Faculty of Engineering and Architecture
Departement ELIS
2015–2016

Recognize Emotion in the brain using EEG Data

by

Andreas DE LILLE

Promotors: Prof. J. DAMBRE
Dr. Ir. P. VAN MIERLO
Assistent: Ir. T. VERHOEVEN

Contents

Nomenclature

ANET Affective Norms for English Text

ANEW Affective Norms for English Words

BCI Brain Computer Interface

CSEA Center for the Study of Emotion and Attention

DEAP Dataset for Emotion Analysis using Physiological Signals

EEG Electroencephalography

ERP Event Related Potential

Fm Frontal Midline

IADS International Affective Digital Sounds

IAPS International Affective Picture System

MEG magnetoencephalography

SAM Self-Assessment Manikin

SAM self-assessment manikins

.

1

Introduction

This chapter describes the context of the thesis, starting with brain computer interfaces(BCI), before defining some BCI basics. After that, the P300 speller and P300 paradigm are introduced. Before the need for an emotionally aware P300 speller is justified, the basic process of emotion in the brain is explained.

1.1 Brain computer interfaces

A Brain Computer Interface (BCI), creates a direct neural link from the brain to the computer[1], that tries to recognize patterns and based on the extracted information, performs actions. A BCI removes to need for physical actions, i.e. typing or moving a mouse, for the transfer of information. The neural link provided by the BCI is made of two important components. The first component is the extraction component, which extract brain signals from the brain. The second component is the computer that interprets signals and performs actions based on the outcome.

1.1.1 Electroencephalography (EEG)

Different technologies exist to analyze the brain, the most convenient method is via Electroencephalography (EEG), since it is a non-invasive method. Non-invasive methods, in contrast to invasive methods require no surgery; they simply measure electrical activity using electrodes placed on the scalp.

The electrical activity in a brain is caused when an incoming signal arrives in a neuron. This triggers some sodium ions to move inside the cell, which in turn, causes a voltage rise[2]. When this increase in voltage reaches a threshold, an action potential is triggered in the form of a wave of electrical discharge that travels to neighboring neurons. When this reaction occurs simultaneously in a lot of neurons, the change in electrical potential becomes significantly, making it visible to the EEG surface electrodes. EEG can thus only capture synchronized activity of many, many neurons.

Signals originating from the cortex, close to the skull, are most visible, while signals originating deeper in the brain cannot be observed directly. Even for signals originating close to the cortex, EEG is far from precise as the bone between the the cortex and electrodes distorts the signal.

Additionally other artifacts like eye and muscle movement add a lot a noise to the signal, that has to be removed. Even though the noise is persistent and has very low spatial resolution, EEG data can provide significant insight into the electrical activity of the cortex while offering excellent temporal resolution[3].

Note that EEG records electrical activity, other methods like magnetoencephalography (MEG) measure brain activity using magnetic fields. Since MEG is more prone to noise from external magnetic signals, i.e. the earth's magnetic field and electromagnetic communication, a magnetic shielded room is required, making this method very expensive and not mobile.

EEG uses electrodes that are placed on the scalp to measure the electrical activity. To ensure that experiments are replicable, standards for locations of electrodes have been developed. One of these systems is the 10/20 system, an internationally recognized methods to describe location of scalp electrodes[4]. The numbers 10 and 20 refer to the distances between the electrodes, which are either 10% or 20% of the total front-back or left-right distance of the skull. Each site is identified with a letter that determines the lobe and hemisphere location.

- **F:** Frontal
- **T:** Temporal
- **C:** Central
- **P:** Parietal
- **O:** Occipital

Note that no central lobe exists; the C letter is only used for identification purposes. The letter z indicates that the electrode is placed on the central line. Even numbers are use for the right hemisphere, while odd numbers are used for the left hemisphere. A picture of a 23 channel 10/20 system is added below for clarification. Note that some experiments may use more channels than shown in figure 1.1, but they all follow the same naming convention.

