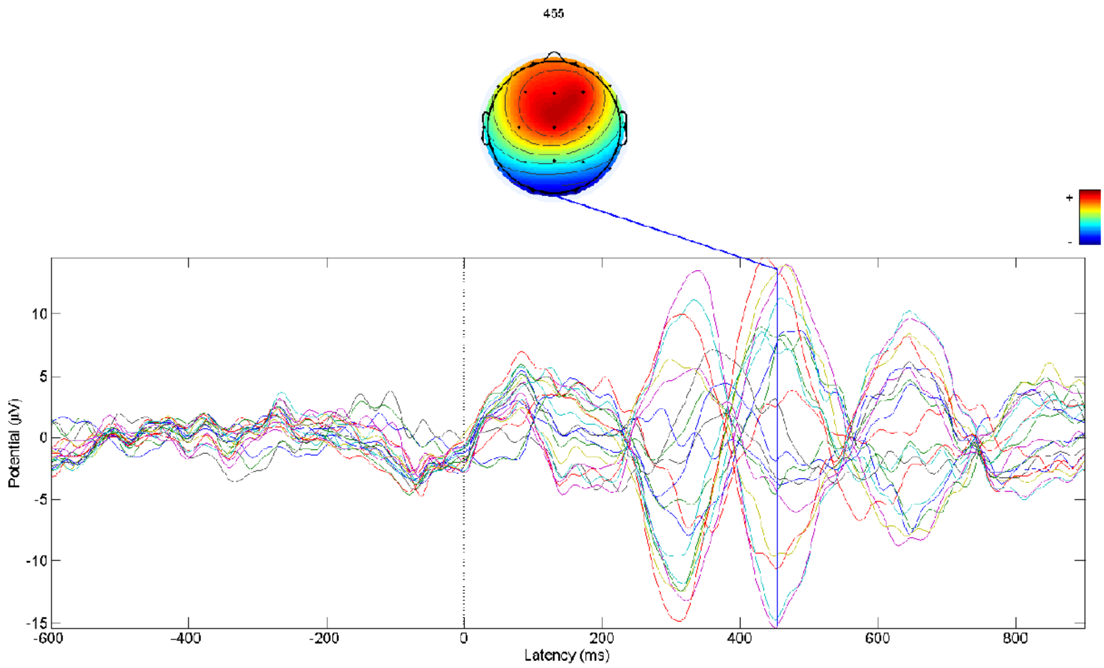


Figure D.1: Example ERP response for different channels

At the moments where the peaks occur, 300 and 450ms, there are some channels that have a high positive peak, and some channels that have a strong negative peak. In figure D.2, you can find a spatial visualization of the EEG signals. Blue areas have a very low voltage, where red areas have a very high voltage. The frontal and central areas have an opposite sign as the parietal and occipital areas. This is probably the result of the reference electrode that is located near the frontal electrodes.

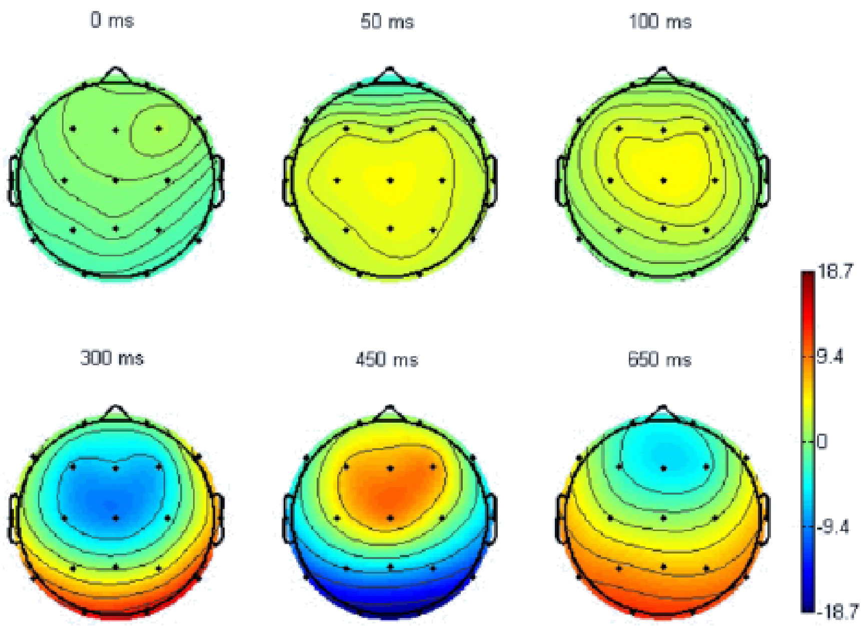


Figure D.2: Spatial division of EEG signals at different moments

Until now, we had all the signals for different persons were averaged together. But we also have to know whether these ERP peaks are visible for the subjects separately. When investigating these signals, there are 19 out of 24 subjects that show a negative peak around 300ms on the Fz electrode. However, the absolute value of the peak is not the same for the different subjects. When doing the same investigation for the Pz channel, we find a positive peak around 400ms post-stimulus for 16 out of 24 subjects. These peaks are less apparent than the negative peaks on the Fz electrode, and there was more noise present in the signal.

From these preliminary experiments, we see that the ERPs are very well visible in the EEG data. However, as predicted by others, clear responses are not visible in every single sample. In order to get some good results, multiple samples have to be averaged.

Even though many papers suggest that the P300 ERP should be most apparent in the parietal and occipital lobes, it is not very clear from these measurements. The Fz channel also provides good values.

D.2.2 EEG, ERPs and reaction time

From the experiments mentioned in section D.2.1, we also know the reaction time of the subject, between the stimulus and the first action (touching the navigation system). We could expect that this reaction time would provide a measurement for the workload of the subject. For that reason, I tried to find a correlation between the reaction time and the ERP properties, in order to see whether the ERP properties give some information about the workload of the subject. Additionally, we tried to correlate the reaction time with some other EEG properties.

ERPs and reaction time

The reaction time that was measured in this experiment, consists of two components: the first part from the stimulus until the time of first response and a second part of entering the city name into the device. The first one we call ‘reaction_time’, the second is called ’enter_time’.

The ERPs were computed from the EEG data. The data was first bandpass filtered between 1Hz and 30Hz, in order to remove a lot of noise. After that, samples with apparent artifacts (e.g. eye blinking or extreme values) were removed by visual inspection. For extraction of the ERPs, the values from the three parietal channels (P3, Pz and P4) are averaged. We find the ERPs by searching the EEG signal for peaks within predefined time-windows. The N2 is sought for between 250 and 350 milliseconds after the stimulus, the P3 is sought for between 350 and 600ms after the stimulus. The search algorithm simply looks for the highest amplitude within that window, and uses that one as the ERP.

