Visual Protocols and Mental Tasks For Communication Brain Computer Interfaces

By Mathew Salvaris

A thesis submitted for the degree of Doctor of Philosophy

School of Computer Science & Electronic Engineering

University of Essex

Date of Conferment

Contents

Chapter 1

Introduction

Humans have always had a need to communicate, this need has been further serviced by technological advancements such as mobile phones and the Internet. The desire to communicate has driven many technologies beyond the point originally perceived by there creators. Communication between humans is necessary in order to convey ideas, thoughts and emotions. These are achieved through myo-actuated output pathways which present themselves as gestures, facial expressions, art, written or spoken language. Modern communication methods rely mainly on verbal or written statements. This is because human machine interfacing is not as multifaceted and diverse as human to human interfacing. In order for someone to interface with a computer they rely on the keyboard and mouse and through a series of interaction are either able to communicate with other people with the computer as a mediator. HCI is a large field in computer science with its goal to make the Human computer interaction as easy and intuitive as possible. Although this area has come a long way the interface still requires a certain amount of learning and the interfaces are not always ideal for all computer related tasks. A far more intuitive and efficient interfacing system would be one that allowed direct communication between the human brain and the machine. In current literature such interfacing is referred to as Brain Computer Interfacing(BCI) or Brain Machine Interfacing(BMI). Currently the main focus of such systems is on providing communication means for people who are severely paralysed with no other method of communicating, such as in the latter stages of ALS. One of the reasons for the focus being such is that BCI systems are much slower than conventional modes of communication and therefore only realistically suitable as a last resort. This dissertation explores various ideas in order to enhance the accuracy and speed of BCIs for communication. BCIs are not currently in everyday use, although many devices have been seen targeted at the general populace and marketed as BCI devices. Unfortunately these devices are not true BCIs since they mainly rely on muscular activity and not brain neuro-electric activity. BCI devices currently are extremely slow when compared to the standard methods of computer interaction(Mouse, Keyboard). Furthermore much of the equipment necessary for BCI use is prohibitively expensive for the general user.

1 Outline of Thesis

Chapter 2

Brain Computer Interfaces

1 Introduction

In this chapter the term brain-computer interfaces is elucidated with special focus on how it can be achieved and the methodologies often employed. In section 2 a brief overview is given into what constitutes a BCI. The section also expands into the possible targte audience for BCIs at this moment as well as some future possibilities. The most common brain signals observed using EEG are outlined in section 3. The subsequent section gives an overview of the usual methods employed to monitor cognitive behaviour with an emphasis on non-invasive methods. In the final section the most prominent and well known EEG based BCIs methodologies are explored. A detailed and exhaustive review of all BCI systems would be impossible and impractical, due to the ever increasing activity in the field of BCI. For detailed reviews please see [1] [2] [3][4][5][6] [7][8] [9].

2 What is BCI?

As mentioned in the previous section BCIs are systems that allow direct communication between the brain and a computer [2]. Although brain-computer interfaces experiments have been carried out on many different animal species when talking about BCIs here they will be referring to systems designed for the human brain. BCIs usually consist of a device that detects specific physiological phenomena associated with cognitive functions, a computer that translates these detected physiological phenomena and interprets them using signal processing and machine learning techniques. These interpretations can be later used to affect the real world via various robotic devices or for communication. The last decade has seen a sharp increase in the amount of BCI research, primarily driven by the growing awareness of the needs and problems of people with disabilities [2][10] [1]). A further contributing factor to the popularity of BCI research is the better understading of brain functions and the ever decreasing cost of computing power and necessarary brain monitoring devices (primarily EEG).

Currently the target groups for BCIs consist of mainly sufferers of various neurological diseases such as spinal cord injury, stroke and motor neuron diseases [10]. Such patients may only be able to control artificial devices through the use of the limited motor functions still available to them to convey their wishes and interact with the world around them. Patients with residual muscular movement are often refered to as locked in. If there is no voluntary muscular movement they are refered to as completely locked in. The most comonly refered to motor neuron disease in BCI literature is amyotropic lateral sclerosis (ALS). ALS is a progressive motor neuron disease that affects the central nervous system as well as the first and second motor neurons. As the disease progresses the patients become increasingly physically impaired. The disease also causes the brain to atrophy with the highest reduction in grey matter over the right-hemispheric primary motor cortex and left-hemispheric medial frontal gyrus. The disease in the final stages may lead to loss of speech and sussesion of unaided brething. Patients in these stages are only able to survive via assisted breathing (tracheotomy). Often such patients may render the locked in or completely locked in state. Patients often express the desire to commuincate as the desease progresses, something that could be considered important for qulity of life.

Although BCIs main focus is on the disambled this hasn't stoped the persuit of incorporating the extra information gleaned through the insight of cognitive processes in other areas. In military and comercial applications the focus has been on assesing emotion, alertness, congnitive workload, task incolvement and concentration. Albeit their use under such conditions may not be interpreted as true BCIs since they do not provide a concious output pathway for the subjects. In the world of computer games the use of BCIs is also becoming more popular. With companies advertising complete game control using BCIs(Emotiv Systems and NeuroSky) to others trying to assess players psychological responses to the games (THQ is using a BCI developed by EmSense to test players' physiological responses to Frontlines, a military-themed shooter game.). Currently the concesnsus within the BCI community is that non-invasive BCIs are not fast enough to directly control fast paced games and invasive BCIs can not be justified for strictly entertainment purposes.

BCI research is highly interdisciplenary with various fields contributing to it:

- Psychology: Providing interfaces and stimuli that maximise the desired user response and minimise potential confounds.
- Neurophysiology: Understainding of the underlinning neuronal responses and how to best localise the desired features.
- Electrical Engineering and Material Science: To design and create electrodes and other brain monitoring hardware that are more desirable, cheaper and easier to use.

- Signal processing: Developing methods to extract relevant information from the raw data often provided by brain monitoring hardware.
- Machine Learning: Be able to infer from the extracted information about the various mental states.
- Software Engineering and Human Computer Interaction: To design interfaces and software tools to quickly and easily create near optimum BCI interfaces(with recommendations drawn from psychology).

In summary BCI is an exciting and highly interdisciplinary field, with the possibilities for improvement arising from any number of fields. The future of BCI will truly depend if the incremental improvemetns offered by the cooperation of these fields leads to a system that is user friendly, robust and practical.

3 Neurophysiological Signals of BCI

In the following section the various rhythmic and event related phenomena observed in EEG will be discussed. This is to provide a brief overview of EEG and the broad cognitive processes associated with the various rhythms and event-related potentials. For a detailed overview of EEG and the broader neurobiology of the brain please see [11][12].

3.1 Rhythmic Brain Activity

EEG rhythms are defined as regularly recurring waveforms of similar shape and duration [11]. They were the first brain signals discovered by Hans Berger [11]. They have been found in the EEG recordings of humans as well as animals. Although the detailed mechanisms that govern these EEG rhythms have only been elucidated relatively recently (over last 40 years). Mainly due to the description of the ionic and electroresponsive properties of various types of individual cells as well as the collective oscillations of large neuronal popultations[11].

During various stages of consciousness the human brain produces different rhythmic brain waves. These rhythms are affected by various actions (moving a limb) and or cognitive tasks/states (counting, navigation, etc.). There have been many proposed brain rhythms but the most recognised have been delta(δ) , theta (θ), alpha (α), beta (β), mu (μ) and gamma (γ). Certain cognitive states are often described by differing brain rhythms over specific brain regions [11][12].

3.1.1 Delta Rhythms

EEG waves under 3.5 Hz are classified as delta rhythms. These rhythms can be found in adults during deep sleep and irregular delta rhythm activity can be seen in infants [12][13].

3.1.2 Theta Rhythms

Theta rhythms are most frequently seen in infants and children. Theta rhythms are only seen in adults during states of drowsiness and sleep. Theta rhythms are found in the 4 and 7.5 Hz frequency band[12, 11].

3.1.3 Alpha Rhythms

Alpha rhythms belong to the frequency band of 8 to 13Hz and occur in adults while awake over the posterior region of the head with higher voltage over the occipital electrodes, although due to volume conduction it is visible in almost all electrodes. Amplitude varies but is mostly below 60 $\mu Vinadults.Mosteasilyobservedinsubjectswiththeireyesclosedandunderconditionsofp$

3.1.4 Mu Rhythms

The mu rhythm has similar amplitude and frequency to the alpha rhythm but is topologically and physiologically different from the alpha rhythm. Mu stand for motor since this rhythm is strongly correlated to functions of the motor cortex. The mu rhythm is not visually discernable in the EEG of every mature subject. The mu rhythm can be blocked by movement, thought about movement, readiness to move and tactile stimulation [14, 12][11]. Central mu is most easily detected at C3 and C4 (in the standard 1020 system) located over the precentral gyrus. The blocking effect is bilateral but more pronounced on the region contralateral to the site of movement, the effect apperas prior to the onset of muscular contraction [15].

3.1.5 Beta Rhythms

Beta rhythms are often found in the 13 to 30 Hz frequency band, although any rhythm of 13 Hz or above could be considered as Beta, recent readoption of the term gamma to indicate brain rhythms above 30 Hz precludes this[11]. Beta rhythms are mainly found over the frontal and central regions of the brain. Central beta rhythm are believed to be related to the mu rhythm and similarly to it can be blocked by motor activity and tactile stimulation [14, 12][11].

3.1.6 Gamma Rhythms

The gamma range although a topic of debated it is widely considered to be between 30 and 60 Hz. 'The old term gamma for the designation of faster beta activity used to be all but forgotten until its resurrection around 1990.'[11]. This renewed interest in the higher EEG frequencies has also peaked the interest of many BCI researches. Much of the available literature associates gamma activity with motor activity [16] [17] although unsuprisingly other congitive associations have been found [18] [19][20].

3.1.7 Event related synchronisation and event related desynchronisation

The amplitude of mu rhythmic brain activity can be found to decrease prior to hand or foot movements and increase after the termination of these movements, these two efects are known as event related desynchronisation and event-related synchronisation. Similar effects have been reported for the imaginary completion of the aforementioned movements [11].

Event related synchronisation is the amplification of certain EEG oscillatory activity due to an internally or externally paced event[15].

Event related desychronisation is the attenuation of certain EEG oscillatory activity due to an internally or externally paced event[21].

ERD and ERS is not only observed in the mu rhythm but also in the other rhythms mentioned in the previous sections. In order to measure an ERD or an ERS the power of a certain frequency band is calculated before and after a certain event over a number of EEG trials. The event can be externally or internally paced. The power is then measured in percentage relative to the power of the reference interval. The reference interval is usually a short interval before the event. The ERS is the power increase and the ERD the power decrease relative to the reference interval.

3.2 Event Related Potentials

Event related potentials occur in the brain in response to certain stimuli or events. The ERP components are divided into two groups exogenous (or early) which occur up to 100 ms after the specific even/stimulus and endogenous (or late) which occur after 100 ms have passed from the specific event/stimulus. The exogenous components are mainly dependant on the stimulus characteristics such as loudness and brightness, where as the endogenous components are mainly influenced by a variety of psychological variables [11]. The amplitudes of ERPs is very small in comparison to the ongoing EEG, consequently event locked signal averaging is required for their observation. The labeling of the ERP peaks and troughs (sometimes refered to as components) is carried out by denoting the polarity with a letter P for positive or N for negative. The differing deflections are then either labeled by their charcteristick peak latencies (e.g. N100, P300, N400), order of appearance (e.g. P1, N1, P2 etc.), or their peak latencies in a specific experiment(e.g. P250, P385). The same label may be used across modalities (visual, auditory, etc.) to denote the same or different neurophysiological phenomena. The most common of these ERPs are reviewed below.

3.2.1 P300

The P300 is a ERP which is exhibited when an infrequent or oddball event is presented and the user is asked to react to it mentaly (counting) or physicaly (pressing a button). The P300 refers to the most positive peak that occurs 250 to 600 ms post stimulus. The exact functional significance of the P300 has yet to be determined [18][22][23]. This is further exasperated by the fact that the P300 is not isolated to any one region of the brain [24, 25] and also the elicited P300's differ between each individual and each stimulus. Some of the factors that affect the P300 are fatigue, changes in food, nicotine, and alcohol consumption, ageing, disease, inattention to stimuli [26] inability to detect or categorise stimulus event and probability of event whether local [27]or global [26, 28]. The stimulus required to exhibit a P300 can be either visual, auditory or somatosensory. This indicates that the P300 does not simply reflect sensory processing of incoming stimuli, but rather reflects higher cognitive processes that may be involved in evaluating the stimulus and updating one's view of the world [26, 28].

3.2.2 RP

The readiness potential (RP) or Bereitschaftspotential is an ERP that precedes a movement. The RP can start upto 1.5 s before movement onset over the motor areas. About 700 to 500 ms before movement onset the topography of activation becomes increasing lateralised to the contralateral side of the intended movement. This is termed lateralised readiness potential (LRP). Similar ERPs have also been identified during imagined movement [29]. In order to seperate the RP/LRP from the imagined movement ERP it is often refered to as movement related potential (MRP), essentially they have many of the same properties. MRPs have also been detected during fast reflective movement, albeit with a shorter duration. BP amplitude has been found to be positively correlated with the force of the intended movement [30][31]. Although psychological variables also seem to affect the BP amplitude [32]. The interesting property of the RP to be detected before the movement offers a great opertunity to design systems that react before the user is able to physically manipulate the controls.

3.2.3 VEP

During visual stimulation, evoked potential can be recorded from the visual cortex in the occipital lobe, these evoked potentials are know as visually evoked potentials (VEPs). VEPs reflect, at least to some extent, the electrophysiological mechanisms underlying the processing of visual information in the brain. The evoked poentials seem to respond to changes in visual stimuli as static stimuli present in the visual field do not have any apperent effect on the EEG activity[33]. VEPs are detected for stimulus mdulations of less than 2 Hz, where as modulation rates of 6 Hz and above induce steady state visually evoked potentials [34]. As the subject focuses their gaze on the stimulus, the amplitude of the SSVEP increases at the frequency of the target as well as the second and third harmonics[35] [36]. The stimulus parameters such as repetition rate and contrast affect the amplitude and phase of the induced SSVEP.

4 Brain Monitoring Hardware

A large variety of brain monitoring hardware exist that exploit various observed phenomena that occur during congnitive processing. These devices can be grouped in severall ways, as well as each having their own advantages and disadvantages. A common grouping used for the devices is invasive and non-invasive. Invasive techniques are the electrocortigogram (ECoG) and intracortical recordings using single electrodes or arrays of electrodes. Non-invasive techniques are electroencephalography (EEG), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), positron emission tomography (PET) and functional near infrared spectroscopy (fNIRS). Invasive methods are often considered less desirable than non-invasise metods especially since current invasive electrodes only allow recording for a limited time. Although with the increased invasiveness comes an increase in the quality of the signal, enabling a degree of control currently unatainable by non-inasive techniwues. Therefor these methods can not be ignored for disabled patients. The non-invasive group can be furter split into methods that rely on neuronal (EEG and MEG) or vascular blood-flow (PET, fMRI and fNIRS) activity. In the follwing section I will first introduce the two invasive methods mentioned. Then the non-invasive methods will be described starting with the methods that rely on the brains vascular activity and secondly the methods that rely on the neuronal activity.

4.1 Invasive methods

Invasive methods are neuronal recording methods that have to placed underneath the scull and therefor require surgery to be placed. Such methods would not be suitable for casual users but may be necessary for disabled subjects. The two most common methods are ECoG that measures the neuronal activity on the cortical surface and intracortical recordings that record the neuronal activity intracortically from within the cortex. Since surgery is required the issues of infection and long term stability arise [37]. Invasive methods currently provide far greter information transfer rates than non-invasive methods, but the afformentioned drawbacks may limit their use to the patients with the gravest necessity for them and greatest possible gain from them.

4.1.1 ECoGs

record the electrical activity of the cortex with the placement of electrode grids or strips either subduraly or epiduraly. The electrodes require a craniotomy for implantation and therefor the method is considered invasive. Although it is often sighted as less invasive than intracortical recordings. Furthermore the use of strips rather than grids may further aleviate the invasivenes of the procedure since they can be inserted via small holes in the scalp. ECoG offers a far higher spatial resolution and broader bandwidth than EEG. ECoG relies on the neuronal activity and therefore rpovides ms temporal resolution. ECoG since implanted beneath the skull is less prone to EMG and EOG artifacts and provides a higher amplitude signal. ECoG are not currently implanted strictly for their use in BCI, but are comonly use in the localisation of epileptogenic brain tissue prior to surgical removal. Therefor the placements of the ECoG electrodes is not always optimum for BCI applications. Recently in BCI competition IV data from 3 patients with implanted ECoG grids was provided. The tasks was to calculate the finger tranjectories on one hand simply by using the information provided by the ECoG. The winner of the competition was able to achieve an average correlation index across all subjects of 0.46 between the predicted and actual finger locations.currently impossible through non-invasive methods.

4.1.2 Intracortical recordings

can be accomplished with singular electrodes or multiple electrodes (arrays). The electrode tips are placed in close proximity to the signal source and are therefore able to cpature the action potentials of individual neurons as well as local field potentials. Intracortical recordings are highly invasive due to there placement within the brain matter. Often the electrodes move relatively to the target neuron or neuron groupings causing scaring and deterioration of the recordings over time. Despite these drawbacks succesfull control of prosthetic devices from intracortical recordings have been achieved in both primates and humans.

4.2 Non - Invasive methods

As mentioned previously the two most common methods of monitoring congnitive processing non-invasively is either through electrical activity of the neuronal populations or the vascular activity induced by the neuronal activity.

4.2.1 Positron Emission Topography (PET)

is a technique that relies on radioactively marked chemical substances to observe the functioning of the desired organ. In brain imaging a the chemical substance marked is glucose, due to the brains dependancy on it for energy. Areas of greater neuronal activity require more glucose and therefore should appear more distinct in a PET scan. PET is technically demanding and may pose a small risk to the patient due to the radioactivity of the marked substance. Furthermore since it depends on the vasculatr activity its temporal resolution is in the range of seconds. Although PET scans offer a high spatial resolution.

4.2.2 Functional magnetic resonance imaging (fMRI)

relies on the blood oxygen level dependent (BOLD) response to monitor the congitive activity of the human brain. The difference in magnetic susceptibility between oxygenated and deoxygenated blood leads to magnetic field variation which are in turn detected by the fMRI scanner. As well as being very expensive, the fMRI relies on superconductors and is therefore very large and only available in hospitals or dedicated laboratories. Compared to to other non-invasive techniques fMRI offers an extremely high spatial resolution but a poor temporal resoltuion. Although with the advent of new acquisition techniques, greater computational power and improved algorithms the temporal resolution has improved considerably.

4.2.3 Functional near infra-red spectrosopy

offers a comparable spatial resolution to fMRI although it is limited to the cortical areas. It is similarly dependant on the BOLD response and measures it based on the reflectance, scattering and absorption of near infra-red light of oxygenated and deoxygenated blood. It is cheaper and more portable than fMRI, therefore more obtainable for research. Funcitonal NIRS has been succesfully used to detect brain activation response to motor movement and imagery. Due to its reliance on vascular activity similarly to fMRI it suffers from a low temporal resolution.

4.2.4 Electroencephalography (EEG)

depends on the electrical activity of large populations of cortical neurons. The sum of this activity can be measured as voltage fluctuations by electrodes placed on the surface of the head. Electroencephalography has been used for a number of years and advancements in amplificaiton digitisation, computational power and electrode design have contributed to its widespread use in BCI today. EEG offers a reasonable spatial resolution, an excellent temporal resolution comparatively cheaply. Due to these benefits it is considered by most BCI researches as the only brain monitoring hardware practical for BCI thus far.

4.2.5 Magnetoencephalography (MEG)

similarly to EEG measures the electrical activity of neuronal populations through the weak magnetic field induced by the electrical current. Similarly to fMRI, MEG relies on superconducators so is not portable and can only be used in a laboratory setting. Furthermore MEG is highly susseptible to movement artifacts and requires a shielded room to obtain recordings of adequate quality. MEG though offers a higher spatial resolution than EEG and higher signal to noise ratio. MEG is more suited for explorative stidies of source localisation in order to agment EEG rather than a feasable BCI device.

5 Ways to achieve BCI

In this section the various rhythms and ERPs associated with motor action/imagination, congitive processing and attention modulation used in BCIs are briefly described. All BCIs described in this section were realised through the use of EEG, since EEG is currently the most common method for achieving cognitively induced control in humans as well as being the method used in this work. The BCIs described in this section can be split into two broad groups, those that are regulated by BCI users and those that are elicited by somatosensory stimulation. The self regulated BCIs require the user to modify their mental state in order to achieve control. Examples of mental states used are imagination of limb movement (left and right arm), mental arithmetic, navigation and auditory recall. The choice of mental task used for BCI contrl is often dependant on the subjects ability to control said mental state. Conversly induced BCIs relie on the modulation of attention to the various stimuli in order to achieve control. They are therefore reliant on exogenous events to achieve control. This is often cited as a drawback of induced BCIs, but in general they offer higher information transfer rates than self regulated BCIs, do not require training and have been shown to work for disabled subjects. Furthermore even though self regulated BCIs do not require exgenous stimulation they often depend on visual feedback in order for the user to achieve a reasonable degree of control.

Many new and interesting BCIs such [38] and jason farquar golden codes graz

5.1 SSVEP

In SSVEP BCI's the subject is presented with a number of oscillating visual stimuli (usually four). When the desired symbol has been differentiated from the others the visual stimulation caused by differential symbol oscillations or colourings are detected as increased oscillatory activation at the occipital areas of the brain. The oscillatory activation are at the stimulus fundemetnal frequency as well as harmonics and subharmonics. Examples of such BCIs can be found in [39] and [33].

Recently there has been increased interest in SSVEP BCIs due to their ease of use, lack of training and high accuracy rates. Classification accuracies of more than 90 percent correct are often reported for SSVEP BCIs. The SSVEP BCIs are considered dependent BCIs since they have been said to require intact gaze and are therefor unsuitable for patients with restricted eye movement [2]. Recent evidence sugest that intact gaze is not necessary to ac hieve control but a reduction in accuracy should be expected [40] [41] . Furthermore Allison et al. demonstrated that simple attention modulation was enough for some subjects to differentiate between two overlapping SSVEP stimuli [42].

5.2 P300

The majority of research conducted utilising the P300 use the protocol pioneered by Donchin & Farwell. The protocol requires that a user be placed in front of a screen which displays a 6x6 grid which contains a number of symbols (characters or words)[43, 44]. In other papers different grid sizes and symbols have been used[45, 46,]. While the user observes the grid containing the symbols the rows and columns flash randomly for a given period. As the user observes the desired symbol when the row or column the symbol belongs to flashes the user is supposed increment a mental counter. The flashing of the desired row/column and the triggered mental event of counting causes the P300 to be elicited in the recorded EEG. Other protocols have been suggested such as the one created Polikoff et al. [47] where instead of selecting a character a direction is chosen from N,S,W or E. A more detailed review of the current BCI protocols can be found in section.

5.3 Sensorimotor Rhythms

The Mu rhythm is concentrated in the primary motor cortical areas and display's EEG activity in the 8-12Hz band when the person is not engaged in producing motor output [48].Movement or preparation for movement is typically accompanied by a decrease in mu rhythms, particularly contralateral to the movement. If these changes are time locked to specific events and the event causes desynchronization of the activity, the changes are referred to as Event Related Desynchronization (ERD), otherwise if an event causes synchronization, this is known as Event Related Synchronization (ERS). Examples of BCI's that use Mu rhythms are Wolpaw [49](The Wadsworth BCI), Pfurtscheller [50] (The Graz BCI), Kostav and Polak [51] Pineda et al. [52] and Penny et al. [53].

5.4 SCP

SCP's are very low frequency components with potential shifts developing over 0.5 to 10 seconds[2]. Examples of SCP BCI's are ones developed by Birbaumer [54, 55] where the user over a period of time learns to control the amplitude of their SCP's and through this control is able to move a cursor or carry out various selections. The difference of this BCI to the P300 and VEP BCI's is that considerable time training is required in order to carry out satisfactory control of this BCI medium.

Chapter 3

Signal Processing and Machine Learning in BCI Literature

In the following section a brief overview of the signal processing, feature extraction and machine learning methods most commonly used in BCI as well as ones employed here. The methods used are mainly focused around the preprocessing and classification of ERPs such as the P300. Since all the work presented here is based on ERPs and giving a broader overview of the various feature extraction and machine learning methods is considered unecessary in light of that fact. Before the EEG signal is classified it has to be filtered and the relevant information extracted. The extraction of relevant information is necessary to reduce the dimensionality of the data in order to partialy aleviate the curse of dimensionality problem. The most common methodologies used for preprocessing/feature extraction are describe in section. In section the various machine learning methods used will be described.

1 Preprocessing

Many classification method are unable to extract the relevant information if the dimensionality of the data is high in comparison to the number of examples provided. This is refered to as the curse of dimensionality and can be somwhat aleviated by removing spurious information and retaining the discriminative information. EEG data is highly dimensional data especially with the advent of 128 and 256 electrode setups with high smapling frequencies (1024 Hz). There are two ways by which the dimensionality of the data can be reduced. One is using domain knowledge to extract the desired neurophsyiological features. e.g. time periods, frequencies and scalp locations that provide most discriminative features. Temporal filtering methods such as infinite impulse response filter (IIR) finite impulse response filter (FIR), fourier based filter (FFT) and discrete wavelet transform (DWT) will briefly described. As well as other common methods eployed in dimensionality reduction for P300 classification (downsampling and window averaging). Spatial filtering methods such as bipolar, common average reference (CAR) and Laplace filtering will also be described. The second method of extracting features could be to use advanced machine learning techniques to extract relevant features without using a priori knowledge of the domain. Such methods are principal component analysis (PCA)and independant component analysis (ICA). Scoring criterions such as Fischer Score DBI or r2 may also be used to eliminate features. Finally search methodologies such as SFFS and RFE can also be used in conjuction with classifiers to select features. Both methodologies suffer from drawbacks, for the strictly neurophysiology based methods suffer from the high trial and subject variability. On the other hand the machine learning methods may also suffer from the high trial variability especially when the number of samples is low, which is commonly the case with BCI. Ideal methods would generalise to the neurophysiology based locations but specialise to each subject based on the information gained from the machine learning methods.

1.1 Finite impulse response (FIR)

filters are a type of digital filter. The filter is named so since the output settles to zero within a finite number of samples. The FIR filter s also known as a feed-forward since it does not recieve any input from its output. If b and K are the filter coefficients and length of filter coefficients respectively, x the input and y the output, the formula for an FIR filter would look very much like:

$$y(t) = \sum_{k=0}^{K} b(k)x(t-k) \tag{3.1}$$

The FIR filter have the following properties: It is causal. It is inhrently stable since it requires no feedback. It is linear. It is also time invariant.

The main disadvantage of FIR filters is that they usually require greater computational power to IIR filter of similar performance.

Differing kinds of filters exist these are; lowpass, highpass, bandpass, bandstop and notch. Lowpass filters attenuate frequencies higher than the specified frequency. Highpass filters attenuate frequencies lower than a specified frequency. Bandpass filters allow frequencies within a certain range to pass through them. Bandstop filter are the opposite of bandpass filters, in that they atenuate frequencies within a certain frequncy range, notch filter are a type of bandstop filter with a narrow frequency range.

1.2 Infinite impulse response (IIR)

is a form of analog or digital filter that allows feed-back, making its impulse response non-zero over an infinite length of time. IIR filters tend to be fast and computationally cheap, but with poorer bandpass filtering and stability characteristics than FIR filters. Using Butterworth, Chebyshev or eliptic methods the filter coefficients a and b are calculated from the desired frequency range and order.

$$y(t) = \sum_{k=0}^{K_b} b(k)x(t-k) + \sum_{k=1}^{K_a} a(k)x(t-k) \tag{3.2}$$

1.3 Fourier based filter

is another way of achieving digital filtering of a signal. The fast Fourier transform (FFT) of the input signal is calculated, representing the input signal in the frequency domain. The frequency doamin represents the signal as a list of amplitudes and phase angles that correspond to a set of harmonically related sinusoids. The filtered signal is obtained by applying a suitable weighting to the frequency components. The filtered signal is then transformed from the frequency domain to the temporal domain by applying the inverse fast Fourier transform (IFFT).