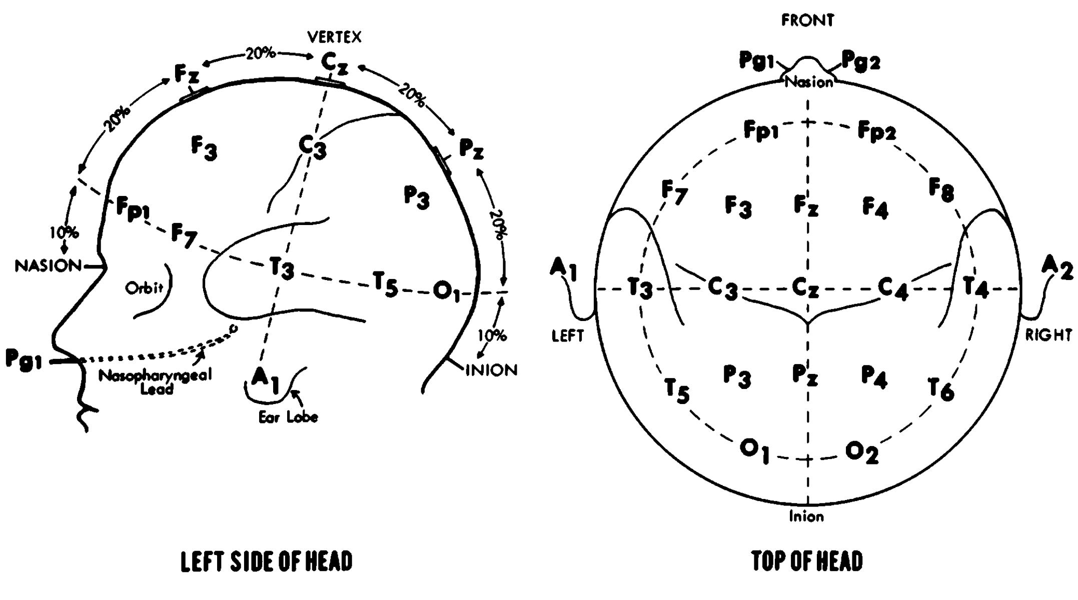

Figure 1.1: The electrode placement of a 23 channel system[5].

Two different types of EEG channels exist, monopolar and dipolar. A monopolar channel records the potential difference of a signal, compared to a neutral electrode, usually connected to an ear lobe of mastoid. A bipolar channel is obtained by subtracting two monopolar EEG signals, which improves SNR by removing shared artifacts[6].

In the frequency domain, brain waves are usually split up into different bands[7, 8]. These wavebands are:

1. **Alpha:** 8-13Hz, indicate how relaxed and/or inactive the brain is.
2. **Beta:** 13-30HZ, indicate a more active and focused state of mind.
3. **Gamma:** 30-50Hz, relate to simultaneous processing of information from different brain areas.
4. **Delta:** 0-4hz, these waves are generated during dreamless sleep and meditation.
5. **theta:** 4-8Hz, occur during dreaming.

Most muscle and eye artifacts have a frequency around 1.2Hz. Artificats caused by nearby power lines, have a frequency around 50Hz[2]. To remove most of this noise, a bandpass filter is usually applied to filter out frequencies below 4Hz and above 40-45Hz.

Event related potentials

An Event related potential (ERP), is a measured brain response to an event, measured by EEG or MEG. The ERP nomenclature usually starts with a letter that indicates the polarity: the P corresponds to a positive polarity, while the N indicates a negative polarity. The number indicates the mean latency, measured between the ERP and the stimulus, which might may variate significantly between subjects.

The most important one is the P300 wave, which is usually elicited using the oddball paradigm; when a low probability target item occurs between high probability items. It consists of two components, the P3a with a latency of 240ms and the P3b with a latency of 350 ms[9]. The later component, P3b only occurs when the subject actively counted either the targeted or more frequent stimuli.