Figure D.3 shows the first results of this approach. The top part shows the reaction time, the second part shows the ERP latencies and the lower part shows the ERP amplitudes. The values are grouped by and averaged for all subjects. Those subjects are on the horizontal axis.

With this information I did not find any more information on which direction I had to search for. Therefore I computed the correlation coefficients for the different variables, where I took all samples separately. The results are shown in table D.1. For the red values, $p < 0.01$, for the blue values $p < 0.05$. As expected, the values for the enter_time and total_time are very much correlated, because the reaction_time is negligible as compared to the enter_time. The rest of the ERP properties seam not to be very much correlated with reaction_time. The best correlation is found between p3 amplitude and reaction_time.

EEG band power and reaction time

Besides the ERP properties, EEG signals also provide information on its own. EEG signals can be divided into several frequency bands, where different bands have different meaning. Important bands are the delta (1-4Hz) and alpha (9-12Hz) bands. It is known from personal experience that the power in the delta band should increase, and the power in the alpha band should decrease when the workload rises. This relationship should be

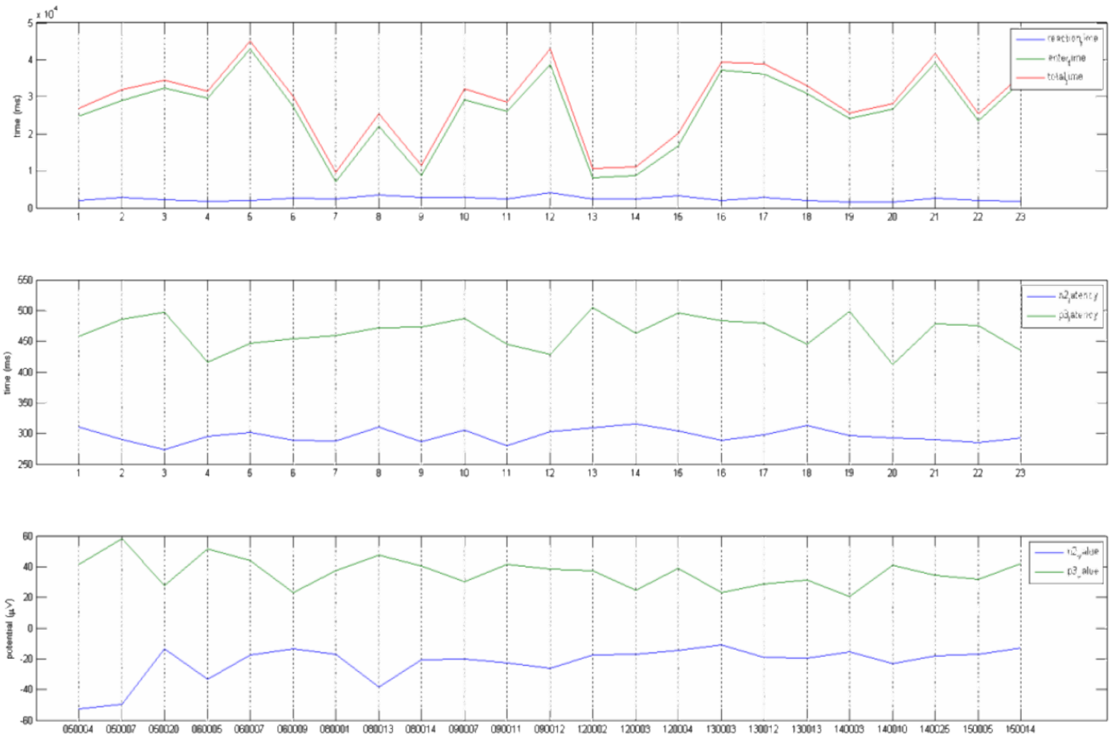


Figure D.3: Reaction times and ERP properties for all subjects

Table D.1: Correlation between reaction time and ERP properties

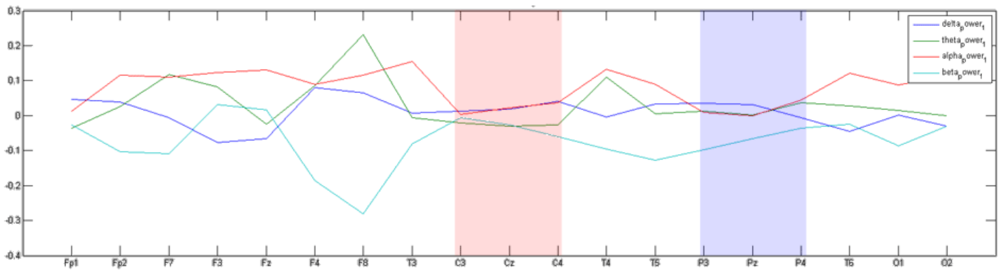
	reaction time	enter time	total time	n2 latency	n2 value	p3 latency	p3 value
reaction time	1.00						
enter time	-0.06	1.00					
total time	-0.01	1.00	1.00				
n2 latency	0.10	-0.08	-0.07	1.00			
n2 value	-0.12	-0.07	-0.08	-0.05	1.00		
p3 latency	0.09	-0.11	-0.10	0.13	0.22	1.00	
p3 value	0.14	-0.02	-0.01	0.02	-0.33	-0.10	1.00

most visible in the occipital and parietal lobes of the brain. Because of this assumption, we tried to find a correlation between the reaction time and the EEG band powers.

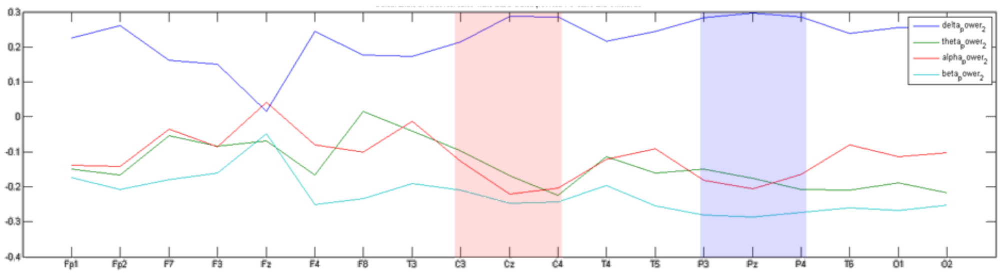
The EEG data was available for 10 seconds around the stimulus, 5 seconds before and 5 seconds after. From the EEG data, the band power was computed for the delta (1-4Hz), theta (5-8Hz), Alpha (9-12Hz) and Beta (15-30Hz) bands. This band power was computed for two time windows, one before the stimulus (-3500 to -1000 ms) and one after the stimulus (1500 to 4500ms). This way, the data would not be affected by the hearing of the stimulus, or the ERPs. The band power is computed for all channels separately.