1.4 Wavelets

are short duration oscillating amplitude functions of time, which are relatively localised in time and frequnecy. A neuroelectric waveform can be decomposed using wavelet analysis with a high degree of flexibility over the resolutotion, providing excellent localisation of the events and neuroelectric components in time, space and scale. This high degree of control of resolution is a highly desirable property for a preprocessing method for BCIs, as demonstrated by the number of publications utilising wavelets.

The wavelet decomposes the neuroelectric waveform into wavelet coefficients in a similar manner to the STFT would decompose it into its Fourier componenets. Similarly to the way the inverse fourier transform can use the Fourier components to reconstruct the neuroelectric wavefor, the inverse wavelet transform can be used to recompose the original neuroelectric waveform from the wavelet coefficients. The WT decomposes the waveform through succesive iterations of scaling and translation of the wavelet and calculation of the correlation coefficients at those various scales and translations. This would result in large coefficients when the wavelet closely matches the shape of the processed waveform.

The continuous wavelet transform (CWT) scales and translated the basic wavelet shpe by very small increments and computes the wavelet coefficient at each step. This as can be imagined generates a large number of coefficients in comparison to the nuymber of samples in the original neuroelectric waveform. This tim-scale plot of the the wavelet coefficients can be very usefull in visuaklsing localised transient events such as ERPs, but a large amount of information contained in these coeffiecients may be redundant for classification. In this matter the CWT is not very efficient, the information at the closely spaced scales and time poitns is highly correlated. A more efficient and computationaly simpler wavelet analysis is the discrete wavelet transform (DWT) introdcued by Mallat (1989). The DWT procides a non-redundant, highly efficient wavelet representation that can be implementred with a simple recursive filter scheme. Also unlike the CWT the number of coefficients produced by the DWT are as many as the samples within the original neuroelectric waveform, without the loss of any information. So similarly to the CWT the DWT permits perfect reconstruction of the original neuroelectric waveform by usig its inverse. A number of wavelets were tested throughout this work with a particular focus on the debauchies 4 and 5 and coif3. Mostly for the use in other peoples work

1.5 Downsampling

is a simple process by which the number of samples that make up the waveform are reduced. This is accomplished suually by selecting the waveform components at constant intervals.

So if the original signal was to be downsampled by a factor of 2 then every other sample would be selected. Before the waveform is downsampled it must be lowpass filtered at or below half the new sampling frequency in order to avoid aliasing. This method of feature reduction is computationally simple and as long as the necesary precautions have been taken quite succesful. One of the drawbacks is that the sampling frequency can not be reduced to twice the frequency of the desired ERPs highest frequency, this may result in a feature vector that still contains redundant information.

1.6 Window averaging

is again a simple process by which the number of samples in the neuroelectric waveform are reduced. Here a window of a certain size is used to succesively average all samples within the reach of that window. Then similarly to the downsampling method the waveform compnents at constant intervals are selected. The windowing is equivalent to lowpass filtering of the signal. This method and the downsamlping method are very similar and may produce similar results, ofcourse with similar drawbacks. This method is quite popular in P300 BCIs.

1.7 Spatial filtering

refers to methods used to filter the signal comming from the electrodes using other electrodes or groups of electrodes. The signal seen at each individual electrode is not just the neurological activity at that location, but may contain external electromagnetic interference and interference from neighbouring cortical activations. To overcome this the electrodes can be referenced to other electrodes either on the scalp or placed in other locations i.e mastoids, earlobe. In a monopolar montage, each electrode is referenced with a common reference. This common reference may be a single or an average of two or more electrodes placed at sites whose activity is pressumed to be unaffected by brain activity i.e earlobe or mastoid. Averagesof two electrodes is often prefered since it offers greater reliability and is able to more reliably detect asymetric potentials. The use of two single elctrodes where one electrode is referenced (subtracted) from another is termed bipolar filtering. By calculating the difference between the electrodes the common part in the electrodes is filtered out and the desired signal remains. In bipolar montages, the electrodes are grouped in non-overlapping pairs, with potentials recorded between each pair. Another method of achieving a referenced signal is common average reference (CAR). In this method all or a subset of EEG channels are averaged, then the average is used as the reference for all the EEG channels. Let C be a subset of all channels and n_c the number of channels in the subset, s be the unreferenced signal from each channel and $\hat{s}$ is the referenced signal. Then:

$$\hat{s_j} = s_j - \frac{1}{n_c} \sum_{i \in C} s_i \tag{3.3}$$

for all channels j. As mentioned previously the whole set of channels is used to calculate the average reference signal, so set C would conatin all channels. In Laplace filtering each electrode is reference to the average of its neighbouring electrodes. So for each channel C a neighbrouhood C_j is defined, where $n_c j$ is the number of electrodes in the neighbourhood. Again the referenced signal is $\hat{s}$ and the unreferenced s. Then:

$$\hat{s_j} = s_j - \frac{1}{n_c j} \sum_{i \in C_j} s_i \tag{3.4}$$

The definition of neighbouring electrodes may differ between techniques and generaly depends on the nature of the EEG cap used, as well as the cortical activation localisation desired.

Although CAR has recieved alot of attention in BCI literature, especially with the use of mu and beta rhythms, it was decided taht monopolar reference to averaged earlobe reference would be better for this work, whose main focus is the P300. This is because 'spatial filters best suited for mu and beta rhythms, which are relatively localized, would probably not be the best choice for measurement of SCPs or P300s, which are more broadly distributed over the scalp'. Furthermore Krusienski found that 'CAR was no different from ear reference for the best channel set'.

1.8 Principal component analysis (PCA)

seeks a linear projection that best represetnts the data in a least-squares sense[**?**]. This is accomplished by maximising the variance of the projected data.

The PCA carries out the feature selection in an unsupervised manner, this means that the feature selected may capture the variance of the data but may not be optimal for class separation. Furthrmore noisy features may be included when they contain no relevant information. This feature extraction method is completely statically based and does not use any prior domain knowledge, which may not be advisable under certain conditions.

1.9 ICA

PCA makes the assumption that the mixed signals are orthogonal, which if violated then the algorithm will fail to seperate the data correctly.

1.10 Feature Ranking

A relatively simple method of selecting feature subsets is to rank them using a scoring function and select the desired number of features from the highest ranked features. Feature ranking methods are simple, easily scalable and with good empirical success [56]. Examples of scoring functions used in BCI are; Pearsons correlation or its square, which in linear regression is the coefficient of determanation.

$$r = \frac{\sigma_i(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{(\sigma_i(y_i - \bar{y})^2)(\sigma_i(x_i - \bar{x})^2)}} \tag{3.5}$$

Using r^2 as a feature ranking criterion forces selection of features that most adhere to the linear separation of the classes. There are an number of in BCI that use correlation or variants of it such as biserial or point biserial correlation. Fisher's score is another example of a ranking criterion that could be used

$$s_i = \frac{|\mu_2^{(i)} - \mu_1^{(i)}}{\sigma_2^{(i)} + \sigma_1^{(i)}} \tag{3.6}$$

Another simple criterion is the discriminability measure used by Glassman and defined as:

$$d = \frac{|\mu_2 - \mu_1}{\sqrt{\sigma_2^2 + \sigma_2^2}} \tag{3.7}$$

The most usefull features are the ones for which the difference between the means is lsarge relative to the standard deviations. In the case of two class classificaiton Fisher's criterion, r^2 and similar criteria are closely related(see, e.g., et al., 1999, Tusher et al., 2001, Hastie et al., 2001). It is important to stress that correlation criteria and the other methods mentioned are only able to detect linear dependencies between variable and target. Furthermore the high scored features may be often highly correlated, therefore redundant and combinations of seemingly unimportant features may yield better classificaiton rate. For this reason more complex multivariate feature selection methods are preffered and scoring functions are mainly used as a visualisation aid in BCI. For a further discussion of feature selection methods with merrits and pitfalls please refer to Guyon.

1.11 Sequential floating forward selection (SFFS)

is a filter method of feature selection first suggested by Pudil et al. It is a popular feature selection method in BCI due to its balance of selection performance and computational demand. The SFFS method accomplishes its task by iteratively selecting combinations of features based on some measure (cross-validatoin accuracy is often used). After each new feature is added to the selected feature vector, and once the selected feature vector has achieved a certain size, smaller combinations of the selected feature vector are also tested. If any of these are found to perform better, the previous selected feature vector is replaced and the process continues. The stopping criteria can be the number of iterations completed, the total number of features in the feature vector or performance. Such methods are inherently time consuming, but provide good results by reducing or removing irrelevant features from the feature vector.

1.12 Recursive feature elimination (RFE)

is an embedded feature selection method first proposed by Guyon et al. In an iterative process an SVM is trained, its weight vector examined and the feature or subset of features with the lowest sum squared weight is eliminated. A similar variant first described by Lal et al. eliminates sets of features belonging to a channel at a time and as a consequence is refered to as recursive channel elimination. A number of BCIs that utilise SVMs have used RCE to eliminate unecesary channels, although it is sometimes hard to decern how such a method may aid a highly robust classifier such as the SVM. A drawback of such a method is that everytime a feature is eliminated the model parameters have to be estimated, often by cross-validation, making this an honerous and slow procedure. For more details please see.

2 Classification

There exist a large variety of classifiers which fit the data utilising various methodologies. The classifiers used and presented here can be split into two groups; discriminative and generative. Discriminative classifiers are able to larn frmo and classify data by defining a loss function. The loss function is used as an optimisation criterion during the learning process. A number of optimisation functions exists but only two will be presented here, Fisher's linear discriminant (FLD) and support vector machine (SVM). Generative classifiers use assumbptions about the probability distribution to estimate te necessary parameters in order to minimise misclassification risk. Only one such classifier type will be presented here, linear discriminant analysis (LDA).

2.1 Linear classification

can be simply described as a projection of the input determined by w. If we have n d-dimensional samples with n1 in subset D1 labeled w1 and n2 in subset D2 labeled w2. So from equation corresponding from a set of n samples y is divided into subsets Y1 and Y2. Therefor yi is the projection of the correcsponding xi onto a line determined by w. In order to achieve the best classification the projection onto the line should provide the maximum separation between the two classes. In other words the clusters of the projected input yi should be maximaly distinct with no or minimum overlap. A brief description on how is classifier achieves this is presented below.

2.1.1 Linear discriminant analysis (LDA)

is a simple linear generative classifier that uses the class probabilities for classification. The generalised classification rule used in the LDA is to assign an example to a group with the higherst conditional probability, this is referred to as the Bayes rule (3.8). If there are g groups, the Bayes' rule is to assign the example to group i where:

$$P(i|\boldsymbol{x}) > P(j|\boldsymbol{x}), \quad for\ \forall\ j \neq i \tag{3.8}$$

Given the features $\boldsymbol{x}$, $P(i|\boldsymbol{x})$ is the probability that an example belongs to group i. Determining $P(i|\boldsymbol{x})$ is very difficult, what is more feasible is to estimate the probability of getting a set of features $\boldsymbol{x}$ given that the object comes from group i, ($P(\boldsymbol{x}|i)$). Bayes theorem states:

$$P(i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|i)P(i)}{\sum_{\forall} P(\boldsymbol{x}|j)P(j)} \tag{3.9}$$

$P(i)$ and $P(j)$ are the prior probabilities of groups i and j respectively. The prior probabilities are either assumed to be equal for all groups or are based on the number of examples in each group. Instead of using $P(\boldsymbol{x}|i)$, which would require a large amount of data to get the relative frequencies of each group for each feature. The distributions are assumed and the probability is theoreticaly derived. The assumptions made by the LDA are that the conditional probability density functions are normally distributed and the class coverances are identical. Using these assumtption the linear discriminant function is derived:

$$f_i(x_k) = x_k^T \Sigma^{-1} \mu_i - \tfrac{1}{2} \mu_i^T \Sigma^{-1} \mu_i + log(P_i) \tag{3.10}$$

The example k belongs to the group with the maximum f_i value.

$$G(x_k) = \underset{i}{\operatorname{argmax}} \left[f_i(x_k) \right] \tag{3.11}$$

The LDA relies on the estimates of the mean and covariance data. With high dimensional data and few examples, as is the case with EEG, this estimation can be very imprecise. This can lead to loss of generalisation. The generalisation performance can be improved by introducing a reqularisation parameter γ (proposed for QDA by Friedman). The regularisation parameter increases the lower eigenvalues and decreases the higher eigenvalues of the estimated covariance matrix.

2.1.2 Support vector machines (SVMs)

use a discriminant hyperplane to classify data into one of the two classes. They maximize the margin from the nearest training points in order to find the optimal hyperplane with the best generalization capabilities. SVMs use a regularization parameter C that enables them to cope with outliers, by permiting errors on the training set. Therefore choosing an optimal value for the regularization parameter, often by cross-validation, leads to better generalization and avoids over-fitting.

An SVM such as the one described is known as a linear SVM since it classifies data by using linear boundaries.

The separating hyperplane ensures that

$$y_i(w^t x_i + b) \geq 1 for i \in 1 \ldots N \tag{3.12}$$

The goal of the support vector machine is to find the separating hyperplane with the largest margin, since it assumed that the larger the margin the better the generalisation of the classifier. The margin is defined as the positive distance from the decision hyperplane. The distance from the decision hyperplane to the transformed pattern x is:

$$\frac{|y(w^t x + b)|}{\|w\|} \tag{3.13}$$

Therefore assuming that a positive margin γ exists:

$$\frac{y_i(w^t x_i + b)}{\|w\|} \geq \gamma for i \in 1 \ldots N \tag{3.14}$$

The solution vector can be scaled arbitrarily and still preserve the hyperplane so to ensure uniqueness the constraing $\gamma\|w\| = 1$. So the goal is to find the weight vector w that maximises the margin γ. Besides solving equations the solution also requires to minimize $\|w\|^2$. If the patterns are lineraly seperable the above stated optimisation works fine, al too often through patterns are not completely linearly seperable. Such a situtaion would make a solution that adheres to the afformentioned contraints impossible. To deal with the non-linearly seperable patterns the constraints are relaxed. Therefor the margin of sepeartion is said to be soft if a data point (x_i, y_i) violates equation. This violation can happen in two ways: The data point is located inside the region of separation but on the correct side of the decision surface. This would still lead to correct classification. The data point falls on the wrong side of the decision hyperplane, leading to incorrect classification.

A soft margin is implemented by incorporating slack variables ξ_i, that measure the deviation of a data point from the ideal condition of pattern seperability and a regularisation constant C. The regularisation parameter C determins the scaling of the ξ penalty parameters, therefor it controls the tradeoff between the model complexity (number of SVMs), and the number of nonseperable patterns. Therefore the margin optimisation equations become the following:

$$y_i(w^t x_i + b) \geq 1 - \xi_i for i \in 1 \dots N \tag{3.15}$$

$$\xi_i \geq 0 for i \in 1 \dots N \tag{3.16}$$

by minimising w through

$$\tfrac{1}{2}\|w\|^2 + C\sum_i \xi_i \tag{3.17}$$

SVMs are also able to create nonlinear decision boundaries through the use of the 'kernel trick', which allows the SVM to map the data onto another space of higher dimensionality by using a kernel function. A number of different kernels exist, but the most popular in BCI literature is the gaussian radial basis function kernel [**?**, **?**, **?**]. The use of the RBF kernel adds another parameter that needs to be tuned through cross-validation, i.e., γ(kernel bandwidth - see equation 3.18).

$$K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2) \tag{3.18}$$

It is important to note that the rbf kernel becoms gaussian if:

$$\gamma = \tfrac{1}{\sigma^2} \tag{3.19}$$

with σ being the parameter to be tuned.

SVMs are known to be insensitive to overtraining and to the curse of dimensionality, allowing them to have good generalization properties [**?**, **?**]. This has lead to ther succesfull use in a number of BCIs P300 data (Kaper et al., 2004; Thulasidas et al., 2006), motor imagery data (Schlgl et al., 2005), and data from other cognitive tasks (Garrett et al., 2003). All of this of course comes at the expense of computational cost and execution speed. The toolbox used to train and test the SVM was LIBSVM[**?**].

2.1.3 Fisher linear discriminant (FLD)

The FLD, or Fisher's LDA, is a linear classifier which, much like the SVM, aims to use hyperplanes to separate the data representing the two classes [**?**].

The hyperplane direction aims to maximise the difference between the projected means while minimising the within class scatter.

$$m_i = \tfrac{1}{n_i} \sum_{x \in D_i} x \tag{3.20}$$

The distance between the projected means $\hat{m}_1$ and $\hat{m}_2$ is calculated using equation.

$$|\hat{m}_1 - \hat{m}_2| = |w^t(m_1 - m_2)| \tag{3.21}$$

Where m_1 and m_2 are the sample means of the two classes, w^t is the weight vector. By scaling w we obtain a large ifference between the projected means, this alone does not lead to good separation. We require theseparation of the projected means to be large relative to some measure ofthe standard deviation of both classes. So the criterion function that maximises the between class scatter while minimising the within class scatter is:

$$j(w) = \frac{|\hat{m}_1 - \hat{m}_2|^2}{\hat{s_1^2} + \hat{s_2^2}} \tag{3.22}$$

where $\hat(m_i)$ the d-dimensional mens is given by

$$\hat(m_i) = \frac{1}{n_i} \sum_{x \in D_i} w^t x \tag{3.23}$$

and the scatter

$$\hat{s_i^2} = \sum_{y \in Y_i} (y - \hat{m}_i)^2 \tag{3.24}$$

To obrain $j()$ as a function of w we first define the scatter matrices of each class S_i as

$$\hat{S}_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^t \tag{3.25}$$

then the within-class scatter S_w is

$$\hat{S_w} = S_1 + S_2 \tag{3.26}$$

The between class scatter S_B can be easily calculated as

$$\hat{S_B} = (m_1 - m_2)(m_1 - m_2)^t \tag{3.27}$$

The criterion $J()$ can be wrtten in terms of S_B and S_W as

$$J(w) = \frac{w^t S_B w}{w^t S_w w} \tag{3.28}$$

This expression is known as the generalised Rayleigh quotient.

The vector w that maximises $J()$ must satisfy

$$S_B w = \lambda S_w w \tag{3.29}$$

for eigenvalues λ. So one can determine the maximum J by calculating the generalised eigenvalues λ and eigenvectors w between S_B and S_w and choosing the highest one with the corresponding eigenvector. If S_w is nonsingular and therefore invertible we can obtain

$$S_w^{-1} S_B w = \lambda w \tag{3.30}$$

Since the scale factor for w is inconsiquential to $J()$. The solution that optimises $J()$ is:

$$w = S_w^{-1}(m_1 - m_2) \tag{3.31}$$

if the number of training examples is smaller than the input vector dimesions, the within class scatter matrix Sw becomes singular and can not be inverted. A solution to this is to regularise the FLD. This can be accomplished by replacing equation 3.31 with the following:

$$w = (S_w + CI)^{-1}(m_1 - m_2) \tag{3.32}$$

where $C \leq 0$ is a requlärisation constant that needs to be estimated. For $C = 0$ the FLD is unrequlärised.

The weight vector w has been calculated but the linear discriminant still requires the threshold b to seperate the projected data into the two clases. This may be done by fitting a univariate gaussian distribution to the projections of the classes and choose b where the posteriors in the distribution are equal.

The linear nature of the classifier offers some advantages and disadvantages. The main advantages of the FLD are its computational and conceptual simplicity. Its lack of regularization and its linearity are often cited as its main disadvantages [**?**], although as highlighted here regularisation is possible, it is not widely used in BCI (except for). Despite this, FLD has been successfully applied to a number of BCI problems [**?**] (Bostanov, 2004; Kaper, 2006)(Lalor et al., 2005)(Blankertz et al., 2002;,krusienski$_{c}omparison_{2}006.Whencomparedtomanyotherdif$

2.1.4 Stepwise linear discriminant analysis (SWDA)

may be considered an embedded method that combines forward and backwards feature selection with FLD [**?**] (Draper and Smith, 1981). This classifier is quite popular in P300 BCI but not in other BCIs[**?**]. The input features are weighted using ordinary least-squares regression to predict the target class label. The model starts with non initial features, the most statistically significant input feature for predicting the target label (p-value ¡ 0.1) is added to the discriminant function. After each new entry to the feature vector of the discriminant funciton, a backward stepwise analysis is performed to remove the least significant features (p-value¿0.15). The process is repeated until the stopping criteria are met, which may be a maximum number of features or no more features satisfy the entry or removal criteria.

3 Linear vs. non-linear classification

Linear classifiers can often be outperformed by non-linear classifiers if the data is not linearly seperable. On the other hand even if the data is linarly seperable non-linear classifiers can approximate linear classifiers, since linear classifiers can be considered as a special case of non-linear classifier. As an example if we were to use an SVM with an RBF kernel, the parameter tuning of the C and gamma would include gamma values sufficiently small to consider the kernel process as linear. BCI literature has many examples where non-linear methods are matched or outperfromed by their linear counterparts. There are a number of reasons for choosing linear classification algorithms if they perform just as well as non-linear ones:

Linear classifiers are computationaly simpler and faster than non-linear methods.

Linear classifiers often have few or no parameters to tune. Making the training time significantly shorter.

With the increasing complexity of a classifier the possibility of overfitting the data also increases. Resulting in poor perfromance on unseen data (test data). This is often dependant on the tuning of the classifier parameters, which as mentioned in the previous point takes time.

The wighting learned by the classifier can be used as a discriminability measure of the input vector (as done by the RFE method mentioned earlier). This information can also be used to determine channel locations, time point and frequency bands that are important to classificaiton. The information can then later coroborate or augment known neurophsyiological phenomena, or provide a greater insight into the subject specific cognitive processes.

On-line adaptive classifiers are highly desirable in BCI due to the non-stationary nature of EEG. Linear classifiers are easier and quicker to adapt and train than non-linear classifiers. Ensemble of linear classifiers have been succesfully used to classify BCI data.

With all the afformeniotend benefits it would be easy to discard non-linear classifiers in favour of linear ones. This though would be a mistake since if the data is non-linear linear classifiers will fail and non-linear ones will succeed. In most of the work carried out here besides study 1 linear classifiers were chosen in favour of non-linear ones. This was due to the proven track record of linear classifiers in the established feature domain, as well as perliminary experimentation carried out by myself. The Fisher linear discriminant was chosen since it is almost identical to LDA except in the estimation of the bias (which does not affect classification as will be seen later). The FLD in many of the datasets used here perfroms better than the SVM and is computational much simpler and faster to train (due to lack of tuning parameters). The FLD is used without regularisation in order to minimise training time.

4 Validation

In order for classifier performance to have a comparable meaning its perfromance must be estimated on unseen data (test data). This is because the training error is meaningless since it does not give a measure of the generalisation ability of the classifier. In example, if a classifier such as an SVM is trained with the incorrect tuning parameters it can achieve 100% classification on the training set but on a test set the perfromance is 50% or worse. The reson for this is that the classifier has overtrained and although according to the trainnig erro should provide excellent classification it has not generalised well resulting in poor perfroamcen on unseen data. So in order to estimate the generalisation performance one has to use validation. In validation the training data is split into a training and validation set. The training set is used to train the classifier and the validation set used to estimate its classification performance. The afformetnioed steps have to be carried out a number of times in order to eliminate possible classification over or under estimatetions due to the variances in the features in the training set. The simplest method is the leave-one-out validation method. Where the classifier is trained on all data except one data point and its performance evaluated on that data point. This is then repeated for all data points. The mean error is then a good approximatino of the generalisation error. Leave-one-out validation can take some time so a simpler quicker and equaly as effective method is k-fold cross validation. Here the data is randomly split into k subsets of usually equal size. The classifier is trained on k-1 subsets and the generalisation error for that split is estimated on the remainning subset. This is repeated for all subsets. If the subsets are split so that they contain the same number of each class in each of the subsets the split is termed to be stratified. Numerous other variations on these validation methods exist each with benfits and drawbacks. It is usually best to match the validation method to the amount of data available as well as the nature of the data. The perfromance of the classifier can vary significantly over time due to the non-stationary nature of EEG. This can be a sever problem when validating classifiers in BCI, since it often means fetures selected and classifiers used in one session may not be optimal for the next. This is a major drawback of all validation strategies, since they do not account for the temporal causality that may be aparent in the data and therefor assume taht all data are derived from the same distribution, which is not the case. An alternative method of validation would be to keep the chronological order of the data so that if 5 sessions were carried out then data from the previous session or sessions would be used to train the classifier which would then be tested on a succesive session. Resulting in features that are more stationary across sessions being selected by the classifier. None the less the drawback of this techinque is in order to obtain a reasonably accurate estimation of perfromance a large number of sessions would have to be recorded, failing this the technique would rarely provide a robust estimate of generalisation perfromance. For some classfiers reqgularisation and other model parameters need to estimtated. This can be done a number of ways, but the most robust was is to carry out cross-validation on the training set to find the optimum parameters. Then these parameters are used to train the classifier on the whole training set and tested on the test set. This in essence is two cross-validation loops and inner loop for parameter selection and an outer loop for estimation of generalisation performance. This can be very time consuming. That is why there was a preference for classifiers with no parameter tuning, since a suable BCI system should be available in a matter of minutes not hours. Of course with the ever increasing pace of available computational power this may no longer be a consideration.

5 Robustification

EEG data can be easilly contaminated with artifacts, either by the subject through EMG, EOG and motion or though exteranl sources of electromagnetic interfereance. These artifacts can correspond to what is know in machine learning as outliers, features in the input data that are unrelated to the underlining distribution and therefore can not be predicted by it. BCI systems can be further confounded by individual difference in cognitive performance of tasks, ill undestood effects of stimuli and the non stationarity of EEG. The work conducted here was carried out in a controlled environment where external sources of electromagnetic interference were limited. The subjects were also told not to move if possible and limit blinking. Although numerous steps were taken to ensure near optimal recording during the epxeriments yhroughout the studies presented here artifact removal techniques were not used. Although strong outliers in this setting should not exists. The non-stationarity of EEG was not directly addressed but rather incorporated as an unavoidable fact of BCI.

6 Subjects

Whenever possible a wide range of subjects were utilised, but still the subject pool used was very biased towards male adults in their early twenties till late twenties. With above average cognitive ability. This is a common skewe in BCI research conducted at Universities and has been addressed by studies carried out on patients. The results from those studies indicate that results often presented by the Universities on the operational performance of BCIs are somrwhat overated and the performance of real patients is considerably lower but mostly comparable. This issue is not addressed here but is certainly considered when decisions of improvement and Design of BCIs is explored. Finally the effects of stimuli on the cognitve tastes are often explored is experimental psychology and neurology, but until recently have been mostly ignored in BCI. This is a wide ranging topic with a number of potential advantages. A large amount of the work presented here will be to try and adress this issue.

7 Hardware

Chapter 4

The P300 and Associated Brain-Computer Interfaces

1 Introduction

2 The P300

As discussed briefly in section 3.2 there are two types of event-related potentials, exogenous and endogenous. The latency, amplitude and topography of exogenous ERPs are mainly dependant on the physical characteristics of the stimulus. Endogenous ERPs are produced later and depend primarily on the context of the stimuli, thus they are more likely to represent the concious processing of the stimuli. The P300 is an endogenous ERP whose discovery has lead to a vast amount of research in the neuroscientific and medical communities. The P300 has often been used as a method to assess the neural underpinings of cognition [18]. The P300 was first discovered by Sutton et al. [58] and since then a vast number of studies have try to elucidate the psychological and neurophysiological meaning of the P300 by observing the changes to the waveform by varying the stimuli. Charateristics of the P300 have been linked to gender, age, substance dependance and brain diseases. For a more detailed review please see [59], and [60, 18].

The P300 can be evoked using different paradigms and stimulus modalities. The most common stimulus modalities used are visual and auditory due to their practicality, although other stimulus modalities may also be used (tactile, olfactory, etc.). The paradigm most often employed to elicit a P300 is the oddball paradigm and its variants. The oddball paradigm consists of two stimuli, the standard and the target. These stimuli are presented in a random stimulus sequence, with the standard stimulus appearing more often than the target. The subject is asked to respond mentally or physically to the target stimulus and ignore the non-target (standard) stimulus. The response is necesary to evoke a clear P300 since passive stimulus processing generally produce smaller P300 amplitudes than active tasks [61, 18]. An exmaple of such a paradig would be one where the subject is asked to press a button to, or count the number of target tones (1000 Hz) in a random sequence of 1000 Hz and 2000 Hz tones. Thus eliciting a target-P300 during the presentation of the 1000 Hz (target) tone.