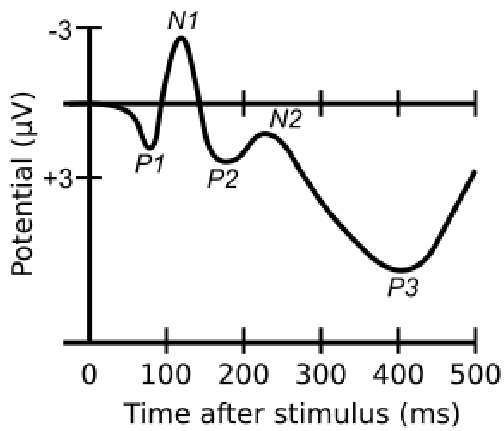

Figure 1.2: The Different ERP linked to an oddball paradigm, found at[10].

1.1.2 The P300 speller

The P300 speller is an active topic of research that uses EEG data to enable persons with the locked in syndrome to communicate[11]. The basic version uses a six by six grid of characters, each row and column is flashed in a random order while the subject silently counts the number of flashes of a certain character, as shown in figure 1.3. This procedure, where a train of stimuli with some infrequent occurring target stimuli is applied, is called the oddball paradigm[12]. It is known that this technique triggers an increase in the potential difference in the EEG around the parietal lobe. This ERP occurs +- 300 milliseconds after the stimuli is flashed, hence its name, the P300 waveform[13]. The presence or absence of the P300 waveform is used by the P300 speller to determine what character the subject was focusing on, which basically allows the subject to spell text.

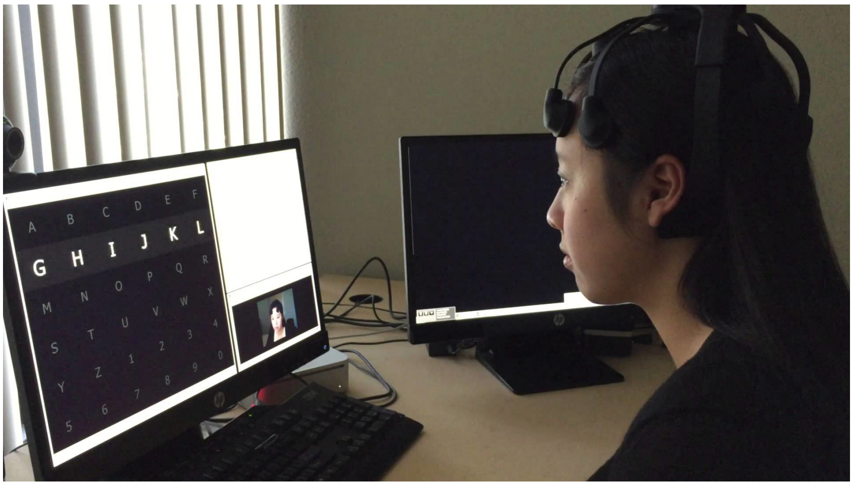

Figure 1.3: Different parts of the P300 speller, found at [14].

To improve the spelling time many improvements and research has been done. Language models were used to predict the word based on the first characters, which enabled great speedups[1], classifiers were compared and tested on both healthy[15] and unhealthy subjects[13]. Since many unhealthy subjects might have an impaired vision or eye movement, tactile[16] and auditory[17] spellers have been developed to circumvent this problem.

To improve accuracy, common problems such as adjacency distraction, when a subject is distracted from a neighboring flash, and double flashes, when the target row and column are flashed close after each other, were avoided using new randomized paradigms[12]. Other input layouts like the T9 interface P300 speller have also been developed[18].

To further speedup the spelling, error potentials were explored. Error potentials are triggered when the user becomes aware of an erroneous action[19], i.e. when a wrong character is selected. When an Error potential is detected, the character is usually changed to the second most probable character according to the P300 decoding[20], which is the most viable character.

The basic P300 speller needs a calibration period before it can be used, when a healthy subject makes a mistake during calibration, he can communicate this and the problem can be resolved. This is not the case for a patient, who has no other means of communication than the P300 speller. Having wrongly labeled data during calibration can lead to severe problems. The unsupervised speller as proposed in [21] solves this problem by removing the need for a calibration procedure. The speller works with expectation maximization and has an undemanding linear classification backend and achieves good results, after a warm-up period when the system is adapting to the given condition.