To start, we evaluated the variances of the different variables. If the variance is high, it is very difficult to find some correlation between multiple variables. These values appear not to be exceptional, so we can continue analyzing the data.

After looking at the variance, we looked at the correlation of the EEG band powers, for different bands, with the reaction time. Results can be seen in figure D.4. The correlation of the band powers after the stimulus is much more correlated with reaction time, than the band powers before the stimulus. This was expected, because the band powers after the stimulus reflect something of the brain activity during the task. The image also gives some evidence that there is indeed some correlation with reaction time. The effect is best seen at the central (C3, Cz and C4, marked red) and parietal (P3, Pz and P4, marked blue) channels. This is most apparent for delta and beta bands.



(a) Correlation between reaction time and EEG band powers before the stimulus



(b) Correlation between reaction time and EEG band powers after the stimulus

Figure D.4: Correlation between reaction time and EEG power bands for different EEG channels before (a) and after (b) the stimulus

When looking at the marked channels, the correlation between the reaction time and the different band powers is significant. All p values are smaller than 0.01 (except for

Table D.2: Correlation between EEG band power and reaction time, for different EEG channels

	C3	Cz	C4	P3	Pz	P4
δ power	0.21	0.29	0.29	0.28	0.30	0.29
θ power	-0.10	-0.17	-0.22	-0.15	-0.18	-0.21
α power	-0.13	-0.22	-0.20	-0.18	-0.21	-0.16
β power	-0.21	-0.25	-0.24	-0.28	-0.29	-0.27

the theta band at C3, which is 0.035). The correlation is shown in table D.2

From these results, I continued experimenting with the central and parietal channels. I made a linear regression model to fit the band power data to the reaction times. We did this for all 6 mentioned channels. The best result was found for channel Pz. The regression model found using the a robust fitting method ('robustfit' in matlab), produces much better results than the standard regression function (depicted in red). However, the error of this prediction is still very high, and can not be used for practical purposes. The prediction could be improved by other nonlinear methods, but I did not have enough time to investigate those methods.

Some investigators state that different levels of attention and workload may be better distinguished by EEG power band ratios, instead of by the single power bands [40, 60]. However, there ratios, such as beta / (alpha + theta) or delta / beta produces only slightly better results, in terms of the root mean squared error (RMSE).

Where the prediction of the reaction time is not performing so well for all samples together, it might work if we try to predict it per subject. That way, subject-specific properties and specialties can be taken into account, which might produce a better result.

Unfortunately, the correlations of the reaction time with the EEG band power for most subjects are not statistically significant. Only for 2 out of 11 subjects the correlations are very high, and this has probably troubled the results for all subjects combined. If all band powers of the central and parietal channels are taken into account, a good predictor might be made for some subjects. It will be very different per subject, and new ones will have to be made for all subjects. This limits the use of these results. However, in some cases this might be a good way to go.

EEG band power ratios pre- and post-stimulus

Another thing I looked at is the ratio between the band powers before and after the stimulus. It is expected that the delta power should rise and alpha power should decrease, because the user will be more active after the stimulus. Thus I expected the ratio of delta power to be higher than 1 and the ratio of alpha power to be lower than 1. It might be possible to use these ratios as a predictor for the reaction time.

As can be seen from figure D.5, the normalized band power for the delta band is quite steady (ratio = 1) , but the other bands are decreasing heavily. Even though the effects are most clear for the F7, F8 and T4 electrodes, we still use the central and parietal

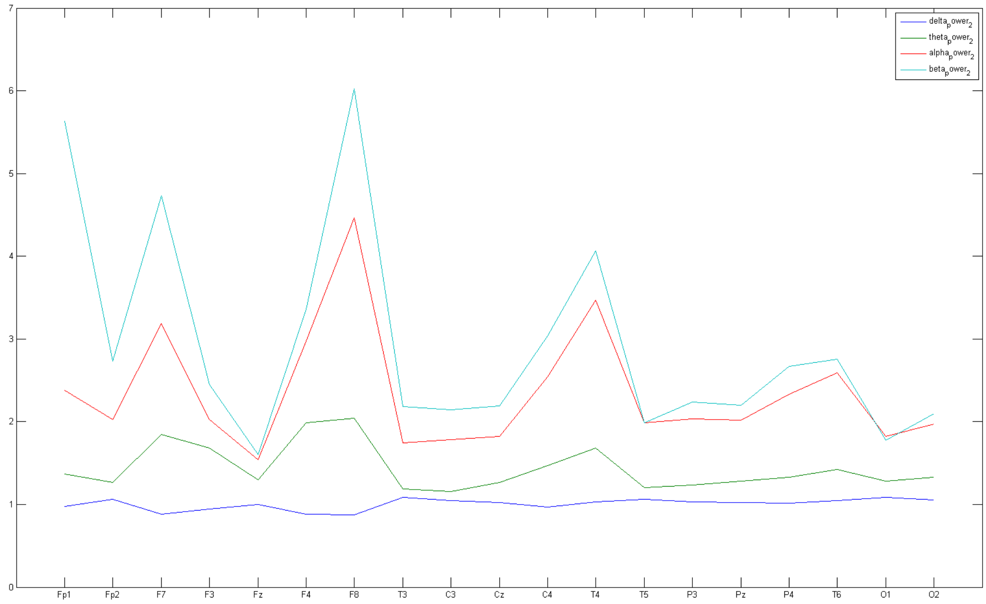


Figure D.5: Ratio of the EEG band powers pre- and post-stimulus per channel

electrodes, because they have proved to show the best correlation with the reaction time, and those are the electrodes that should give some good results.

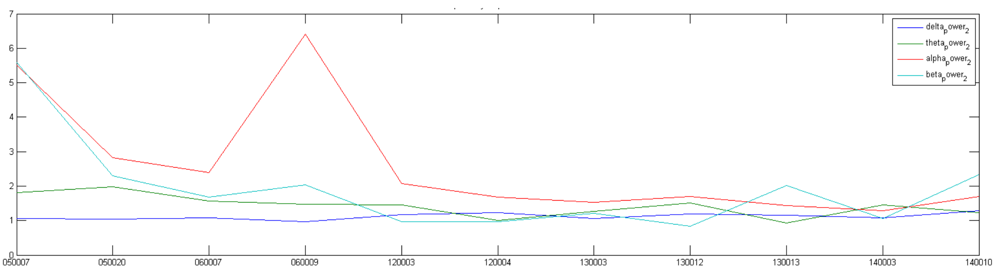


Figure D.6: Ratio of the EEG band powers pre- and post-stimulus per subject

As depicted in figure D.6, the ratio in alpha band power is high for all subjects. Extreme values are found for 2 subjects, but even if those values are excluded, the alpha band power ratio is still clearly above 1 for all subjects. The beta band power is more fluctuating, whereas the delta band power is also quite steady. Theta band power is also mostly above one.