In the three-stimulus oddball paradigm a further stimulus is added, often refered to as the distracter stimulus. The distracter stimulus appears infrequently in the sequence of target and non-target stimuli. The distracter stimulus is often novel and the subject is instructed to ignore it. Variation of the novelty and context of the distracter stimulus also affect the topography of the ERP produced by the distracter stimuli, this will be discussed later. In the single-stimulus paradigm only the target stimulus is presented in the stimulus sequence. Similarly to the other two paradigms described the subject must react to the presence of the target stimulus either mentally or physically.

The P300 is named so due to its positive peak amplitude at around 300 ms post stimulus. The P300 is also refered to as the P3 or late positive component (LPC) [18]. If an infrequenct distinct tone is presented in a series of frequent tones, similar to the standard oddball paradigm, and the stimuli are simply observed, then the infrequent tone will produce a positive going waveform with a short peak latency and central/parietal maximum amplitude distribution. This componenet was named P3a [62]. The same paradigm but with a task assigned to the infrequent stimuuli prodecues a task relevent potential, P3b. The P3a is not exclusive to the adutitory modality since it has been observed in the visual modality [63]. A number of ERPs that appear related to the P3a can be produced by modifying the stimulus context and properties of distracter stimuli inserted into a target/standard sequence (three-stimulus oddball paradigm) . The novelty P300 has a frontal/central amplitude distribution, a short latency and repidly habituates [64]. The novelty P300 is elicited when perceptually novel distracteers occur in a series of more typical stimuli. It is also present across modalities [65, 66] and habituates quickly [**?**].

The no-go P300 is elicited if in the three-stimulus paradigm with the distracter being a non-novel repeated stimulus. Subject still only respond to the target stimulus and not the distracter [67]. The topographic distribution of the no-go P300 is somewhat more central than the P3b [68]. The no-go P300 has been linked to response inhibition mechanisms, especially when motor responses are used for target stimuli [69, 70].

It has been suggested that the P3a, novelty and no-go P300 are likely to be variants of the same ERP that vary in their scalp topography as a function of the attnetional and task demands [18]. To summarise the P3a seem to indicate attentional focus for stimulus evaluation, where as the P3b occurs when the attentional resource activations promote context updating and subsequent memmory storage [71] [18].

The P3b is also influence by a number of stimulus properties, some of these are discussed below:

Target probability: Initial studies carried out on the P300 elucidated the role of target stimulus probability and task relevance. The observation was that peak amplitude of the target P300 (P3b) was inversely related to the probability of the target stimulus in the random sequence. Therefore higher peak amplitude P3b waves were evoked with lower target stimulus probabilities than with higher target stimulus probabilities. The relationship between target probability and P300 amplitude holds true for a wide range of probabilities [72] and across a number of modalities (auditory, visual and somatosensory)[73] **??**[74]. Besides the global target probability, i.e the target stimulus occurs 10% of the time in the whole paradigm, local probabilities also affect the P300 amplitude, i.e the number of non-target events that occur before target event [74] [75]. Many P300 stimulus paradigms use this as a guidline very often opting for low target stimulus probabilites in order to elicit consistent target P300s.

Inter-stimulus interval (ISI) Another factor that has been found to affect the amplitude of the P3b is the inter-stymulus interaval. The inter-stimulus interval is in essence the time between two consequtive stimuli. Several studies have indicated that shorter ISIs evoke smaller ampltidue P300 than longer ISIs [76] [77] [78]. It has been suggested that ISI has a greater effect on P300 amplitude than target probability, this is exemplified by the observation that target probability has little to no affect on P300 amplitude for ISIs of 6s or longer [76] [27] [79].

Target-to-Target Interval (TTI) Target-to-target interval is defined as the time taken between two consequetive target stimuli in the stimulus sequence [80] [81]. Decreasing target probability, increasing non-target sequence length and increasing the ISI all increase TTI. Therefor the interactions of ISI and target-stimulus probability can be simply interpreted as the effect of TTI. With longer TTIs providing larger P3b amplitudes than shorter TTIs. This theory is supported by P300 findings from single-stimulus paradigms. Comparable P300 components are obtained from both single-stimulus and oddball tasks across a range of task variables and stimuli [82] [61] [83] [84]. Similar waveforms, topographic distributions, and dipole coordinates have been found between both tasks during source localisation studies of the P300 [85] [86]. Thus, even when the global and local probability of the target stimulus is unitary, the time between the target events is the primary determinant of P3b amplitude. This effect may be similar to the observed affet of ISI on sensory ERPs e.g., [87] [88] [89] where the ERP is modulated by a ”recovery cylce”.

Attention The amplitude of the P3b is very much dependant on the subjects attention to the task. Consequently smaller P300 amplitudes are observed during passive stimulus processing than active stimulus processing [61, 18]. The P3a on the other hand remains largely unaffected by modulation of attention, thus it is often observed even if the subject choses to ignore the stimulus.

Habituation The P3a has been found to habituate rapidly [64] [90], the P3b on the other hand seems to show littel or no habituation affect.

Task difficulty As the task difficulty increase so does the latency of the P3b peak, with the amplitude having the inverse relationship. However, if the task demands, even though increased, are still within the subjects capabilities the P300 amplitude may increase as the subject devotes more resources to the task. The task difficulty is often modulated by the discriminability of the target from the standatd tones, with less distinct tones being more difficult than distinct tones. In contrast the P3a amplitude increases as the task difficulty increases.

Although the P300 and its variants have been well studied the underlining congnitive processess that create and modulate have not been yet well defined. The P300 is affected by a wide variety of psychological and physical parameters, this often makes the task of consistently obtaining a clear P300 under varying conditions and across a number of subjects a hard task. Therefore it is important to try to keep all of the above observations in mind when designing and using P300 based protocols.

3 Brain-Computer Interfaces Utilising P300

The premise of the P300 BCI is that the user modulates their attention to the targets by carrying out a covert or overt task to the desired stimuli, while igonring others. This should induce a target-P300 (P3b) whenever the desired stimulus is present, thus the target can be inferred from the time locked responses to the random stimulation patterns. More specifically, a P300 BCI would be based the following procedure. The user is presented with the possible commands, the user decides which command they wish to select. The stimuli are modulated in a random order and the user tries to concetrate on the stimulus that corresponds to their desired command. During the presentation of the stimuli the users EEG is recorded and later analysed. The analysis procedure will contain some pre-procesing and importantly a classification algorithm. The goal of the analysis is to infer from the EEG which stimulus the user had chosen as a target. The stimuli can be visual, auditory or somatosensory, a number of BCI paradigms will be presented below with a focus on visual P300 paradigms.

3.1 P300 Matrix Speller

The P300 matrix speller was created by Farwell and Donchin and was the first P300-based BCI [44]. In their BCI a 6 by 6 matrix of letters, words and other symbols was presented on a computer screen. All the rows or all the columns were flashed in a random order and after a brief pause the orientation (rows or columns)that had not been flashed were also flashed. The subject would pay attention to a specific symbol and count the number of times the row or column intersecting that symbol flashed. The flashes of the row and column that contained the target symbol were the target stimuli and subsequently evoked a target-P300 response. The other rows and columns that were not targets were ignorred and susequently did not evoke a target-P300 response. Therefore the target symbol could simply be infered by identifyin which row and column inviked a target-P300 responses. The Farwell and Donchin matrix speller is a well known and subsequently commonly used BCI protocol. A number of extensions, modifications and investigations have been carried out on the P300 speller paradigm. The most frequent modification are the stymulus-onset asynchrony (SOA) and the number of symbols in the matrix. In the original paper by Farwell and Donchin two SOAs (in the manuscript they are refered to as inter-stymulus interval) were tried, 125 ms and 500 ms. The results indicated better results with the longer SOA. Allison and Pineda tried SOAs of 125 ms 250 ms and 500 ms and found larger P300 amplitudes for the longest SOAs [91]. In contrast Meinnicke et al. [92] and Sellers et al. [93] reported higher accuracies with shorter SOAs. Menicke et al. tried SOAs of 150 ms, 300 ms and 500 ms where as Sellers et al. tried 175 ms and 350 ms. It is unclear why even through Allison and Pineda reported higher P300 amplitudes for longer SOAs both Sellers et al. and Menicke et al. achieve higher classification rates using shorter SOAs.

The effects of the differing number of elements in a matrix were also explored. The first to do so was Allison and Pineda, who tested square matrices with 16, 64 and 144 elements [45]. The outcome of the study was that P300 amplitude increased as the size of the matrix increased. In contrast the work conducted by Sellers et al. [93] indicated higher accuracies for the smaller matrix size. The square matrix sizes tested were 9 and 36 symbols. Other interesting modification to the Farwell and Donchin speller include the modification of flash patterns by Allison and Pineda [91], the flashing of singular items instead of row columns by Guan (guan_2004) and use of different stimulus modulation by Martens et al. [94]. All of the aformentioned papers have opened avenues for further exploration into improving the performance of the P300 BCI. Although as highlighted by the disparity in the conclusions obtained by some of the studies further investigation into these modifications and others is of paramount importance.

3.2 Cursor Control

Besides spelling matrices another interesting use of the P300 has been in cursor control. This was first explored by Polikoff et al. The implementation consisted of a fixation cross with target arms in the north, south, east and west directions displayed on a computer screen [47]. The targets were crosses at the ends of the arms. Two stimulus conditions were tested one where the crosses were replaced by asterisk or null-stimulus. The user payed attention to a particular orientation and counted the number of times the cross was replaced by either an asterisk or the null-stimulus. Cursor control was not implemented in the study but offline analysis showed that it would be possible. An interesting study into the possibilities of using P300 for cursor control was produced by our group [95]. Here no binary decisions were made to the presence or absence of a P300 but the direction was determined by a filter that uses wavelet features and is optimised by an evolutionary algorithm. The results were encouraging although they could not be compared to earlier work by Polikoff et al. due to the absence of binary classification.

3.3 P300 in other stimulus modalities

The P300 is not confined strictly to visual stimuli but can be also induced by auditory and tactile stimuli. It is often possible that patients may be unable to direct gaze, adjust focus or perform eye blinnks precluding them from using visualy based BCIs. Subsequently other modalities have to be explored. The most popular alternative modality is the suditory stimulation with a number of recent publications [96][97][98][99].

The other modality is somatosensory where the user pays attention to vibrotactile stimulation around the torso. The preliminary study published by Brouwer and van Erp showed encouragin results of vibrotactile P300 BCIs (Graz Conf page 280).

Both modalities have not recieved the same kind of attention the visual sensory modality is recieved therefore a number of questions still remain, such as how fast and accurate can such systems be and whether the target audience would truly benefit from this diferent sensory modality.

3.4 P300 and ALS

The main goal of BCIs is to provide a communication and control channel to patients who can not use the normal output pathaways [2]. Over the last few years a number of publication has provided encouraging results as to accomplishing that end. Sellers and Donchin utilised a 4 element square matrix with the four stimuli being YES, NO,PASS and END [99]. The stimuli were presented auditorily, visually or in both modalities. The results of the study showed that the majority of the disabled subjects were abled to communicate using the BCI as well as the able bodied subjects tested in the study. Piccione et al. utilised a visual paradigm where the stimuli were flashing arrows presented at the periphery of the visual display [100]. The goal of the paradigm was to move a cursor based on the direction of the arrow stimuli. The study used 5 severly handicapped and seven able-bodied subjects. The results of the study were that the handicapped subjects were able to control the BCI, albeit with lower accuracy than the able bodied subjects. Hoffman et al. tested a visual paradigm with 6 stimuli (choices) that allowed five severly disabled ALS subjects and 4 healthy subjects to control various devices, a window and a door [101]. In general the able bodied subjects were able to achieve higher communication rates than the disabled subjects. Nevertheless the disabled subjects were able to use the system with a high degree of success. Nijboer et al tested the classical Farwell and Donchin protocol composed of a square matrix of 36 or 49 elements on 8 ALS patients [102]. The study showed that the ALS patients could use the P300-based BCI for text writing and the performance did not degrade over a period of weeks and months. All of the results are highly encouraging for P300-BCIs not least because of the performance achieved. Over the next few years the exact limitations of the different stimulus modalities will have to be determined more rigourusly concerning this very important target audience.

4 Classification Schemes for P300-Based BCIs

In order to interpret the users interaction with the BCI interface through the EEG recorded some steps are necessary to preprocess classify and interpret the resutsl. The input to the classifier is usually featrues extracted from the EEG by some preprocessing method. The EEG is usually segmented into epochs that are time locked to the stimulus. The epochs are then preprocessed to reduce the amount of redundant a spurious information. A classifier then classifies the data and outputs a score for each epoch. The score signifies to what degree a P300 is detected in the example epoch tested. The scores are then aggregated over a number of trials and the aggregate score together with the stimulus identities and temporal information is used to determine the target stimulus. In section the aggregation procedure is described. Then in section the most common preprocessing and classification methods utilised in P300 BCI are briefly presented.

4.1 Trial Aggregation Procedures

If we abstract the classification procedure for the time being and say that it simply gives us a score, there is still the question of how the scores from the various epochs and therefore stimuli should be aggregated. The aggregation procedure depends heavily on the stimulus method used. Here we will focus on how the score aggregation is carried out in matrix visual stimuli, much like the protocol first proposed by Farwell and Donchin. Let us say that our speller matrix is made up of 36 symbols arrnaged in a 6 by 6 square. Each of the rows and columns of the matrix has to be modulated (usually intensified) once within a trial. Therefore each trial is composed of 12 intensifications. Each of the intesnifications correspond to an epoch extracted from the EEG. As mentioned earlier once preprocessed and classified these epochs are given a score. The row and column with the highest score are the ones selected and the symbol that is present at their intersection the target symbol. Simply using one trial to determine the desired symbol, depending on the SOA, provides very high information transfer rates, but unacceptale accuracy. Consequently the scores for the differing rows and columns are combined over multiple trials to improve the accuracy (see equations 4.1 and 4.2) at the cost of information transfer rates.

$$\text{predicted column} = \underset{col}{\operatorname{argmax}} \left[\sum_{j_{col}=1}^{J} w \cdot f(x_{j_{col}}) \right] \tag{4.1}$$

$$\text{predicted row} = \underset{row}{\operatorname{argmax}} \left[\sum_{j_{row}=1}^{J} w \cdot f(x_{j_{row}}) \right] \tag{4.2}$$

The process described above is the most common but within this basic framework further nuasense exist such as; The aggregation of the row and column scores does not happen seperately but are combined to provide indivisual scores for each of the symbols in the matrix [43]. In theory there should be no apreciable difference between the aggregation procedures. In the afformetnioedn procedure the rows and columns within each trial are flashed in a random manner meaning there is no segregation between column or row intensifications. Other implementations include flashing all rows and then all columns (or viceaversa) with or without a pause in between. Serby et al. found that the subjects in their study provided better results for row classification than column classification therefor in their online procedure would tend to repeat more trials for the column stimuli than the row stimuli. In order to make up for this imbalance in classification. No empirical evidence has been put forward to suggest which of these flashing pattern methods is superior. Another issue of variance between matrix speller implementations is the number of trials used before a decision is made. A number of studies have kept the number of trials constant similarly to the original farwell and donchin matrix speller, with differing trial lengths ofcourse. More recently though there has been evidence that usign variable number of trials dependant on some criterion provides high an accuracy with a faster information transfer rate (refered to as adaptive stopping by Hoffman) [103] [104] [105]. Furthermoe seperating the number of trials used for the differing matrix orientations (rows and columns) may confer an even grater advantage [103].

4.2 Pre-processing and Classification Algorithms

One of the most active areas in P300 BCIs and BCIs in general has been pre-processing and classification. In part due to the importance of pre-processing and classification in BCI paradigms, but also due to the popularity of the BCI competitions. As mentioned in the previous section the accuracy of the target symbol selection is dependant on the aggregations of the trial scores. With the scores determined by a classification method, which indicate the presence of a target-P300 in the epoch or not. Discrimination between the P300 and non-P300 epochs is a non-trivial matter due to the small amplitude of the P300 wave in comparison to the ongoing EEG and the high inter-subject variability of the P300s, latency, topography and amplitude. This is further exacebated by the high dimensionality of the EEG data due to the large number of channels and high sampling frequencies. The usual approach taken to solve the classification problem of P300-BCIs it to gather sufficient data from the subject. Carry out a predefined pre-processing methodology, which will reduce the amount of redundant features. Train a supervised machine learning algorithm to solve the binary classification problem of the presence or absence of target-P300 from the extracted epochs. Then this trained classifier in combination with the pre-processing method can be used to classify unseen data from that subject.

A common approach used is to select a subset of electrode locations where one expects there to be a high P300 amplitude. The raw EEG from each of the electrodes is segmented from stimulus onset into epochs and then bandpass filtered and downsampled. The filtered and downsampled epochs from the selected electrodes are combined into a feature vector and fed into a machine learning algorithm [43]. The filtering and downsampling can be accomplished a number of ways with one of the most common being using a moving average window filter [106] [101] [107]. The parameters used for the moving window average filter depend on the original smapling frequency and the desired sampling frequency.

In the method described by Kapper et al. 10 channels were used consisting Fz, Cz, Pz, Oz, C3, C4, P3, P4, PO7 and PO8 [108]. The data was filtered between 0.5 and 30 Hz. The epochs extracted from each of the channels were 600 ms. It is important to note that the data was not downsampled.

Thulasidas et al. used a total of 25 channels out of 64 with the slection being around the midline, C3, C4 and two distant parietal locations P7 and P8 [107]. The data was low-pass filtered using, as described by them "an optimal cut-off frequency". The data was then downsampled using a moving-average window. PCA was later applied to reduce the 25 channels to 20. In addition to the downsampled signals, estimamtes of the time-derivatives of the signals were used. This improved the classification accuracy according to Thulasidas et al. [107].

The wavelet transform is an excellent decompositiion tool for ERP analysis due to its excellent joint time-frequency resolution. Donchin et al used db4 DWT in combination with an SWDA classifier and found it supperior to just using the standard filtering and downsampling [43]. Glassman created a unique wavelet (SNAP) and used it for feature extraction for P300 and sensory motor rhythm BCIs with reasonable success in both [109]. Bostanov used the CWT with a mexican hat wavelet to transform the extracted epochs into wavelet coefficients [110]. A t-test was used, during training, to identify the coefficients for which there was small intra-class variance but large inter-class variance. The coefficients selected were classified using an LDA. The method pioneered by Bostanov was among the best performaing entries for the P300 and SCP datasets in BCI competition 2003. For more information on wavelets please see section , Samar et al. [111] and [112].

Xu et al. for BCI competition 2003 bandpass filtered the data between 2 and 8 Hz [113]. The data was the epoched into epochs of 650 ms from stimulus onset. The epochs with the same stimulus code were then averaged together. The number of channels was reduced from 64 to 22 by using PCA. Later ICA was used to compute the independant components from the selected channels on the training data and an automated method is used to determine the most relevant features. Lenhardt et al. extracted 800 ms epochs from the ongoing EEG starting at stimulus onset. The data was later bandpass filtered between 1 and 9 Hz [105]. PCA was later applied and the top 100 principal componenets were chosen for classifciation. The number of principal conmpornents was decided through educated guessing. As mentioned earlier Thulasidas et al. used PCA to simply reduce the number of channels from the originally selected 25 to 20 [107]. Makeig et al. were the first to apply ICA to the analysis of ERPs [114][115]. Similar to PCA ICA is a blind source seperation algorithm that makes certain assumptions about the nature of the data. ICA is considered supperior to PCA snce it does not rely on the dataq being orthogonal. Serby et al. extracted epochs of 500 ms in duration strting 100 ms after stimulus onset [103]. Three channels were used Cz,Pz and Fz. The data was later lowpass filtered at 6 Hz. Then the independant components are computerd from the training daata using ICA. The componenets that most represent the target-P300 were determined manually for each subject. Piccione et al. recorded EEG from 4 channels along the midline (Fz, Cz, Pz and Oz) [100]. The data was filtered between 0.15 and 30 Hz, epochs started 500 ms before stimulus onset and ended 1000 ms after stimulus onset. ICA was the used to computer the independant componenets from the extracted epochs. The independant components that most represented the P300 ERP were selected using a fuzzy method.

Once the preprocessing has been carried out the preprocessed epochs need to be given a score to indicate the presence or absence of a target-P300. A popula choice is the SVM. The SVM is a very powerfull machine learning algorithm ith some very desirable properties. The details of the SVM are discusse in section. They have been succesfully used by Kaper et al. (The winner of BCI competition 2003) Thulasidas et al. and Rakotomamonjy et al.(winner 2004) [108] [107] [116] [117]. Thulasidas et al. and Kaper et al. both used non-linear SVM with a Gaussian Kernel. Rakotomammonjy et al. used a number of linear SVMs in an ensemble, with the channels selected for each SVM throghu RCE. For details on RCE and specifically RFE please refer to section.

Simple linear classifiers such as the FLD and LDA have also been used with great success [101][118][105]. Krusieski et al. compared a number of classifiers including an SVM with a Gaussian kernel, a linear SVM, an FLD and SWDA. The results concluded that the FLD and SWDA provided the best performance and that the non-linearity of the Gaussian SVM hindered perfroamcne rather than aid it. Hoffman et al. also created the BLDA and compared it to the FLD and concluded that the FLD and BLDA provided similar perfroomance if the number of features was kept low, but the BLDA outperformed the FLD as the number of features increased beyond a certain threshold. Lenhardt et al. were able to achieve an average transfer rate of 29.35 bits/min with an accuracy of 87.5% using an LDA for classificaiont.

Stepwise linear discriminant analysis is an extension of FLD which incororates forward and backward feature selection [106]. This method has been used by a number of publications and was found to be one of the best performing classification methods by Krusienski et al. [118].

As shown by the plethora of preprocessing and classificaiotn methods described above there are certainly many ways by which one can accomplish P300 classificaiont. Many of the above methods have only been quantified on proprietory datasets with no comparisons to other popular classificaiont methods. That is why the BCI competition data sets are so important since they provide a benchmark by which all classification and preprocessing methodologies can be assesed. Another alternative is to make the data used in the study easily available to others so that they might test their own classification methods on them. Hoffman et al. are an excellent example of this, since not only have they provided the data but also the code for the classifier they designed [101]. The preprocessing feature selection methods can be broadly placed into two categories, ones that rely on prior knowledge and those that use machine learning to identify the relevant features. Either category has benefits and drawbacks, but the most succefull algorithms should in theory incorporate both. For detailed review of preprocessing methods for BCIs please see [119], and for a detailed review of classification methods for BCI please see [120].

5 BCI Systems Evaluation

Making a reasonable comparison between the wide variety of P300-based BCI protocols can be a very hard if not impossible. This is due to the different P300 vs non-P300 classification methods, stimulus presentations, trial aggregation procedures and protocol timings. Nevertheless, a number of key propertises of a BCI system will be described and metrics provided where possible.

Information transfer rate is an important aspect of any BCI system since it indicates the expected communication speed achievable by said system. In P300 BCIs the ITR depends on the SOA (ISI), classification accuracy, number of stimuli and other temporal properties of the protocol. The ITR, sometimes refered to as btrate, is measured in bits/min and is the amount of information (bits) that can be transfered in a certain amount of time [121]. The ITR is calculated using the following formula:

Where N is number of different commands a user can send, p is the probability that the command is correctly recognised and t is the time in seconds necessary to send a single command. Although the metric is very informative the bitrate provided should be considered as mostly a theoretical value. As can be seen by equation the bitrate mainly depends on the number of commands possible, the time taken to send a single command and the accuracy of a single command. In paradigms where the number of commands and the time taken are exactly the same or similar, the accuracy is the only point of variance. Therefore the accuracy of the single trial classification can be used as the performance metric. The amount of accuracy necessary the number of commands available and the time taken (or number of trial) to perform an action will depend on the demands and needs of the user and task. At this moment in time all of the afformentioned properties need to be balanced, often improvements in one are to the detriment of the rest. Other than mettrics that relate to the performance of the BCI further properties are also important in order to make the system applicable to severly handicapped users. One important factor that was until recently largely overlooked was the often limited cognitive capacity of such users [10] [99] [100] [101] [122]. One very good example has been the hunt for ever shorter SOAs in Farwell and Donchin protocol without the consideration of cognitive load [92] Guan 2004 [93]. Such systems may perform admerably when used by Ph.D students, but when used by the disabled subjects found to be considerably less satisfactory. Problems such as reduced congnitive agility, attention and visual acuity all need to be considered when new BCIs are designed and tested. Ideally any newly designed BCI system needs to be tested on handicapped subjects if that is its intended audience.

If the system is to be outside the lab issues of practicality come into play. The time required to prepare and place the electrodes is one very important issue, with many manufactures proposing dry electrodes.Although their use by the broader BCI scientific community has yet to be approaved. Subsequently systems that use large numbers of electrodes may be less prefarable due to the time required to setup than systems with fewer electrodes. Ofcourse a balance has to be struck between the number of electrodes necessary and the accuracy/ITR achievable by the system. Another property of the designed system that must also be considered is the time and effort required to train the classifier. Systems with long training data gathering sessions and computationaly expensive classifier training would be less than desirable. Consequently properties of the classification algorithm such as amount of data necessary and training time are nearly just as important as the accuracy the classifier finally obtains. Also of importance is the computational complexity of the algorithm during execution since this may also prevent its adoption. So far the number of pre-processing and classifciation methods proposed have been used in a laboratory environment, therefore such things as parameter tuning and general intervention are considered acceptable. Outside of the laboratory such specialist knowledge will not be available to the user therefore necesary user intervention should be limited if not avoided completely. Another trate of BCI systems is that some require a period in which the subject has to train to use the BCI.In BCIs such as the P300 this is not required, although a period of adjustment to the task may show a slight improvement over the first few minutes. It is unclear whether over time subjects performance improves, not neccessarily during one session, since fatigue will often mask any improvement. Although work published so far indicates no real improvement or perfroamcne drop over time. A system that allowed a subject to use a system reliably and quickly without training and then over time through the subjects performance of the repetative task the BCI system could incorporate further features in order to further increase performance. Such systems would most likely incorporate two types of BCI modalities something that has not previously been done, but has been suggested by Wolpaw et al. [2].

6 Conclusion

Chapter 5

Visual modifications on the P300 speller BCI paradigm

1 Introduction

The P300 BCI protocol can be split into the following three subproblems: *a)* protocol design *b)* signal processing/feature section *c)* classifier training. Many attempts have been made to improve the transfer rates and accuracy of this BCI protocol by improving results in each of these subproblems [117, 120, 119, 43, 103, 123, 109, 118]. Although modification to matrix element dimensions, flash patterns, stymulus-onset asychrony (SOA) and inter-stymulus interval (ISI) have been explored [45, 91, 93, 92]include s martens , the effects of inter-symbol distance, symbol size and differing foreground and background colours have not.

The work presented here aims to tackle subproblem a, in other words, to ascertain whether the simple modifications to the visual protocol carried out in these experiments will provide classification differences between them and what these differences will be.

2 Methods

It is important to keep in mind that the methods for this experiment were chosen in order to allow comparability between the various visual protocols with as little bias from the classifiers and the preprocessing methods used. Therefore, many state of the art methods such as discreet wavelet transforms and recursive channel elimination have been ignored in favour of simplicity and comparability [109, 117].

2.1 Signal Processing

The EEG data were collected with a Biosemi ActiveTwo system at a sampling rate of 512 Hz. A total of 66 channels were used, with two of the channels placed on the earlobes as references. The channels were referenced to a linked ear reference after data collection. The data were bandpass filtered at 0.1-30 Hz using an 8th order Butterworth filter. Out of the 64 channels, 8 were selected. These were Fz, Cz, Pz, Oz, P3, P4, PO7, PO8, since they have been shown to provide good classification performance [106](see figure 5.1 for illustration). Following each intensification, 800 ms segments of data were extracted from each of the chosen channels. Similarly to [118] the segments were then moving-average filtered and decimated by a factor of 19. The preprocessing technique used resulted in a feature vector of 168 (i.e. (512 * 0.8)/19 samples * 8 channels) elements.