1.2 Emotion recognition

How does one define emotion? Psychology makes a clear distinction between physiological behavior and the conscious experience of an emotion, called expression[2]. The expression consists of many parts, including the facial expression, body language and voice concern. Unlike expression, the physiological aspect of an emotion, e.g. heart rate, skin conductance and pupil dilation, is much harder to control. To really know one's emotions, it seems, one has to research the physiological aspect of the emotion. One possibility for this is analysis of brain activity via Electroencephalography[22].

1.2.1 Emotion in the brain

Before emotions can be recognized, a classification model is needed. A common model to classify emotions is the bipolar arousal-valence model[2, 23], that places emotions in a two dimensional space. The main advantage of using a multidimensional model, is that all emotions are modeled in its space, even when no particular discrete label can be used to define the current feeling. Figure 1.4 shows the mapping of different emotions for this model.

Even though arousal and valence describe emotion quite well, a third dimension can also be added. The new model than has three dimensions: arousal, valence, dominance and liking. Arousal indicates how active a person is and ranges from inactive, bored to active, excited. The valence indicates if the emotion is perceived as positive or negative. The third dimension, the dominance, indicates how strong the emotional feeling was and ranges from a weak feeling to an empower, overwhelming feeling. The dominance can help to filter out samples of strong feelings, since feelings with low dominance are less likely to show significant effects.

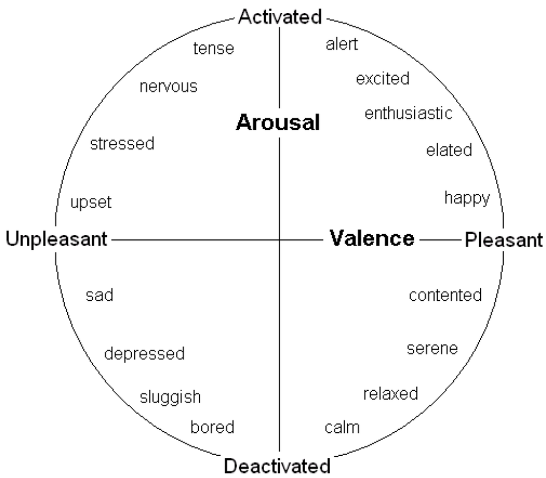

Figure 1.4: The arousal - valence model maps emotions in a two dimensional plane.

Determining valence

The most known and used feature is the frontal asymmetry of the alpha power[3]. The right hemisphere is generally speaking, more active during negative emotion than the left hemisphere which is in turn more active during positive emotions[23, 22]. The asymmetry is given for L and R being the Left and Right frontal alpha powers as:

$$Asymmetry = \frac{L-R}{L+R}$$

Computing the spectral power of the alpha band is possible via e.g. the fast Fourier transform or wavelet transform.

It is also possible to include beta waves in the process. High alpha rates correspond with an inactive brain, while high beta waves with an active brain. Looking for an increase in beta activity and a decrease in alpha activity at one side, while the other hemisphere should show an increase in alpha waves and a decrease of beta waves as indication that is becomes less active, offers an insight to the frontal asymmetry and thus the valence[2].

Another feature is the frontal midline (Fm) theta power, that is shown to increase with increasing pleasant ratings for audio stimuli[24].

Determining arousal

Arousal can be determined in several ways. A first methods for the arousal recognition uses only alpha and beta bands. the Alpha band corresponds with a relaxed state and is often connected to brain inactivity[2, 23]. Beta waves, on the other hand, are an indication that the brain is a more active state or has a higher level of arousal. Combining these two parameters gives the beta/alpha ratio as an indication of the arousal level.

Other methods are based on the EEG coherence across the prefrontal and posterior beta oscillations, which is known to increase when high arousal images are viewed. Additionally, the gamma power is said to increase with arousal with a delay of 500ms[3].

Datasets

One of the most used datasets is the Dataset for Emotion Analysis using Physiological Signals (DEAP) dataset[25]. This dataset contains EEG samples at 512 Hz of 32 persons each viewing 40 videos. A preprocessed version of this dataset, that is down sampled and has EOG removal will be used extensively during this thesis.