The correlation values (see table D.3) show that is should be best to try predicting the reaction time with the delta power band ratio. Using this ratio at the parietal and occipital channels, we achieve an even worse result than we got for the delta/beta measure as introduced earlier in this paragraph.

Table D.3: Correlation between reaction time and band power ratios. Red values indicate p ¡ 0.01 and blue values indicate p ¡ 0.05

	C3	Cz	C4	P3	Pz	P4	O1	O2
δ power	0.13	0.14	0.09	0.13	0.15	0.17	0.16	0.19
θ power	-0.04	-0.08	-0.09	-0.05	-0.03	-0.07	-0.09	-0.05
α power	0.01	0.00	-0.08	-0.01	0.01	-0.03	-0.03	-0.02
β power	-0.05	-0.07	-0.09	-0.07	-0.08	-0.10	-0.03	-0.05

D.2.3 ERPs and drowsiness

From earlier research [58] it was known that there are differences in the EEG signals for people that are awake on the one hand and drowsy people on the other hand. Also the ERP were said to be different. Because of the drowsiness experiments performed here, at the laboratory of systems reliability, it was interesting to look after these differences. Unfortunately, most of the data from earlier experiments were lost due to a computer crash. For that reason, I could only take the data from one subject into account, which is clearly not enough to draw serious conclusions, but might give a direction for further research.

The experiment that was done, consisted of two times a ride in the driving simulator, once when the driver was fresh, and one time when the driver had not slept for 24 hours. Along the track were some green traffic lights. If a traffic light turned red, the driver had to stop immediately. Sometimes the traffic light turned off, and at those moments the driver did not have to respond at all. This happened less often than the red lights. These changes of the traffic lights were taken as stimuli, and EEG signals were measured, from which the ERPs were derived.

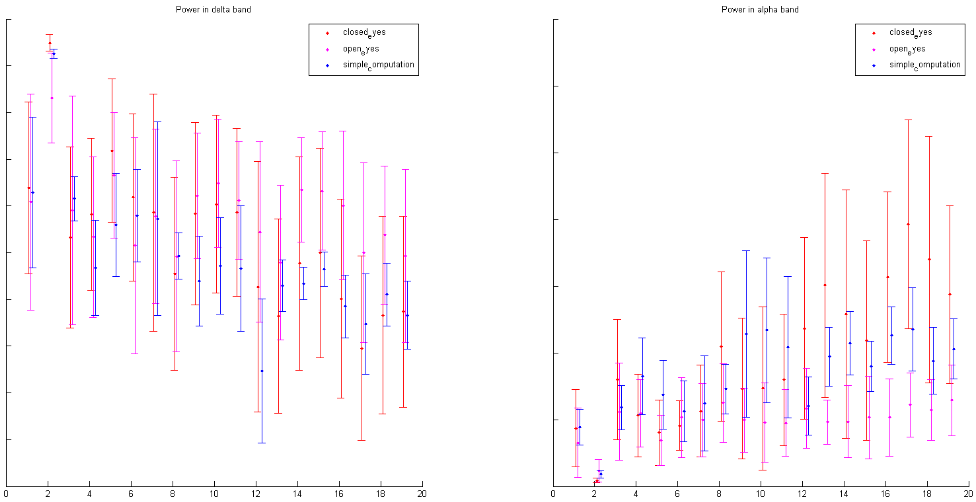
Although the 'standard' measures of reaction time and lane and speed deviation showed a clear distinction between the fresh and the drowsy part, it is not clearly found from the ERP properties. The amplitudes and latencies do not seem to differ that much, for different freshness or different signals, although it was expected to see a clear difference. Most of the differences in mean are non-significant. However, it seems to be that the 'none' signals show lower latencies and higher values on both ERPs than in case of a red signal. This is actually in line with expectations, because the 'none' signal was presented less frequently. For the fresh-drowsy combinations, the differences are even less precise. The values of the ERPs have overall closer-to-zero values in the drowsy case.

D.2.4 Stability of ERPs

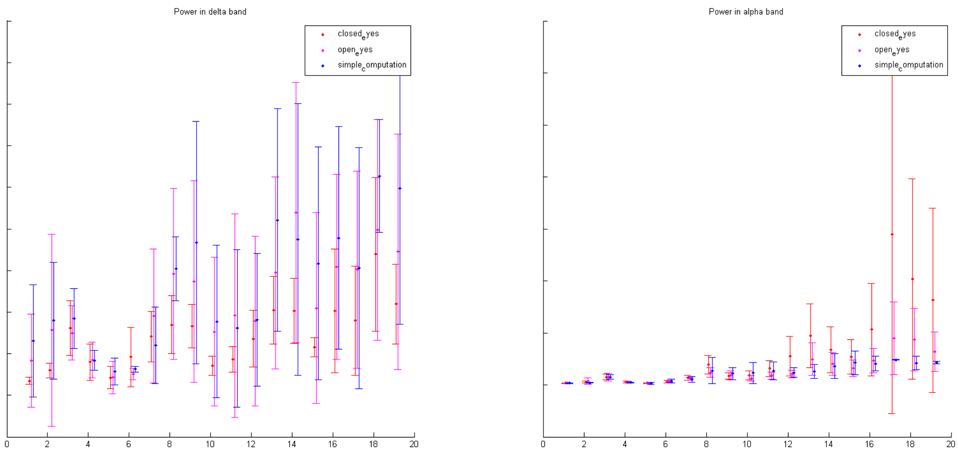
In order to find some measurement for the workload of some system from EEG signals, the EEG signals should not be very different, for different measurements. That is, the signal should be somehow stable. If we take for example the alpha band power as a measure, it should not be very high the one moment, and very high the next moment, when measuring the same subject under similar circumstances.

Some experts on this field have stated that the EEG signal is stable for an healthy adult person, during the adult part of his life (say 20 60 years). However, they do not say in what way the signal is stable. It could be always the same level, or always produces similar artifacts, or maybe have always the same ratio of band powers. It is also possible that the signals are well classifiable by visual inspection, whereas it is much more difficult to do it with a computer. As we do not know this, we are not able to find a good measurement from EEG signals.

During this experiment I have tried to find in what way the EEG signal is stable, and whether we can use some measures to differentiate between two states. I have used the 'open eyes' and 'closed eyes' states, measured before each measurement, because they are very clearly distinguishable by visual inspection. As a third state, we have the subject doing some computations.