Different channels to the ones used may have provided better results, but as mentioned previously it was not the purpose of this work to try and achieve the highest classification accuracy scores, but rather to have a setup that was, as much as possible, comparable across subjects. None the less the channels selected have been shown to yield high classification performance in the P300 speller paradigm [118, 106].

2.2 Abbreviations

In order to save space and make figures more readable, abbreviations have been introduced throughout. They can be seen in 5.1.

2.3 Experimental Protocols

A total of eight subjects were used (6 males and 2 female between the ages of 19-28, average age 22.3), all able bodied, with each subject carrying out a series of six different experiments.

The experiments undertaken in this study were as follows (the first item is considered the original Farwell & Donchin protocol, although the timing parameters used were different):

1. Black Background: Where a grid of white characters is displayed on a black background.

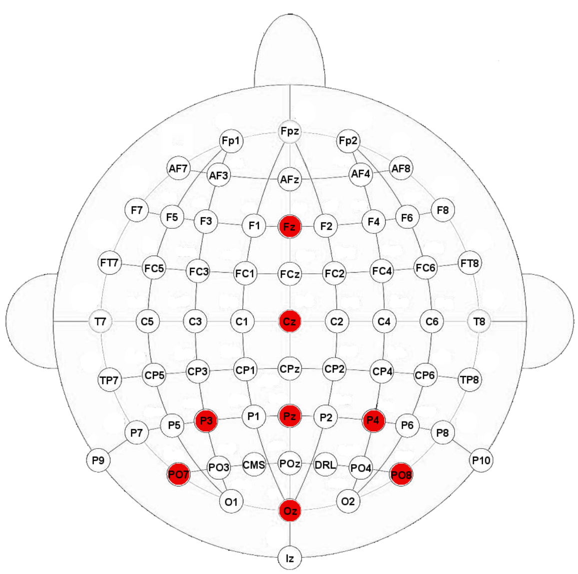


Figure 5.1: BioSemi cap layout. Selected channels are highlighted in red.

Table 5.1: Table of Abbreviations

Abbreviation	Full Name	Description
WB	White Background	Visual protocol with the white background.
BB	Black Background	Visual protocol with the Black background.
LISD	Large Inter-symbol Distance	Visual protocol with the largest distance between symbols/characters.
SISD	Small Inter-symbol Distance	Visual protocol with the smallest distance between symbols/characters.
SSS	Small Symbol Size	Visual protocol with the smallest symbol/characters.
LSS	Large Symbol Size	Visual protocol with the largest symbols/characters.
DT	Distance Top	Distance of the top edge of first row to the top edge of the screen.
DB	Distance Bottom	Distance of the bottom edge of the last row to the bottom edge of the screen.
DR	Distance Right	Distance of the right edge of the right most column from the right edge of the screen.
DL	Distance Left	Distance of the left edge of the left most column to the left side of the screen.
CX	Character X	Width of the character D.
CY	Character Y	Height of the character D.
ICX	Inter-character X	Horizontal distance between characters. Measured from the center of one character to the center of the next.
ICY	Inter-character Y	Vertical distance between characters. Measured from the center of one character to the center of the next.
BX	Board X	Horizontal distance from the left edge of the first column to the right edge of the last column.
BY	Board Y	Vertical distance from the top edge of the first row to the bottom edge of the last row.

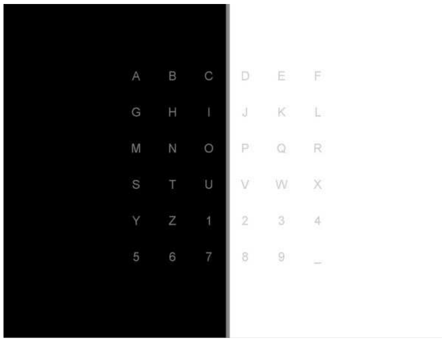
2. White Background: Where a grid of black characters is displayed on a white background.

3. Large Symbol Size: Where a similar setup to the Black Background is used except that the size of the symbol/characters is increased.

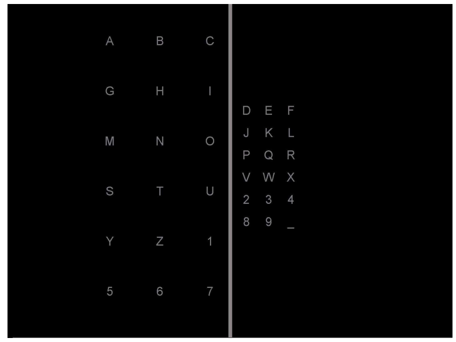
4. Small Symbol Size: Where a similar setup to the Black Background is used except that the size of the symbol/characters is decreased.

5. Large Inter-Symbol Distance: Where a similar setup to the Black Background is used except that the distance between the symbols is increased.

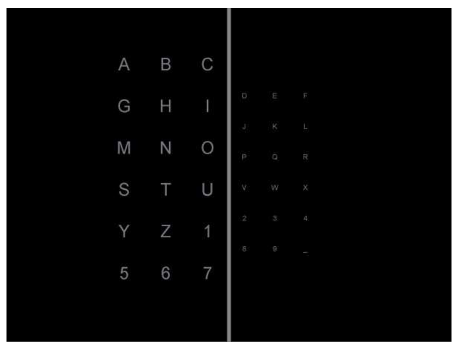
6. Small Inter-Symbol Distance: Where a similar setup to the Black Background is used



(a)



(b)



(c)

Figure 5.2: Samples of the visual protocols: (a) White and black background visual protocols (b) Large and small inter symbol distance visual protocols (c) Large and small symbol size visual protocols.

except that the distance between the symbols is decreased.

Samples of the visual protocols can be seen in 5.2. Each visual protocol would take up the whole screen, the half screen representations depicted in 5.2 are just for brevity.

For all visual protocols except WB, the background colour in the 8-bit **RGB** model was (0, 0, 0)(*black*), the default character colour was (120,120,120)(*dark grey*) and the highlighted row/column colour was (240,240,240)(*light grey*). For WB, the background colour was (255,255,255)(*white*), the default character colour was (200,200,200)(*grey*) and the highlighted row/column colour was(0,0,0)(*black*) (see 5.2). The metrics used for each visual protocol can be seen in 5.2 and 5.3. Please refer to 5.1 for an explanation of the abbreviations used.

Table 5.2: Metric values for each visual protocol (cm)(see 5.1 for explanation of abbreviations).

Board metric	WB	BB	LISD	SISD	SSS	LSS
Distance Top	4.5	4.5	0.8	8.3	6.4	2.7
Distance Bottom	5.1	5.1	1.6	9	6.8	3.8
Distance Right	8.7	8.7	5	12.5	10.3	7.2
Distance Left	8.7	8.7	5	12.5	10.3	7.2
Character X	0.7	0.7	0.7	0.7	0.4	1
Character Y	0.8	0.8	0.8	0.8	0.5	1.3
Inter-Character X	3.8	3.8	5.3	2.3	3.2	4.3
Inter-Character Y	3.8	3.8	5.3	2.3	3.2	4.3
Board X	20.1	20.1	27.3	12.6	16.5	23.4
Board Y	20.1	20.1	27.3	12.6	16.5	23.4

The order in which each subject executed the experiments was randomized. In the majority of cases, the 6 experiments were carried out over 2-3 days (2-3 experiments each day). This was to ensure the subject would be able to provide each experiment with their full attention. The only subject for whom this was not done was subject 5, who carried out all six experiments in one day. This was due to the subjects time constraints. During the running of the experiments, the subjects had 5 minutes to relax before the next experiment. If the subjects required more time, this was given.

From each experiment, a total of 39 characters were spelt. These characters spelt out

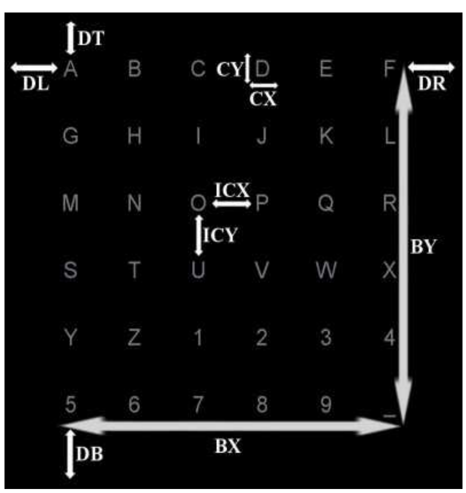


Figure 5.3: Size metrics for visual protocols.

the sentence "THE_QUICK_BROWN_FOX_JUMPS_OVER_LAZY_DOG ". This sentence was chosen since it includes all the letters in the alphabet. Each subject was presented the whole sequence of letters on a piece of paper during setup and was allowed to retain it during the experiment. During the experiment no feedback was given to the subjects as to whether they performed the task well or not.

The subjects were seated at 1 meter from a 19 inch TFT screen. Great care was taken to ensure that the subject was relaxed and alert throughout the whole experiment. The stymulus-onset asychrony (SOA) was 300 ms with an inter-stymulus interval of 150 ms or if the definition provided by Farwell and Donchin [44] is used a inter-stimulus interval (ISI) of 300 ms, with a stimulus intensification period of 150 ms. Each character epoch consisted of 10 trials, with each trial consisting of 12 flashes of each row or column. The rows and columns were flashed in a random manner. During each trial, all of the rows and columns were flashed once. In each character epoch, the user would focus on one character in the sentence. The sentence was spelt out in the order shown in the previous paragraph. After each character epoch, there was a pause of 2500 ms in which the subject located the next desired symbol on the board. After 500 ms into the pause, an auditory cue was presented to indicate to the user that they must move on to the next character if they have not already started to do so. This results in the user selecting a character every 38.5 seconds. For a graphical representation of the protocol see 5.4.

2.4 Classification

The classifiers used in this work were a support vector machine (SVM) with radial basis function (RBF) kernel and a Fisher's linear discriminant (FLD). The process and data distributions used to train both classifiers were identical except that parameter tuning was carried out for the SVM, which was unnecessary for the FLD. The presence of absence of the target P300 from the chosen EEG features is a binary classification problem and the

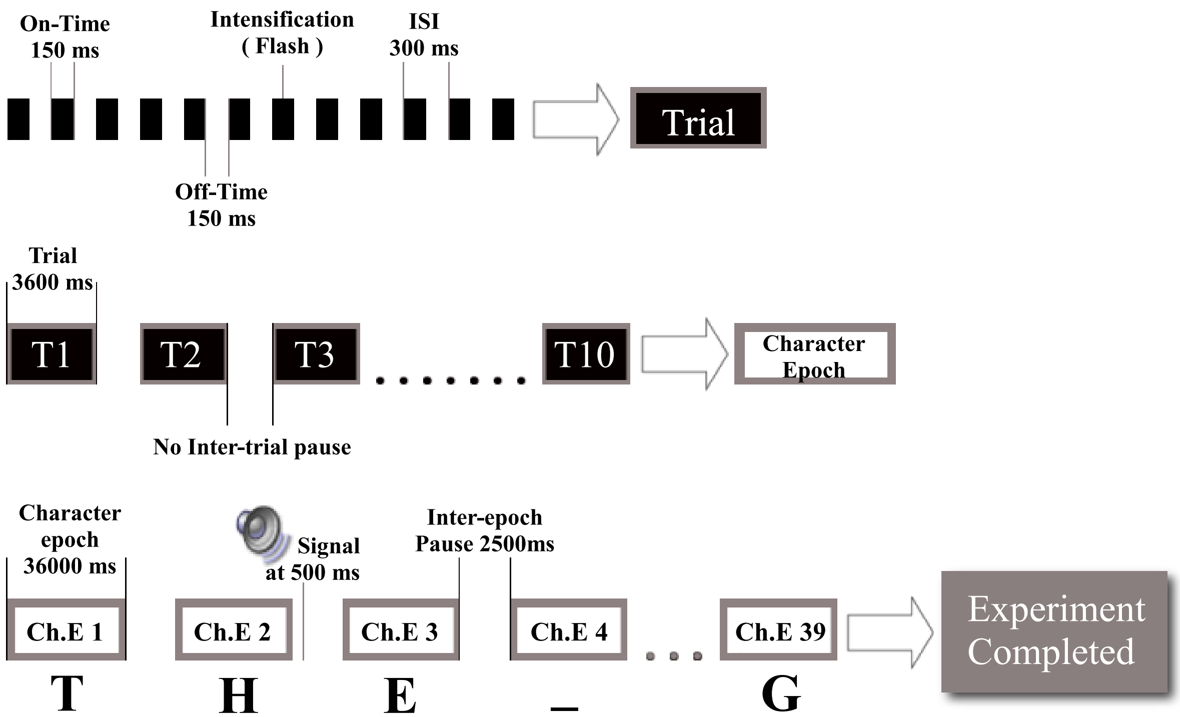


Figure 5.4: Experiment Protocol - The trial is made up of 12 sub-trials, with ISIs of 300 ms and intensification period of 150 ms. Each epoch is made up of a series of 10 trials which last 3.6 seconds, with no inter-trial pause. Thus the whole experiment is made up of 39 character epochs. There is a 2.5 second pause between each character epoch.

decision hyper-plane of the discriminant function is defined by equation 5.1.

$$w \cdot f(x) + b = 0 \tag{5.1}$$

Where $f(\cdot)$ is the transformation function, x the feature vector, w is a vector of classification weights and b the bias. For the FLD $f(\cdot) = x$, but for the RBF SVM $f(\cdot)$ is the kernel transformation shown in equation 3.18. From this binary classification the desired row and column are selected through equations 5.3 and 5.2. Where J is the number of trials. The selection of the row/column is equivalent to selecting the response that strongly represents the characteristic of the target P300, as defined by the training data, by selecting the feature vector with the largest positive distance from the trained hyper-plane. This makes the bias term of the equation 5.1 irrelevant for the row/column selection. The character that appears at the intersection of the predicted row and column in the matrix is the one chosen.

$$\text{predicted column} = \underset{col}{\operatorname{argmax}} \left[\sum_{j_{col}=1}^{J} w \cdot f(x_{j_{col}}) \right] \tag{5.2}$$

$$\text{predicted row} = \underset{row}{\operatorname{argmax}} \left[\sum_{j_{row}=1}^{J} w \cdot f(x_{j_{row}}) \right] \tag{5.3}$$

The reasons for choosing the FLD and RBF SVM were so that a comparison could be made between classifiers with and without regularization, linear and non-linear. Furthermore, their popularity in BCI makes their use here even more pertinent [120, 119].

2.4.1 Support vector machine

The SVM used was a non-linear SVM with an RBF kernel (see section for details). The classification performance was determined by 3 fold cross-validation, with each split containing 26 character epochs for training and 13 character epochs for testing. The data were normalized to mean zero and unit variance, the parameters estimated for the normalization were dtermined by the training set in each cross-validation iteration. The parameters for the SVM were found by five-fold cross-validation for each outer cross-validation. The performance of each validation set was determined by equation 5.4 taken from [116], where fp, tp and fn are the number of false positives the number of true positives and the number of false negatives respectively.

$$C_{cs} = \frac{tp}{tp + fp + fn} \tag{5.4}$$

The reason for using this performance criterion and not simply the classification performance is the fact that it does not take into account the number of true negative examples, which is important for unbalanced datasets, since its omission helps the parameter selection procedure focus on parameters that give positive scores to target examples.

2.4.2 Fisher's linear discriminant

As mentioned the other classifier used was the FLD which is described in detail in section. The classifier did not have any model tuning parameters, so no inner cross-validation loop was necessary. The classifier perfromance was estimated via 3 fold cross-validation.

3 Results

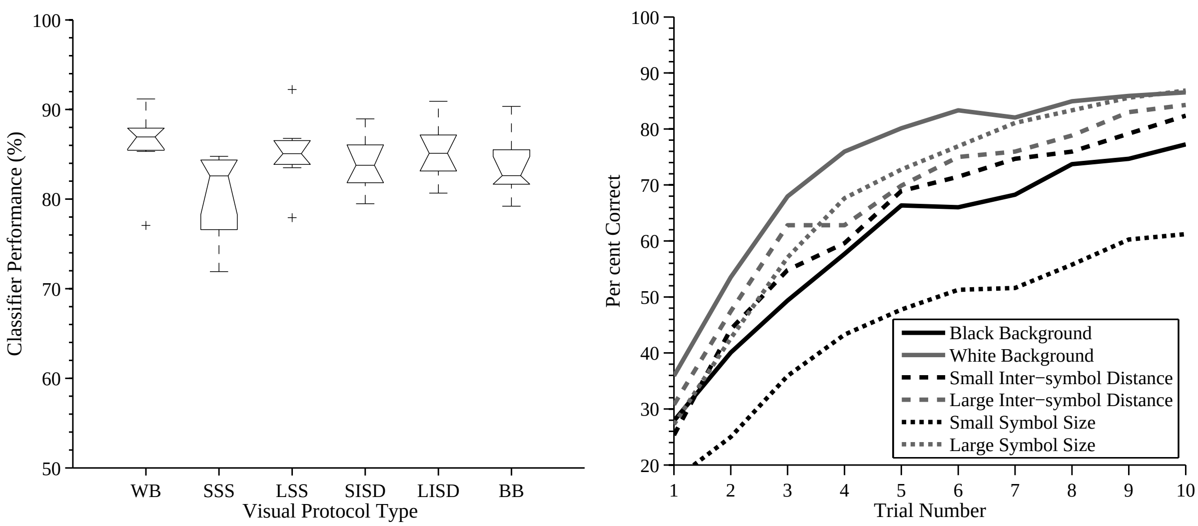
3.1 Subjective choice of visual protocol

At the end of the experiments each subject was asked to declare whether they had a preference for a particular protocol or whether they all seemed similar. The answers were collated into the list shown below:

- Subjects 2,3,5,6,7 and 8 declared that no visual protocol was preferable.
- Subject 1 declared a preference for the White Background visual protocol.
- Subject 4 declared a preference for the Small Inter-symbol Distance visual protocol.

3.2 Support vector machine classification performance

A box-plot of P300 vs No-P300 classifier performance for each visual protocol across all subjects can be seen in 5.5a. The corresponding character classification performance can be seen in 5.5b.



(a) (b)

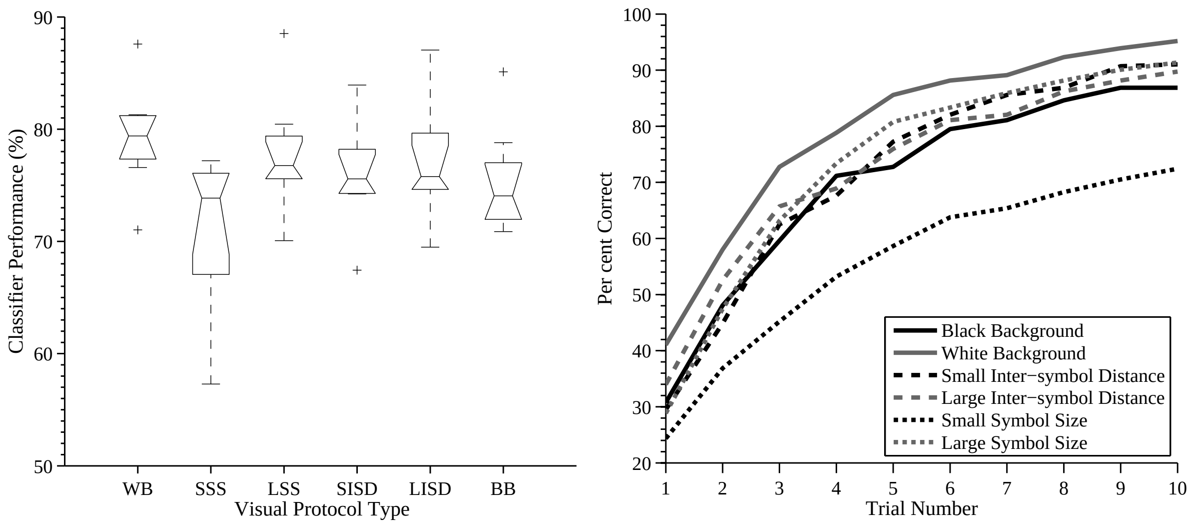
Figure 5.5: SVM accuracy results for all subjects: (a) Mean SVM performance for P300 vs No-P300 epochs (b) Mean character classification performance of SVM.

An explorative study was carried out on the classifier performance to assess the normality and homogeneity of variance of the results. The Kolmogorov-Smirnov test revealed a significant ($p < 0.05$) departure from a normal distribution for SSS. The Shapiro-Wilk test also showed ($p < 0.05$) a departure from normal distribution for SSS as well as WB and LSS. Levene's test was carried out to test the homogeneity of variance, the test result was non significant indicating the variances could be considered to be homogeneous. Friedman's test was used to assess the statistical significance of the classification performance results. The classification performance of the SVM did significantly change over the 6 visual protocols tested ($\chi^2(5) = 13.925, p < 0.05$). Wilcoxon tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a

0.0033 level of significance. Using the SVM, the only two visual protocols that showed statistically significant difference in performance between them were WB and SSS, T=0, r=-0.63).

3.3 Fisher's linear discriminant classification performance

The classification performance of the FLD can be seen in figures 7.3 and 5.6b.



(a) (b)

Figure 5.6: FLD accuracy results for all subjects: (a) Mean FLD performance for P300 vs No-P300 epochs (b) Mean character classification performance of FLD.

Similarly to the SVM results the FLD results were subjected to an explorative study to assess the normality and homogeneity of variance of the data. Both the Shapiro-Wilk and Kolmogorov-Smirnov tests revealed a significant ($p < 0.05$) departure from normal distribution for SSS. Where as the Kolmogorov-Smirnov test showed significant ($p < 0.05$) result for LSS as well. Levene's test did not provide a significant result indicating the variances could be considered to be homogeneous. Friedman's test revealed that the classification performance of the FLD did significantly change over the 6 visual protocols tested ($\chi^2(5) = 13.07$, $p < 0.05$). Multiple pairwise comparison were carried out using

Wilcoxon tests with Bonferroni correction, so all effects are reported at a 0.0033 level of significance. Similarly to the results obtained for the SVM, the two visual protocols that had significantly different performance from each other were WB and SSS, T=0, r=-0.63.

4 Discussion

The statistical test results indicate that the only visual protocols that substantially differ in performance are WB and SSS. This could be interpreted as the P300 classification is unaffected by the majority of the visual modification made here. We believe this to be incorrect since the results presented show that certain protocols (WB, LSS) consistently outperform other visual protocols. The results presented in figures 5.5a and 7.3 show box-plots of the classifier performance for the RBF SVM and the FLD across the differing visual protocols. The highest median value in both situations is recorded by the WB visual protocol. What is of note is the large variability in the range of the data in the visual protocols and the number of outliers. The character classification results presented in figures 5.5b and 5.6b similarly to the box-plots, show the superior performance of the WB visual protocol. A more accurate conclusion would be that the differences in performance obtained by the various visual protocols are less prominent than the effects of individual variation, habituation, fatigue and memory load on P300 classification performance. The negative effects of fatigue may have been accentuated by the memory load of the experimental protocol. The increase in memory load was caused by the constant requirement from the subject to remember their current position in the 39 character sequence. Both memory load and fatigue have shown to adversly affect P300 amplitude. Although the presentation of the target symbol before each character epoch was dismissed in favour of the chosen method, in order to limit possible perceptual confounds, the results indicate that the method used may be less preferable.

The P300 vs No-P300 classification performance of the RBF SVM is better than the FLD, but the character classification performance of the FLD is better than the RBF SVM's. The most likely reason is that the RBF SVM trained to classify the majority of the negative (do not contain target P300) samples correctly, which make up 83.3% of the data. This would lead to higher classification rates than the FLD in simple classifier performance but ultimately lead to worse character classification performance. Furthermore the non-linear nature of the RBF SVM did not seem to provide any classification advantage over the FLD, which is consistent with the findings by Krusienski et al. [118].

4.1 Conclusion

From the results presented, it is apparent that the choices made in visual protocols are reflected in the classification obtained. Evidently, the results presented here are limited and the variation between visual protocols is small, but the indication is that greater accuracies can be achieved by simple subject dependent choice of visual protocols. It is also important to note that the subjective choice of protocol did not correlate with the best performance. Furthermore, most changes in accuracy did not seem to be classifier dependent. The classification difference between the visual protocols only reached statistical significance between the white background and small symbol size visual protocols. However, the results showed that the white background consistently outperformed all other visual protocols.

Chapter 6

FLD and DWT for BCI competition P300 data

1 Introduction

A vital part of identifying the P300 in the ongoing EEG is the pre-processing used. Although powerfull classifiers such as the SVM are able to provide good results with minimal pre-processing, simpler classifiers are often able to perform just as well or better with adequate pre-processing of the data. There are a number of preprocessing methods utilised throughout P300 BCI literature, some of these are: ICA, PCA, downsampling, averaging window, DWT. The most comonly used is the averaging window method. In this method the data is simply filtered to the desired range (0.1 - 10 Hz) and then averaged and downsampled by the desired averaging window size. The size of the averaging window is dpendant on the original sampling frequency and the desired downsampled frequency.

The simple downsampling method just entails filtering the signal to the desired frequency range and the downsampling it by the desired factor. Again the desired downsampling factor is dependant on the original sampling frequency.

The DWT method simply uses a wavelet to extract the wavelet coefficients from the desired decomposition level. The desired decomposition level is defined by the original sampling frequency and the desired frequency range. Wavelet techniques can optimise the analysis of time-varying nonstationary signals such as EEG by providing excellent joint time-frequency resolution. Their use in signal processing and specifically the study of ERP is widespread.

A number of methods have been explored for the classification of the P300 data. The most prominent of them being the use of an SVM or ensemble of SVMs. The purpose of the explorative study was to determine whether a simple linear classifier could with adequate preprocessing perform just as well as the computationally expensive SVM methods. There are a number of linear classifiers to choose from, the most prominent in BCI research are the LDA, FLD and SWDA. The FLD and LDA often arrive at a similar solution and as will be explained later their use in the classification of the target row/column could be interchangeable without any observable difference in performance. The SWDA is a slightly more complex method than the FLD or LDA, which incorporates a feature selection method within its training procedure (see section for details). The decision to use the FLD was determined by its succcesfull use in the classification of P300 data demostrated in the literature, its similarity to the LDA and it being the basis for the SWDA.

Out of the preprocessing methods available DWT was chosen. Its choice was due to its excellent joint time-frequency resolution as well as its wide spread use in ERP analysis. Besides the choices made for the preprocessing method and classifier used a number of other mehodology selections needed to be made, such as: which channels to use, how to overcome the issue of imbalanced data, what is the optimum wavelet and decomposition level. A number of analysis were set up in order to answer these questions as well as address the question first posed of whether a linear classifier with adequate pre-processing of the data can prove to be just as effective as more complex classification methods.

1.1 BCI Paradigm

In the protocol used to collect the data the subjects were presented with a 6x6 matrix containing a total of 36 symbols. Each row and column of the matrix was individually intensified in a random order and once within a trial. For a letter to be selected the user must pay attention to the row and column that intersect at the desired symbol. So within a trial 2 out of the 12 row/column intensifications should contain an increased amplitude P300 response. The SOA used was 175 ms with an intensification period of 100 ms.

1.2 Data

The data as provided by the BCI competition II and III website was sampled at 240 Hz at 64 electrode sites[124, 125]. Data set IIb from BCI competition II contained data collected from one subject. The data set was made up of 42 training symbols and 31 test symbols, each composed of 15 trials. Out of the 42 available training samples only 39 were used, since the last set of training data contained an error in the event cue information. BCI competition III data set II contained data collected from 2 subjects. The data from each subject was comprised of 85 training symbols and 100 test symbols, with 15 trials per symbol.

2 Method 1

The first method implemented was based on the work presented by Kaper et al. and so used the same channels. The aim was to try and get the largest amount of salient features without increasing the classifier input vector too much. The results obtained by this method for the BCI competition 2003 data were published at the 4th International BCI conference in Graz.

2.1 Signal Processing

The data as provided by the BCI competition II website was sampled at 240Hz at 64 electrode sites[124]. The data were later lowpass filtered at 30Hz and highpass filtered at 0.1Hz using a 8th order Butterworth filter. The channels chosen were: **Fz, Cz, Pz, Oz, C3, C4, P3, P4, PO7, PO8**, due to their successful use with the BCI competition II and other P300 data sets[108][106][118]. The data was then downsampled to 60Hz.