Furthermore, the center for the study of emotion and attention (CSEA), by the university of Florida made several visual datasets available:

- **IAPS:** International Affective Picture System
  This dataset consists of a large set op emotional stimuli in the form of color photographs.
- **IADS:** International Affective Digital Sounds
  This dataset consists of acoustic emotional stimuli designed for investigation and research of emotion and attention.
- **ANEW:** Affective Norms for English Words
  This dataset provides a set of emotional ratings for a large number of english words.
- **ANET:** Affective Norms for English Text
  This set provides normative ratings of emotion for a large set of brief texts.
- **SAM:** Self-Assessment Manikin
  A non-verbal pictorial assessment technique that measures the pleasure, arousal and dominance associated with a person's affective reaction to a wide variety of stimuli.

The stimuli from these sets are often used in experiments to trigger emotions[2, 26, 23, 22].

1.2.2 Benefits of creating an emotionally aware P300 speller

Emotions play a major role in non-verbal communication, are quite complex and essential to understand human behavior. The ability to recognize emotions will improve the ability of computers to understand human interaction[27] and are likely to improve the P300 speller's accuracy.

To improve the detection of error potentials, emotion can be used. It is expected that when a person makes a lot of mistakes, his/her emotional state will change to a less happy, more frustrated state. Making the speller emotionally aware, could improve the detection of error potentials.

Research with visual stimuli on healthy subjects, show that emotion has an effect on the auditory P300 wave[28]. Both the P300 peak amplitude and area was highest when viewing neutral pictures and descended further, in decreasing order, for sadness, anger and pleasure. The amplitudes were significantly lower at both Fz and C3 positions than Pz and Oz. The latency of the P300 ERP speller was shortest or neutrality and in increasing order longer for pleasure, anger and sadness. Interestingly, the P300 amplitudes were significantly larger for woman than for men. Additionally women showed lower amplitudes for pleasure than neutrality or sadness while man showed smaller amplitudes for pleasure then either neutral or sadness.

Contrary to what subjects might think, the P300 speller is unable to read the mind and know what a person is thinking about[11]. The P300 speller provides no more than a means of communication that the subject can use. Should he chose to ignore the instructions and focus his attention elsewhere, then the recordings become useless. Knowing how the subject feels, can help him communicate more humane on one hand, while providing more insight for ethical issues, on the other hand, e.g. "How does the subject think about the P300 speller recording and analyzing his brain activity?". Information about the subject's emotional state can help answer some of these ethical questions.

1.3 Goal of the thesis

This thesis aims to improve the performance of the P300 speller by making the speller emotionally aware. An emotional aware speller is expected to yield better performance, since the P300 wave is affected by the emotion. Furthermore, the detection of error potentials can be combined with the emotional state, since the emotional state is expected to change with increasing errors.

More concrete, the main goal is to recognize emotions in the arousal-valence model, using the DEAP dataset. First the emotions of a single person should be recognized, since the features are known to differ from person to person. Later, the model will be adjusted so that it is capable of detecting emotion of different persons. Once the emotion recognition is able to classify the emotions with decent accuracy, it will be integrated in the P300 speller, which should give additional accuracy and aid in the error potential detection. The expected results are:

- Being able to recognize emotions of a single person
- Being able to recognize emotions across different persons
- Improved accuracy for the P300 speller
- Improved error potential recognition for the P300 speller

Additionally the gained information for emotion recognition can be used for other ethical research studies, to answer ethical questions about the usage of BCI on patients.

2

A first look at the data

2.1 The DEAP dataset

This thesis uses the DEAP dataset[25], a dataset for emotion analysis that is publicly available for academic research. This dataset contains EEG recordings of 32 participants, each watching 40 one minute excerpts of music videos. Each video was rated individually by each person on valence, arousal, dominance and liking. The first three ratings correspond to the valence, arousal and dominance space of an emotion. The liking component indicates how much the person liked the video excerpt and should not be confused with the valence component. The liking measure inquires about the participants' tastes and not their feelings, i.e. a person can like a video that triggers angry or sad emotions. The liking rates are neglected, since they are not part of the emotion space.