(a) Mean values and standard deviation for subject 050007



(b) Mean values and standard deviation for subject 050007, one year later

Figure D.7: Mean values and standard deviations for normalized delta and alpha band power, for three different states.

In the first stage, I looked at simple measurements of the EEG band power. Figure

D.7(a) shows one measurement, where figure D.7(b) shows the same measurements, but measured one year later. The horizontal axis shows the electrodes. These values are taken from 1 subject, so it might not be representative for the other subjects, but this only is for a preliminary case.

The first thing we notice in figure D.7(a) is that the alpha band power is much lower, but also seems to be more stable. Especially for the ‘open eyes’ and ‘closed eyes’ states, the delta band power has a very high standard deviation. For the computation state, however, the band power seems to be much more stable for both bands, and for the delta band it is generally lower than the other states.

For the high numbered electrodes (parietal and occipital lobes), the alpha band shows a distinction between open and closed eyes. Actually, this was expected, because with closed eyes, the basic rhythm dominates and the EEG signal shows high values in the alpha band.

Unfortunately, comparison with the second measurement of this same person (see figure D.7(b)) shows very different results. From this picture, the distinction between open and closed eyes in the alpha band is less distinct. Also, the deviation for the computation state in the delta band is much higher in this measurement.

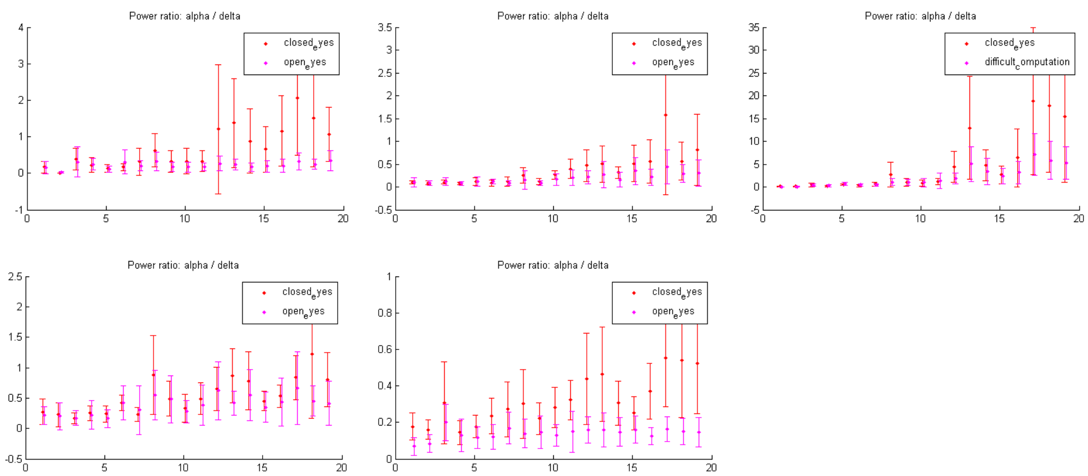


Figure D.8: α / δ band power ratio for 5 subjects on 19 electrodes

In order to find some more or less stable feature of the EEG signal, we computed the ratio of the powers of the different bands (see figure D.8). From visual inspection, the alpha / delta ratio (or the other way around) seems to provide the best way to make a distinction between the two states of open and closed eyes. The distinction is not perfect, but it might be a good approximation.

However, the signals are not really stable. Especially the last person shows great deviations in the ratio. And the first two images show data from the same person, but measured a year after. Those pictures have the same scale, though, so the measurements are in the same order of magnitude.

Based on the α / δ band power ratio, I tried to predict the state of a subject. If the

Table D.4: Percentage of good classification for the simple predictor

	subj. 1	subj. 2	subj. 3	subj. 4	subj. 5
3 states: 'open eyes', 'closed eyes' and 'simple computation'	66,7%	60%	61,8%	56,5%	58,8%
2 states: 'open eyes', 'closed eyes'	80,8%	64,7%	69,2%	66,7%	53,3%

signal is more or less stable and the states are separable by some EEG feature, it should be possible to find some way to predict the state, based on that EEG feature. We used a very simple predictor. It assumes the distribution of the values within a class (state) is normal. So for each given value of the α / δ ratio, we computed the probability of belonging to a class, based on the mean and standard deviation. The class which it most probably belonged to, it was assigned to. Results of this prediction is shown in table D.4. The results are not very convincing, but with more measurements, more or better properties and a better classifier, the figures could improve.

D.2.5 ERPs and workload

In order to test whether and how to use the ERPs as a measurement for the mental workload on a person, we have done an experiment, specifically designed for this purpose. This experiment consisted of driving in a simulator. During the ride, there were traffic lights every 250 meters. The traffic lights were normally green, but when the traffic light turned red, the driver had to stop immediately. This was the event for which the ERPs are measured. The driver also had to do a secondary task, computing a sum of 6 numbers. This computation was done several times, so that we have measurements of ERPs when the driver was busy computing, and also when the driver was not busy computing.

Unfortunately, because there was not enough time, and there were only few people available, we could only do this experiment for 2 persons. But just as for the drowsiness experiment, this might give some direction for future research.

The first results show that the n2 peaks are higher (more negative) and p3 peaks are lower for the baseline task, as compared to the computation task. This is depicted in figure D.9. These values are computed by averaging the EEG data for all samples, and computing the ERP properties for this averaged data for different channels separately. The vertical bar in all subplots shows the standard deviation. The effect in latencies is not the same for both subjects, but the effect in the amplitude is.

However, if we do not average the EEG signals for multiple samples, and find the ERPs for each sample individually, we get much worse results, as can be seen in figure D.10. To get this picture, we computed the ERP properties for all samples separately. The differences between the two tasks are not very sharp anymore, and the standard deviations are much larger. This is in line with other research, that states that ERPs are hard to be found in single samples, but better in averages of different samples.