2.2 DWT processing and FLD training

After the data had been preprocessed a period of 700ms post stimulus was selected from each of the 10 chosen channels. This meant that each epoch was represented by a 10x42 sample matrix. Each 42 sample vector was then decomposed using single level wavelet decomposition, resulting in an approximation and detail vector. The approximation vector was chosen and the detail discarded. Wavelet used was coif3. This resulted in the epoch being represented by a 10x21 matrix. The matrix was then reshaped to make a 210 element vector. This 210 element vector would be the input to the FLD classifier. At each trial only two out of the 12 epochs should contain a target P300. Unfortunately this results in the data being unbalanced (there are more negative samples than positive). This is often solved in two ways, either by subsampling or oversampling. Although subsampling has been used successfully by a previous entrant [108] the method favoured here is oversampling due to the inclusion of all the data available. Before the data is classified it is normalised to zero mean and unit variance. The classifier is then trained to distinguish into which of the two classes the epoch belongs to. These two classes are whether a P300 exists or whether it does not.

2.3 Test data classification

At each epoch the 210 element feature vector is the input to the FLD and the predicted class the output (1 for P300 present 2 for P300 not present). The data is usually too noisy for the classifier to be able to correctly identify the chosen row/column from one trial alone. Multiple trials are combined by summing the discriminant function values for each row and column at each trial (see equations 5.2 and 5.3). After the set number of trials the row and column with the largest values are the ones selected.

2.4 Results

BCI Competition II For each of the subdistribution the accuracies achieved on the test set were 86.61%, 88.78%, 80.75%, 80% and 80.39%. Since no method was used to try and predetermine the performance of the subdistribution the performance of he subsampling method was taken as the mean, which was 83.31%. The oversampling method achieved an 88.4% correct classification rate for distinguishing between the presence or absence of the P300 in the epoch. It was also able to achieve 100% classification of the test set after only 5 trials per symbol spelt (figure **??**). For a lower number of trials, the error increases.

BCI Competition III For subject A the mean performance using the subdistribution method was 66.54% for P300 vs No-P300 classification and 83% chracter classification at 15 trials. The oversampling method achieved 68.64% classification performance for P300 vs No-P300 and 84% character classification at 15 trials. For subject B the mean performance using the subdistribution method was 72.34% for P300 vs No-P300 classification and 92.4% character classification at 15 trials. The oversampling method achieved 73.34% classification performance for P300 vs No-P300 and 93% character classification at 15 trials.

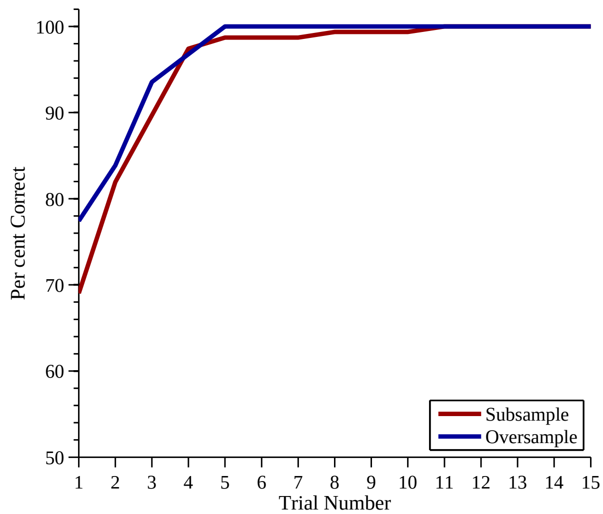
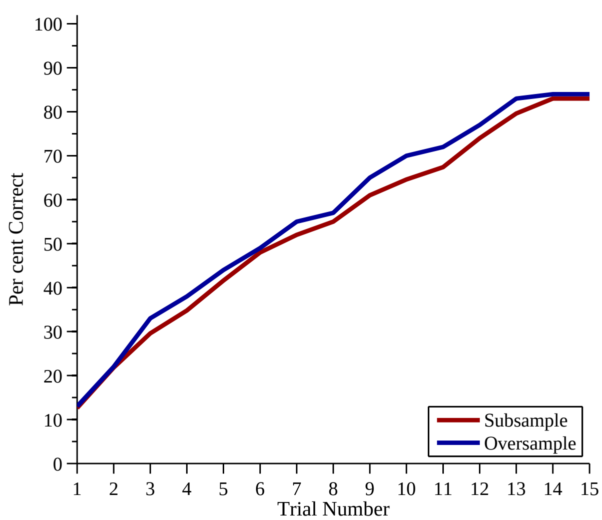
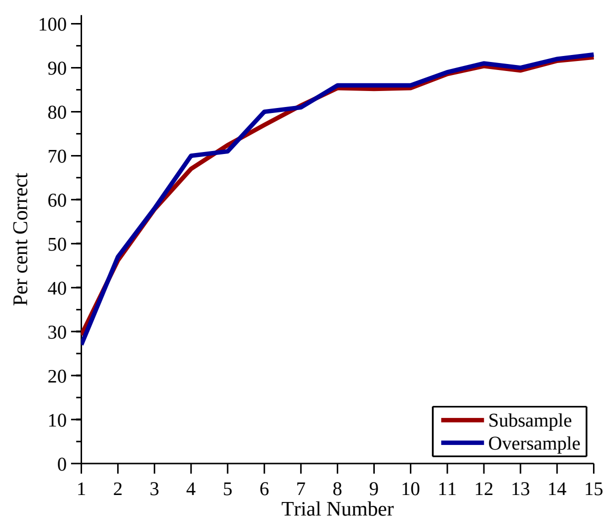


Figure 6.1: BCI competition 2003.



(a) BCI competition 2004 Subject A.



(b) BCI competition 2004 Subject B.

Figure 6.2

2.5 Discussion

The P300 BCI paradigm provides some of the highest achievable information transfer rates using non-invasive BCI. The results achieved using the FLD and DWT combination are as good as the results presented by the winner of the BCI competition II data set IIb, when comparing the number of trials required to achieve 100% classification[126][108] (Figure ??). Oversampling proved to be a better choice than subsampling when used in conjuction with an FLD. Rakotomamonjy et.al [116], using the same 10 channels used in this study, achieve a worse result (6 trials for 100% accuracy), at a much greater computational cost.

For BCI competition III (2004) the results would have ranked this method 5th. The performance of this method is less than satisfactory and is most likely due to the large variability of the data and the susceptibility of the classifier to outliers.

It is important to emphasise the following points about the present method:

1. The feature vector was 210 samples long.
2. The same 10 channels are used as in Kaper et.al [108].
3. No channel selection method was used.
4. The classifier used was very simple and fast.
5. The classifier required no parameter tuning.
6. Only one classifier was used to classify whole test set.

3 Method 2

The second method was implemented due to the shortcommings of the first method concerning the classification accuracy achieved for the BCI competition 2004 data. This method set to optimise the wavelet used, the level to which the data is decomposed and the size of epoch to extract. A paper in a similar vein to the aim of this work was published by Fazel-Rezai et al. Although this paper predated my previous work it had not come to my attention until after the publication at Graz. In their work a number of classifiers were tested, among them the FLD. They used the same channel set, but instead of DWT for pre processing they used PCA. They reported as obtaining 100% charcter classification for the BCI competition 2003 data after 4 trials. It was initialially assumed that PCA provided better feature selection for the P300 ERP than the DWT, but what was also noteable about the methodology used was that they did not mention how they overcame the imbalanced data. So it was presumed that they did not rectify the imbalance of the data. So a further investigation into the effect of not rectifying the data was also carried out. The assesment of the suitability of the epoch duration, wavelet and decomposition level were carried out in two ways. The first method relied on ten fold cross validation of the data with the assesment criteria being the test value since the character classification accuracy could not be used since the character epochs in the training set of BCI competition 2003 and 2004 could not be evenly split into 10 sets. The second method specified 10 characters should always be in the test set and preceded to split the training data a number of times according to that criterion. So for the BCI competition 2003 data there were 3 training and testing distributions and for BCI competition 2004 there were 8.

3.1 Signal Processing

Similarly to the first method the data was bandpass filtered 0.1-30 Hz by an 8th order Butterworth filter. Two sets of channels were chosen **Fz, Cz, Pz, Oz, C3, C4, P3, P4, PO7, PO8** and **Fz, Cz, Pz, Oz, P3, P4, PO7, PO8**. This was in an attempt to reduce the size of the input vector to the classifer. The data was not downsampled and remained at the original sampling frequency of 240 Hz.

3.2 DWT processing and FLD training

As mentioned in section, this method set to optimise the length of the extracted epochs, the wavelet used as well as the decomposition level. This was to be done by cross validation on the training data. The epoch length were varied from 700 to 1000 in 50 ms increments. The wavelets tested were db3 ,db4 ,db5 ,db6 ,sym2 ,sym3 ,sym4 ,coif2 ,coif3 and coif4. The level of wavelet decomposition used was varied by wmaxlev-1 till wmaxlev+1. These were tested in conjuction with the two channel sets mentioned in section.

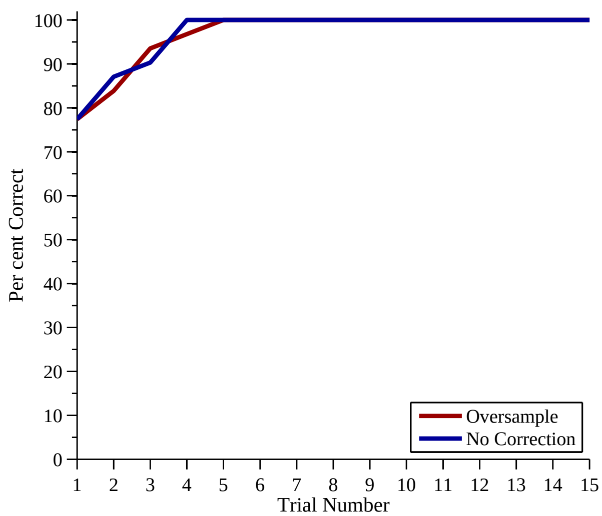
3.3 Results

The three figures in 6.3 show the difference between oversampling and not applying any correction for the imbalanced data for the BCI competition 2003 and 2004 datasets. The pre-processing method and channels used is the one presented in section.

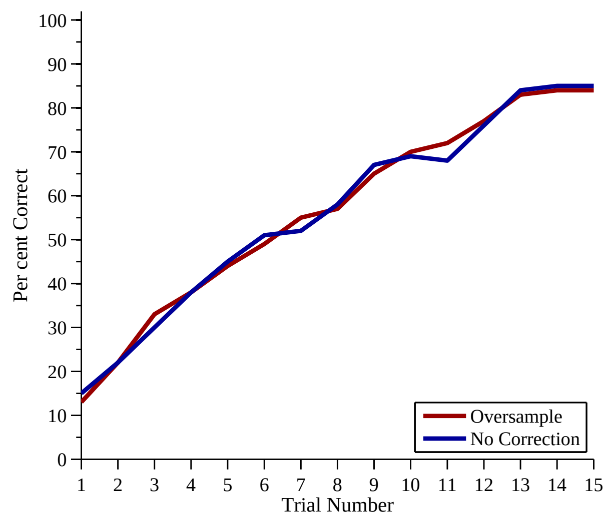
It was also important to see how the oversampling affects the distribution in comparison to not carrying out any modification on the class distribution. In figures the distribution of the distance from the seperating hyperplane of the positive (containing target P300) and negative (not containing distributions) test examples are plotted for the classifier preprocessing methods.

The top three preprocessing configurations can be seen in the tables below. The ordering is based on the tstval. The tables also display the mean chracter classification over all the trials. Figures show the test data character accuracy of the three top performing pre-processing methods for each of the datasets.

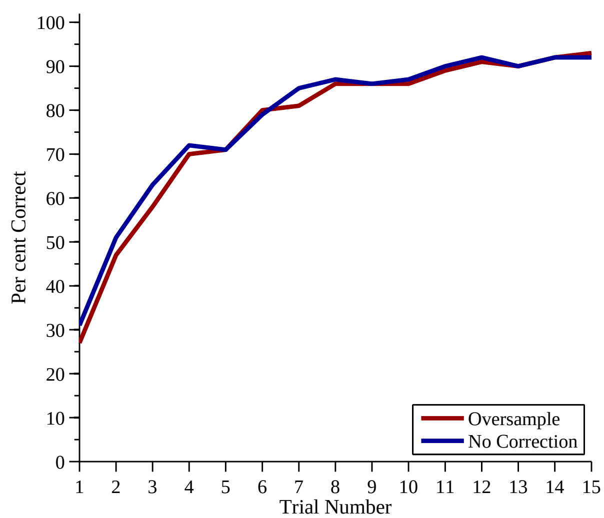
Below are the tables for the best performing pre-processing configuration based on the character accuracy on the variable crossvalidation on the training data. The results for the character classificaittion perfromance on the test data can be seen in figures.



(a) BCI competition 2003.

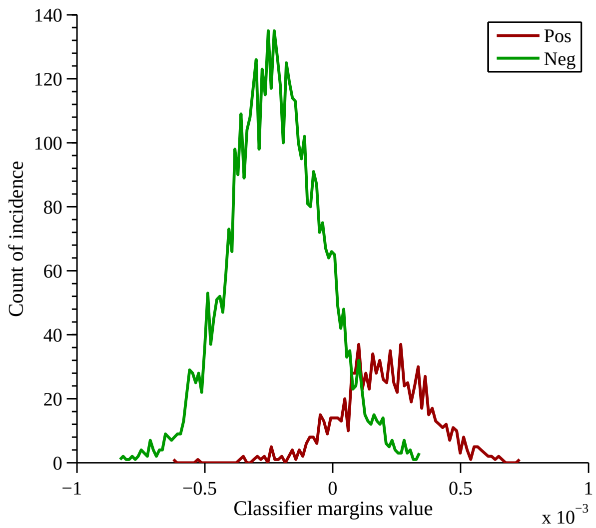


(b) BCI competition 2004 Subject A.

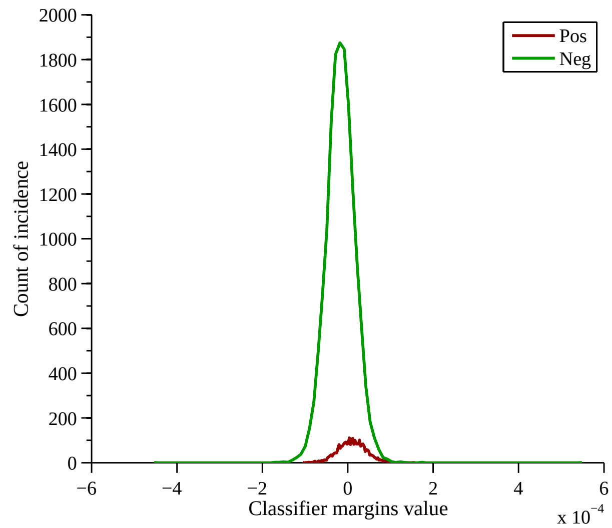


(c) BCI competition 2004 Subject B.

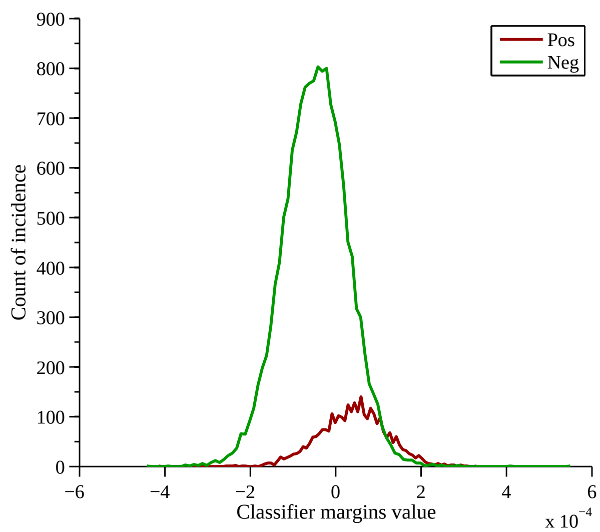
Figure 6.3



(a) BCI competition 2003.

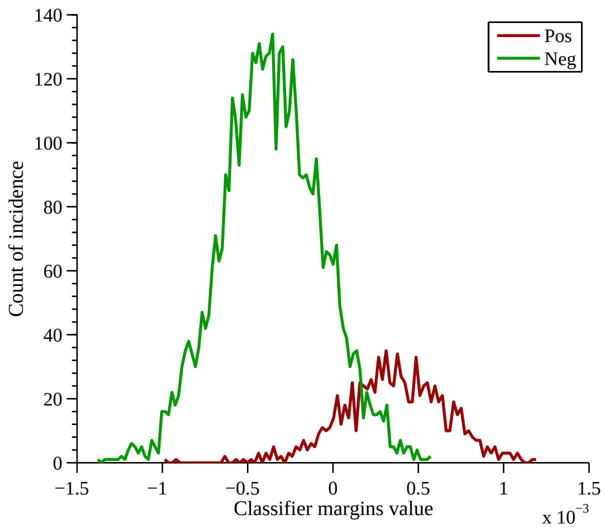


(b) BCI competition 2004 Subject A.

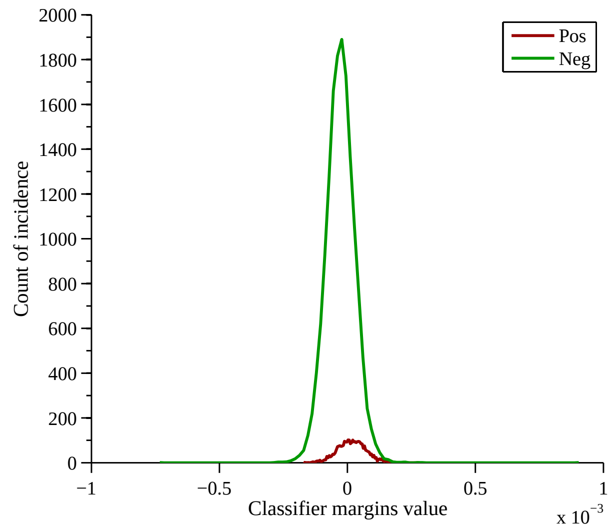


(c) BCI competition 2004 Subject B.

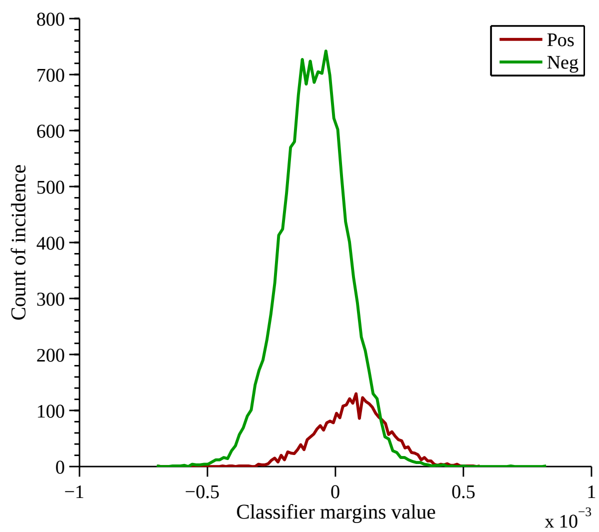
Figure 6.4: Oversampling of training data.



(a) BCI competition 2003.

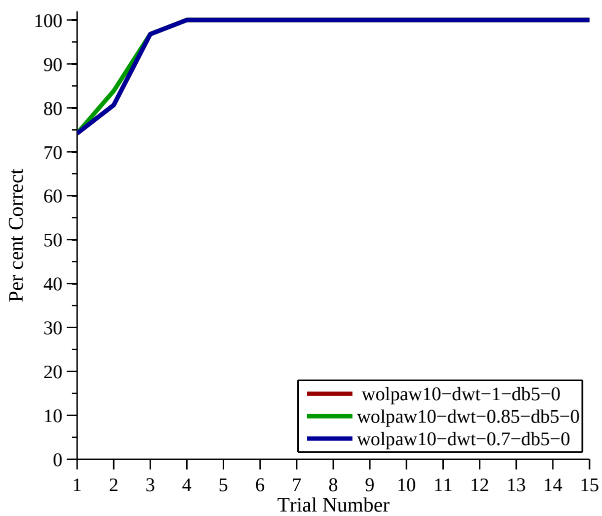


(b) BCI competition 2004 Subject A.

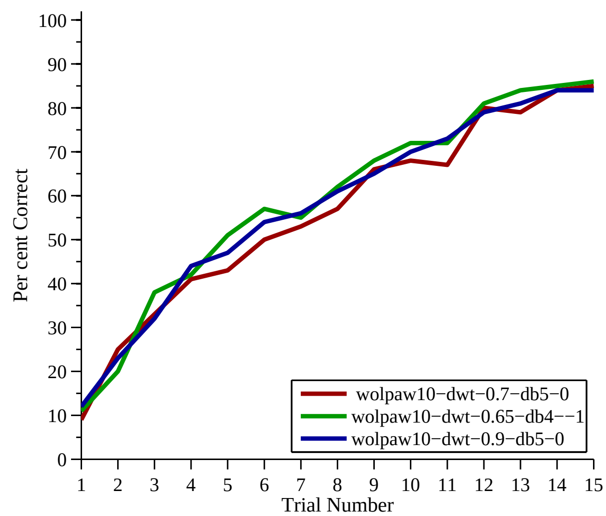


(c) BCI competition 2004 Subject B.

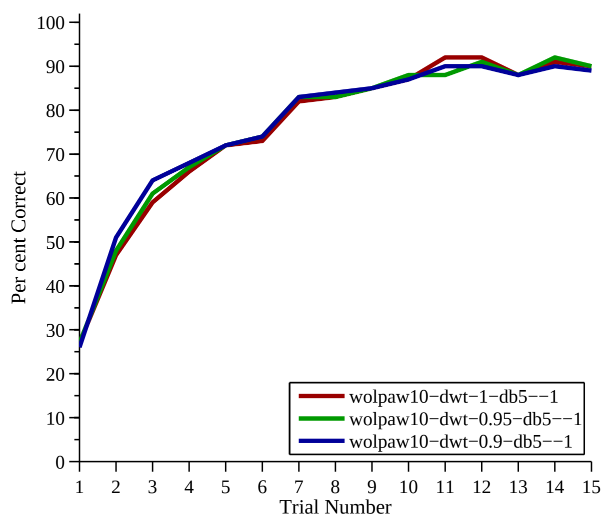
Figure 6.5: No modification on training data class distributions.



(a) BCI competition 2003.

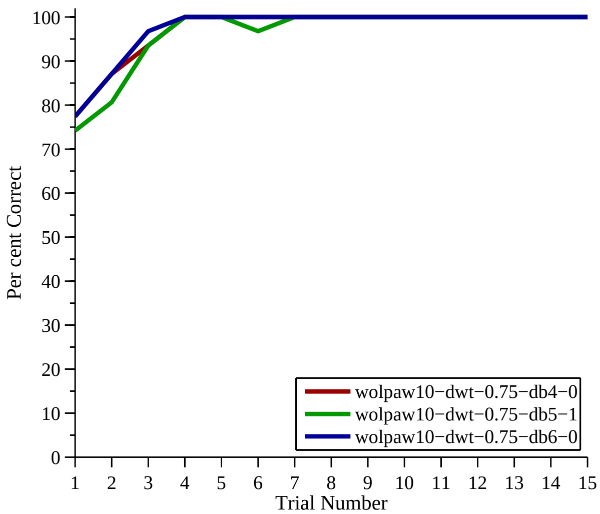


(b) BCI competition 2004 Subject A.

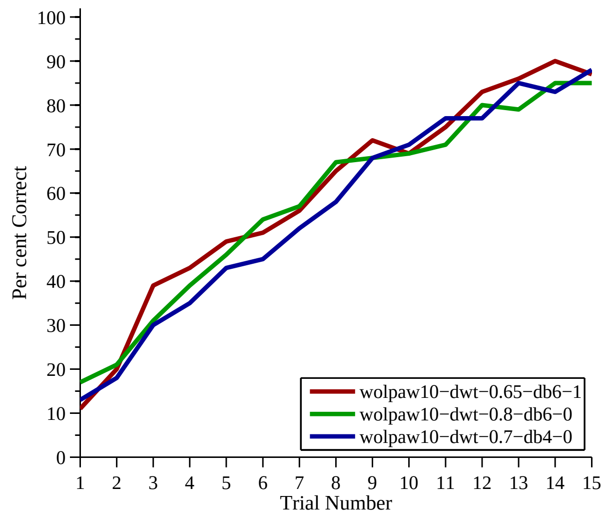


(c) BCI competition 2004 Subject B.

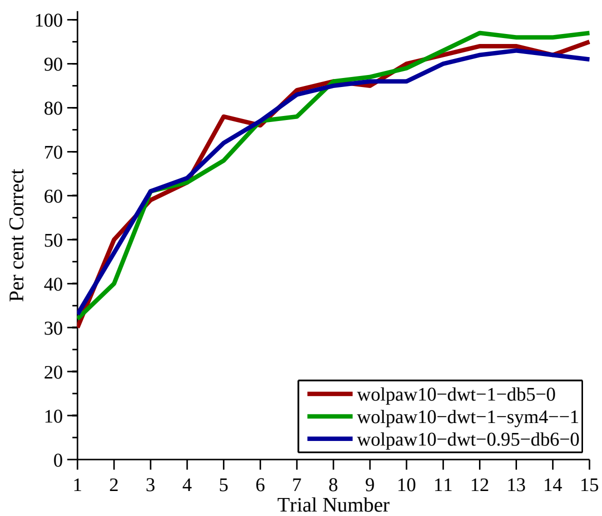
Figure 6.6: No modification on training data class distributions.



(a) BCI competition 2003.



(b) BCI competition 2004 Subject A.



(c) BCI competition 2004 Subject B.

Figure 6.7: No modification on training data class distributions.

3.4 Discussion

The effect of not modifying the class ditribution on the training data in comparison to oversampling was to decrease the number of trials required to obtain 100% character classification on the BCI competition 2003 data. This makes the FLD+DWT method just as succesfull as the FLD+PCA method reported in . Overall the difference between oversampling and not modifying the training data distributions is very small, with a slight advantage towards no modification. By observing the distributions in figures it is evident which data sets will achieve the highest classification performance and which the lowerst. The oversampling of the BCI competition 2003 data seems to reduce the range of the hyperplane seperation values where as for both of the BCI competition 2004 datasets the oversampling increases the range of hyperplane separation. Overall though the distributions look very similar between oversampling and no correction, as might be expected.

The results provided by the pre-processing search based on the tstval did not provide a consistent singular configuration that performed best across all data sets. The general concesus though was that using the wolpaw10 channel set, an epoch duration of 800 ms and a debauchies wavelet between 4 and 5. Therefore a reasonable choice based on the results would be a 800 ms epoch decomposed to wmaxlev using a db5 wavelet.

The second method used to asses the suitability of the preprocessing method did not provide any more conclusive or consistent results. The general trend was to the use of the wolpaw10 channel configuration in conjuction with 800ms epoch. The most commonly selected wavelets were debauchies between 4 and 6.

Although the best configurations for each data set differed as expected due to the differences is the P300 shape and duration there did seem to be a common trend. The best performing channel set was always the 10 channel set, the epoch duration varied around 800ms and the best choice of wavelet were usually from the debauchies family of wavelets.

Comparing the results obtained by the two parameter selection procedures shown in figures and . Indicate that for the BCI competition 2003 dataset the tstval 10-fold crossvalidation results provided slightly more consistent results. Although the difference between them is very small. For thetstval method the top selected preprocessing method provides 58% classification at 5 trials. At 10 trials the chracter classifcation is 88%. For the character classification method the top preprocessing method provides a character classification method of 63.5% at 5 trials and 93% at 15 trials.

This would place this method 2nd overall which is a distinct improvement from 5th using the method described in section. Still this method does not outperform the method presented by rakoto

4 Method 3

The second method presented in section showed that by selecting the optimum parameters through crossvalidation on the training set can significantly impove the results. Althgouh the method provided state of the art results for BCI competition 2003 data it did not beat the top entrant for BCI competition 2004. In this method the preprocessing steps will remain the same, the channels used will be selected by sequential forward floating serach (SFFS). Furthermore the classification will not happen through the use of one classifier alone but from a group of classifiers in an ensemble. This is because with EEG the signal to noise ratio is quite low and often random or unrelated perturbations in the EEG can cause disruption in the classifier. This can be addressed to some extent through the use of a classifier with a regularization parameter or by using a number of classifiers in an ensemble. Each classifier in the ensemble is trained on a specific partition of the training data. Each of the classifiers is then presented with the test data and the output of each classifier in the ensemble is either used as a vote to determine the class or their scores summed and the value of the summation determines the class. In this method each classifier in the ensemble will be trained on a portion of the data and SFFS will be carried out using the remainder of the training data to optimise the selected channels. The number of classiers in each the ensemble will be 6 for the BCI competition 2003 data and 10 for the BCI competition 2004 data. The number of classifiers was determined by the amount of data in each training set as well as an estimation of the data required to adequately train the FLD.

