DECODING BRAIN STATES USING FUNCTIONAL BRAIN IMAGING TECHNIQUES

BY MARIA PEIFER

A thesis submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Professor Laleh Najafizadeh

and approved by

_____

_____

_____

New Brunswick, New Jersey

October, 2015

ABSTRACT OF THE THESIS

Decoding Brain States Using Functional Brain Imaging Techniques

by Maria Peifer

Thesis Director: Professor Laleh Najafizadeh

Non-invasive neuroimaging techniques provide safe methods for investigating the functionality of the brain. Functional near infrared spectroscopy (fNIRS) is a non invasive brain imaging method, which uses light in the near infrared range to measure the changes in concentration of cerebral hemoglobin. Electroencephalography (EEG) is a noninvasive brain imaging technique that measures regional cortical activity by measuring the potential difference at various points on the surface of the scalp. In this work the two brain imaging techniques are used to decode brain states, using a paradigm for three conditions: rest, motor and motor imagery.

The first part of the study attempts the classification of motor and motor imagery by using least square support vector machine (LS-SVM) with a radial basis function kernel. The data was recorded using functional near infrared spectroscopy. All preprocessing methods are selected to be possible for execution in a real-time setting. The first goal was to determine the optimal window length and starting point for the extraction of features. Once the optimal window length was established, two feature selection methods were compared: Fisher discriminant ratio (FDR) and the combined method, which uses FDR and K-means. Reducing the number of features improved the classification time with negligible impact on the classification accuracy.

The second part of the study uses a LS-SVM with a linear kernel to perform two classifications on EEG data: rest and motor imagery, and rest and motor. The average power of the frequency band between $10\ Hz - 14\ Hz$ was used to extract features from each channel. The two feature selection methods previously mentioned were compared. As expected the combined method produced better results.

Acknowledgements

I would like to take this opportunity to thank my adviser, my professors, and my colleagues, who have shaped my path and without whom my research would not have been possible.

First, I would like to acknowledge my adviser, Dr. Laleh Najafizadeh. She has always motivated me to improve and excel and has always encouraged me to follow my research interest. I am tremendously thankful for her support during my graduate studies.

I want to acknowledge the two professors that have had an enormous influence on my research interest. I am grateful to Dr. Athina Petropulu for introducing me to the field of signal processing, and to Dr. Kristin Dana for taking the initiative to teach a course that introduced me to machine learning.

I would also like to thank Li Zhu for his collaboration and help in data collection.

Lastly, I would like to thank my family for their encouragement and support: my parents who have always stressed the importance of education and instilled in me a thirst for knowledge from a young age, my sister, and my husband who have supported me unconditionally.

Dedication

This is dedicated to my family and friends. Many thanks to my wonderful husband, Michael for encouraging and believing in me. I am especially grateful to my parents, Marian and Cornelia Oprea. I appreciate all the sacrifices they have made for my education. A special thanks goes to my sister Luciana, who has taught me to read at a young age, probably one of the most crucial elements in my education. I have really appreciated the support of my in-laws, John and Beverly Peifer, for welcoming me into their family and encouraging my studies. I would also like to address a big thanks to my godparents Dan and Simona Voic. They have been very supportive throughout the process and always have good advice to share.

Table of Contents

Chapter 1

Introduction

The motivation of this study was to develop methods for detecting brain states with application in a brain computer interface (BCI). BCI can enable severely disabled persons to interact and control their environment, and therefore, can significantly improve their quality of life [76]. An efficient BCI must meet two requirements: it must have great accuracy and it must have fast detection. Therefore, accuracy of detection as well as real-time capabilities have been considered in this study. Two brain imaging techniques were used: near infrared spectroscopy (NIRS) and electroencephalography (EEG).

Near infrared spectroscopy is an emerging technology that uses light in the near infrared range to indirectly measure brain activity. Changes in cerebral hemoglobin concentration are calculated using the measurements of the intensity of the emitted and detected light [73]. Compared to other brain imaging modalities such as electroencephalography (EEG) or magnetic resonance imaging (MRI), it is less prone to artifacts due to movement,and has a short preparation time [56].

EEG is a noninvasive neuroimaging technique that measures the potential at the surface of the scalp. While some systems have a lengthy setup time, due to the gel that needs to be added to each electrode, there are emerging technologies that use dry electrodes with similar performance [63]. EEG has great spatial resolution, with high sampling frequencies, which makes it ideal for BCI use.

In the first part of this study the detection of brain states using near infrared spectroscopy (NIRS) was attempted. Classification of motor imagery and motor was done using least square support vector machine with a radial basis function kernel.

Different time windows were used to extract features to find the optimal window length. Using the optimal window length, as determined by the accuracy of classification, from the full feature set, two feature selection methods were compared: Fisher discriminant ratio (FDR) and a combined method using FDR and K-means. The best classification was achieved using a time window of $0.4s$ at the onset of the block with 36 features.

In the second part of the study Electroencephalography (EEG) is used for detecting brain states during three conditions: rest, motor and motor imagery. Classification of rest and motor, and rest and motor imagery was attempted using least square support vector machine and the average powers of the frequency band between $10Hz$ and $14Hz$ of each channel as the features. Two methods of feature selection were compared: Fisher discriminant ratio (FDR) and a combined method using FDR and K-means. The combined method achieves better performance.

The second chapter introduces the background and the principles of the brain imaging techniques. The third chapter introduces basic concepts of machine learning as well as feature selection methods. The fourth chapter summarizes the current literature in decoding brain states using neuroimaging techniques for the application in BCI. The fifth and sixth chapter discuss the work done using two neuroimaging techniques. The final chapter summarizes the conclusions and discusses future work.

Chapter 2

Neuroimaging Techniques

2.1 Introduction

This chapter describes the types of signals that can be recorded as measures of brain functions, as well as functional imaging techniques.The first part of the chapter introduces how neuronal activation can be recorded directly or indirectly through the changes in hemodynamic concentrations. Tasks, intended to represent paradigms of specific functions, can be configured using a block design, event-related design or a mixed design.

The second part of the chapter introduces two imaging techniques, near infrared spectroscopy and electroencephalography, and their advantages and disadvantages. Near infrared spectroscopy, described in Section 2.4, uses light in the near infrared range to measure brain activity. Electroencephalography measures the electrical potential difference at the surface of the scalp to some reference point.

2.2 Measures of Brain activity

2.2.1 Neuronal Activity

The human brain is part of the central nervous system and has many functions including: control of body temperature, heart rate, breathing, processing sensory information, such as visual or olfactory information, movement, working memory and emotions. It contains neurons which are electrically excitable cells used for processing and transmission of information. With tens of billions of neurons connected by trillions of transition points, the brain is a complex structures with a processing power which far exceeds current supercomputers [3, 25].

The electrical signal used by neurons to communicate is called an action potential. This is a change in membrane potential caused by exchange of sodium and potassium ions through the membrane of the neuron with an amplitude of 100 mV [31]. The brain is able to decode information based on the path the action potential takes and not on the shape since action potentials are an all or none event.

2.2.2 Hemodynamic Response

Neuronal activation causes changes in the concentration of hemoglobin, both oxygenated and deoxygenated[47]. In preparation of the increased oxygen consumption during activation, there occurs an increase in the local cerebral blood flow (CBF) and the concentration of oxygenated hemoglobin (HbO_2) and a decrease in the concentration of deoxygenated hemoglobin (Hb) [47]. The increased oxygen consumption causes the concentration of oxygenated hemoglobin to drop and the concentration of deoxygenated hemoglobin to increase (Figure 2.4) [47].

2.3 Brain Imaging Experiments

2.3.1 Selection Factors

Structural and Functional Brain Imaging

Brain imaging falls into two major categories: structural and functional. Structural imaging is focused on the physical aspect of the brain and can detect abnormalities in the structure of the brain. Functional imaging is focused on the functions of the human brain and areas that are activated during specific tasks. In this type of imaging the signal is recorded over a period of time.

Resolution: Spatial and Temporal

Spatial resolution is the property that describes how sensitive a measurement is to differences in activation over a small distance. More measurements per unit of volume, result in a better spatial resolution. Temporal resolution is a measure of how sensitive

a recording is to changes in activation over a short period of time. A method with a higher temporal resolution will have the capability to record more samples per second.

Ease of Use

Ease of use can determine the applications for brain imaging methods. There are invasive techniques, such as electrocorticography [40], which necessitate a device to be implanted in the head and non-invasive methods, which take measurements at the surface of the scalp. Methods that do not disturb the subjects natural environment are preferred because they are assumed to be less distracting from the task and thus contribute to less noise.

The preparation time of each subject differs between imaging methods. Some methods, like certain electroencephalography systems require a cap with electrodes to be placed on the head and each electrode to be filled with gel in order to reduce electrode-scalp impedance [21].

The size of the equipment can greatly affect the range of applications and subjects that it can be used for. Equipment which is heavy or large in size might not be suitable to record brain activity from small children or elderly people.

2.3.2 Task Design

During a brain imaging experiment a subject is asked to perform a task which represents a paradigm of a specific function such as working memory, motor function, oddball detection or pattern recognition. There are three types of experimental design: block design, event-related design, and mixed design, shown in Figure 2.1. In a block design trials are presented consecutively within each block in order to increase the activation power of the signal. This is repeated across multiple blocks, however blocks of different type are often interspersed in order to reduce the subject's anticipation of following stimulus [80]. Event-related design has single trials followed by rest periods. The inter-trial interval (ITI) can be the same for all trials, as in the case of periodic single trial design, or it can differ from trial to trial, as in the case of jittered single trial design. The length of the ITI is dependent on the type of imaging method used. The

mixed design is a combination of block and event-related design and has blocks with varying ITIs.

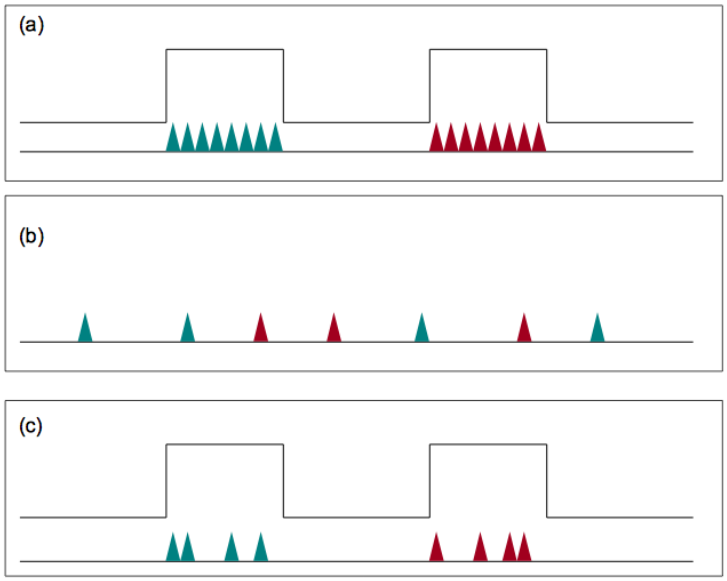

Figure 2.1: The three types of task design: (a) block design, (b) event-related design, (c) mixed design. The triangles of different colors represent different types of trials.