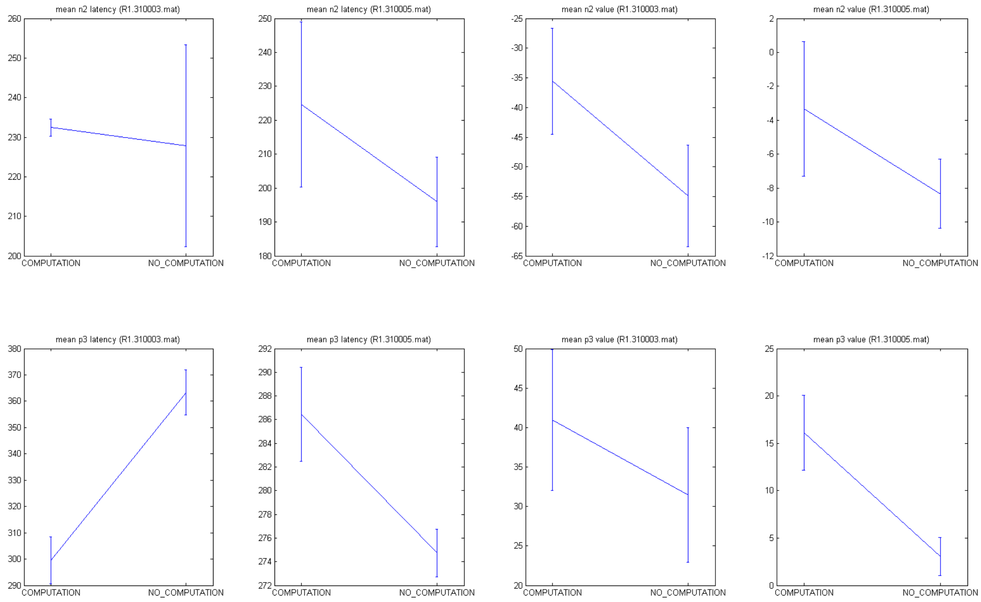


Figure D.9: Mean and standard deviation for ERP properties on different channels, where all samples are averaged (channels: Cz, P3, Pz, P4, O1 and O2)

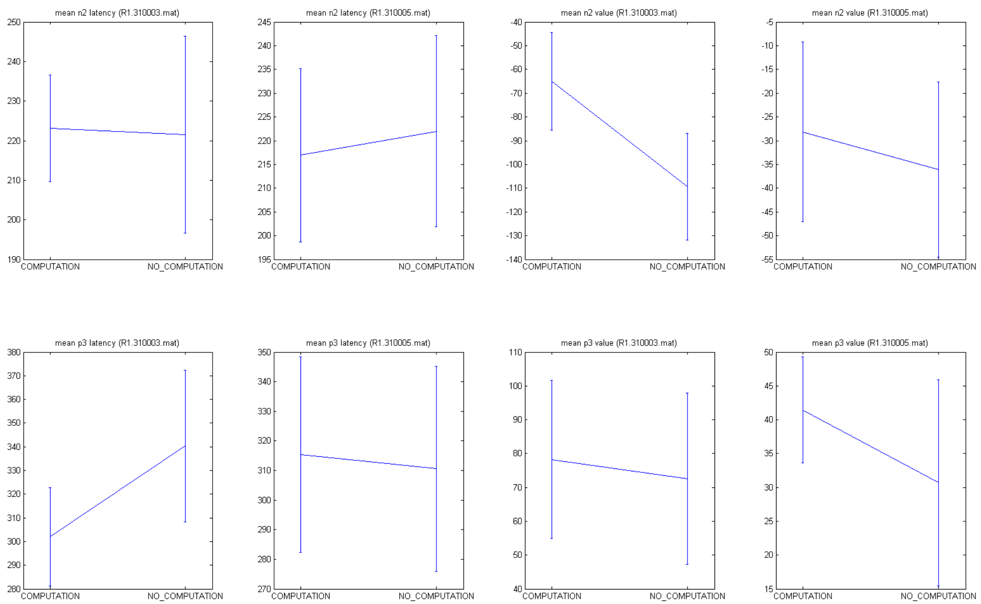


Figure D.10: Mean and standard deviation for ERP properties on different samples, where all channels are averaged. Extreme values are filtered out. (channels: Cz, P3, Pz, P4, O1 and O2)

Another possibility was taking into account other properties of the ERPs. Lelekov-Boissard et al. [39] suggests the use of the difference between p3 and n2 amplitudes. I also computed these properties for the available samples as can be seen in figure D.11. This also does not give a clear distinction between the two tasks.

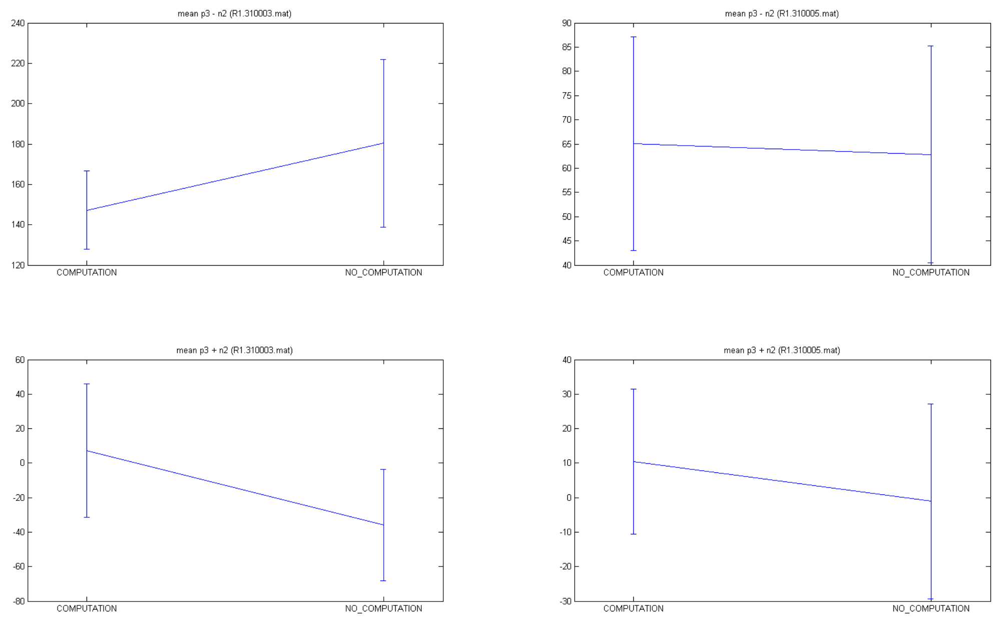


Figure D.11: Combined ERP properties, averaged over the samples (channels: Cz, P3, Pz, P4, O1 and O2)

D.3 Conclusions

After all these experiments, one thing has become clear: working with EEG is highly subjective, and very difficult to draw conclusions for. It is also seen in literature, that one paper finds some results in their EEG data, and other papers say the opposite. This is one of the major drawbacks of using EEG or ERPs for measurements. Because of that, it would have not been fair to have too high expectations from this research project.