4.1 DWT processing and FLD training

From the previous analysis the epoch period and wavelet that provided the best results overall were 800 ms and db5. The data was first segmented into epochs that started at stimulus onset and lasted 800 ms. From the 800 ms epochs the 4th level DWT coefficients were computed, which resulted in 20 coefficients per channel. The training data was then split into 6 stratified segments for BCI competition II data and 10 stratified segments for BCI competition III data. SFFS was performed on the training data in order to select the optimum channels for each segment. For each data segment an FLD was trained using the channels selected through SFFS. Each training data segment was normalized to zero mean and unit variance. The test data was also normalized using the mean and variance of the training data segment before being classified by the respective FLD.

4.2 Test data classification

At each epoch, the feature vector (x) of the wavelet coefficients from each of the selected channels, is the input to each FLD in the ensemble. The score of each of the classifiers (k) is summed to provide the cumulative score for that particular epoch. The data is usually too noisy for a classifier to correctly identify the chosen row/column from one trial alone. Multiple trials are combined by summing the ensemble cumulative scores for each row and column at each trial (j is the number of trials). After a set number of trials the row and column with the largest scores are the ones selected (see equations).

$$\text{predicted column} = \underset{columns}{\text{argmax}} \left[\sum_{j_{col}=1}^{J} \sum_{k=1}^{K} w_k \cdot f_k(x_{j_{col}}) \right] \tag{6.1}$$

$$\text{predicted row} = \underset{row}{\text{argmax}} \left[\sum_{j_{row}=1}^{J} \sum_{k=1}^{K} w_k w_k \cdot f_k(x_{j_{row}}) \right] \tag{6.2}$$

5 Results

As can be seen in Table 6.1 and Figure**??** the proposed preprocessing and classification method achieves 100% classification at 4 trials for the BCI competition II data, which is comparable to the results presented in [116]. Furthermore the accuracy achieved by the proposed method for subject B, in the BCI competition III, is greater than the one reported in [117] for 5 and 15 repetitions (The way it was assessed during the competition). Unfortunately the performance of the proposed method for subject A in the BCI competition III data set is considerably worse than in [117](see Table 6.2). This resulted in a mean performance at 5 trials of 71.5% compared for 73.5% and 95% compared to 96.5% at 15 trials. Although overall the ensemble method performs slightly worse for the BCI competition III data set it is still superior to the 2nd ranked algorithm on the BCI competition III website (55% at 5 trials and 90.5% at 15 trials)[**?**]. It is important to highlight that the increase in accuracy between the proposed method and the method implemented by Rakotomamonjy and Guigue [117] comes at a high computational overhead both during training and execution.

The optimum number of channels selected for each FLD in the ensemble can be seen in Tables 6.3 and 6.4. For subject B the number of channels chosen per FLD in the ensemble closely resembles and in some cases is more than the number of channels chosen for each

Table 6.1: Number of correctly classified symbols for BCI competition II data set IIb

	Number of Trials							
Classifier	1	2	3	4	5	10	13	15
EFLD	17	25	28	31	31	31	31	31
ESVM	27	29	30	31	31	31	31	31

Table 6.2: Number of correctly classified symbols for BCI competition III data set II

		Number of Trials							
Sub.	Clas.	1	2	3	4	5	10	13	15
A	EFLD	18	34	44	53	65	82	87	93
	ESVM	16	32	52	60	72	83	94	97
B	EFLD	44	58	68	77	78	93	93	97
	ESVM	35	53	62	68	75	91	96	96
Mean	EFLD	31	46	56	65	71.5	87.5	90	95
	ESVM	25.5	42.5	57	64	73.5	87	95	96.5

SVM in [117]. For subject A this is not the case, with the method proposed in [117] using on average nearly twice as many channels as the FLD ensemble method shown here. Possibly the stopping criteria used for the SFFS method may stop it too early resulting in suboptimal channels selected. Another possibility is that the inclusion of further channels causes the FLD to be affected by the curse of dimennsionality something the SVM is able to overcome. Finally the method used is a forward searching algorithm where as the method used by Rakotomammonjy is a backward searching algorithm so some differences in channels selected might be expected.

The top ranked channels for each data set are shown in Table 6.5. The most frequently selected channels by the SFFS method are, as a majority, in agreement with the channels selected by the method employed by Rakotomamonjy et al.[117, 116]. Electrodes in the central and parietal region such as POz CPz and Pz are present in all data sets. Furthermore PO7 and PO8 are in the top 4 for both subject A and subject B in the BCI competition III data set II.

Table 6.3: Optimum number of channels selected by SFFS for each of the FLDs in the ensemble for BCI competition III data set II

	Classifier index									
Subject	1	2	3	4	5	6	7	8	9	10
A	15	16	19	18	18	20	19	16	19	19
B	13	13	17	17	15	18	10	18	19	18

Table 6.4: Optimum number of channels selected by SFFS for each of the FLDs in the ensemble for BCI competition II data set IIb

Classifier index					
1	2	3	4	5	6
17	13	13	16	11	13

As a final investigation an FLD ensemble for each data set was trained using 8 prefixed channels and the test results contrasted to those obtained using SFFS channel selection (see Table 6.6). The 8 prefixed channels used are the same as the ones used in [118]. The results for BCI competition III are as expected with the ensemble FLD in conjunction with SFFS outperforming the ensemble FLD using the 8 prefixed channels. The results for BCI competition II though are somewhat surprising, with the ensemble FLD with the 8 prefixed channels outperforming the method with SFFS channel selection. A closer inspection into the training and testing accuracies revealed a possible reason for this. The FLDs that were trained using the channels selected through SFFS achieved higher accuracies when tested on the remainder of the training data rather than the test data. The inverse was true for the FLDs that used the prefixed channels. This highlights one of the possible issues with using channel selection methods with EEG data. Not all the channels selected were ones conventionally associated with the P300, but were very much tuned to classifying the remainder of the training data. This coupled with the small amount of training data available and the transient nature of EEG, lead to worse performance on the test data with channel selection than with the 8 prefixed channels. Still it is encouraging that the

Table 6.5: Top ten ranked channels across classifier ensemble for BCI competition II and III Data. Channels are ranked in descending order.

Data set	Ranked Channels									
BCI Comp. II	T8	P8	CPz	POz	Iz	FC4	Fp1	F1	F6	Pz
BCI Comp. III Sub. A	Pz	PO8	P7	PO7	CP3	T9	P6	POz	O1	FC1
BCI Comp. III Sub. B	PO7	PO8	O1	Cpz	POz	Iz	C3	Pz	P2	PO4

Table 6.6: Number of correctly classified symbols for Ensemble FLD with SFFS selected channels or 8 preselected channels for BCI competition II and III data sets IIb and II

	Number of Trials						
Sub.	Algorithm	1	2	3	4	5	15
BCI Comp III data set II							
A	EFLD with SFFS	18	34	44	53	65	93
	EFLD with 8 prefixed chan.	17	28	33	41	52	84
B	EFLD with SFFS	44	58	68	77	78	97
	EFLD with 8 prefixed chan.	34	46	57	69	75	92
BCI Comp II data set IIb							
	EFLD with SFFS	17	25	28	31	31	31
	EFLD with 8 prefixed chan.	24	27	31	31	31	31

ensemble FLD method was able to classify all 31 test characters correctly with only 3 trials.

6 Discussion

As demonstrated here and in [117] the use of classifier ensembles for the classification of the P300 have been shown to be very successful. The use of the FLD in an ensemble, in conjunction with DWT for preprocessing, successfully achieved a high degree of accuracy, albeit not as successful as the SVM method presented in [117]. The computational expense of the parameter tuning and channel selection of the Rakotomamonjy and Guigue algorithm [117] is certainly higher than the method proposed here. Although as noted in [117] the performance of a single SVM without channel selection matches the classification performance of the Rakotomamonjy and Guigue method at 15 trials (although performs worse at 5 trials). During the investigation it was noticed that channel selection had a far greater impact on the classification performance of the FLD than on the SVM results reported in [117]. The use of channel selection for the SVM does not seem as necessary as it is for the FLD. This is most likely due to the fact that the SVM is able to manage outliers better due to the tuning of the regularization parameter as well as being able to better deal with the curse of dimensionality problem. Although as highlighted by the BCI competition II results shown in Table 6.6, the use of channel selection for FLDs may sometimes confer no advantages in comparison to prefixed channels. The computational overhead of the SFFS method increases the training time considerably. A more a efficient channel selection procedure would greatly reduce the training time. In turn the reduction in training time may allow for continuous adaptation during the course of its use. This would be accomplished by replacing classifiers in the ensemble during the course of the experiment. This may confer greater performance stability over time.

Chapter 7

Classification Effects of Real and Imaginary Movement Selective Attention Tasks on a P300-Based Brain-Computer Interface

1 Introduction

In P300 protocols the subject often responds to the desired stimulus overtly (e.g., pressing a button) or covertly (e.g., mental counting of stimulus events). The effects of the task on the target and non-target P300 responses have been extensively explored, but results are rarely in agreement. Pfefferbaum et al.[67] and Starr et al.[127] found no effect of tasks on the P300 amplitude while Polich[128] , Barrett et al.[129] and Salisbury et al.[130, 69] report larger P300 amplitudes with covert rather than overt responses. As for the NoGo P300 response, Burle et al.[131] agrees with Pfefferbaum et al.[67] describing similar effects on the NoGo P300 in both overt and covert tasks, albeit with larger P300 amplitude for the overt task. Still, Nakata et al.[132] and Bruin and Wijers[133] found no frontocentral increase for NoGo relative to Go in their Count condition. Smith et al.[70] found the P300 response to Go trials to be similar between counting and button pressing, with the amplitude of the P300 response greater for the button pressing trials. Furthermore, the NoGo P300 increase for Press relative to Count trials was localized in the central region contralateral to the responding hand, with a larger amplitude reported for the button press task.

A well documented effect on the P300 amplitude is its inverse relationship with target-stimulus probability[28, 44, 79]. This has been established for a wide range of target probability and stimulus modality manipulations[75, 72, 73]. Even though stimulus probability is an important factor in P300 amplitude, the interstimulus interval (ISI) has also been found to affect P300 magnitude[77]. With smaller P300 amplitudes reported for shorter ISIs in comparison to longer ISIs. Gonsalvez et al. [134, 80, 81] have reported that both effects are better interpreted or modulated by the Target-to-Target Interval (TTI). TTI is the time between a given target and the preceding target. Therefore it is affected by both the target-stimulus probability and ISI. Decreasing global or local probability increases TTI. Variation in the oddball task ISI governs the TTI and is of greater consequence to the P300 amplitude relative to probability[81].

The goal of this work is to investigate whether the modification of the standard selective attention task to imaginary or real movement will improve results. To this aim we concentrate on differences between tasks as reflected in classification performance. The addition of movement-related potentials (MRPs) to the P300 ERP would be of benefit where discrimination between the target and non-target ERP is increased. The methods used to preprocess and classify the data are based on the assumption that the only ERP of use for discrimination is the P300 ERP, and are optimized accordingly.

In this study, we investigated the effects of combining a P300-based approach with both imaginary and executed motor selective attention tasks on the classification performance using a modified four-directional BCI protocol based on a previous protocol developed by our group. It is important to note that in this four-directional BCI protocol the stimuli were present close to the center of the screen in order to minimize eye movement, in contrast to the Citi et al.[95] protocol which had the stimuli present round the edges of the screen.

2 Methods

2.1 Data Acquisition and Signal Processing

The EEG data were collected with a BioSemi ActiveTwo system at a sampling rate of 512 Hz. A total of 72 channels were used. 64 of the 72 channels were placed on the scalp (using the Biosemi ABC layout based on the 10/20 setup) and the others were placed in the following locations: Two electrodes placed on either earlobe for reference and two more were placed on the left and right arm, respectively. For 3 subjects two electrodes were placed beneath the chin. For the remaining subjects one electrode was removed from beneath the chin and placed next to the left eye. The two final electrodes were placed on the left and right leg respectively. The above configuration allowed the monitoring of electromyography (EMG) generated by the subject. The first three subjects were the most experienced BCI subjects and their control over blinking was much better than less experienced subjects tested later. During the initial test trials with inexperienced subjects it was determined that monitoring of blinking might be necessary for possible artifact removal later. Unfortunately the acquisition device used only allowed 8 external electrodes, therefore a single electrode was moved from beneath the chin to the left side of the eye. This allowed for monitoring of possible blink EMG without compromising recording of tongue movement EMG. Results were unaffected as no artifact correction was applied. Data were then filtered between 0.1 and 30Hz using an 8th order Butterworth bandpass filter.

Channels used were *Fz Cz Pz Oz P3 P4 PO7 PO8*, as used by Krusienski et al.[118]. The same channels were used for each of the experimental tasks. Whilst acknowledging that a recursive channel selection procedure may have provided increased classification accuracy, and inclusion of additional relevant features, we concentrate on whether modification of the standard selective attention task improves results under 'standard' conditions.

2.2 Experimental Protocols

A total of 10 subjects were used (7 males and 3 females between the ages of 21 and 30), all able-bodied, except for subject 7 who was blind in the right eye. Of the 10 subjects, 5 had used BCIs before. Of those 5, 3 had prior experience with imaginary movement based BCIs before, one of whom also having previous experience with the novel protocol used. Each subject carried out a total of 6 experiments in random order.

Experiments had either numerical counting as the standard P300-based task, or target stimulus synchronized real or imaginary movement. The only stipulation restricting the random order of the experiments was that a real movement experiment needed to be carried out before an imaginary one in order to allow the subject to gain some somatosensory perception of the movement. No feedback was provided to the subject for any of the experiments. In the end, each subject carried out 2 numerical counting experiments, 2 real movement experiments and 2 imaginary movement experiments. Each subject carried out all the experiments on the same day. Below is a brief explanation of each of the task protocols.

2.2.1 Mental Counting

The subject was presented with a screen with four stimuli close to the center of the screen, all in a dark red color placed on a black background. After a quick attention sound (Windows ding.wav, duration = 800 ms; the protocol did not pause for the duration of the

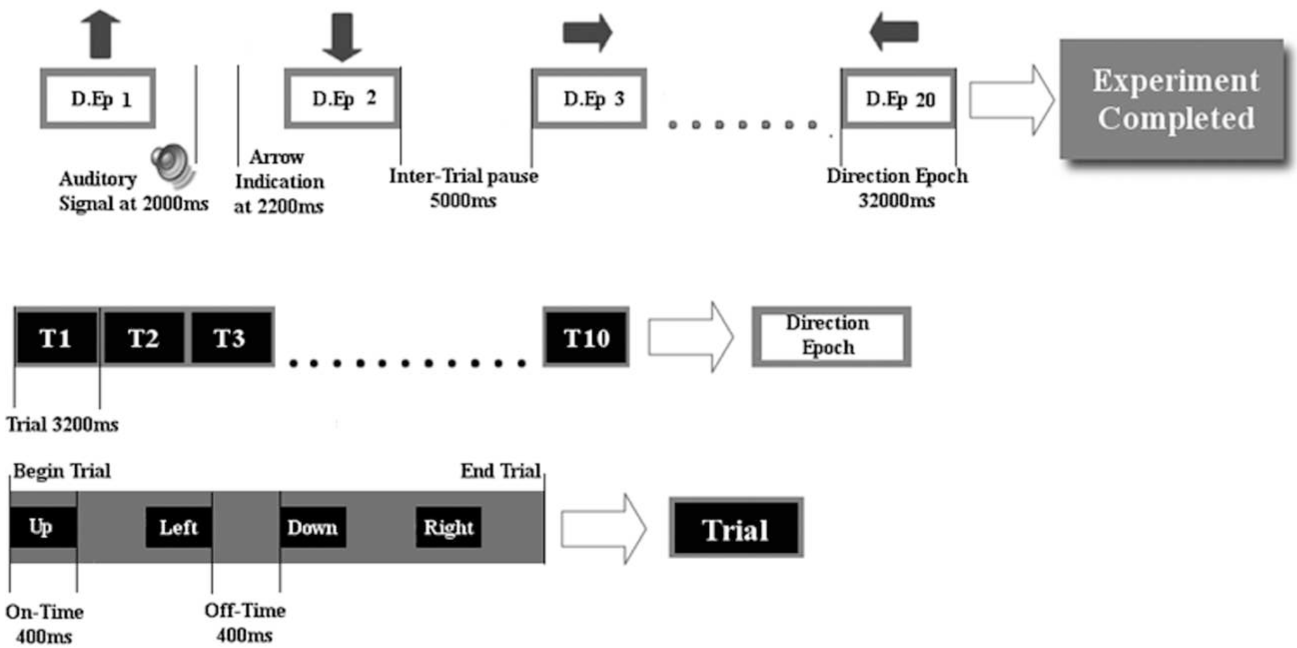


Figure 7.1: Each experiment is made up of 20 direction epochs. At t=2 seconds the direction epoch is proceeded by an auditory signal. At t=2.2 seconds an arrow appears indicating the desired direction. At t=4 the arrow disappears. At t=5 the direction epoch begins. Each direction epoch is made up of 10 trials each lasting 3.2 seconds. During the 10 trials the subject pays attention solely to the stimuli indicated by the arrow before commencing the direction epoch. Each trial is composed of 4 sub-trials with and ISI of 800ms and a stimulus on-time of 400ms.

sound) and a 200 ms pause, a red arrow appeared indicating to which direction the subject should pay attention (Down, Up ,Left or Right). After a period of 1.8 seconds, the arrow disappeared and, after another second, each of the four stimuli flashed in pseudo-random order for a total of 10 trials, with each trial containing a single flash in each direction. The 10 trials were referred to as direction epochs since during those trials the subject would attend a single direction. After a 2 second pause the process would be repeated again with the same or new direction. See Figure7.1 for a representation of the protocol.

Participants were instructed to keep a mental note of the number of flashes that occurred only in the direction previously indicated by the arrow (oddball paradigm). The stymulus-onset asychrony was 800 ms with a 'stimulus on' time (inter-stymulus interval) of 400 ms and a TTI of either 1.6, 2.4 or 3.2 seconds (randomly selected). This was done by preventing any of the stimuli from flashing twice consecutively. Long SOAs were chosen in order to allow the subject sufficient time in the real movement task to execute the desired movement. SOA was kept consistent across the tasks in order to assess classification performance without the possible confounds of shorter TTIs. Within each experiment 5 direction epochs are performed for each of the 4 direction types. For each experimental grouping 400 trials (10 trials $*$ 5 direction epoch $*$ 4 directions $*$ 2 experimental runs) were therefore recorded for each P300 task type (counting, real movement, imaginary movement).

2.2.2 Real movement

The real movement experiments were conducted in exactly the same manner as the numerical counting experiments except that subjects were instructed to carry out a real movement in sync to the stimulus of interest (i.e., the blinking target). Movements consisted of calf extension of both legs for the down stimulus, tongue protrusion for up, right hand-wrist extension for right and left hand-wrist extension for left. Movement was only carried out for the direction indicated by the arrow at the beginning of the epoch and only when the visual stimulus in that particular orientation flashed.

2.2.3 Imaginary movement

The imaginary movement experiments were the same as the real movement experiments except that subjects were instructed to perform motor imagery of said movements rather than executing them.

2.3 Preprocessing

Prior to being passed to the classifier data was preprocessed to reduce dimensionality and extract the most salient features. Channels used were *Fz Cz Pz Oz P3 P4 PO7 PO8*, since they were shown to be very effective for the discrimination of the P300 ERP [118]. Channels were fixed across all tasks and subjects. From these 8 channels, 800 ms periods post stimulus were extracted. At 512Hz each of these periods were 410 samples long. Utilizing all available data would produce a 3280 element feature vector (410 samples $*$ 8 channels). To reduce the size of the feature vector the 5th level 'db4' wavelet coefficients were extracted from each 410 sample vector. For each 410 sample vector 19 features were extracted, producing a vector of 152 features (19 coefficients * 8 channels). Data were normalized to mean zero and unit variance. No spectral estimation methods were used. Furthermore classification was based solely upon data from stimulus onset through 800 ms post stimulus.

2.4 Classification

Classification was done using single stimulus/response trials (as opposed to approaches that require several trials for classification). Extracted and decomposed 800ms post stimulus segments were classified by an FLD. After the desired number of trials the output of the classifier was summed for each of the four directions and the direction with the highest value was the chosen direction. The accuracy was determined through ten-fold cross validation.

$$\text{predicted direction} = \underset{direction}{\text{argmax}} \left[\sum_{j_{dir}=1}^{J} w \cdot f(x_{j_{dir}}) \right] \tag{7.1}$$

$$\tag{7.2}$$

Where j is the number of trials, x_j is the feature vector.

3 Results

3.1 Subjective choice of task

All subjects except subject 8 declared the following order of preference (i.e., comfort and ease of use, from most to least preferred) for the experimental protocols:

1. Real Movement

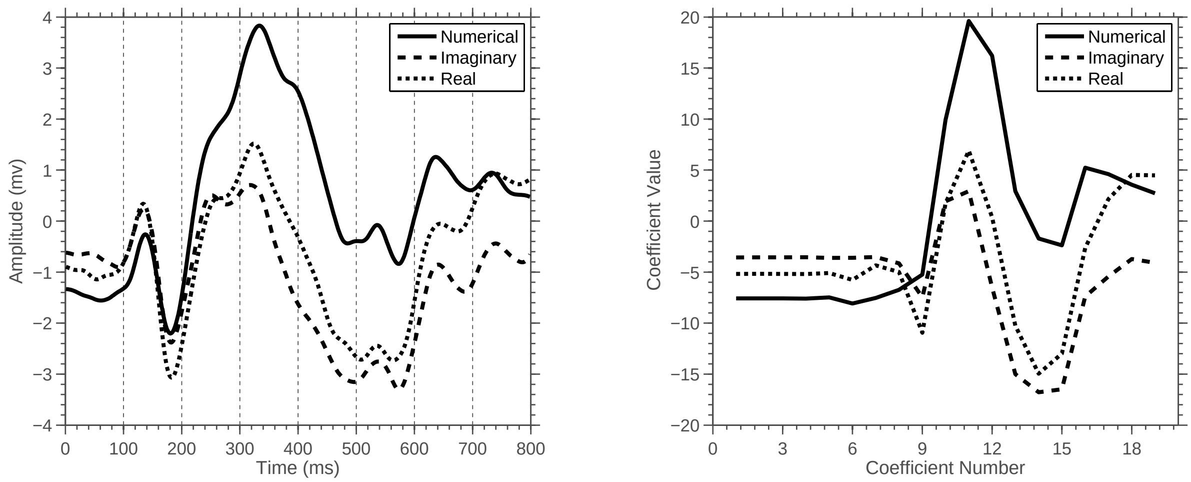
2. Counting

3. Imaginary Movement

Subject 8 had much more experience with the novel protocol, with many hours of use prior to the recordings used in this study. Consistently, subject 8 was able to achieve the highest accuracy in the P300 imaginary movement task with an average epoch classification of 94%. Although subject 8 stated that the real movement task was the easiest to perform the imaginary and counting tasks could be executed quicker. All other subjects found the imaginary movement taxing but slightly more engaging than the counting.

3.2 Does the protocol generate a P300?

In order to assess whether the protocol generated a P300 ERP the averages of the 8 chosen channels for the 800 ms post-stimulus target epochs were plotted. These can be seen in Figure7.2a. The highest amplitude P300 was achieved by the numerical counting task, with the real and imaginary movement tasks having a much smaller peak. This is in agreement with the findings of Polich[128], Barrett et al.[129] and Salisbury et al.[130, 69]. It also interesting to note the greater negativity of the real and imaginary movement tasks at



(a) (b)

Figure 7.2: Averages of target epochs: 7.2a Average of target epochs over the eight chosen channels for the three tasks. 7.2b Average of wavelet coefficients used of the target epochs over the eight chosen channels for the three tasks.

500 ms, possibly an indication of the interaction of movement-related potentials with the ERP as reported in [130]. In Figure7.2b the average of the wavelet coefficients for the 800 ms target epochs across the 8 chosen channels can be seen. It is evident that the chosen coefficients reflect the activity of the target ERP.

3.3 P300 vs No-P300 classification

P300 vs. No-P300 classification refers to the single epoch classification of whether or not a P300 wave is exhibited in the 800 ms sample (across 8 channels) input to the FLD. The results presented are from ten-fold crossvalidation.

3.3.1 Fisher's linear discriminant classification

The results for P300 vs. No-P300 classification for the FLD are illustrated in Figure7.3. From Figure7.3, it is evident that the task that produces the best classification performance is the real movement task. Overall the counting task shows poorest levels of performance.

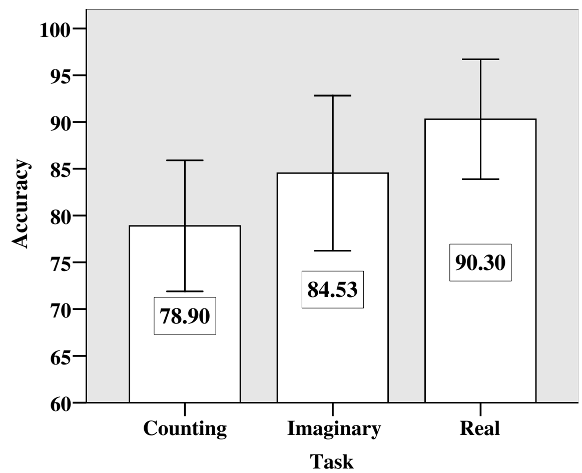


Figure 7.3: Mean accuracy results for FLD P300 vs No-P300 epochs across all subject for each attention protocol *(error bars: ±2 standard deviations)*.

The imaginary movement task showing improvement upon counting without reaching the result produced by the real movement task. It is important to note that the large variance seen in Figure7.3 is most likely due to the variability of the inter-subject performance. Also, external sources of interference further exacerbate this issue. Examples of such sources are subject fatigue and habituation, drying of electrode gel and possible fluctuating external electromagnetic interference.

To assess the differences in the performance across the tasks, we performed a repeated measurement analysis of variance with task as the within subject factor and classification accuracy as the dependent variable. Classification accuracy varied significantly ($F(2,18) = 35.9$, $p < 0.000001$) between the three tasks. Pairwise comparisons (with Bonferroni correction) indicate that the means for each task type were significantly different (at the overall 0.05 level). The p-value for the pairwise comparison of real movement (M=90.3, SD=3.2) and numerical counting (M=78.9, SD=3.5) was below 0.00005. For the real movement and imaginary movement (M=84.53, SD=4.15) $p < 0.005$. Finally for imaginary movement and numerical counting $p < 0.008$. To summarize, statistically significant improvements are found upon the standard numerical counting task in both real and imaginary movement conditions; with the real movement task showing significant advantage over that using motor imagery. It is encouraging that even subjects who had never previously used an imaginary movement BCIs were able, without any training, to obtain results statistically distinct from those for the counting task.

3.3.2 Possible EMG Effects on Classification

One question arising from the high real movement classification results was whether higher classification rates were using information derived from EMG present in the channels chosen. In order to see whether there were stimulus locked EMG artifacts present in the channels, grand averages were taken of all channels including the EMG channels for the 4 different movements across all the tasks and for each subject. The raw time series as well as grand averages were then visually inspected. It is important to note that the aim was to find time locked EMG artifacts that may aid classification rather than transient EMG artifacts which often lead to degradation of performance. Both the counting and imaginary movement tasks were, as far as could be determined, free from time locked EMG. Furthermore, the EMG channels did not seem to contain time locked EMG (transient EMG and periodical ECG were present on some EMG channels).

For the real movement, the only artifact that was noticeably present in the recorded data was that of the tongue movement (up direction). This was not only present as high amplitude, high frequency artifacts but also as a low frequency drift across a majority of the electrodes. Rather than remove the trials contaminated with EMG, the contribution of EMG to the classification was assessed by calculating the accuracies across all tasks again, but without the up direction epochs.