2.4 Functional Near Infrared Spectroscopy

2.4.1 The Basics

Neural activation in the brain is followed by a hemodynamic response consisting of a change in blood oxygenation and cerebral blood flow [8]. Therefore neural activation can be indirectly monitored by recording the changes in blood oxygenation. Biological tissue has a low optical absorbance for light in the near infrared range, however the chromophore hemoglobin has a relatively high absorbance both in its oxygenated and deoxygenated form. The absorption of oxygenated hemoglobin (Oxy Hb), deoxygenated hemoglobin (Deoxy Hb), and water are shown in Figure 2.2. It can be seen that if light were to be transmitted at a wavelength between 700 nm and 900 nm absorption will mainly be due to the presence of Oxy Hb and Deoxy Hb.

Functional near infrared spectroscopy (fNIRS) is an emerging functional brain imaging technology, which uses light in the near infrared range (700 nm to 1000 nm) to

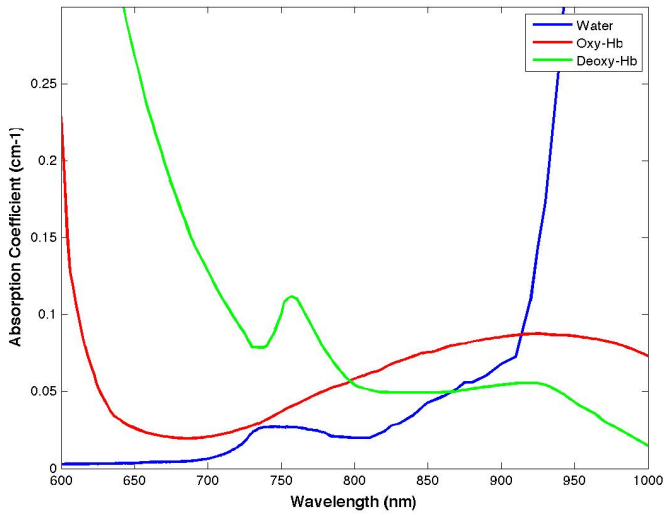

Figure 2.2: Absorption spectra for Oxygenated and Deoxygenated Hemoglobin in the Near-Infrared range. In lower frequencies deoxy-hemoglobin has higher absorption coefficient, while in higher frequencies oxy-hemoglobin has a higher absorption coefficient.

detect changes in cerebral concentration of oxygenated and deoxygenated hemoglobin related to brain activities [1, 6, 73].

During NIRS experiments, an array of detectors and low power NIR light sources are placed on the head. The light entering at a source position and exiting the head at a detector position samples a diffuse, banana-shaped volume along this path that is dependent on the distance between source and detector. Due to the low optical absorption of biological tissue at NIR wavelengths, NIR light can penetrate deep enough to sample the outer 1.5 - 2 cm of the head through skin and skull and reach the outer approximately 5 - 10 mm of the brain tissue, as shown in Figure 2.3.

2.4.2 Properties of NIRS

NIRS has good spatial resolution, and records a signal from localized activity [79]. It can detect activation on a cortical area with a depth of 1 cm to 3 cm from the surface, depending on the distance between the optodes. However, NIRS does have some spatial limitations. Unlike magnetic resonance imaging (MRI), it has a limited effective imaging depth and can only detect activation at the cortical area [56].

The temporal resolution of NIRS is better than brain imaging methods like fMRI,

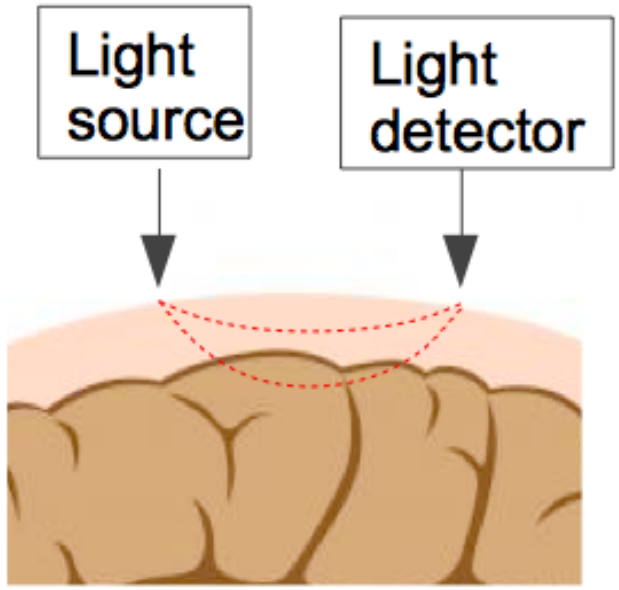

Figure 2.3: Propagation path of NIR light through scalp, skull and brain.

positron emission tomography (PET), and magnetoencephalography (MEG). Its recording rate can range from 2 Hz to 250 Hz [68]. The hemodynamic response follows the neural activation and peaks about $(5-6)$ s after the stimulus onset [10]. A typical hemodynamic response is shown in Figure 2.4. In order to fully capture activation, a time window of about 10 seconds is required [10].

NIRS, compared to Electroencephalography (EEG), is less prone to artifacts due to movement, which makes it ideal for measuring brain activity during everyday life activities and even freely moving animals [68]. Since changes in hemoglobin concentrations are measured, the recorded signal will have oscillations due to the heart beat. This signal has a frequency of 1 Hz and an amplitude that is lower than the peak of hemodynamic response. Motion of the sources and detectors on the hair can cause a drift in the hemodynamic response [70]. In order to remove noise due to heart beat and the baseline drift, bandpass filter can be applied with a bandwith of 0.01 Hz – 0.7 Hz. It has a short preparation time and is relatively cheap [70]. NIRS devices can be made small and portable and light which makes them suitable for brain imaging of infants and elderly population [48, 7].

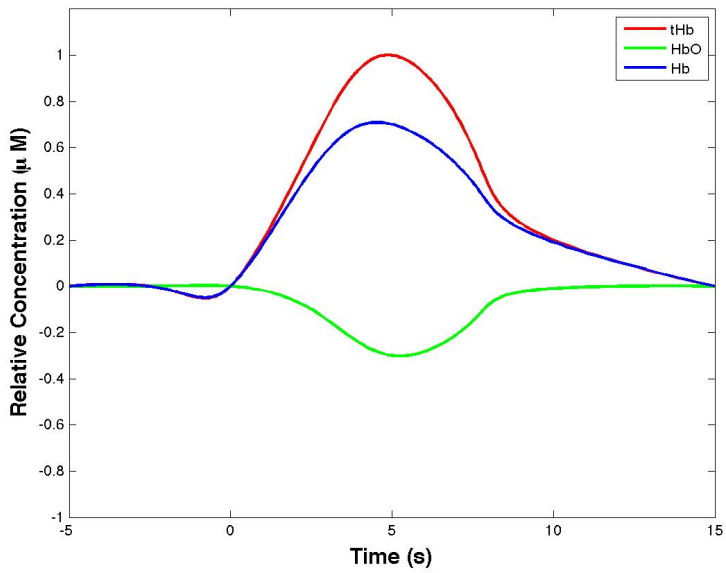

Figure 2.4: A typical hemodynamic response signal recorded using NIRS. [HbO] : oxyhemoglobin concentration, [Hb] : deoxyhemoglobin concentration and [tHb] : total hemoglobin concentration.

2.4.3 Modified Beer-Lambert Law

NIRS measures the light intensity at the detectors. The modified Beer-Lambert law can be used for the conversion of the raw data to oxygenated and deoxygenated hemoglobin. The law is based on the assumptions that the absorption is homogeneous across the illuminated area and that the scattering loss is constant over time [34]. From the raw data, attenuation can be calculated by using (2.1) [34].

$$A = ln\frac{I_{src}}{I_{det}} = L\mu_a + G \tag{2.1}$$

In (2.1) A is the attenuation, I_{src} is the incident light intensity, I_{det} is the detected light intensity, L is the total mean pathlength, μ_a is the absorption coefficient and G is a geometry dependent factor due to scattering, which is considered to be constant over time. To eliminate the unknown geometry dependent factor, it is common to evaluate the changes in attenuation relative to an initial state, as illustrated in (2.2).

$$\Delta A = ln\frac{I_{det_1}}{I_{det_2}} = L\Delta\mu_a \tag{2.2}$$

For each wavelength, the change in the absorption coefficient is a linear combination of the changes in the underlying concentrations of two chromophores: oxygenated hemoglobin and deoxygenated hemoglobin. This is illustrated in equation (2.3). α represents the molar absorption coefficient, which is dependent on the wavelength of the light (see Figure 2.2).

$$\Delta\mu_a = \alpha_{HbO_2}\Delta C_{HbO_2} + \alpha_{Hb}\Delta C_{Hb} \tag{2.3}$$

Since α is wavelength dependent, by emitting light at two different wavelengths, and solving the system of equations, ΔC for both chromophores can be estimated (2.4).

$$\begin{bmatrix} \Delta C_{HbO_2} \\ \Delta C_{Hb} \end{bmatrix} = \begin{bmatrix} \frac{-\alpha_{Hb}^{\lambda_2}}{\alpha_{Hb}^{\lambda_1}\alpha_{HbO_2}^{\lambda_2} - \alpha_{Hb}^{\lambda_2}\alpha_{HbO_2}^{\lambda_1}} & \frac{\alpha_{Hb}^{\lambda_1}}{\alpha_{Hb}^{\lambda_1}\alpha_{HbO_2}^{\lambda_2} - \alpha_{Hb}^{\lambda_2}\alpha_{HbO_2}^{\lambda_1}} \\ \frac{\alpha_{HbO_2}^{\lambda_2}}{\alpha_{Hb}^{\lambda_1}\alpha_{HbO_2}^{\lambda_2} - \alpha_{Hb}^{\lambda_2}\alpha_{HbO_2}^{\lambda_1}} & \frac{-\alpha_{HbO_2}^{\lambda_1}}{\alpha_{Hb}^{\lambda_1}\alpha_{HbO_2}^{\lambda_2} - \alpha_{Hb}^{\lambda_2}\alpha_{HbO_2}^{\lambda_1}} \end{bmatrix} \begin{bmatrix} \Delta A^{\lambda_1} \cdot (L^{\lambda_1})^{-1} \\ \Delta A^{\lambda_2} \cdot (L^{\lambda_2})^{-1} \end{bmatrix} \tag{2.4}$$

The mean pathlength factor is dependent on the distance from the source to the detector, as shown in equation (2.5). In this equation, DPF represents the differential pathlength factor and d represents the distance between source and detector.

$$L = DPF \times d \tag{2.5}$$

Differential pathlength factor cannot be computed and is different across subjects, therefore it was taken from literature [13].