From the first experiment (finding ERPs in EEG signals) it was found that the ERPs are apparent in EEG data for most people, when you have enough samples. It is very hard to use ERPs from only one sample, because EEG signals are also influenced by lots of other thoughts. When using different samples, the noise of other thoughts is effectively removed, if we assume the other thoughts are independent and have zero mean EEG values. It is however not perfectly clear on which lobes the ERPs are best measured. Good results seem to be found on the parietal and occipital lobes. This is probably the result of the use of visual stimuli, because the visual cortex is located in the occipital lobe [30]

Because of the subjectivity of the EEG signals, different subjects should also be treated as such. Moreover, I have not found evidence for the stability of the EEG signal in time. The first experiment (for example figure 3) shows the differences in amplitude and latency of the ERPs for different subjects. The other experiments also show great fluctuations in results between subjects. However, there are several cases where different subjects' EEG signals show similar behavior (see for example the research on P300 ERP). These cases should be used to investigate further use of EEG and ERPs for workload measurement.

It has not been possible for me to find a good way to measure workload using EEG and ERPs, because we had not enough time and available subjects. We only did measurements for two subjects, and the results of both subject differ significantly. However, results from other research have been promising (see for example [52]), so continuing on investigating this relationship is a promising way. When more subjects are measured, it might be possible to get some better results.

Bibliography

[1] L.I. Aftanas, N.V. Reva, A.A Varlamov, S.V. Pavlov, and V.P. Makhnev. Analysis of evoked EEG synchronization and desynchronization in conditions of emotional activation in humans: Temporal and topographic characteristics. *Neuroscience and Behavioral Physiology*, 34(8):859–867, 2004.

[2] M.W. Bhatti, Y. Wang, and L. Guan. A neural network approach for human emotion recognition in speech. In *ISCAS '04. Proceedings of the 2004 International Symposium on Circuits and Systems*, volume 2, pages II– 181–4, 2004.

[3] P. Bouchner. *Driving Simulators for HMI Research.* PhD thesis, Czech Technical University, 2007.

[4] M. M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin (SAM) and the semantic differential. *Journal of Experimental Psychiatry & Behavior Therapy*, 25(1):49–59, 1994.

[5] G. Bush, P. Luu, and M.I. Posner. Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences*, 4:215–222, 2000.

[6] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 205–211, New York, NY, USA, 2004. ACM Press.

[7] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun. Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals. Technical Report 05.02, Computer Vision Group, Computing Science Center, University of Geneva, 2005.

[8] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[9] J. Cheng. Could EEGs have prevented clippy? microsoft taps brain scan for ui work. [hyphens]http://arstechnica.com/news.ars/post/20071015-could-eegs-have-prevented-clippy-microsoft-taps-brain-scan-for-ui-work.html, 2008. Visited on January 21, 2008.

[10] A. Choppin. EEG-based human interface for disabled individuals: Emotion expression with neural networks. Master's thesis, Tokyo Institute of Technology, 2000.

[11] R. Cowie, E. Douglas-Cowie, N. Taspatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor. Emotion recognition in human-computer interaction. *Signal Processing Magazine, IEEE*, 18:32–80, 2001.

[12] F. Dellaert, T. Polzin, and A. Waibel. Recognizing emotion in speech. In *ICSLP 96. Proceedings*, volume 3, pages 1970–1973, October 1996.

[13] A. Delorme and S. Makeig. EEGLab: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134:9–21, 2003.

[14] A. Delorme, S. Makeig, and T. Sejnowski. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. In *Proceedings of the Third International ICA Conference*, 2001.

[15] Z. Dharmawan. Analysis of computer games player stress level using EEG data. Master's thesis, Delft University of Technology, 2007.

[16] J. Dien. Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, 30(1):34–43, 1998.

[17] P. Ekman. Basic emotions. In T. Dalgleish and M. Power, editors, *Handbook of cognition and emotion*, pages 45–60. Sussex, U.K.: John Wiley & Sons, Ltd., 1999.

[18] Emotiv. Emotiv website. `http://www.emotiv.com/`, 2008. Visited on January 21, 2008.

[19] B. Fasel and J. Luettin. Automatic facial expression analysis: A survey. *Pattern recognition*, 36:148–275, 2003.

[20] T. Felzer, R. Fischer, T. Groensfelder, and R. Nordmann. Alternative control system for operating a pc using intentional muscle contractions only. In *Proceedings of the 2005 CSUN Conf. on Technology and Persons with Disabilities*, 2005.

[21] Gizmag. The first commercially available brain computer interface. `http://www.gizmag.com/go/6971/`, 2007. Visited on January 21, 2008.

[22] g.tec. g.bcisys bci system. `http://www.gtec.at/products/g.BCIsys/bci.htm`, 2007. Visited on May 7, 2007.

[23] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[24] A. Haag, S. Goronzy, P. Schaich, and J. Williams. Emotion recognition using bio-sensors: First steps towards an automatic system. *Lecture Notes in Computer Science*, 3068:36–48, 2004.

[25] D. Hagemann, E. Naumann, and J.F. Thayer. The quest for the EEG reference revisited: A glance from brain asymmetry research. *Psychophysiology*, 38:847857, 2001.

[26] P. He, G. Wilson, and C. Russell. Removal of ocular artifacts from electro-encephalogram by adaptive filtering. *Medical and Biological Engineering and Computing*, 42:407–412, 2004.

[27] W. Heller, J.B. Nitschke, and D.L. Lindsay. Neuropsychological correlates of arousal in self-reported emotion. *Neuroscience letters*, 11(4):383–402, 1997.

[28] B. Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1970.

[29] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. `http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf`, 2007. Last updated: December 20, 2007.

[30] D.H. Hubel. *Eye, brain, and vision.* New York: Scientific American Books, 1988. `http://hubel.med.harvard.edu/index.html`.

[31] N. Ille, P. Berg, and M. Scherg. Artifact correction of the ongoing EEG using spatial based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, 19(2):113–124, 2002.

[32] B.W. Jervis, E.C. Ifeachor, and E.M. Allen. The removal of ocular artefacts from the electroencephalogram: a review. *Medical & biological engineering & computing*, 26:2–12, 1988.

[33] C.A. Joyce, I.F Gorodnitsky, and M. Kuta. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41:313–325, 2004.

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

[35] P.R. Kleinginna Jr. and A.M. Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5:345–379, 1981.

[36] M.B. Kostyunina and M.A. Kulikov. Frequency characteristics of EEG spectra in the emotions. *Neuroscience and Behavioral Physiology*, 26(4):340–343, 1996.

[37] C.M. Krause, V. Viemerö, A. Rosenqvist, L. Sillanmäki, and T. Aström. Relative electroencephalographic desynchronization and synchronization in humans to emotional film content: an analysis of the 4-6, 6-8, 8-10, and 10-12 hz frequency bands. *Neuroscience letters*, 286(1):9–11, 2000.

[38] P.J. Lang, M.M. Bradley, and B.N. Cuthbert. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report Technical Report A-6, University of Florida, Gainesville, FL, 2005.