The classification results obtained through the removal of the 'up' target direction were slightly different from the results which were inclusive. Overall results remained the same, changes across the three differing tasks were found not to be statistically significant. The contribution of EMG cannot be discounted completely, but its contribution was found to be minimal. Furthermore, the purpose of the inclusion of the real movement task into the experiment was to assess the possible ceiling of classification results using this novel protocol and not as a valid BCI task as many potential users may not have this ability.

3.4 Direction classification

In this novel protocol direction was selected after a certain number of trials. The number of trials chosen per direction epoch was 10. This resulted in a particular direction being selected every 32 seconds. The results for all subjects for the three different tasks can be seen in Figure7.4.

It is evident that across all subjects the real movement task performs best, as seen in the classification performance of 87.5% for single trial (3.2 seconds), reaching 99.5% after 4 trials (12.8 seconds) and reaching 100% at 7 trials. The imaginary movement task has a single trial (3.2 seconds) classification performance of 83%, reaching 99.5% at 5 trials (16 seconds), reaching a maximum of 99.75% at 6 trials. The worst performing task was the counting task, with a single trial classification performance of 70%, reaching 98.2% after 10 trials (32 seconds). It is important to note that only one of the subjects had used this novel protocol before and only 3 of the 10 had used imaginary movement BCIs before. For the imaginary movement task over 95% accuracy was reached in 3 trials (9.6 seconds). Furthermore, for single trial, the average classification accuracy exceeds 80% for the imaginary movement task.

3.4.1 Experienced vs Inexperienced

As a final analysis, the results for subjects who had prior experience performing motor imagery for BCI experiments were compared to those who did not. The results can be seen in Fig. 7.5.

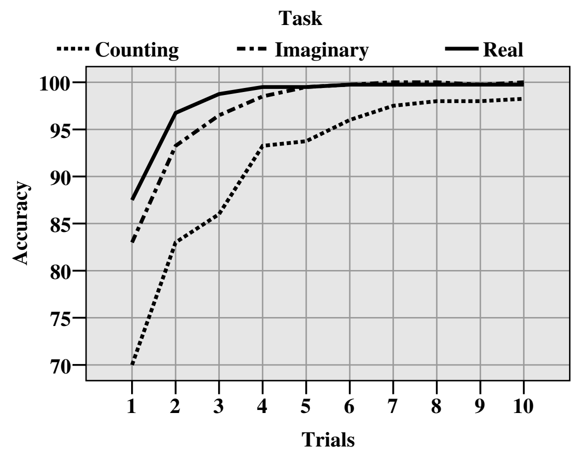
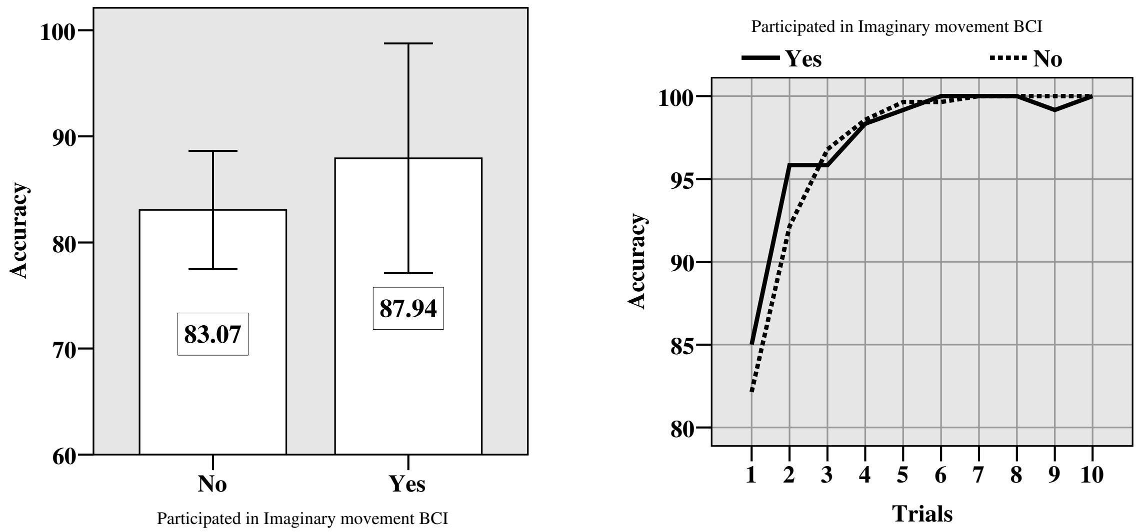


Figure 7.4: Mean accuracy across all subjects over trials.

Figure7.5a suggests that the performance for subjects who had previously used imaginary movement BCIs was superior to that of subjects who had not. However, the same conclusions cannot be drawn from the results presented in Figure7.5b. The performance of experienced subjects for the first two trials is better than that for inexperienced ones (with the experienced reaching over 95% in two trials (6.4 seconds)). Beyond this, the results become less clear, with inexperienced subjects on occasion marginally outperforming the experienced. Close inspection of the results revealed that the drops in accuracy and the irregular shape of the experienced users graph were due to the inconsistent performance of subject 6. This may indicate lapses in concentration or faults at the electrodes. A two-sample T-test was carried out to ascertain whether the differences in the mean FLD accuracy were statistically different. The means of the FLD accuracy were found not to be significantly different ($t(8)=1.946$, $p=0.087$). Even though the statistical test shows the performance means of the experienced vs inexperienced not to be different, it would not be prudent to say that experience plays no role in the classification performance. The reasons for this are: a) the number of subjects belonging to either group were not equal b) the performance of one of the three experienced subjects was somewhat irregular, possibly



(a) (b)

Figure 7.5: FLD accuracy results: 7.5a Epoch FLD accuracy for imaginary movement - comparing subjects that had participated in imaginary movement experiment before and subjects that had not. *(error bars: ±2 standard deviations)*. 7.5b FLD accuracy for imaginary movement task across trials - comparing subjects that had participated in imaginary movement experiment before and subjects that had not.

indicating problems with the equipment or fatigue and therefore leading to greater variance in the classification performance; and c) the inexperienced subjects performance may have improved more quickly over the course of the experiments, further confounding the results. In order ascertain the influence of experience a balanced subject set is desirable and a more thorough investigation would be required.

4 Discussion

The real movement task has proved to be significantly better than the imaginary and counting task. In turn, the imaginary movement task is significantly better than the counting task. Increased accuracy in the real movement task could be attributed to: a) the addition of movement related potentials in the target P300 epochs; b) the increase in NoGo P300 positivity in the non-target epochs as reported by Burle et al.[131]; c) the increase in subject engagement in the task and the immediate movement feedback, as well as the lessening of memory load. The imaginary movement task should portray similar effects in its ERPs, with movement related potentials present in the target P300 trials and increased NoGo P300 positivity in the non-target trials in comparison to the counting task. The possible reasons why overall the imaginary movement task classification performance is not the same as for the real movement are the following: a) the movement related potentials will be of lesser amplitude in the imaginary than the real movement task; b) increase in P3 positivity will not be as pronounced in the imagined movement task as found by Burle et al.[131]; and c) memory loading of the imagined movement task will be greater than the real movement task and possibly the counting task (as highlighted by the subjective task preference). The increase in task complexity may yield advantages to some subjects, while being a disadvantage to others (as memory loading increases beyond individual threshold), most markedly between the subjects that have done imaginary movement BCIs before and those that had not. However, it might be relevant to note that two subjects out of the ten in the imaginary movement task approached and even surpassed the accuracy obtained by the real movement task. Both those subjects had used imaginary movement BCIs before.

The importance of the classification superiority yielded by the movement related tasks is further amplified by the fact that the channels and methods used were optimized for the counting P300 selective attention task and that the majority of the subjects had not been trained to use imaginary movement based BCIs. It may be possible to improve the results further by carrying out some or all of the following:

1. Using a channel selection method to select the optimal channel configurations for each subject and task.

2. Incorporating other features (i.e. sensory-motor rhythms) to increase the discrim-

inability of the imaginary movement task.

3. Using separate classifier for each movement.

4. Increasing the epoch time window from 800ms.

5. Use task optimized and subject optimized wavelet.

6. Allow for training sessions with feedback.

7. Use of a regularized and possibly non-linear classifier such as an support vector machine (SVM).

One of the disadvantages of this novel protocol is that it requires large SOAs when compared to standard P300 protocols, to allow the subject to carry out the real and imaginary movements. It is quite possible that the SOA of the counting task could be reduced, but by how much would have to be determined. The inherent reduction of the overall discriminability between the target and non-target epochs due to shorter TTIs as well as the possibility of repetitions blindness and attentional blink would have to be considered [81, 135, 136, 137]. Although the slow presentation rate may be interpreted as a disadvantage, it may also prove to be an advantage as demands on attention could be reduced in comparison to P300 BCIs with very short SOAs Further experiments are also needed in order to determine the degree to which the SOA can be reduced. Another possible disadvantage of the imaginary movement task may be a decrease in the P300 amplitude due to task difficulty in comparison to the counting task (possibly indicated by the subjective preference of task), the P300 becoming smaller as task difficulty / complexity exceeds the attentional resources of the subject [138]. However, moderate increases in task demands within the subject's capabilities should increase P300 as the subject devotes more resources to the task [130].

This protocol differs from other protocols that use imaginary movement in that it does not use SMRs explicitly. Rather it relies on the P300 and possibly movement-related potentials to identify the target. This allows it to have a very short classification period (800 ms) and very quick succession of repeated movements (1.6,2.4 or 3.2 seconds). Furthermore the classification accuracy achieved is much higher than that reported by recent 4-class sensory-motor rhythm (SMR) protocols [139, 140].

In this study four distinct movements were used to represent each of the directions, but a single imagined movement, such as a finger tap, could potentially be used to register whether the stimulus is a target or not. A reduction in the complexity of the movement may lead to a reduction in memory load and infer a greater classification advantage to the imaginary movement task. This has interesting implications for currently used P300 protocols since greater performance may be achieved simply by changing the task from counting to imaginary movement. Furthermore by using imagined movement in combination with an oddball protocol allows for a system that can be used immediately whilst also providing increased performance over time as the subject becomes accustomed to the task. Increased performance would be derived from being able to carry out the imagined movement task with greater reliability and potentially from a reduction in memory load as subjects become more accustomed to motor imagery.

Although the theoretical benefits of replacing four movements with one are compelling, the use of four distinct movement in this protocol may confer a further advantage (that was not exploited here) over using a single movement. By using a different movement for each direction differing sensory-motor activation patterns should be present in each direction epoch. Therefore by monitoring the topography and spectral properties of the SMRs within each direction epoch further relevant information may be passed to the classifier, potentially improving performance. This would not be possible with a singular movement since the SMRs would be similar or the same across the direction epochs. Work is currently being carried out to assess the practical aspects of this implementation.

4.1 Conclusion

The novel approach of replacing the standard selective attention task (target counting) with real / imaginary movement in a 4-class P300-based BCI protocol has yielded interesting results. The inclusion of the extra relevant information from the task has allowed for higher classification rates over the standard mental counting task. This is especially poignant since SMRs were not exploited directly and classification is based on the decomposed 800 ms post stimulus epoch over 8 channels. Further exploration into what factors contribute to higher classification rates and what are the best methods and features to capture this extra information need to be carried out in the future.

Chapter 8

Perceptual Errors in the Farwell & Donchin Matrix Speller

1 Introduction

In the Farwell and Donchin matrix speller the two types of perceptual errors that may affect P300 based protocols are; target-events not being perceive and non-target events perceived as targets. Cinel et al. [137] showed how non-target rows/columns neighboring target rows/columns often generate P300 like potentials, indicating that they may have been perceived as targets. Fazel-Rezai [141] showed that these P300 like potentials often lead to an incorrect choice of symbol and a decrease in the overall accuracy. Repetition blindness (RB) is a perceptual phenomenon that can occur when two identical targets are presented in a stream of non-targets at intervals of less than 500ms [136]. Similarly the attentional blink (AB) can occur when two different target stimuli embedded in a series of non-target stimuli appear in quick succession (less than 500ms between them)[135]. Both RB and AB cause the second target not to be perceived and have been shown to affect the P300 ERP [142]. The possible effects of RB and AB on the Farwell & Donchin protocol have been discussed before [137, 141] but their influence on P300 classification, achieved by means of a state of the art classification algorithm, have not been explored. This work will attempt to address this as well as provide some suggestions for modifications of the Farwell & Donchin P300 matrix speller in order to minimize such errors.

2 Methods

2.1 Pre-Processing and Classification

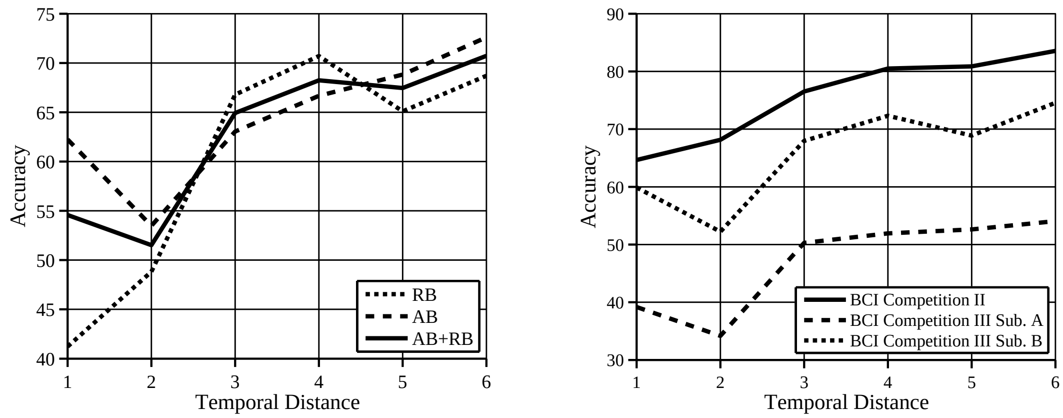
The data used for this investigation come from BCI competition II data set IIb and BCI competition III data set II. The data were sampled at 240Hz at 64 electrode sites[**?**, **?**]. Using an 8th order Butterworth filter the data were lowpass filtered at 30Hz and highpass filtered at 0.1Hz. In combination the data sets contained data from three subjects, one from BCI competition II and two from BCI competition III. The data set IIb was made up of 42 training symbols and 31 test symbols, with each symbol composed of 15 trials. Out of the 42 available training samples only 39 were used, since the last set of training data contained an error in the event cue information. The data from each of the subjects in BCI competition III were comprised of 85 training symbols and 100 test symbols, again with 15 trials per symbol. The data was segmented into 800 ms epochs, taken post-stimulus, and the wavelet coefficients computed using a 4th level db5 discrete-wavelet transform (DWT). A Fisher's linear discriminant (FLD) ensemble was used to classify the test data. Sequential floating forward search (SFFS) was used on the training data to chose the optimal channels for each FLD in the classifier ensemble. The training and test data were normalized to zero mean and unit variance. For more details on the preprocessing and classification procedure please refer to section.

2.2 P300 matrix speller protocol and perceptual errors

As mentioned previously the matrix speller protocol utilizes a 6x6 matrix, comprised of a number of symbols. The symbols that make up the matrix may differ but most often are composed of the 26 letters of the alphabet, numbers from 1 to 9 and a space symbol, as is the case with the experiments presented here. The stymulus-onset asychrony (SOA) used in the experiments was 175 ms with an intensification period of 100 ms.

As mentioned in the previous section the number of trials per symbol of all data sets used was 15. Within each trial 2 intensifications (1 row and 1 column) should elicit a target P300 response and the other 10 should not. Detection of the P300 response and its correspondence to a certain row or column is what enables a P300 BCI system to identify the chosen symbol. Unfortunately due to the relative low amplitude of the P300 response and low signal to noise ratio of EEG it is not currently possible to obtain a high accuracy of symbol selection simply by using a single trial. This is why the P300 BCI paradigm relies on multiple trials in order to boost the accuracy to acceptable levels. Several phenomena affect the amplitude and latency of the P300, the following were explored in this investigation; Attentional Blink (AB) and Repetition Blindness (RB).

The AB can occur when two different target stimuli embedded in a series of non-target stimuli appear in quick succession (less than 500ms between them)[135]. This leads to the first target being correctly identified while the second is not detected at all. In the Farwell & Donchin protocol it might also be possible that an AB may occur if a targets neighboring row or column intensifies before the target, and if the non-target row or column is perceived as a target. The effect of the AB seems more pronounced if a non-target event follows a target event before the mental processing of the target event has been completed (<200ms)[143]. Otherwise the performance of identifying the second target varies according to the temporal proximity to the first target. The effect of AB is also modulated by the discriminability between the target and non-target events [143].



(a) (b)

Figure 8.1: (a) Affect of Attentional Blink and Repetition Blindness on performance of all subjects - Mean Classification accuracy for target events proceeded by a target event or a false positive event. (b) Affect of Attentional Blink and Repetition Blindness on performance of each subject. The temporal distance varied from the target event being directly preceded by another target event or false positive (1) to being preceded by a target or false positive at a temporal distance of 5 intervening non-target events (6).

RB can occur when two identical targets are presented in a stream of non-targets at intervals of less than 500 ms[136]. Similarly to the AB, RB causes the second target event to be missed and is therefore another possible source of perceptual errors in the P300 matrix speller. The majority of studies into the effects of an AB on the P300 ERP found that the P300 component is suppressed during an AB period [144, 145]. Although earlier components such as P1 and N1, which tend to be associated with sensory processing, are generally unaffected [146]. The main reported effects of RB on the P300 are that correctly identified repeated targets evoke more positive ERPs than unidentified repeated targets in the 0-400 ms window [147]. Koivisto and Revonnsuo [142] found that perception is impaired at the same point in time in AB and RB, at around 250 ms post stimulus. Furthermore unrecognized second targets for both repeated and unrepeated stimuli produced ERPs similar to non-target events. It was also reported in [142] that RB has a larger affect on subjective visibility than AB.

In the BCI competition II and III data sets IIb and II respectively the row and columns were randomly interleaved (some P300 matrix speller implementations may have the all rows intensify and then all columns or vice-versa). This and the random nature of the presentation order can lead to two targets appearing within 175 ms of each other, which is sufficiently short for both AB and RB to occur.

The aforementioned perceptual errors will be investigated by classifying the test data and comparing the predicted class labels provided by the classification algorithm to the true class labels. It is expected that the closer two target events are to each other in time the higher the classification error of the second target event. If the miscladdified non-target stimuli are truly P300 related they are expected to cause a detrimental affect on the classification of succeeding target events that occur within a 500 ms window.

3 Results

In Figure8.1a the dotted line is the percentage of target events (based on test class label) that were correctly classified and were preceded, at the various temporal distances, by a target event. The accuracy fluctuations at the various temporal distances could be interpreted as an RB effect on the classification, since the two targets within a character epoch are always the same symbol. The dashed line is the percentage of target events that were correctly classified and were preceded by a false positive. The fluctuations in performance can be attributed to AB since the two events would be focused over different sets of symbols. The solid line is the percentage of target events that were correctly classified and were preceded either by a target event or a false positive event. Therefore representing the AB+RB effect on the target event classification performance. With the AB the second or possibly third event after the false positive should be most affected [135]. As can be seen the accuracy of the solid line representing the AB+RB effect on accuracy at a temporal distance of 1 is improved in comparison to the RB effect, but at position 2 the accuracy is nearly as poor as the RB effect alone. The subject from BCI competition II does not seem to display a consistent AB effect where the second or possibly third lag (temporal distance) produce worse results than lag 1 (see Figure8.1b). The effect is more noticeable in the subjects from BCI competition III. The RB effect is noticeable across all subjects. It is also important to highlight at this point that the identification of possible AB and RB within the protocol relies on the classification methods incorrectly classifying an event as not containing a target P300 when it should. Although the literature supports the view that unattended target events due to AB or RB induce ERPs similar to non-target events [142], the possibility of a target event being correctly classified when it should not still remains.

After identifying the FP events that could possibly cause an AB the analysis of their spatial location in comparison to the target row/column was carried out. In [137] and [141] it was found that non-target rows/columns neighbouring target rows/columns produced P300 like ERPs and often had a detrimental effect on the classification method. In Figure8.2 the classification performance of the classifier is shown with regard to the number of non-target events that are correctly classified in relation to their spatial proximity to the target row/column. A spatial distance of 1 indicates that the event occurred next to the target row or column. As a general trend the accuracy increases for events that occur further away from the target row/column. The percentage of correctly classified target events that are preceded by FP events at various temporal and spatial distances are shown in Figure8.3. Both subjects A and B from BCI competition III data s et II seem to show a accuracy performance dip at temporal distances of 2 for all the spatial distances. The results for BCI competition II are less clear, although for spatial distance 1,2 and 4 the performance dip does seem to happen but is most pronounced at the temporal distance of 3. The dip in performance at temporal lags 2 and 3 is consistent with results for AB reported in the literature [143].

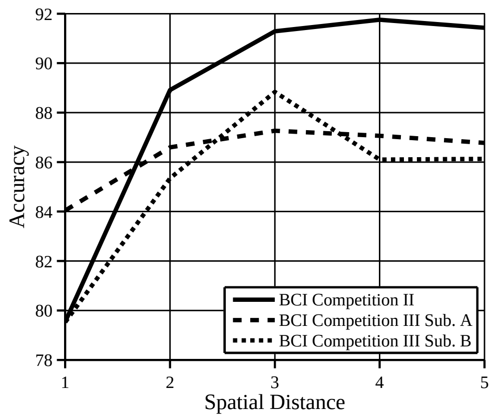


Figure 8.2: Percentage of correctly classified non-target events as a function of spatial distance from target row or column, for each subject.

4 Discussion

The results indicate a strong probability of AB and RB perceptual errors occurring in the Farwell & Donchin protocol (see Figure8.1a). Unfortunately the results were not consistent across all subjects (see Figure8.1b) and were not as pronounced when compared to studies done using RSVP [143, 142].

As expected the incidence of FP increases as the spatial distance between non-target and target event decreases (see Figure8.2). It was also shown that the FP events cause a somewhat consistent AB effect (see Figure8.3). This indicates that the FP is not simply due to random pertubations in the EEG but an attentional effect which affects later target evetns. What was suprising is that the overall accuracy for spatial distance of 1 was higher than that for spatial distances of 3 and 4. It was expected that the closer, spatially, the non-target events were to the target event the harder they would be to distinguish from the target event and therefore lead to a greater AB effect [143].

Relying on the classifier results to identify possible perceptual errors may be partly to blame for a number of the inconsistencies. Furthermore the SOA used for the P300 protocol is longer than what is used in many RSVP experiments, leading to a diminished effect in the earlier lags. The effects of the RB and AB were most noticeable for subject A and B from BCI competition III. The subject from BCI competition II was affected by the spatial proximity of the non-target events and showed a consistent RB effect. Although the AB effect was inconsistent across the spatial distances (Figure8.3c). This may be attributed to the high classification accuracy and small amount of available data in comparison to the data from BCI competition III. Martens et al. showed that some subjects do not produce the phenomenon of attentional blink. These individuals also tend to have shorter P300 latencies (indicating faster simulus processing) and larger P300 amplitudes. Therefore it might be expected that the best performing subject does not show any disernable AB effect.

Numerous studies by Gonsalvez et al. have shown that a primary determinant in P300 amplitude is the target-to-target interval (TTI), with longer target-to-target intervals yielding larger P300 responses than shorter intervals [134, 80, 81]. Subsequently some of the degradation in performance may also be attributed to the short TTIs.

It is clear though that temporal and in conjunction spatial proximity to target events lead to classification performance degradation. One solution would be to increase the SOA, but this would slow down the theoretical bit rate achievable by the BCI system. Another possible solution would be to decrease the SOA and have the rows and columns intensify separately with a 500 ms pause in between. This would increase the length of the shortest TTI (therefore increasing P300 amplitude) as well as reducing the effects of the RB and certain cases of AB perceptual errors. Although the increase in the shortest TTI may negate or possibly eliminate the effects of RB by decreasing the SOA the effects of AB due to FP may be exacerbated. This may be avoided by increasing the visual discriminability of the events, since this has shown to decrease AB [143]. The effects of AB may be further abated by arranging the order of the row/column intensifications in such a way as to limit the spatial proximity of succeeding events.

5 Conclusion

Through the results presented the possible occurrence of AB and RB in the Farwell & Donchin protocol has been confirmed. Their affect on classification is noticeable and the overall performance of the protocol may be improved especially for low performing subjects by eliminating or at least limiting their occurrence. Further investigations need to be carried out with a larger number of subjects to assess the true extent of the affect these perceptual errors have on the P300 matrix speller performance and whether the use of it as the standard P300 speller protocol should persist. Finally the validity of the identification of events as being affected by AB or RB should be tested by using other classifiers.

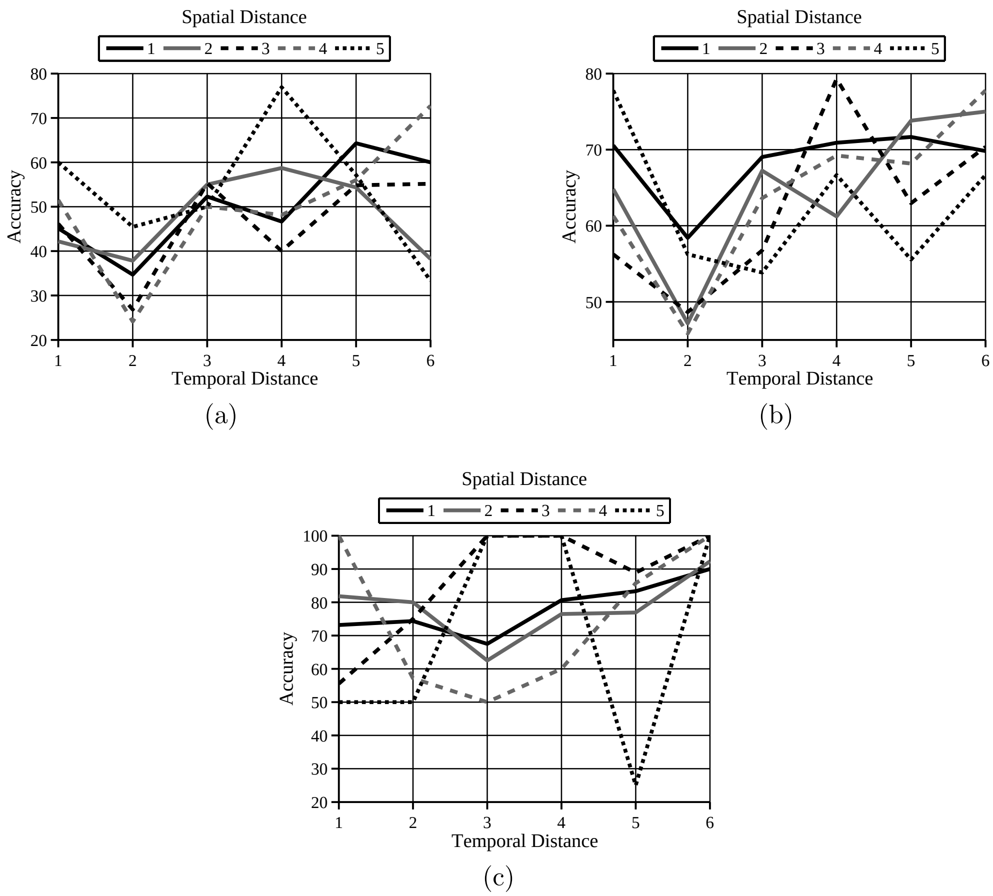


Figure 8.3: Percentage of correctly classified target events that follow false positives at various temporal lags. The spatial distances indicate how far the non-target rows/columns that were misclassified were from the target row/column. With 1 indicating the non-target row/column was next to the target row/column and 5 that they were seperated by 4 rows/columns. (a) BCI competition III Subject A (b) BCI competition III Subject B (c) BCI competition II.

References

[1] A Kübler, B Kotchoubey, J Kaiser, J R Wolpaw, and N Birbaumer. Brain-computer communication: unlocking the locked in. *Psychological Bulletin*, 127(3):358–375, May 2001. PMID: 11393301.