2.5 Electroencephalography

2.5.1 The Basics

Electroencephalography (EEG) is one of the oldest non-invasive brain imaging methods [63] which records the signal from the postsynaptic activity of several neurons. In a typical EEG experiment, electrodes are placed on the subjects head and the potential difference of each electrode with respect to a common reference is measured. The most common electrode arrangement is the 10-20 system, shown in Figure 2.5 [45]. This

arrangement has 21 electrodes placed according to four reference points: the nasion, the inion and the preaurical point, next to the ear, on each side of the head. The nasion is the intersection of the frontal bone with the two nasal bones in the human skull and is located at the dip between the forehead and the nose. The inion is the most prominent part of the bone at the back of the skull. The intersection of the axis formed by connecting the anion to the nasion and the axis that connects the two preaurical points is the center of the head, where electrode Cz is placed. The distance between nasion and inion is measured and divided into 10 %, 20%, 20%, 20% , 20%, 10% of that distance [26]. The electrodes are placed at those points.

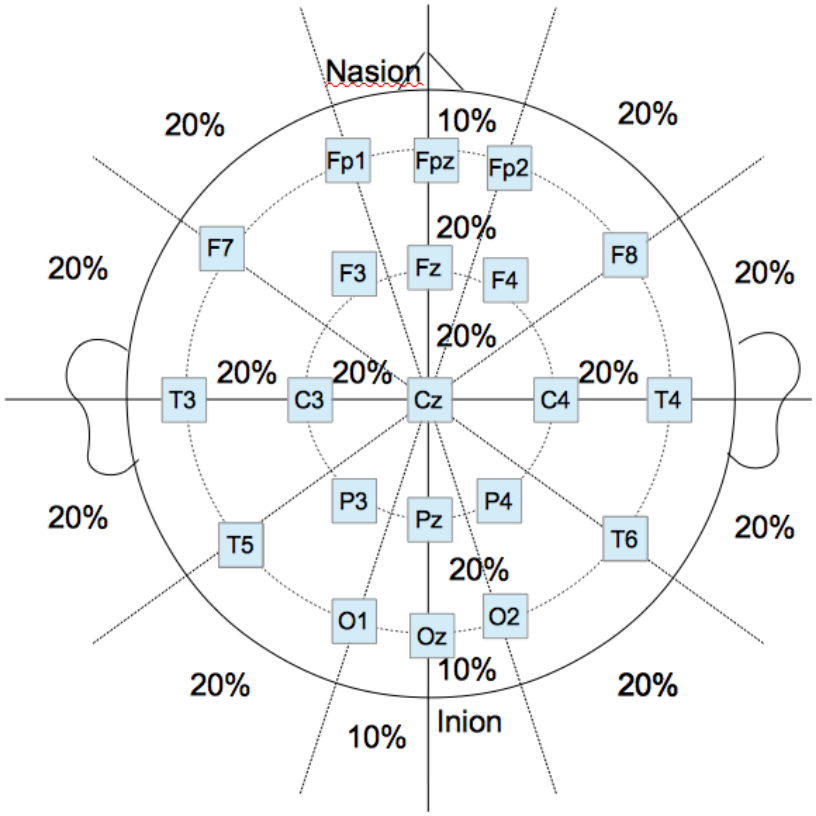

Figure 2.5: Illustration of the 10-20 system. Fp: frontal polar, F: frontal, C: central, P: parietal, O: occipital

2.5.2 Properties of EEG

EEG measures electrical potential at the surface of the scalp and, therefore, suffers from inter-trial variability and spacial mixing: different electrical sources across distances of over 5 cm are superimposed onto the scalp [51]. The localization of the neuronal generations of the EEG signal is an inverse problem. There have been proposed several methods for solving the inverse problem, i.e. minimum norm (MN) [24, 54], weighted minimum norm (WMN)[54] and low resolution brain electromagnetic tomography (LORETA)[53, 54], however, because of the nature of the problem, the solution is not unique, and therefore, there is no perfect solution [54]. Hence, EEG has poor spacial resolution.

EEG has a great temporal resolution, with sampling rates between 250 Hz and 1 kHz. The postsynaptic potential has a duration of 10 ms to 100 ms, which makes EEG the non-invasive imaging method of choice for analysis requiring good temporal resolution.

EEG is prone to noise due to movement, however there have been done successful recordings using EEG during movement [12]. Additional sources of artifacts can be attributed to movement of eyeballs and eyelids, tension of the muscles in the head and the neck, and electrical activity generated by the heart [30]. To minimize these artifacts, one option is to optimize the recording by choosing the right task, environment and electrode location [14]. Signal that is heavily affected by artifact can be removed altogether from the study, however this may cause bias in the selection of observations. Temporal filters can remove artifacts given that the noise is outside the frequency range of interest. Spatial filters can be can also be used to increase the signal-to-noise ratio [46]. Ocular artifacts can be measured and removed after the recording [14]. Artifacts can be removed by using independent component analysis (ICA) [30]. ICA attempts to identify the source signals that have been mixed to obtain the recorded signal. It works under the assumptions that the original source signals are independent, were linearly mixed, and that the number source signals is equal or less than the number of recorded signals [30].

Equipment is usually small and portable. Electrode caps are fairly comfortable, however, some devices require each electrode on the cap to be filled with gel, in preparation of the experiment . Hence, some EEG devices have a lengthy setup time which makes it less attractive for everyday use.

2.5.3 Analysis of EEG

EEG signal are generally analyzed in the time domain and in the frequency domain. In time domain EEG signals are often analyzed by their event related potential (ERP). In the frequency domain different oscillations are inspected.

Event related potential is a measure of the electrical response of the cortex to a cognitive, sensory or motor event induced by a stimulus. It is a result of a large number of action potentials which are linked to the event [67]. Event related potentials have a small amplitude ($1 - 30$ μV), however, unlike other aspects of the EEG, they are time-locked to the event and are assumed to be constant across samples [14]. Through the process of averaging the signal to noise ratio of the ERP is improved.

ERPs can be evaluated via three metrics: amplitude, latency and scalp distribution [67]. The amplitude measures the extent of neural activity in response to the event. The latency quantifies the timing of the activation and is the time interval between the stimulus onset and the time point at which the peak occurs. The scalp distribution is the pattern of the voltage gradient of a component over the scalp at any instant of time [67].

ERP components can be named after their polarity (Figure 2.6), positive or negative going, and their timing [77]. C1 is the first component following the stimulus and has a $(50 - 70)$ ms latency. This component is attributed to sensory processing and can be either positive or negative. P1 is a positive component with a latency of $(90 - 100)$ ms, which is associated with sensory and perceptual processing [77, 65]. N1 is a negative component with a latency of $(170 - 200)$ ms, which has been attributed to perceptual processing, expert recognition, and visual discrimination [77, 74]. The negative component N2 has a latency of $(225 - 250)$ ms and has been attributed to

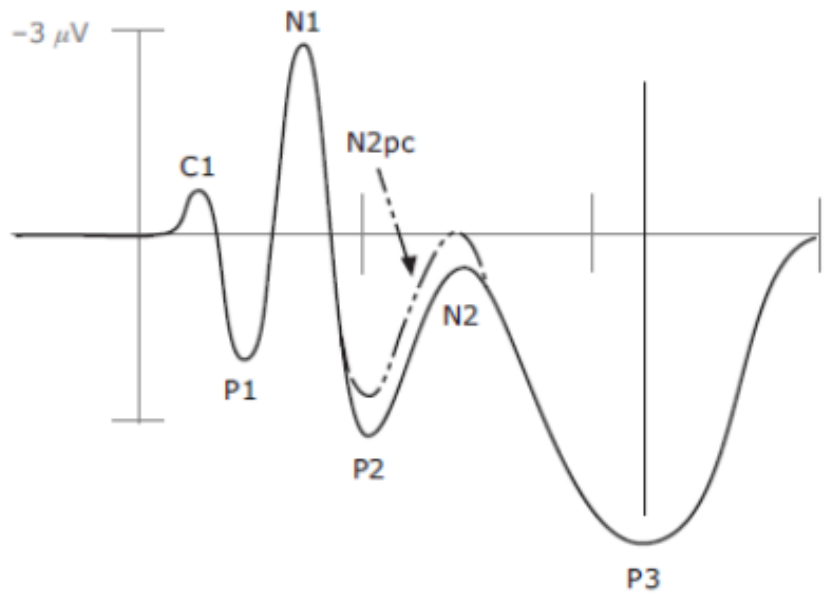

Figure 2.6: A model of an event related potential [77].

object recognition and categorization [77, 15, 66]. One of the most studied component is the P3 component (P300), with a latency of about 300 ms, it has been attributed to many processes: stimulus evaluation time, categorization, context (working memory) updating, cognitive load, maintenance in visual, working memory, and response preparation [77, 62, 57, 71].

Analysis of EEG in the frequency domain, relates to studying brain rhythms. In healthy adults, different brain rhythms are associated with different states, such as wakefulness and sleep [67]. These brain rhythms are associated with five frequency bands: alpha (α), beta (β), theta (θ), delta (δ), and gamma (γ) [67].

Delta waves have the lowest frequency band, $(0.5 - 4)$ Hz and are primarily associated with sleep and are sometimes present when awake. Artifacts due to muscle movement from the jaw and neck lie in the same frequency band, however, these usually have larger amplitudes [67]. Theta waves lie in the frequency range of $(4-8)$ Hz. These are associated with a state that is between consciousness and sleep, creative inspiration and deep meditation. The waves are present in childhood and infancy and are often use in developmental studies.

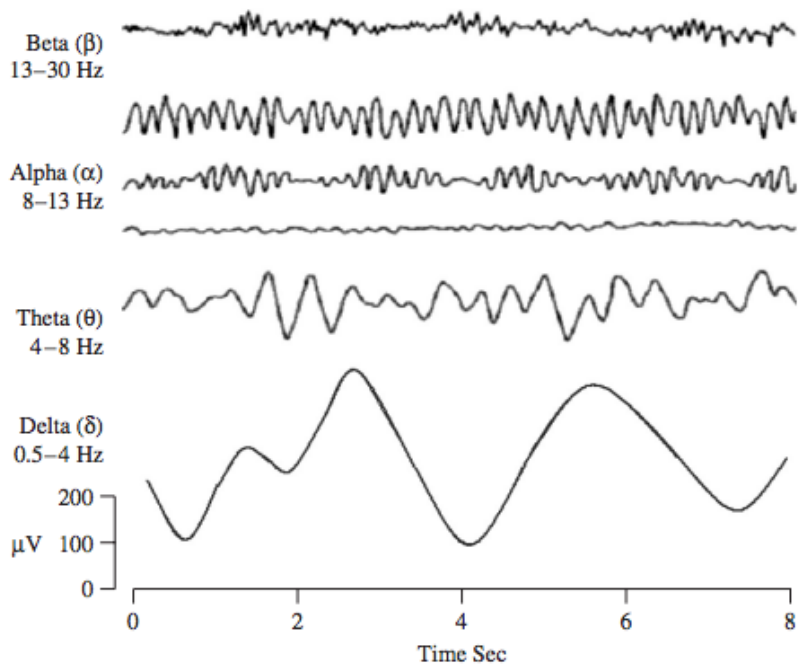

Figure 2.7: The four most prominent frequency bands from highest to lowest [67].