For assessment of these scales, the self-assessment manikins (SAM), were used[25]. SAM visualizes the valence, arousal and dominance scale with pictures, each picture corresponds to a discrete value. The user can click between the different figures, which makes the scales continuous.All dimension are given by a float between 1 and 9. A preprocessing step scaled and translated these value to ensure they range between 0 and 1. Zero corresponds to the minimum of the scale and one corresponds or the maximum value of the scale.

The used SAM figures are shown in Figure 2.1. The first row gives the valence scale, ranging from sad to happy. The second row shows the arousal scale, ranging from bored to excited. The last row represents the different dominance levels. The left figure represents a submissive emotion, while the right figure corresponds with a dominant feeling.

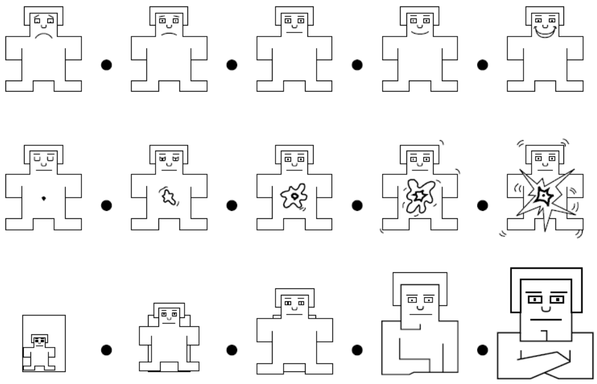

Figure 2.1: The images used for the SAM[25].

To further inspect the distribution of the user ratings and whether or not the data is balanced, the average for each emotion dimension (valence, arousal and dominance), was determined using all videos of all persons. These can be seen in Table 2.1. A uniform distribution would give a value of 0.5. The averages of the DEAP are a little above 0.5, which gives a first indication that the data is balanced.

	Valence	**Arousal**	**Dominance**
value	0.532	0.520	0.548

Table 2.1: The average value of each component.

To further inspect the data, each component was divided in 8 ranges or bins. Next all user ratings are placed in their corresponding bin. Then the percentage of movies for each bin was calculated. This shown in Figure 2.2, 2.3 and 2.4 for valence, arousal and dominance respectively. It is clear that even though all components have most of their weight in the higher bins, the data is more or less balanced. Therefor no further measurements to balance the data, e.g. extra penalty for one class, was taken.

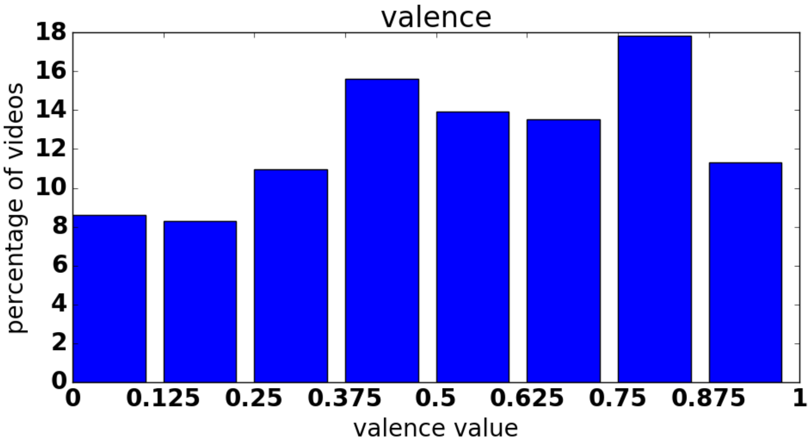

Figure 2.2: The distribution of the valence values.