[39] T. Lelekov-Boissard, M. Duraz, C. Percher, S. Robert, I. Aillerie, M.P. Bruyas, R. Michelet, P. Deleurence, A. Chapon, L. Garcia-Larrea, and J.L. Magnon. Evaluation of the driver's attention on a driving simulator using brain evoked potentials analysis. In *DSC 2004*, 2004.

[40] J.F. Lubar. Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders. *Applied Psychophysiology and Biofeedback*, 16(3):201–225, 1991.

[41] S. Makeig, A.J. Bell, T.-P. Jung, and T.J. Sejnowski. Independent component analysis of electroencephalographic data. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 145–151. The MIT Press, 1996.

[42] O.H. Mowrer. *Learning Theory and Behavior.* Wiley, New York, 1960.

[43] F.C. Murphy, I. Nimmo-Smith, and A.D. Lawrence. Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, 3:207–233, 2003.

[44] T. Musha, Y Terasaki, H.A. Haque, and G.A. Ivanitsky. Feature extraction from EEGs associated with emotions. *Artificial Life and Robotics*, 1(1):15–19, 1997.

[45] N. Nicolaou and S.J. Nasuto. Temporal independent component analysis for automatic artefact removal from EEG. In *Proceedings of the 2nd International Conference on Medical Signal and Information Processing*, pages 5–8, 2004.

[46] N. Päivinen, S. Lammi, A. Pitkänen, J. Nissinen, M. Penttonen, and T. Grönfors. Epileptic seizure detection: A nonlinear viewpoint. *Computer Methods and Programs in Biomedicine*, 79(2):151–159, 2005.

[47] M Pantic and L.J.M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 22, pages 1424–1445, December 2000.

[48] M Pantic and L.J.M. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. In *Proceedings of the IEEE*, volume 91, pages 1370–1390, September 2003.

[49] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: Criteria of max-dependency, max-relevance and min-redundancy. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 27, pages 1226–1238, 2005. Implementation available at `http://research.janelia.org/peng/proj/mRMR/index.htm`.

[50] R. Plutchik. *The Emotions: Facts, Theories and a New Model.* New York: Random-House, 1962.

[51] L.J. Prinzel, R. Parasuraman, F.G. Freeman, M.W. Scerbo, P.J. Mikulka, and A.T. Pope. Three experiments examining the use of electroencephalogram, event-related potentials, and heart-rate variability for real-time human-centered adaptive automation design. Technical Report TP-2003212442, NASA Langley Research Centre, Hampton, Virginia, 2003.

[52] M.E. Rakauskas, N.J. Ward, E.M. Bernat, M. Cadwallader, C.J. Patrick, and D. de Waard. Psychophysiological measures of driver distraction and workload while intoxicated. In *3rd International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 2005.

[53] B. Reeves and C. Nass. *The media equation; how people treat computers,television, and new media like real people and places.* Stanford : CSLI, 1996.

[54] M. Rohál'ová, P. Sykacek, M. Koska, and G. Dorffner. Detection of the EEG artifacts by the means of the (extended) kalman filter. *Measurement Science Review*, 1(1):59–62, 2001.

[55] A. Savran, K. Ciftci, G. Chanel, J.C. Mota, L.H. Viet, B. Sankur, L. Akarun, A. Caplier, and M. Rombaut. Emotion detection in the loop from brain signals and facial images. `http://www.enterface.net/results/`, 2006. Visited on September, 26, 2007.

[56] F. Schiffer, M.H. Teicher, C. Anderson, A. Tomoda, A. Polcari, C.P. Navalta, and S.L. Andersen. Determination of hemispheric emotional valence in individual subjects: A new approach with research and therapeutic implications. *Behavioral and Brain Functions*, 3:207–233, 2007.

[57] A. Schlögl, P. Anderer, S.J. Roberts, M. Prezenger, and G. Pfurtscheller. Artefact detection in sleep EEG by the use of kalman filtering. In *Proceedings of the EMBEC 99, Part II*, pages 1648–1649, 1999.

[58] A. Sekine, Y. Niiyama, O. Kutsuzawa, and T. Shimizu. A negative component superimposed on event-related potentials during light drowsiness. *Psychiatry and Clinical Neurosciences*, 55(5):473–478, 2001.

[59] N.V. Shemyakina and S.G. Danko. Influence of the emotional perception of a signal on the electroencephalographic correlates of creative activity. *Human Physiology*, 30(2):145–151, 2004.

[60] B. Streitberg, J. Rhmel, W.M. Herrmann, and S. Kubicki. Comstat rule for vigilance classification based on spontaneous EEG activity. *Neuropsychobiology*, 17:105–117, 1987.

[61] H. Tijms. *Understanding Probability: Chance Rules in Everyday Life*. Cambridge University Press, 2007. Available online at `http://books.google.com/books?id=Ua-_5Ga4QF8C`.

[62] J. Villegas. Definition of emotion. `http://www.ciadvertising.org/studies/student/97_fall/theory/emotion/ea/definition.html`, 1997. Visited on February, 28, 2007.

[63] M. Vourkas, G. Papadourakis, and S. Micheloyannis. Use of ann and hjorth parameters in mental-task discrimination. In *First International Conference on Advances in Medical Signal and Information Processing*, pages 327–332, 2000.

[64] C. Wang, C. Guan, and H. Zhang. P300 brain-computer interface design for communication and control applications. In *Proceedings of 2005 IEEE Engineering in Medicine and Biology 27th annual conference*, volume 5, 2005.

[65] D.L. Weber. Scalp current density and source current modelling. `http://dnl.ucsf.edu/users/dweber/dweber_docs/eeg_scd.html`, 2001. Visited on September, 19, 2007.

[66] P.D Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. In *IEEE Trans. Audio Electroacoustics*, volume AU-15, pages 70–73, 1967.

[67] M. Wessel. Pioneering research into brain computer interfaces. Master's thesis, Delft University of Technology, 2006.

[68] Wikipedia. Nyquist rate - wikipedia, the free encyclopedia. `http://en.wikipedia.org/wiki/Nyquist_rate`, 2008. Visited on January, 21, 2008.

[69] H. Zhang. The optimality of naive bayes. In *Proceedings of the Seventeenth Florida Artificial Intelligence Research Society Conference*, pages 562–567. The AAAI Press, 2004.

[70] L. Zhukov, D. Weinstein, and C. Johnson. Independent component analysis for EEG source localization. *Engineering in Medicine and Biology Magazine, IEEE*, 19(3):87–96, 2000.

[71] A. Ziehe and K.-R. Mller. Tdsep - an efficient algorithm for blind separation using time structure. In *ICANN'98*, pages 675–680, Skovde, 1998.