[2] J R Wolpaw, N Birbaumer, D J McFarland, G Pfurtscheller, and T M Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 113(6):767–91, June 2002. PMID: 12048038.

[3] Andrea Kübler and Nicola Neumann. Brain-computer interfaces–the key for the conscious brain locked into a paralyzed body. *Progress in Brain Research*, 150:513–525, 2005. PMID: 16186045.

[4] Brendan Z Allison, Elizabeth Winter Wolpaw, and Jonathan R Wolpaw. Brain-computer interface systems: progress and prospects. *Expert Review of Medical Devices*, 4(4):463–74, July 2007. PMID: 17605682.

[5] Mikhail A Lebedev and Miguel A L Nicolelis. Brain-machine interfaces: past, present and future. *Trends in Neurosciences*, 29(9):536–546, September 2006. PMID: 16859758.

[6] S G Mason, A Bashashati, M Fatourechi, K F Navarro, and G E Birch. A comprehensive survey of brain interface technology designs. *Annals of Biomedical Engineering*, 35(2):137–169, February 2007. PMID: 17115262.

[7] Niels Birbaumer and Leonardo G Cohen. Brain-computer interfaces: communication and restoration of movement in paralysis. *The Journal of Physiology*, 579(Pt 3):621–636, March 2007. PMID: 17234696.

[8] Niels Birbaumer, Ander Ramos Murguialday, and Leonardo Cohen. Brain-computer interface in paralysis. *Current Opinion in Neurology*, 21(6):634–638, December 2008. PMID: 18989104.

[9] G. Dornhege, J del R. Mill\'an, T Hinterberger, D J McFarland, and K R Müller. *Toward Brain-Computer Interfacing (Neural Information Processing)*. MIT Press, 1 edition, September 2007.

[10] Niels Birbaumer. Brain-computer-interface research: Coming of age. *Clinical Neurophysiology*, 117(3):479–483, March 2006.

[11] Ernst Niedermeyer and F.H.Lopes Da Silva. *Electroencephalography: Basic Principles, Clinical Applications and Related Fields.* Lippincott Williams and Wilkins, 5th revised edition edition, November 2004.

[12] Mark F. Bear, Barry W. Connors, and Michael A. Paradiso. *Neuroscience: Exploring the Brain.* Lippincott Williams & Wilkins,US, 2Rev ed edition, November 2000.

[13] Bruce J. Fisch, Rainer Spehlmann, and Bruce J. Fisch. *Fisch and Spehlmann's EEG Primer.* 1999.

[14] G Pfurtscheller. Functional topography during sensorimotor activation studied with event-related desynchronization mapping. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, 6(1):75–84, 1989. PMID: 2915031.

[15] Gert Pfurtscheller and F.H. Lopes da Silva. *Event-Related Desynchronization: Event-related Desynchronization Vol 6.* Elsevier, rev sub edition, November 1999.

[16] G Pfurtscheller, D Flotzinger, and C Neuper. Differentiation between finger, toe and tongue movement in man based on 40 hz EEG. *Electroencephalography and Clinical Neurophysiology*, 90(6):456–460, June 1994. PMID: 7515789.

[17] Ramaswamy Palaniappan. Utilizing gamma band to improve mental task based brain-computer interface design. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 14(3):299–303, September 2006. PMID: 17009489.

[18] J Polich. Updating p300: an integrative theory of p3a and p3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 118(10):2128–48, October 2007. PMID: 17573239.

[19] Kyung Hwan Kim, Jin Yoon, Ja Hyun Kim, and Ki-Young Jung. Changes in gamma-band power and phase synchronization with the difficulty of a visual oddball task. *Brain Research*, 1236:105–112, October 2008.

[20] Daria Osipova, Dora Hermes, and Ole Jensen. Gamma power is phase-locked to posterior alpha activity. *PLoS ONE*, 3(12):e3990, 2008. PMID: 19098986.

[21] G Pfurtscheller and A Aranibar. Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalography and Clinical Neurophysiology*, 46(2):138–146, February 1979. PMID: 86421.

[22] Rolf Verleger. P3b: towards some decision about memory. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 119(4):968–970, April 2008. PMID: 18222107.

[23] Rolf Verleger, Piotr Jaskowski, and Edmund Wascher. Evidence for an integrative role of p3b in linking reaction to perception. *Journal of Psychophysiology 2005*, 19(3):165–181, 2005.

[24] D S Ruchkin, R Johnson, H Canoune, and W Ritter. Short-term memory storage and retention: an event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, 76(5):419–39, November 1990. PMID: 1699736.

[25] R Johnson. A triarchic model of p300 amplitude. *Psychophysiology*, 23(4):367–84, July 1986. PMID: 3774922.

[26] J Polich and C Margala. P300 and probability: comparison of oddball and single-stimulus paradigms. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 25(2):169–76, February 1997. PMID: 9101341.

[27] J Polich. Probability and inter-stimulus interval effects on the p300 from auditory stimuli. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 10(2):163–70, December 1990. PMID: 2272863.

[28] E Donchin and M. Coles. Is the p300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11:357–374, 1988.

[29] R Beisteiner, P Höllinger, G Lindinger, W Lang, and A Berthoz. Mental representations of movements. brain potentials associated with imagination of hand movements. *Electroencephalography and Clinical Neurophysiology*, 96(2):183–193, March 1995. PMID: 7535223.

[30] W Becker and R Kristeva. Cerebral potentials prior to various force deployments. *Progress in Brain Research*, 54:189–194, 1980. PMID: 7220915.

[31] J T Wilke and R W Lansing. Variations in the motor potential with force exerted during voluntary arm movements in man. *Electroencephalography and Clinical Neurophysiology*, 35(3):259–265, September 1973. PMID: 4126176.

[32] Bridgitte Rockstroh. *Slow Cortical Potentials and Behaviour*. Urban & Schwarzenberg, 2 edition, December 1989.

[33] Ming Cheng, Xiaorong Gao, Shangkai Gao, and Dingfeng Xu. Design and implementation of a brain-computer interface with high transfer rates. *IEEE Transactions on Bio-Medical Engineering*, 49(10):1181–6, October 2002. PMID: 12374343.

[34] Xiaorong Gao, Dingfeng Xu, Ming Cheng, and Shangkai Gao. A BCI-based environmental controller for the motion-disabled. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 11(2):137–140, June 2003. PMID: 12899256.

[35] Yijun Wang, Ruiping Wang, Xiaorong Gao, Bo Hong, and Shangkai Gao. A practical VEP-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 14(2):234–239, June 2006. PMID: 16792302.

[36] Gernot R Müller-Putz, Reinhold Scherer, Christian Brauneis, and Gert Pfurtscheller. Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components. *Journal of Neural Engineering*, 2(4):123–130, December 2005. PMID: 16317236.

[37] Leigh R Hochberg, Mijail D Serruya, Gerhard M Friehs, Jon A Mukand, Maryam Saleh, Abraham H Caplan, Almut Branner, David Chen, Richard D Penn, and John P Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164–171, July 2006. PMID: 16838014.

[38] Marcel van Gerven and Ole Jensen. Attention modulations of posterior alpha as a control signal for two-dimensional brain-computer interfaces. *Journal of Neuroscience Methods*, 179(1):78–84, April 2009. PMID: 19428515.

[39] M Middendorf, G McMillan, G Calhoun, and K S Jones. Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Transactions on Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):211–4, June 2000. PMID: 10896190.

[40] S P Kelly, E Lalor, C Finucane, and R B Reilly. A comparison of covert and overt attention as a control option in a steady-state visual evoked potential-based brain computer interface. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 7:4725–4728, 2004. PMID: 17271364.

[41] Simon P Kelly, Edmund C Lalor, Ciarán Finucane, Gary McDarby, and Richard B Reilly. Visual spatial attention control in an independent brain-computer interface. *IEEE Transactions on Bio-Medical Engineering*, 52(9):1588–1596, September 2005. PMID: 16189972.

[42] Brendan Z. Allison, Dennis J. McFarland, Gerwin Schalk, Shi Dong Zheng, Melody Moore Jackson, and Jonathan R. Wolpaw. Towards an independent brain-computer interface using steady state visual evoked potentials. *Clinical Neurophysiology*, 119(2):399–408, February 2008.

[43] E Donchin, K M Spencer, and R Wijesinghe. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):174–9, June 2000. PMID: 10896179.

[44] L A Farwell and E Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–23, December 1988. PMID: 2461285.

[45] B.Z. Allison and J.A. Pineda. ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):110–113, 2003.

[46] Holger Hill and Markus Raab. Analyzing a complex visuomotor tracking task with brain-electrical event related potentials. *Human Movement Science*, 24(1):1–30, February 2005. PMID: 15949581.

[47] J B Polikoff, H T Bunnell, and W J Borkowski. Toward a p300-based computer interface. In *RESNA '95 Annual Conference and RESNAPRESS and Arlington Va.*, pages 178–180, 1995.

[48] B. J. Fisch. *Fisch and Spehlmann's EEG Primer: Basic Principles of Digital and Analog EEG*. Elsevier, 3 edition, December 1999.

[49] J R Wolpaw, D J McFarland, G W Neat, and C A Forneris. An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3):252–9, March 1991. PMID: 1707798.

[50] G Pfurtscheller, C Neuper, C Guger, W Harkam, H Ramoser, A Schlögl, B Obermaier, and M Pregenzer. Current trends in graz Brain-Computer interface (BCI) research. *IEEE Transactions on Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):216–9, June 2000. PMID: 10896192.

[51] A Kostov and M Polak. Parallel man-machine training in development of EEG-based cursor control. *IEEE Transactions on Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):203–5, June 2000. PMID: 10896187.

[52] J A Pineda, B Z Allison, and A Vankov. The effects of self-movement, observation, and imagination on mu rhythms and readiness potentials (RP's): toward a brain-computer interface (BCI). *IEEE Transactions on Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):219–22, June 2000. PMID: 10896193.

[53] W D Penny, S J Roberts, E A Curran, and M J Stokes. EEG-based communication: a pattern recognition approach. *IEEE Transactions on Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):214–5, June 2000. PMID: 10896191.

[54] N Birbaumer, T Elbert, A G Canavan, and B Rockstroh. Slow potentials of the cerebral cortex and behavior. *Physiological Reviews*, 70(1):1–41, 1990. PMID: 2404287.

[55] N Birbaumer, N Ghanayim, T Hinterberger, I Iversen, B Kotchoubey, A Kübler, J Perelmouter, E Taub, and H Flor. A spelling device for the paralysed. *Nature*, 398(6725):297–8, March 1999. PMID: 10192330.

[56] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, 2003.

[57] V Franc and V Hlavac. Statistical pattern recognition toolbox for matlab. Research Report 2004.08, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, June 2004.

[58] S Sutton, M Braren, J Zubin, and E R John. Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, 150(700):1187–1188, November 1965. PMID: 5852977.

[59] E Donchin. Presidential address, 1980. Surprise!...Surprise? *Psychophysiology*, 18(5):493–513, September 1981. PMID: 7280146.

[60] Sander Nieuwenhuis, Gary Aston-Jones, and Jonathan D Cohen. Decision making, the p3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131(4):510–532, July 2005. PMID: 16060800.

[61] R Mertens and J Polich. P300 from a single-stimulus paradigm: passive versus active tasks and stimulus modality. *Electroencephalography and Clinical Neurophysiology*, 104(6):488–497, November 1997. PMID: 9402891.

[62] N K Squires, K C Squires, and S A Hillyard. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38(4):387–401, April 1975. PMID: 46819.

[63] Y W Jeon and J Polich. P3a from a passive visual stimulus task. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 112(12):2202–2208, December 2001. PMID: 11738190.

[64] E Courchesne, S A Hillyard, and R Galambos. Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, 39(2):131–143, August 1975. PMID: 50210.

[65] M Fabiani, D Friedman, and J C Cheng. Individual differences in p3 scalp distribution in older adults, and their relationship to frontal lobe function. *Psychophysiology*, 35(6):698–708, November 1998. PMID: 9844431.

[66] S Yamaguchi and R T Knight. P300 generation by novel somatosensory stimuli. *Electroencephalography and Clinical Neurophysiology*, 78(1):50–55, 1991. PMID: 1701715.

[67] A Pfefferbaum, J M Ford, B J Weller, and B S Kopell. ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology*, 60(5):423–34, May 1985. PMID: 2580694.

[68] J Katayama and J Polich. Stimulus context determines p3a and p3b. *Psychophysiology*, 35(1):23–33, 1998. PMID: 9499703.

[69] D F Salisbury, C B Griggs, M E Shenton, and R W McCarley. The NoGo p300 'anteriorization' effect and response inhibition. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 115(7):1550–8, July 2004. PMID: 15203056.

[70] J L Smith, S J Johnstone, and R J Barry. Movement-related potentials in the Go/NoGo task: The p3 reflects both cognitive and motor inhibition. *Clinical Neurophysiology*, 119(3):704–714, March 2008.

[71] A Kok. On the utility of p3 amplitude as a measure of processing capacity. *Psychophysiology*, 38(3):557–577, May 2001. PMID: 11352145.

[72] C C Duncan-Johnson and E Donchin. On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology*, 14(5):456–67, September 1977. PMID: 905483.

[73] J Polich, T Brock, and M W Geisler. P300 from auditory and somatosensory stimuli: probability and inter-stimulus interval. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 11(2):219–23, August 1991. PMID: 1748598.

[74] K C Squires and E Donchin. Beyond averaging: the use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, 41(5):449–59, November 1976. PMID: 61848.

[75] R Johnson and E Donchin. Sequential expectancies and decision making in a changing environment: an electrophysiological approach. *Psychophysiology*, 19(2):183–200, March 1982. PMID: 7071298.

[76] P G Fitzgerald and T W Picton. Temporal and sequential probability in evoked potential studies. *Canadian Journal of Psychology*, 35(2):188–200, June 1981. PMID: 7248851.

[77] D L Woods and E Courchesne. The recovery functions of auditory event-related potentials during split-second discriminations. *Electroencephalography and Clinical Neurophysiology*, 65(4):304–15, July 1986. PMID: 2424743.

[78] D L Woods, E Courchesne, S A Hillyard, and R Galambos. Recovery cycles of event-related potentials in multiple detection tasks. *Electroencephalography and Clinical Neurophysiology*, 50(5-6):335–347, December 1980. PMID: 6160977.

[79] J Polich. P300, probability, and interstimulus interval. *Psychophysiology*, 27(4):396–403, July 1990. PMID: 2236442.

[80] C L Gonsalvez and J Polich. P300 amplitude is determined by target-to-target interval. *Psychophysiology*, 39(3):388–96, May 2002. PMID: 12212658.

[81] C J Gonsalvez, R J Barry, J A Rushby, and J Polich. Target-to-target interval, intensity, and p300 from an auditory single-stimulus task. *Psychophysiology*, 44(2):245–50, March 2007. PMID: 17343708.

[82] M Cass and J Polich. P300 from a single-stimulus paradigm: auditory intensity and tone frequency effects. *Biological Psychology*, 46(1):51–65, June 1997. PMID: 9255431.

[83] J Polich, S E Eischen, and G E Collins. P300 from a single auditory stimulus. *Electroencephalography and Clinical Neurophysiology*, 92(3):253–261, May 1994. PMID: 7514994.

[84] J Polich and M R Heine. P300 topography and modality effects from a single-stimulus paradigm. *Psychophysiology*, 33(6):747–52, November 1996. PMID: 8961797.

[85] R J Croft, C J Gonsalvez, C Gabriel, and R J Barry. Target-to-target interval versus probability effects on p300 in one- and two-tone tasks. *Psychophysiology*, 40(3):322–8, May 2003. PMID: 12946107.

[86] I M Tarkka and D S Stokic. Source localization of p300 from oddball, single stimulus, and omitted-stimulus paradigms. *Brain Topography*, 11(2):141–151, 1998. PMID: 9880172.

[87] H Davis, T Mast, N Yoshie, and S Zerlin. The slow response of the human cortex to auditory stimuli: recovery process. *Electroencephalography and Clinical Neurophysiology*, 21(2):105–113, August 1966. PMID: 4162003.

[88] J Polich, M Aung, and D J Dalessio. Long latency auditory evoked potentials: intensity, inter-stimulus interval, and habituation. *The Pavlovian Journal of Biological Science*, 23(1):35–40, March 1988. PMID: 3357711.

[89] W T Roth, P L Krainz, J M Ford, J R Tinklenberg, R M Rothbart, and B S Kopell. Parameters of temporal recovery of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology*, 40(6):623–632, June 1976. PMID: 57048.

[90] R T Knight. Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalography and Clinical Neurophysiology*, 59(1):9–20, February 1984. PMID: 6198170.

[91] Brendan Z Allison and Jaime A Pineda. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: implications for a BCI system. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 59(2):127–40, February 2006. PMID: 16054256.

[92] Peter Meinicke, Matthias Kaper, Manfred Heumann, and Helge Ritter. Improving transfer rates in brain computer interfacing: A case study. *Advances In Neural Information Processing Systems 15*, 15:1107—1114, 2003.

[93] Eric W Sellers, Dean J Krusienski, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. A p300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance. *Biological Psychology*, 73(3):242–52, October 2006. PMID: 16860920.

[94] S M M Martens, N J Hill, J Farquhar, and B Schölkopf. Overlap and refractory effects in a brain-computer interface speller based on the visual p300 event-related potential. *Journal of Neural Engineering*, 6(2):026003, April 2009. PMID: 19255462.

[95] Luca Citi, Riccardo Poli, Caterina Cinel, and Francisco Sepulveda. P300-based BCI mouse with genetically-optimized analogue control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 16(1):51–61, February 2008. PMID: 18303806.

[96] A Furdea, S Halder, D J Krusienski, D Bross, F Nijboer, N Birbaumer, and A Kübler. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, 46(3):617–625, May 2009. PMID: 19170946.

[97] Jeremy Hill, Thomas Navin Lal, Karin Bierig, Niels Birbaumer, and Bernhard Schölkopf. An auditory paradigm for Brain–Computer interfaces. http://eprints.pascal-network.org/archive/00002128/, 2005. Motivated by the particular problems involved in communicating with "locked-in" paralysed patients, we aim to develop a brain-computer interface that uses auditory stimuli. We describe a paradigm that allows a user to make a binary decision by focusing attention on one of two concurrent auditory stimulus sequences. Using Support Vector Machine classification and Recursive Channel Elimination on the independent components of averaged event-related potentials, we show that an untrained user's EEG data can

be classified with an encouragingly high level of accuracy. This suggests that it is possible for users to modulate EEG signals in a single trial by the conscious direction of attention, well enough to be useful in BCI.

[98] Shin'ichiro Kanoh, Ko ichiro Miyamoto, and Tatsuo Yoshinobu. A brain-computer interface (BCI) system based on auditory stream segregation. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2008:642–645, 2008. PMID: 19162737.

[99] Eric W Sellers and Emanuel Donchin. A p300-based brain-computer interface: initial tests by ALS patients. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 117(3):538–548, March 2006. PMID: 16461003.

[100] F Piccione, F Giorgi, P Tonin, K Priftis, S Giove, S Silvoni, G Palmas, and F Beverina. P300-based brain computer interface: reliability and performance in healthy and paralysed participants. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 117(3):531–7, March 2006. PMID: 16458069.

[101] Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens. An efficient p300-based brain-computer interface for disabled subjects. *Journal of Neuroscience Methods*, 167(1):115–125, 2008.

[102] F. Nijboer, E.W. Sellers, J. Mellinger, M.A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D.J. Krusienski, T.M. Vaughan, J.R. Wolpaw, N. Birbaumer, and A. Kübler. A p300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119(8):1909–1916, August 2008.

[103] Hilit Serby, Elad Yom-Tov, and Gideon F Inbar. An improved p300-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 13(1):89–98, March 2005. PMID: 15813410.

[104] Ulrich Hoffmann. *BAYESIAN MACHINE LEARNING APPLIED IN A BRAIN-COMPUTER INTERFACE FOR DISABLED USERS*. PhD thesis.

[105] A. Lenhardt, M. Kaper, and H.J. Ritter. An adaptive P300-Based online Brain–Computer interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16(2):121–130, 2008.

[106] D J Krusienski, E W Sellers, D J McFarland, T M Vaughan, and J R Wolpaw. Toward enhanced p300 speller performance. *Journal of Neuroscience Methods*, 167(1):15–21, 2008. PMID: 17822777.

[107] Manoj Thulasidas, Cuntai Guan, and Jiankang Wu. Robust classification of EEG signal for brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, 14(1):24–9, March 2006. PMID: 16562628.

[108] Matthias Kaper, Peter Meinicke, Ulf Grossekathoefer, Thomas Lingner, and Helge Ritter. BCI competition 2003–Data set IIb: support vector machines for the p300 speller paradigm. *IEEE Transactions on Bio-Medical Engineering*, 51(6):1073–6, June 2004. PMID: 15188881.

[109] Elena L Glassman. A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs. *IEEE Transactions on Bio-Medical Engineering*, 52(11):1851–62, November 2005. PMID: 16285389.

[110] Vladimir Bostanov. BCI competition 2003–Data sets ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Transactions on Bio-Medical Engineering*, 51(6):1057–61, June 2004. PMID: 15188878.

[111] V J Samar, A Bopardikar, R Rao, and K Swartz. Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. *Brain and Language*, 66(1):7–60, 1999. PMID: 10080864.

[112] Gilbert Strang and Truong Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press,U.S., 2nd revised edition edition, December 1996.

[113] Neng Xu, Xiaorong Gao, Bo Hong, Xiaobo Miao, Shangkai Gao, and Fusheng Yang. BCI competition 2003–Data set IIb: enhancing p300 wave detection using ICA-based subspace projections for BCI applications. *IEEE Transactions on Bio-Medical Engineering*, 51(6):1067–1072, June 2004. PMID: 15188880.

[114] S Makeig, M Westerfield, T P Jung, J Covington, J Townsend, T J Sejnowski, and E Courchesne. Functionally independent components of the late positive event-related potential during visual spatial attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 19(7):2665–80, April 1999. PMID: 10087080.

[115] S Makeig, M Westerfield, J Townsend, T P Jung, E Courchesne, and T J Sejnowski. Functionally independent components of early event-related potentials in a visual spatial attention task. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 354(1387):1135–44, July 1999. PMID: 10466141.

[116] A. Rakotomamonjy, V. Guigue, G. Mallet, and V. Alvarado. *Ensemble of SVMs for Improving Brain Computer Interface P300 Speller Performances*, pages 45–50. 2005.

[117] Alain Rakotomamonjy and Vincent Guigue. BCI competition III: dataset II- ensemble of SVMs for BCI p300 speller. *IEEE Transactions on Bio-Medical Engineering*, 55(3):1147–54, March 2008. PMID: 18334407.

[118] 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*, 3(4):299–305, December 2006. PMID: 17124334.

[119] A Bashashati, M Fatourechi, R K Ward, and G E Birch. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, 4(2):R32–57, June 2007. PMID: 17409474.

[120] F Lotte, M Congedo, A Lécuyer, F Lamarche, and B Arnaldi. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2):R1–R13, June 2007. PMID: 17409472.

[121] Jonathan R Wolpaw. Brain-computer interfaces (BCIs) for communication and control: a mini-review. *Supplements to Clinical Neurophysiology*, 57:607–13, 2004. PMID: 16106662.

[122] Andrea Kübler, Adrian Furdea, Sebastian Halder, Eva Maria Hammer, Femke Nijboer, and Boris Kotchoubey. A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Annals of the New York Academy of Sciences*, 1157:90–100, March 2009. PMID: 19351359.

[123] U. Hoffmann, G. Garcia, J.-M. Vesin, K. Diserens, and T. Ebrahimi. A boosting approach to p300 detection with application to Brain-Computer interfaces. In *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, pages 97–100, 2005.

[124] B Blankertz. BCI competition III. http://ida.first.fraunhofer.de/projects/bci/competition_iii/.

[125] B Blankertz. BCI competition II. http://ida.first.fraunhofer.de/projects/bci/competition_ii/.

[126] B Blankertz, K Müller, G Curio, T M Vaughan, G Schalk, J R Wolpaw, A Schlögl, C Neuper, G Pfurtscheller, T Hinterberger, M Schröder, and N Birbaumer. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. *IEEE Transactions on Bio-Medical Engineering*, 51(6):1044–51, June 2004. PMID: 15188876.

[127] A Starr, P Sandroni, and H J Michalewski. Readiness to respond in a target detection task: pre- and post-stimulus event-related potentials in normal subjects. *Electroencephalography and Clinical Neurophysiology*, 96(1):76–92, 1995. PMID: 7530191.

[128] J Polich. Response mode and p300 from auditory stimuli. *Biological Psychology*, 25(1):61–71, August 1987. PMID: 3447637.

[129] G Barrett, R Neshige, and H Shibasaki. Human auditory and somatosensory event-related potentials: effects of response condition and age. *Electroencephalography and Clinical Neurophysiology*, 66(4):409–19, April 1987. PMID: 2435521.

[130] D F Salisbury, B Rutherford, M E Shenton, and R W McCarley. Button-pressing affects p300 amplitude and scalp topography. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 112(9):1676–84, September 2001. PMID: 11514251.

[131] B Burle, F Vidal, and M Bonnet. Electroencephalographic nogo potentials in a no-movement context: the case of motor imagery in humans. *Neuroscience Letters*, 360(1-2):77–80, April 2004. PMID: 15082183.

[132] H Nakata, K Inui, Y Nishihira, A Hatta, M Sakamoto, T Kida, T Wasaka, and R Kakigi. Effects of a go/nogo task on event-related potentials following somatosensory stimulation. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 115(2):361–8, February 2004. PMID: 14744578.

[133] K J Bruin and A A Wijers. Inhibition, response mode, and stimulus probability: a comparative event-related potential study. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 113(7):1172–82, July 2002. PMID: 12088714.

[134] C J Gonsalvez, E Gordon, S Grayson, R J Barry, I Lazzaro, and H Bahramali. Is the target-to-target interval a critical determinant of p3 amplitude? *Psychophysiology*, 36(5):643–54, September 1999. PMID: 10442033.

[135] J E Raymond, K L Shapiro, and K M Arnell. Temporary suppression of visual processing in a RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, (18):849–860, 1992.

[136] N G Kanwisher. Repetition blindness: type recognition without token individuation. *Cognition*, 27(2):117–43, November 1987. PMID: 3691023.

[137] C Cinel, R Poli, and C Citi. Possible sources of perceptual errors in p300-based speller paradigm. In *Proceedings of the 2nd International Brain-Computer Interface Workshop and Training Course*, Graz, 2004.

[138] J Isreal, C Wickens, and E Donchin. P300 amplitude changes during a tracking task as a function of continuous variations of tracking difficulty. *Psychophysiology*, 16(16):179–183, 1979.

[139] T Geng and J Q Gan. Towards a virtual 4-class synchronous BCI using motor prediction and one motor imagery. Graz, 2008.

[140] Elisabeth V C Friedrich, Dennis J McFarland, Christa Neuper, Theresa M Vaughan, Peter Brunner, and Jonathan R Wolpaw. A scanning protocol for a sensorimotor rhythm-based brain-computer interface. *Biological Psychology*, August 2008. PMID: 18786603.

[141] R. Fazel-Rezai. Human error in p300 speller paradigm for Brain-Computer interface. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 2516–2519, 2007.

[142] Mika Koivisto and Antti Revonsuo. Comparison of event-related potentials in attentional blink and repetition blindness. *Brain Research*, 1189:115–26, 2008. PMID: 18035339.

[143] M M Chun. Types and tokens in visual processing: a double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology. Human Perception and Performance*, 23(3):738–55, June 1997. PMID: 9180042.

[144] Ralf H Trippe, Johannes Hewig, Claudia Heydel, Holger Hecht, and Wolfgang H R Miltner. Attentional blink to emotional and threatening pictures in spider phobics: electrophysiology and behavior. *Brain Research*, 1148:149–60, May 2007. PMID: 17367765.

[145] E K Vogel, S J Luck, and K L Shapiro. Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology. Human Perception and Performance*, 24(6):1656–74, December 1998. PMID: 9861716.

[146] Edward K Vogel and Steven J Luck. Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, 9(4):739–743, December 2002. PMID: 12613677.

[147] H E Schendan, N G Kanwisher, and M Kutas. Early brain potentials link repetition blindness, priming and novelty detection. *Neuroreport*, 8(8):1943–8, May 1997. PMID: 9223082.