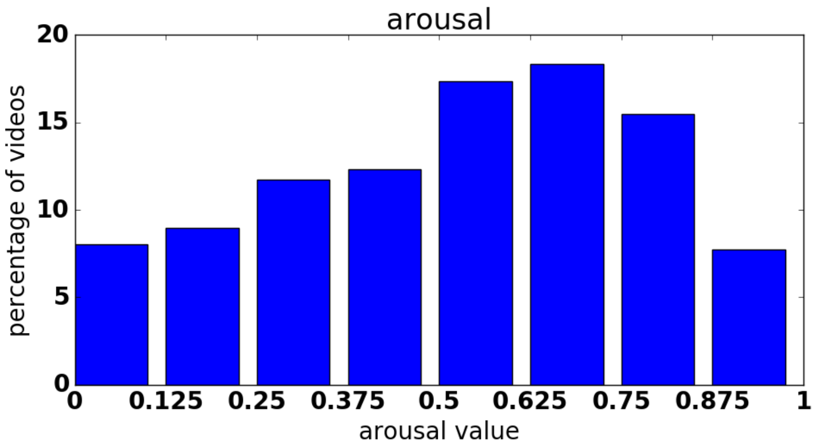

Figure 2.3: The distribution of the arousal values.

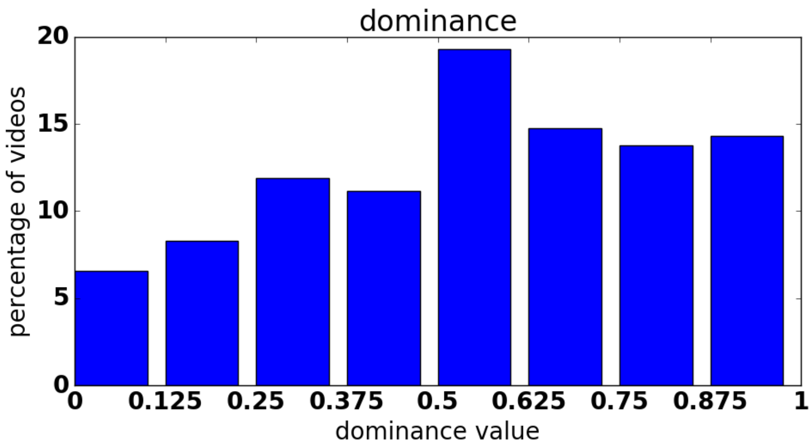

Figure 2.4: The distribution of the dominance values.

2.2 Feature selection

The first step in this thesis is to recognize emotion of a single person. To do this, good features are needed. These features are evaluated using a features selection procedure consisting of the following steps:

1. For each person the data is split in a train and test set, with ratio: 0.75 / 0.25.
2. A linear SVM is trained using the train set and the result is evaluated using the test set.
3. The previous step is repeated for each of the 32 persons in the DEAP set.
4. The average over all results is calculated. Averaging over all persons limits the influence person specific effects on the feature selection, while still using person specific classifiers.

2.2.1 Valence

As discussed in 1.2.1, the most used feature to determine valence is the alpha asymmetry, which is given for L and R being the Left and Right alpha powers as:

$$Asymmetry = \frac{L-R}{L+R}$$

A first attempt will classify valence in two classes: low valence (sad) and high valance (happy). The valences are therefor thresholded at 0.5 everything below is the first class, all valence values above are part of the second class. Valence values close to the threshold are removed as they can potentially confuse the classifier during training. To determine which channels are useful, each left channel is first grouped with its corresponding right channel. For each channel pair the asymmetry is determined and the feature selection procedure described above is executed. The result is a test value averaged over all persons, which is shown in Table 2.2.

Channel Group	Avg Test Accuracy	Channel Group	Avg Test Accuracy
Fp1 - Fp2	38.461	**T7 - T8**	70
AF3 - AF4	71.428	**CP5 -CP6**	50
F3 - F4	44.444	**CP1 - CP2**	55.556
F7 - F8	61.538	**P3 - P4**	53.846
FC5 - FC6	42.857	**P7 - P8**	50
FC1 - FC2	44.444	**PO3 - PO4**	75
C3 - C4	37.5		

Table 2.2: Average Test values for each channel pair

Looking at the results, it is clear that the most promising channels are: AF3-AF4, F7-F8, T7-T8, PO3-PO4. Note that not all channels are located in the frontal regions of the cortex, which is different from what some papers report.

Bibliography

[1] H. Verschore, "A brain-computer interface combined with a language model: the requirements and benefits of a p300 speller," afstudeerwerk, Ghent University, June 2012.