The alpha wave, with a range of $(8-13)\ Hz$, is found in the posterior region of the brain and prominently in the occipital region. This wave is present in human adults during relaxation and relative mental inactivity [50]. Alpha waves can be produced by subjects while they have their eyes closed which is believed to be a state of scanning for images or awaiting the next image [67]. These are reduced when eyes are opened or by some mental concentration.

Beta waves lie in the range $(13-30)\ Hz$ and is found mostly in the frontal and central regions of the brain. These waves are associated with active thinking, active attention, focus on the outside world, or solving concrete problems [67]. Subjects can experienced an increase in the frequency of beta rhythm during high stress or panic.

Frequencies that are above 30 Hz are part of the gamma range. These waves have a small amplitude and are not often encountered. They are sometimes used for detection of certain brain diseases. Increased frequencies of the EEG signal and increased cerebral blood flow are located in the frontocentral area of the brain [67].

The rhythms described are not easily detectable from raw EEG data. Signal processing tools must be used to extract the waveforms from the recorded EEG signal.

2.6 Conclusion

This chapter introduced the basics of functional brain imaging as well as the two functional imaging techniques used in the study presented in this thesis. Near infrared spectroscopy indirectly measures neuronal activity by emitting light and measuring the intensity of the light at the detector. The changes in concentrations of oxygenated and deoxygenated hemoglobin are obtained by using the modified Beer-Lambert law. NIRS is advantageous for applications requiring good spatial resolution. Electroencephalography measures neuronal activity by recording the potential difference at several points on the surface of the brain. This technique has a very good temporal resolution.

Chapter 3

Machine Learning

3.1 Introduction

The first part of this chapter discusses the basics of machine learning. Machine learning constructs algorithms that are used to learn patterns from data and make predictions on future data, such as assignment to a particular class. Classification is divided into several categories: supervised and unsupervised, linear or nonlinear, generative or discriminant, static or dynamic and stable or unstable. Data that is non-linearly separable can be mapped into a linear feature space using kernel methods.

The second part of this chapter introduces various feature selection methods. Subspace projection uses principal component analysis (PCA) to map the data into a lower dimensional feature space. Alternatively the wrapper approach performs the classification and uses the result to eliminate features. The filtering method assigns each feature an index based on the usefulness of that feature. An example of a filtering method, Fisher discriminant ratio, is discussed in more detail.

The final part of this chapter discusses some classification methods such as regression, linear discriminant analysis, ridge regression, support vector machine and least-square support vector machine. The approach to solving the classification problem differs between the type of problem that has a greater number of samples than of features and the type of problem that has a greater number of features than of samples. For each type of problem the optimal solutions of the linear methods described are equivalent.

3.2 The Basics

3.2.1 The Basic Structure of Classification

Classification uses machine learning methods to group data into specific classes [37]. It has two phases: the learning phase and the test phase. In the learning phase the algorithm uses given data to find the discriminant function which will successfully group the test data in their appropriate classes. The data set used for the learning phase is called the training data. In the testing phase, a new data set, different from the one used for training, is used to test the performance of the discriminant function. Data should be chosen randomly for the training and testing set. The process is repeated several times for cross validation.

3.2.2 Types of Learning Methods

There are two types of learning, supervised and unsupervised [37]. In supervised learning the labels of the data in the training set are known, and used to find the discriminant function. In unsupervised learning, the labels of the data in the training set are unknown. Finding the labels of the data in the training set is part of the training phase.

Supervised learning methods can be categorized as linear and nonlinear [37]. Linear learning methods have been preferred in many studies as they take less processing time and are more effective for high dimensional feature vector space. Linear learning methods are also an essential part in some feature reduction methods. Nonlinear learning methods offer more flexibility in finding the decision boundary, however, they do run the risk of over-fitting. Over-fitting can occur when there are vectors that are not representative of the class in the training data set.

Figure 3.1 is an example of a possibly overfitted training data. The points outside of the decision boundary could be outliers due to noise in the data or could just be faulty measurements. However, it is possible that they are truly representative of the class that they belong to. Figure 3.2 shows an example in which the two classes are only

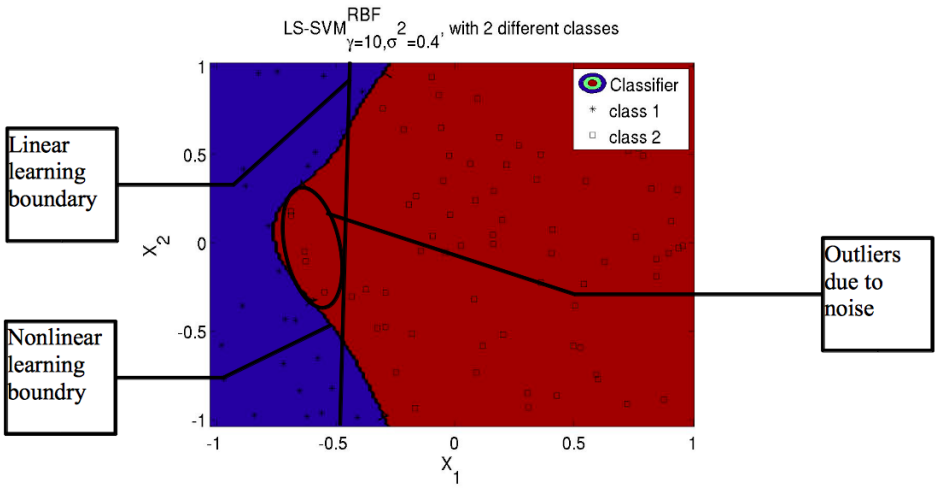

Figure 3.1: Example of possibly over-fitted nonlinear classification.

separable using a nonlinear kernel. In this example the two classes have a clear structure and symmetry. Because the structure is known the problem can be easily made linear by adding a new feature as seen in (3.1) (see Figure 3.3). However, most cases are not as simple as the example presented, and a method which make the problem linear is not evident.

$$X3 = X1 \times |X1| + X2 \times |X2| \tag{3.1}$$

In order to determine which type of learning is best suited for the type of data used in the study, a testing data set is used to test the accuracy of the classification. The testing or validation set is data set aside from the training set, which has known labels for all vectors. This data set is not used for training the classification algorithms, but is used to pick the best algorithm.

Classification methods can be categorized as generative or discriminative. Generative methods classify a vector by calculating the likelihood of it belonging to each class. The class with the highest probability of containing the vector is chosen [42]. Discriminative methods, on the other hand, learn an algorithm for discriminating between the classes [29].

Static classifiers do not include order, temporal or spatial, of the features in the vector during the classification. Dynamic classifiers, like hidden Markov model, take

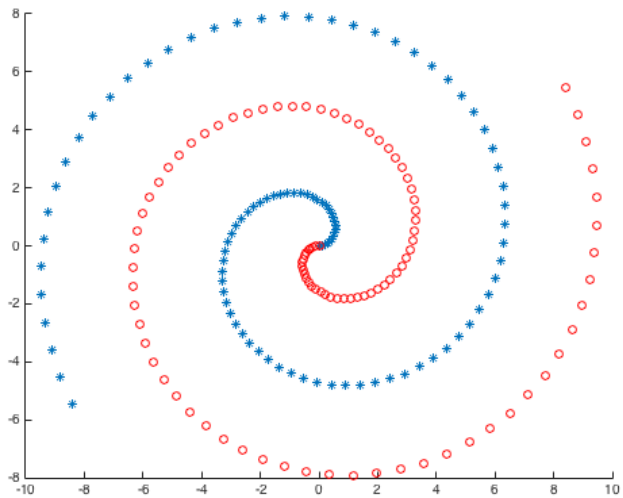

Figure 3.2: A two spiral classification example which can be done with LS-SVM using an RBF kernel.

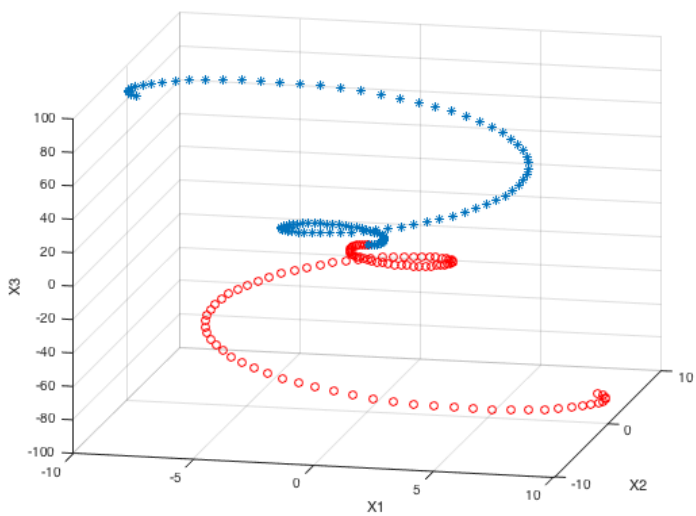

Figure 3.3: The two spiral classification example made linear by adding another feature.

the order of the features into account and can catch temporal or spatial dynamics [42].

Classification algorithms can also be divided into stable and unstable algorithms. The performance of stable classifiers is not affected by small changes in the training data. Due to their low complexity, these methods are less likely to overfit data. Unstable methods are affected by small variations in the training data, due to their high complexity [42].

3.2.3 Kernel methods

Kernel methods map the training data into a new feature space which is then used for classification [69]. This allows classification of data in high dimensional feature spaces with few computations [27]. Classification in the new feature space is linear and thus a nonlinear problem can be made linear with the use of a kernel [27].

The training vectors can be put in an $M \times N$ matrix X, where M is the number of features and N is the number of training vectors. Linear classification attempts to find the decision vector w and the threshold b such that:

$$X^T w + b \cdot e = y, \tag{3.2}$$

where y is the vector that contains the labels of each training vector. All Kernel methods satisfy the learning subspace property (LSP), which states that the optimal solution w is in the subspace of the training vectors space.

$$w = \sum_{n=1}^{N} a_n x_n = Xa \tag{3.3}$$

Using LSP (3.2) becomes:

$$X^T w = X^T X a = Ka \tag{3.4}$$

In (3.4) K is the linear kernel function. A linear classification model can be written in Kernel form. The decision function will change as follows:

$$f(x) = w^T x + b = x^T w + b = \sum_{n=1}^{N} x^T x_n a_n + b = \sum_{n=1}^{N} K(x, x_n) a_n + b \tag{3.5}$$

The kernel function can also be nonlinear. Most commonly used functions are the polynomial kernel function (3.6) and the Gaussian kernel function (Radial Basis function) (3.7) [37]. In (3.6) p is the order of the polynomial, and σ is a kernel parameter [72].

$$K(x, y) = \left(1 + \frac{x \cdot y}{\sigma^2}\right)^p \tag{3.6}$$

$$K(x, y) = \exp\left\{-\frac{||x - y||^2}{2\sigma^2}\right\} \tag{3.7}$$

When a nonlinear kernel is used, the data is mapped to an intrinsic vector space Φ, such that $K = \Phi^T \Phi$. The vectors in the intrinsic space have a greater length than those in the original space, and thus more features are obtained [37, 72]. When the feature space is already large a linear kernel is preferred [72].

3.3 Feature Reduction

3.3.1 Subspace Projection and Feature Selection

Feature reduction is an important part in improving the accuracy of a classifier by minimizing the number of features and removing features that are noisy or redundant. Furthermore, reducing the number of features, reduces the size of w and leads to fewer computations.

The two main techniques for dimensionality reduction of features are subspace projection and feature selection. Reducing the dimension of the features can reduce the computational cost and the risk of over-fitting in cases in which the training set has a vector dimension (M) that is much larger than the number of samples (N) [37]. Subspace projection uses PCA to identify the optimal subspace. It reconstructs the

training data using the principal components with the highest eigenvalues. Feature selection uses either the filtering approach, in which a score is assigned to each feature, or the wrapper approach, which uses a linear classifier and selects features based on their roles in the classification result [23, 35]. The wrapper approach uses the values of the decision vector for each feature as scores. Features with lower scores are considered less effective. Filtering methods include the signed SNR score [22], Fisher discriminant ratio (FDR) [55], Symmetric divergence [44] and two sample t-test.

3.3.2 Fisher Discriminant Ratio

Fisher discriminant ratio compares distance of centroids for each class along each feature with the sum of the variances of each class (3.8) [55].

$$FDR(j) = \frac{(\mu_j^+ - \mu_j^-)^2}{(\sigma_j^+)^2 + (\sigma_j^-)^2} \tag{3.8}$$

In (3.8) μ_j^+ and σ_j^+ are the mean and standard deviation respectively of feature j across all training samples belonging to class +. FDR scores each feature based on how well it separates the two classes. Removing features based on this scoring maximizes inter-class separability. This method works very well for removing features that are not characteristic to the classes.

3.3.3 K-means

K-means is an unsupervised machine learning method which groups data into a selected number of classes. The objective is to minimize the variation within a class, by minimizing the euclidean distance from the centroid of the class to its members. In the initial phase the centroids of the classes are chosen randomly [37]. For every vector x_t the distance to the centroid of each cluster is computed and compared. The vector is assumed to belong to the class k with the smallest distance:

$$k = \underset{j:j=1,\ldots,K}{arg\ min} ||x_t - \mu_j||_2^2, \tag{3.9}$$

where μ_j is the centroid of class j. After a new vector is added to a class the centroid of that class needs to be updated:

$$\mu_k = \frac{1}{N_k} \sum_{x_t in C_k} x_t, \tag{3.10}$$

where N_k is the number of samples in cluster C_k. The process of matching a vector to a class and updating the centroid of that class is repeated until the equilibrium is found and (3.11) is minimized [28]:

$$\min \sum_{k=1}^{K} \sum_{x_t in C_k} ||x_t - \mu_k||_2^2 \tag{3.11}$$

K-means can be applied along the features to group features which have a similar behavior. Features, which belong to the same group are similar and therefore, it is redundant to use both for classification. A features from a class provides little new information from another feature from the same class. Using K-means as a feature selection tool minimizes inter-class similarity between features and thus eliminates redundancy.

FDR and K-means can be combined to reduce features, such that both inter-class separability is maximized and intra-class similarity is minimized. As it will be discussed in chapters 5 and 6, FDR is applied first to all features and the features with the lowest scores are dropped. K-means is used to group the remaining features into k classes, for which k is the desired final number of features. The feature with the highest FDR score from each class is selected for classification.

3.4 Classification Methods

3.4.1 Methods for the overdetermined case

There are several techniques used to solve the classification problem. In the overdetemined case, $N \geq M$, there are more equations than unknowns which means that the solution either exists and is unique or does not exist but an approximation can be found.

Linear regression attempts to find the decision vector w and the threshold b that minimize the least square error [4].

$$E_{LSE}(w,b) = ||\epsilon||^2 = ||(X^T w + b \cdot e - y)||^2 \tag{3.12}$$

Fisher's discriminant analysis (FDA) also known as linear discriminant analysis (LDA) attempts to maximize the signal to noise ratio. The signal is considered the distance between the centroids of the two classes and the noise represents the variance between the cluster. The optimal solutions for linear regression and FDA are equivalent [37].

3.4.2 Methods for the underdetermined case

In the underdetermined, case $M \geq N$, in which there are more unknowns than equations, an additional condition is needed to ensure the uniqueness of the solution. The objective of linear ridge regression is to minimize both the error and the decision vector [37].

$$\min_{w,b} E(w,b) = \min_{w,b} \left\{ \|\epsilon\|^2 + \rho \|w\|^2 \right\} \tag{3.13}$$

Similarly, FDA has been modified for the underdetermined case. The new model is called perturbation discriminant analysis and its optimal solution is equivalent to that of linear ridge regression [37]. Support vector machine(SVM) has the objective of minimizing the decision vector as shown in equations (3.14) and (3.15).

$$\mathbf{min} \, \|w\|^2 \tag{3.14}$$

$$\mathbf{subject\ to} \quad y_i(w^T x_i + b) \leq 1 \tag{3.15}$$

SVM is a supervised binary classification method, that attempts to find the hyperplane that best separates the data into the two classes [16]. SVM only uses the samples that are closest to the separating hyperplane to train the algorithm. These are called support vectors [16].

3.4.3 Least Square Support Vector Machine

Least square support vector machine (LS-SVM) is an extension of the support vector machine model. Its optimal solution is equivalent to that of kernel ridge regression (KRR). When a linear kernel is used with LS-SVM, then the optimal solution is equivalent with the linear ridge regression [37]. In LS-SVM all vectors are considered support vectors and used in classification. The model is as follows:

$$\min_{w,e} \frac{1}{2}||w||^2 + \frac{1}{2}c\sum_{k=1}^{N} e_k^2 \tag{3.16}$$

$$\textbf{subject to} \quad y_k(w^T x + b) = 1 - e_k \quad k = 1, .., N \tag{3.17}$$

The constraints can be derived from the original requirements by adding the possibility of an error e. In order to solve the problem the Lagrangian is taken and maximized. To find the maximum of the Lagrangian the partial derivatives are taken and set to zero [72].

$$\mathcal{L}(w, b, e, \alpha) = \frac{1}{2}w^T w + c\frac{1}{2}\sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N} \alpha_k\{y_k[w^T x_k + b] - 1 + e_k\} \tag{3.18}$$

$$\frac{d\mathcal{L}}{dw} = 0 \Rightarrow w = \sum_{k=1}^{N} \alpha_k y_k x_k, \tag{3.19}$$

$$\frac{d\mathcal{L}}{db} = 0 \Rightarrow \sum_{k=1}^{N} \alpha_k y_k = 0 \tag{3.20}$$

$$\frac{d\mathcal{L}}{de} = 0 \Rightarrow \alpha_k = ce_k, \quad k = 1, ..., N \tag{3.21}$$

$$\frac{d\mathcal{L}}{d\alpha} = 0 \Rightarrow y_k[w^T x_k + b] - 1 + e_k = 0, \quad k = 1, ..., N. \tag{3.22}$$

This creates a set of 2N+2 equations with the same number of unknowns, which has a unique solution. Least squares-support vector machine (LS-SVM) has many advantages over traditional techniques [38]: it requires few training points with large dimensions, also called the small n large p problem; it is fast at classification once the parameters are set after training; it achieves good classification on new data.

3.5 Performance Measure

The most widely used measure to assess the performance of a classification method is accuracy as described in (3.23).

$$\text{accuracy} = \frac{\text{number of vectors which are classified correctly}}{\text{total number of classified vectors}} \tag{3.23}$$

Accuracy gives a good measure of how many vectors are classified correctly, however, it fails to account for bias towards classifying one class over the other. Accuracy of the classification of each individual class can offer more information of the performance. Specificity and sensitivity also used to further asses the performance of the classification of each class.

$$\text{Sensitivity} = \frac{\text{\# of true positives}}{\text{\# of true positives} + \text{\# of false negatives}} \tag{3.24}$$

$$\text{Specificity} = \frac{\text{\# of true negatives}}{\text{\# of true negatives} + \text{\# of false positives}} \tag{3.25}$$

True positives are the vectors that have been correctly classified as belonging to the positive class. False negatives are the vectors that have been falsely classified to the negative class but in fact belong to the positive class. True negatives are vectors which are correctly classified as belonging to the negative class. False positives are vectors which belong to the negative class and have been classified incorrectly.

$$\text{accuracy} = \frac{N^{+}\text{sensitivity} + N^{-}\text{specificity}}{N} \tag{3.26}$$

Accuracy can be obtained from specificity and sensitivity as seen in (3.26). N^{+} is the total number of vectors that belong to the positive class, N^{-} is the total number of vectors that belong to the negative class, and N is the total number of vectors classified. N is the sum of N^{+} and N^{-}.

Ideally specificity will be equal to sensitivity which will make them equal to the accuracy. In this case there isn't any bias towards one class or another. An accuracy is

considered significant, if it is above 75% and has both sensitivity and specificity above 50%.

3.6 Conclusion

This chapter first introduced some basic principles of machine learning and the types of classification methods. Linear classification methods are simpler and are less likely to overfit data. However, nonlinear methods are able to learn more complicated patterns in data.

Feature selection is an important part of classification because it discards redundant or noisy data and reduces the computational complexity. There are several approaches such as subspace projection, filtering and wrapper method for feature reduction. Fisher discriminant ratio was described in more detail in this chapter because it is a feature selection method used in this study. K-means is a nonsupervised classification method, not traditionally used in feature selection, however, it was included in the feature selection methods because it was used for this purpose in this study.