[2] D. O. Bos, "Eeg-based emotion recognition," 2007.

[3] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, "A review on the computational methods for emotional state estimation from the human eeg," *Computational and Mathematical Methods in Medicine*, vol. 2013, no. 573734, p. 13, 2013.

[4] T. C. Technologies, *10/20 System Positioning manual.* Fortis Tower, 2012.

[5] unknown, "Electrode placement," 2015.

[6] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Time-frequency optimization for discrimination between imagination of right and left hand movements based on two bipolar electroencephalography channels," *EURASIP journal on Advances in Signal Processing*, vol. 2014, no. 38, 2014.

[7] K.-E. Ko, Hyun-Chang, and K.-B. Sim, "Emotion recognition using eeg signals with relative power values and bayesian network," *International Journal of Control, Automation, and Systems*, 2009.

[8] Brainworks, "What are brainwaves?," 2015.

[9] N. K. Squires, K. C. Squires, Steven, and A. Hillyard, "Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man," *Electroencephalography and Clinical Neurophysiology*, pp. 387–401, 1975.

[10] Wikipedia, "Event-related potentials," 2015.

[11] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 70, pp. 510–523, 1988.

[12] T. Verhoeven, "Brain-computer interfaces with machine learning: an improved paradigm for the p300 speller," afstudeerwerk, Ghent University, June 2013.

[13] N. V. Manyakov, N. Chumerin, A. Combaz, and M. M. V. Hulle, "Comparison of classification methods for p300 brain computer interface on diabled subjects," *Computational intelligence and neuroscience*, 2011.

[14] Cognionics, "Cognionics dry eeg p300 speller demo," 2015.

[15] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the p300 speller," *Journal of Neural Engineering*, vol. 3, no. 4, p. 299, 2006.

[16] A.-M. Brouwer and J. B. van Erp, "A tactile p300 brain-computer interface," *Frontiers in Neuroscience*, vol. 4, May 2010. doi:10.3389/fnins.2010.00019.

[17] J. Höhne, M. Schreuder, B. Blankertz, and M. Tangermann, "Tow-dimensional auditory p300 speller with predictive text system," *IEEE EMBS*, 2010.

[18] F. Akram, H.-S. Han, H. J. Jeon, K. Park, S.-H. Park, J. Cho, and T.-S. Kim, "An efficient words typing p300-bci system using a modified t9 interface and random forest classifier," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 2251–2254, July 2013.

[19] A. Coone, "A study on different preprocessing and machine learning techniques for the detection of error-potentials in brain-computer interfaces," afstudeerwerk, Ghent university, June 2011.

[20] R. Chavarriaga, A. Sobolewski, and J. d. R. Millán, "Errare machinale est: The use of error-related potentials in brain-machine interfaces," *Frontiers in Neuroscience*, vol. 8, no. 208, 2014.

[21] A. B. model for exploiting application constraints to enable unsupervised training of a P300-based BCI, "Pieter-jan kindermans and david verstraeten and benjamin schrauwen," *Plos ONE*, vol. 7, April 2012.

[22] Y. Lio and O. Sourina, "Eeg databases for emotion recognition," *International Conference on Cyberworlds*, 2013.

[23] Y. Lio, O. Sourina, and M. K. Nguyen, "Real-time eeg based human emotion recognition and visualization," 2010.

[24] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, pp. 209–304, 2007. DOI: 10.1111/j.1469-8986.2007.00497.x.

[25] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," *Affective Computing, IEEE Transactions on*, vol. 3, pp. 18–31, Jan 2012.

[26] M. Technologies, "Emotions by mensia," 2015.

[27] R. W. Picard and J. Klein, "Computers that recognise and respond to user emotion: theoretical and practical applications," vol. Interacting with computers, no. 14, pp. 141–169, 2002.

[28] Y. Morita, K. Morita, M. Yamamoto, Y. Waseda, and H. Maeda, "Effects of facial affect recognition on the auditory {P300} in healthy subjects," *Neuroscience Research*, vol. 41, no. 1, pp. 89 – 95, 2001.