Several linear classification methods are presented in this chapter. A distinction is made between the type of problems that have a greater number of features than of samples and those that have a greater number of samples than features. Most brain imaging studies are of the former type and the appropriate methods must be used. Least square- support vector machine algorithm was chosen in this study because it offers fast classification and has a good performance on data that has few samples and many features.

Chapter 4

Previous Work

4.1 Introduction

This chapter will discuss previous research in decoding brain states. It is mainly focused on the work that is intended to be applied to brain computer interface. In the Near Infrared spectroscopy field this is a new topic. Research has been focused on finding a feature space that improves accuracy and reduces latency. EEG has long been recognized for its potential suitability for a brain computer interface due to its high temporal resolution. In addition to the traditional left and right motor classification, recent research has been extended to multi-class classification [58] as well.

4.2 Common Paradigms

There are two strategies for designing a brain computer interface: the direct approach [9, 10, 49, 70] and the indirect approach [2, 32]. The direct approach to designing a BCI uses motor tasks or other sensory actions, which are then classified to represent the same movement or actions. Most commonly researched classification problems are finger tapping and rest [10], right and left finger tapping [70], right and left motor imagery [49, 70], right and left wrist movement [49], and target selection using motor imagery [9]. The indirect approach uses mental tasks, such as mental arithmetic [2, 32] and mental counting [32]. The mental tasks can be combined with motor tasks in order to add more movement directions, that the BCI can decode, as is done in [32]. The combination is advantageous because mental tasks will cause a hemodynamic response in the prefrontal cortex, while motor tasks will cause a response in the motor cortex [32]. Having activations in different areas of the brain increases separability of the two

classes. A stronger hemodynamic response of a task can also improve its separability from another class. Therefore, there are classification problems that are naturally easier. It has been shown that actual movement has a higher hemodynamic response than motor imagery and therefore will have higher classification results when differentiating between left and right movement [70]. The classification problem of right and left movement has been shown to have good separability due to the fact that the left and right movement cause a hemodynamic response in the right and left hemisphere, respectively [70].

Many brain imaging classification studies have focused on a variety of tasks such as motor [41], motor imagery [39, 58, 75, 60], or on classification of various mental tasks [17]. In addition to the traditional binary left and right motor or motor imagery classification, the decoding of brain states for other types of imagery movement such as tongue [58], both feet[58, 33] and both hands [33] has been attempted. Multi-class motor task classification has in general a poorer performance than a binary classification [33]. Multi-class classification accuracy can be improved by using a larger time segment of data for classification [17].

4.3 Machine Learning using NIRS Data

In recent years near infrared spectroscopy has been used in research for brain computer interface. It is a promising brain imaging method because it is non-invasive and offers good spatial resolution, portability, and affordability [9, 70]. NIRS also has its limitations. It has a poor temporal resolution due to the latency of the hemodynamic [70] which makes its use for real-time applications challenging. The signal to noise ratio can be affected by the hair on the head, as well as movement of the optodes [70].

Feature extraction and feature selection play an important role in the classification. Commonly used features are changes in oxygenated hemoglobin and deoxygenated hemoglobin at every time point in a window [10, 70], and average over a time window [49, 2]. Features can be extracted by using linear regression to fit a line through all the data points (representing oxygenated hemoglobin) in a time window [49]. The length

and the start time of the window can affect the classification accuracy and should be carefully selected [49]. Additional features can be first (4.1) and second order gradient (4.2) [10]. In (4.1) and (4.2) $x(t)$ is the signal at time point t.

$$\nabla x(t) = x(t) - x(t-1) \tag{4.1}$$

$$\nabla^2 x(t) = x(t+1) + x(t-1) - 2x(t) \tag{4.2}$$

Feature selection is often done using parameters that quantify the usefulness of an individual feature [2, 10]. Most commonly used parameters are contrast-to-noise-ratio(CNR) [10], which is the square root of FDR, and mutual information index [2]. Thus it can be determined whether the signal should be obtained from oxygenated hemoglobin only [49] or both oxygenated hemoglobin and deoxygenated hemoglobin [2], a single channel or from multiple channels [10].

The classification method can affect the performance. A linear, discriminative method is most commonly used because it is simple and does not overfit training data [9, 10, 49, 70]. Generative methods, such as hidden Markov model [70] and naive Bayes Parzen window [2] have also been used for classification.

Table 4.1 shows some recent results. A direct comparison of accuracies is informative only within the same study, because many parameters including task, data collection, and feature space are different across studies. Using the average of data points over a time window, generally reduces noise but also throws out some information. This feature space has been shown to produce results that are less accurate than the feature space using all data points in the time window [49]. A feature space, which includes both oxygenated and deoxygenated hemoglobin, shows a slight improvement in accuracy, but this varies across individuals [10]. Including both is especially beneficial when the time window has a short length [10]. The first and second order gradient feature spaces do not improve classification accuracy [10]. In general adding linear combinations of features to a feature space does not change the accuracy, when using a linear classifier.

Table 4.1: Comparison of Previous Classification Methods using NIRS.

Source	**Method**	**Feature Space**	**Task**	**Average Accuracy** %
[10]	linear SVM	2 s time segment of Oxy-Hb	finger tap	83.5
[10]	linear SVM	2 s time segment of Oxy-Hb and deoxy-Hb	finger tap	83.58
[10]	linear SVM	2 s time segment multi channel	finger tap	85.5
[70]	linear SVM	10 s segment of 20 channels	left vs. right finger tap	87.5
[70]	linear SVM	10 s segment of 20 channels	left vs. right motor imagery	73.1
[49]	LDA	5 s averaged segment	right and left wrist motor imagery	77.56
[49]	LDA	5 s linear regression fitted line	right and left wrist motor imagery	87.28
[49]	LDA	10 s averaged segment	right and left wrist motor imagery	73.35
[49]	LDA	10 s linear regression fitted line	right and left wrist motor imagery	83.0
[2]	NBPW	average of the 12 s trial over 2 channels	easy vs hard mental arithmetic	62.0
[2]	NBPW	average of the 12 s trial over 2 channels	easy vs medium mental arithmetic	64.2
[2]	NBPW	average of the 12 s trial over 4 channels	easy vs hard mental arithmetic	69.4
[2]	NBPW	average of the 12 s trial over 4 channels	easy vs medium mental arithmetic	72.2
[2]	NBPW	average of the 12 s trial over 5 channels	easy vs hard mental arithmetic	69.8
[2]	NBPW	average of the 12 s trial over 5 channels	easy vs medium mental arithmetic	72.9

4.4 Machine Learning using EEG Data

The performance of the classification is greatly affected by the features used. Electron configuration and the duration of the trials determine the available information. The choice of electrode configuration is made more difficult by EEG's poor spatial resolution. The location of the neuronal activity cannot be uniquely determined by the EEG recording [61]. One approach is to use a small number of electrodes locally, on the surface of the scalp near the location of the expected neuronal activity [17, 58, 63, 75]. Another frequently used configuration has electrodes evenly spread over the surface of the scalp [33, 41, 63]. However, including 64 channels, evenly placed on the entire scalp, in a configuration can add redundancy. Furthermore, studies have shown that it is sufficient to include between 16 and 20 channels in order to maximize performance with the electrodes partly concentrated over the area of expected neural activity and partly spread over the rest of the surface of the scalp [39].

From the recorded data the feature space can be obtained from the time domain [17, 33], frequency domain [41, 58, 63] or time-frequency domain [75]. Motor imagery causes movement related potential (MRP) as well as event-related desyncronization/synchronization (ERD/ERS) signals [41]. MRP is an event-related potential that can be observed up to 500 ms before movement onset [41]. Decreases and increases of power in certain frequency bands can be observed during the performance of the tasks. This can be caused by an increase or decrease in synchrony of the underlying neuronal populations [59]. The decreases in power are called event-related desynchronization (ERD), and the increase in power is called event-related synchronization (ERS) [59].

The most common method to select features from an EEG signal is the common spatial patterns (CSP) technique [5, 36, 41, 43]. This method creates spatial filters which will maximize the variance of one condition and minimize the variance of the other condition [5]. CSP technique divides the training data set into two sets X_1 and X_2, where X_i is the recorded signal for condition i. The variance is calculated $\Sigma_i = X_i^T X_i$. The spacial filters w are calculated by finding the extremes in (4.3) [42].

$$J(w) = \frac{w^T \Sigma_1 w}{w^T \Sigma_2 w} \tag{4.3}$$

CSP filters work well in differentiating mental states that evoke ERD/ERS effects, because the variance of band-passed filtered signals is equal to the power of the respective frequency band [36]. However, CSP is sensitive to noise and prone to overfitting [43]. An alternative method of feature selection to CSP is discriminate spatial patterns (DSP) [41]. This method has the advantage that it can work for more than two conditions. DSP creates spatial filters that will maximize the projected class means and minimize the classes variance [41].

A large variety of classification methods have been used on EEG signal to detect brain states. Linear methods, such as SVM and LDA, are often used in binary classification problems [41, 63]. However, these binary classification methods can be used pairwise to detect more than two states [17, 33]. Deep learning methods, like neural networks, have been used in [17], but have not been found to be an improvement over simpler methods, when using the same task and features (see Table 4.2).

A summary of some recent research projects is shown in Table 4.2. It can be inferred that on average a setup including 10 – 20 electrodes is needed to obtain a good separability between brain states. Configuration using less electrodes are also possible, however, the duration of the recorded signal may need to be increased. Features extracted from both the time and frequency domain can lead to a better classification accuracy [41]. From [33], it can be noted that the separability of motor imagery of one limb and two limbs (two hands, two feet) is greater than the separability of single-single limb motor imagery or double-double limb motor imagery. This could be because the classification method used the power of the frequency band, which increases with intensity of activity.

Table 4.2: Comparison of Previous Classification Methods using EEG. ERD: event-related desynchronization; CSP: common spatial pattern; DSP: discriminative spatial pattern; MRP: movement related potential. Setup 1: C1-C4, Cz, CP1-CP4, CPz; Setup 2:C1-C4, Cz, FC1-FC4, FCz; Setup 3: F1-F4, FC1-FC4, C1-C4; Setup 4: all electrodes in setup 1 to 3 and electrode at Fz. Setup 5: C3, C4, P4, O1, O2.

Source	**Method**	**Feature Space**	**Task**	**Average Accuracy** %
[58]	supervised K-means using Mahalanobis distance (class covariance)	changes in band power of mu frequencies in C3, Cz, and C4	motor imagery of tongue, left hand, right hand, and feet	56.32
[75]	weighted mean of correlations	ERD on each time-frequency grid point (10 channels setup 1.)	left and right motor imagery	78
[75]	weighted mean of correlations	ERD on each time-frequency grid point (10 channels setup 2)	left and right motor imagery	80
[75]	weighted mean of correlations	ERD on each time-frequency grid point (12 channels setup 3)	left and right motor imagery	79

[75]	weighted mean of correlations	ERD on each time-frequency grid point (20 channels setup 4)	left and right motor imagery	80
[63]	LDA	ERD of the μ and β rhythms (6 electrodes)	left vs. right motor imagery	86.8
[63]	LDA	ERD of the μ and β rhythms (64 electrodes)	left vs. right motor imagery	91.7
[41]	linear SVM	DSP of MRP (64 electrodes)	left and right finger tap	79.93
[41]	linear SVM	CSP of ERD (64 electrodes)	left and right finger tap	79.5
[41]	linear SVM	DSP of MRP and CSP of ERD (64 electrodes)	left and right finger tap	86.26
[17]	SVM, RBF kernel	AR õn 0.5 s segments, setup 5	rest, math, l̃etter, rotate ĩmagery, count	52.3
[17]	Neural Networks	AR õn 0.5 s segments, setup 5	rest, math, l̃etter, rotate ĩmagery, count	52.8
[17]	LDA	AR õn 0.5 s segments, setup 5	rest, math, l̃etter, rotate ĩmagery, count	44.8
[17]	SVM, RBF kernel	AR õn 20 s segments, setup 5	rest, math, l̃etter, rotate ĩmagery, count	72

[17]	Neural Networks	AR õn 20 s segments, setup 5	rest, math, l̃etter, rotate ĩmagery, count	69.4
[17]	LDA	AR õn 20 s segments, setup 5	rest, math, l̃etter, rotate ĩmagery, count	66
[33]	LDA	log-variance of spatially filtered signal, 64 channel	left\right hand imagery	61.24
[33]	LDA	log-variance of spatially filtered signal, 64 channel	left\both hands imagery	77.19
[33]	LDA	log-variance of spatially filtered signal, 64 channel	right\both hands imagery	77.60
[33]	LDA	log-variance of spatially filtered signal, 64 channel	left hand\both feet imagery	80.93
[33]	LDA	log-variance of spatially filtered signal, 64 channel	right hand\both feet imagery	81.96
[33]	LDA	log-variance of spatially filtered signal, 64 channel	both hands\both feet imagery	67.01
[33]	LDA	log-variance of spatially filtered signal, 64 channel	right\left hand\both hands imagery	57.38

[33]	LDA	log-variance of spatially filtered signal, 64 channel	right\left hand\both feet imagery	61.05
[33]	LDA	log-variance of spatially filtered signal, 64 channel	right\left hand\both hands\both feet imagery	51.18
[33]	LDA	log-variance of spatially filtered signal, 64 channel	right\left hand\both hands imagery\rest	58.30
[33]	LDA	log-variance of spatially filtered signal, 64 channel	right\left hand\both feet imagery\rest	58.11

4.5 Conclusion

The problem of decoding brain states for the use of BCI has been researched in many studies. Performance can be improved by making the problem easier by exploiting the strengths of the recording system. NIRS has good spatial resolution, which makes it suitable in decoding brain states that are characterized by activations in different brain areas, such as right vs left hand movement. EEG has great temporal resolution and can detect high frequencies that are indicative of more strenuous tasks. Hence, EEG works well in detecting brain states that are characterized by different intensities.

Linear classifiers are the most common methods used in detecting brain states and have been shown to work well in most classification problems. Brain imaging produces many features with relatively few samples, therefore feature selection is a crucial part of classification.

Chapter 5

Decoding Brain States using Near Infrared Spectroscopy

5.1 Introduction

Near Infrared spectroscopy is a noninvasive brain imaging method that can be used to decode brain states. NIRS is an emerging technology for safe optical imaging, used to detect changes in concentration of oxygenated hemoglobin and deoxygenated hemoglobin [78, 73]. It has certain benefits when compared to other imaging methods such as EEG or MRI, because it is less prone to artifacts due to movement, it has a short preparation time, can be portable and relatively cheap [70]. The technology has a few limitations, including low temporal resolution compared to EEG and a drift due to motion of the optodes.

In this study, the decoding of brain states during motor and motor imagery actions was attempted. Data was recorded using NIRS and preprocessed to remove noise, before it was used in the classification. All processing algorithms used, can be implemented in a real-time setting as would be preferred for a brain computer interface. Time windows of different sizes were chosen as features for classification and performance was compared. To improve processing time even further feature reduction algorithms were applied. Two algorithms, Fisher discriminant ration and combined Fisher discriminant ration and K-means, were used and compared.

5.2 Instrumentation

ETG - 4000 Hitachi system (Figure 5.1) is an optical tomography system approved by the Food and Drug Administration(FDA) for medical uses. Probe configurations of different sizes can be combined to measure the hemodynamic activity on any area of

the scalp. The laser diodes emit light at two wavelength, 695 nm and 830 nm, with a data sampling rate of 10 Hz.

Figure 5.1: Hitachi ETG - 4000

Two 4x4 probe grids were placed on the left and right hemisphere above the motor cortex as presented in Figure 5.2. The 4x4 probe holder alternates between source and detector on each row and column in order to maximize the number of channels. The distance between each source and detector is 3 cm. In order to ensure that the sources and detectors do not move during the experiment, the holder is attached to a cap. The optodes are spring-loaded, which ensures their contact with the scalp, even if the cap is loosely fitted. Due to the nature of the optodes, there is no additional preparation needed other than placing the cap on the head to start the recording.

5.3 Paradigm

Two tasks, which emulate motor activity and imagery motor activity, were created using a block design. The motor activity task consisted of blocks of finger tapping experiments (represented by a mouse click using the index finger), following a stimulus. For the imagery motor task the subject was instructed to only think about the action

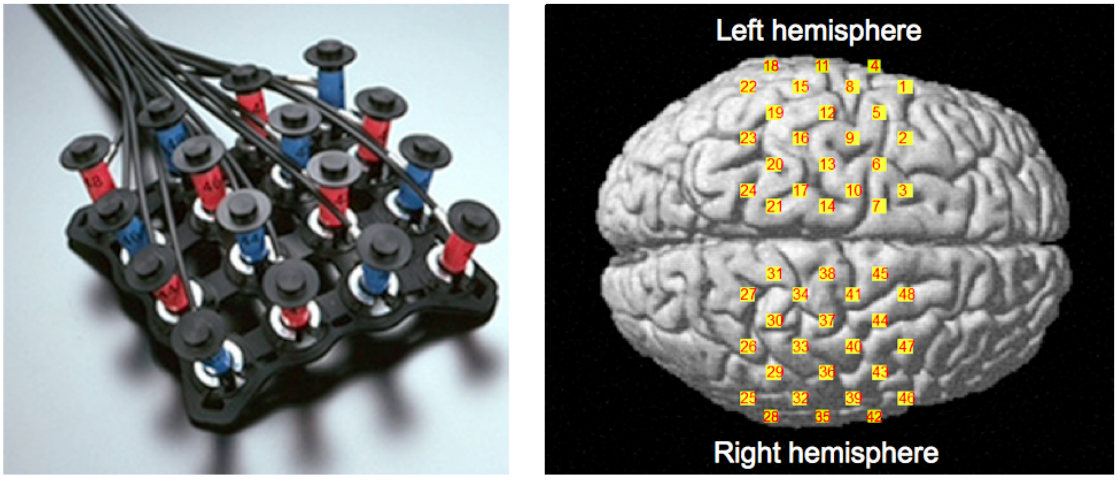

Figure 5.2: Probe configuration on head. (left) 4x4 probe holder. (right) placement on head showing channels.

of clicking.

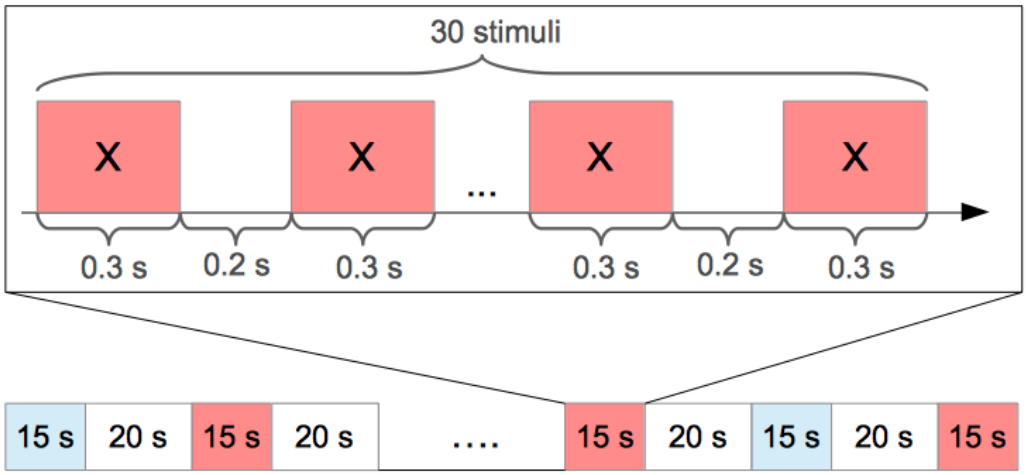

Figure 5.3: Visualization of paradigm. Sequence consisted of 6 trials of each condition: finger tapping and imagery finger tapping. The light blue block is the finger tapping condition, the rose block is the imagery tapping condition, and the white block is the rest period.

The experiment included twelve blocks alternating between actual finger tapping and imagery finger tapping, as presented in Figure 5.3. Each block consisted of 30 trials triggered by a stimulus presented on a screen. Between blocks there was a 20 second rest period. The task was designed using E-Prime. Instructions were presented at the beginning of each block, indicating which type of task, actual or imagery, was

following. Within the block, each stimulus was presented on the screen for 0.3 s followed by an inter-stimulus interval of $0.2s$. Having 30 trials in a block resulted in a block duration of 15 seconds.

The task consisted of six blocks for each of the actual and imagery experiments, therefore, including all six subjects, a total of 72 blocks of data were collected, half of which correspond to actual finger taping experiment.

5.4 Data Collection

Six right-handed subjects with no known history of mental disorders or injuries were recruited for this study. All subjects signed a consent form approved by the Rutgers Institutional Review Board. Using an ETG-4000 Hitachi system, NIRS data was recorded while the subjects performed either actual finger taping or imagery finger tapping experiments.

The sampling rate was 10 Hz. Data was collected from 24 channels on the left hemisphere of the motor cortex, because it is known that more activation will be observed in this hemisphere during right hand finger tapping [70].

5.5 Preprocessing

The preparation of the NIRS data for classification was done in three steps: noise removal, conversion to oxygenated hemoglobin and deoxygenated hemoglobin, and segmentation. The types of artifacts that were targeted during the noise removal step are the baseline drift and heart beat oscillations.

Baseline Correction

NIRS data often has a drift caused by mechanical factors such as probe deflection [52]. To correct the baseline drift, a real-time filtering method was applied to the signal, as seen in (5.1), which aims to remove low frequencies from the signal.

$$y[n] = x[n] - \frac{1}{T} \sum_{k=n}^{n-T+1} x[k] \tag{5.1}$$

In (5.1), a moving average is subtracted from each point. y is the baseline corrected signal, x is the raw signal, and T is the number of time points chosen for the moving average. It is important to choose an T in order to not remove changes in the signal caused hemodynamic concentration change. T needs to be larger or equal to the expected number of time points for activation. In this study T was chosen to be 150, which corresponds to 15 seconds and is equal to the duration of the block.

Removal of Noise caused by Heart Beat

Noise caused by heart beat is a very predictable type of artifact. It is generally an oscillation with a frequency of about 1 Hz. A moving average filter was applied to reduce the noise effects from heart beat. In (5.2), y is the filtered signal and M is the window size. The averaging window duration is $1s$, which corresponds to a window size of 10 points.

$$y[n] = \frac{1}{M} \sum_{k=n}^{n-M+1} x[k] \tag{5.2}$$

While the preprocessing was applied offline, after the signal was recorded, it does not depend on future input and could be implemented in a real-time setting. In future experiments the methods could be used during the recording.

5.6 Data Analysis

The preprocessed data was divided into time segments of equal length used for classification of the two brain states. A baseline, with a duration of 3 s immediately preceding the segment, was set and removed. Baseline removal was done by averaging over the 3 s baseline and subtracting that value from each point in the segment. Time intervals of 0.2 s, 0.4 s, 0.5 s, 1 s, and 2 s were used for segmentation.

The features chosen for classification were the change in concentration of both oxygenated and deoxygenated hemoglobin at each time point in the segment from each channel. The number of features is dependent on the length of the segment but can be calculated using (5.3) as:

$$N_f = n_h \cdot n_{\text{channels}} \cdot f_s \cdot T_{seg} = 480 \cdot T_{seg} \tag{5.3}$$

In the feature number calculation, N_f represents the number of features, n_h is the number of types of hemoglobin used, $n_{channels}$ is the number of channels, f_s is the sampling frequency, and T_{seg} is the duration of the segment. For durations of 0.2 s, 0.4 s, 0.5 s, 1 s, and 2 s the number of features is 96, 192, 240, 480, and 960 respectively.

The classification was implemented both using only the segments immediately following the stimulus and using the segments throughout the block. In the study, which uses only the segments at the beginning of the block, the number of features is much larger than the number of training vectors. Therefore, the number of features were reduced, using the fisher discriminant ratio (FDR), to improve classification accuracy and duration.

Out of the total number of segments half are used for training the classification algorithm and the rest are used for validation. The training vectors were chosen randomly ten times for cross validation. The classification algorithms were implemented using the LS-SVMlab Toolbox [11].

5.7 Results

5.7.1 Preprocessing

The preproccessing was very effective at eliminating drift and reducing the effect of heart beat. Figure 5.4 shows an example of drift removal applied on a channel. The algorithm eliminates the drift but preserves the nature of the signal.

Oscillations in the signal due to the heart beat are removed well and the signal becomes much smoother after the moving average filter is applied. Figure 5.5 shows a 50 second time span in which it can clearly be seen that the signal to noise ration is greatly improved.

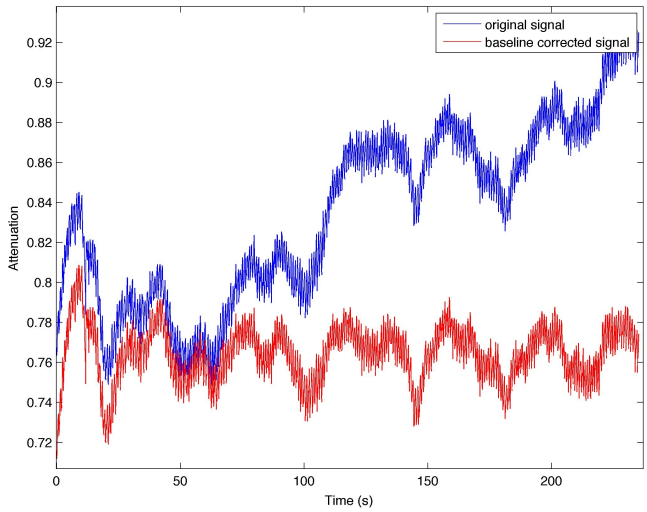

Figure 5.4: Result from baseline correction.

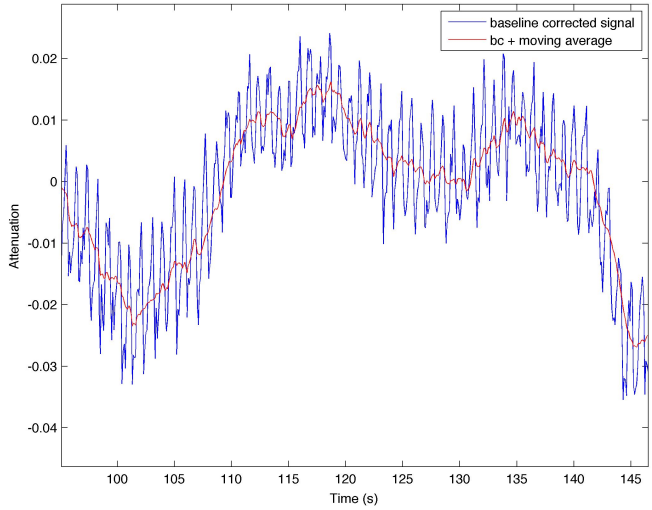

Figure 5.5: Result from reducing noise caused by heart beat.

5.7.2 Classification

It was observed that there are differences in concentrations changes of oxygenated an deoxygenated hemoglobin between real and imagery finger tapping. Time intervals of 0.2, 0.4, 0.5, 1 and 2 seconds were considered for classification both of the first tap after the rest, and of consecutive finger tapping. Results are shown in Table 5.1 and Table 5.3 for classification after the rest, and consecutive tapping, respectively. The classification results show that real finger tapping versus imagery tapping can be classified with an accuracy of 76.39 % when finger tapping is classified after a rest period. Accuracy was slightly improved and classification time was significantly improved by reducing the feature number to 36 features (see table 5.2). Reducing features even more continues to reduce the classification time with impact on the accuracy. Reducing the number of features, using the combined FDR and K-means method, did not improve detection. When the combined method is used to reduce the number of features to 20, classification accuracy drops to 50% (see Table 5.4).

Table 5.3 shows the classification results when segments are taken from the entire block. An accuracy of 63.65 % was achieved when consecutive finger tapping classification, with a segment duration of 0.5 s, is attempted. Table 5.5 compares the results in this study with the results of similar studies. While different studies use different tasks, features and classification methods, the classification accuracies in this study are similar to the classification accuracies of other studies.

Table 5.1: Classification Results of Tapping Following Rest Period

Interval (s)	**Accuracy %**	**Sensitivity %**	**Specificity %**	**Decision Time (s)**
0.2	75	63.89	86.11	0.0672
0.4	76.39	69.44	83.33	0.1513
0.5	72.23	66.67	77.78	0.1816
1	66.66	58.33	75	0.3431
2	63.89	58.33	69.44	0.5349

Table 5.2: Classification Accuracy and Duration, Using a 0.4s Time Segment, Based on Feature Number.

Number of Features	**Accuracy** %	**Detection Time (s)**
192	76.39	0.1513
100	74.17	0.0773
60	75	0.0497
50	74.86	0.0396
36	77.78	0.0316
20	75	0.0180
18	75.28	0.0177

Table 5.3: Classification Results of Consecutive Tapping

Time Interval (s)	**Accuracy** %	**Sensitivity** %	**Specificity** %
0.2	68.2	46.07	90.33
0.4	54.41	56.61	52.2
0.5	63.65	54.89	72.41
1	60.83	55.27	66.38
2	55.11	57.34	52.88

Table 5.4: Comparison of accuracies using two different feature reduction methods

Number of Features	**Accuracy using FDR only** %	**Accuracy using combined** %
36	77.78	77.03
20	75	50

Table 5.5: Comparison of Results with Previous Classification Methods using NIRS.

Source	**Method**	**Feature Space**	**Task**	**Average Accuracy** %
this study	SVM with RBF kernel	0.4 *s*immediately following rest (FDR to reduce features to 36)	motor and motor imagery	77.78
this study	SVM with RBF kernel	0.4 *s*immediately following rest	motor and motor imagery	76.39
[70]	linear SVM	10 *s* segment of 20 channels	left vs. right finger tap	87.5
[70]	linear SVM	10 *s* segment of 20 channels	left vs. right motor imagery	73.1
[49]	LDA	5 *s* averaged segment	right and left wrist motor imagery	77.56
[49]	LDA	5 *s* linear regression fitted line	right and left wrist motor imagery	87.28
[49]	LDA	10 *s* averaged segment	right and left wrist motor imagery	73.35
[49]	LDA	10 *s* linear regression fitted line	right and left wrist motor imagery	83.0

5.8 Conclusion

The results show that the optimal segment for classification of initial finger tapping is 0.4 seconds. This time interval corresponds to only one finger tap. Similarly, the optimal segment duration for classification of consecutive finger tapping does not exceed the time frame of one finger tap. This indicates that one motor action is sufficient to be create changes in hemodynamic response that can be detected. However, consecutive motor actions are more difficult to detect.

Reducing the number of features using FDR improved the classification time and did not have a significant effect on the accuracy. The effective detection time, which is the recording time plus the classification time, was reduced from 0.5513 s to 0.4326 s when 36 features are considered. For this set of features, the detection accuracy was improved which makes FDR a great feature reduction method. The combined feature reduction method led to lower classification accuracies. This method assumes that there are strong similarities between features which lead to a redundancy, which did not appear to be the case for this data set.

Chapter 6

Decoding Brain States using Electroencephalography

6.1 Introduction

Decoding brain states is an important part of the design of a brain computer interface (BCI). Electroencephalography (EEG) is a non-invasive brain imaging method with a great temporal resolution, which makes it promising for real-time applications. Traditionally, EEG active electrodes have a lengthy setup time which makes the technology less attractive for everyday use. However, new improvements in electrode technology can greatly reduce the setup time and make EEG suitable for many applications.

In this study, brain states were decoded using least squares - support vector machine. Data was collected using EEG during rest, motor and imagery motor and the three conditions were pairwise classified. Two methods for feature reduction, Fisher discriminant ration (FDR) and a combined method using FDR and K-means, were compared. The latter method had a better performance due to similarities in signals between EEG channels.

6.2 Instrumentation

The BrainAmp standard amplifier is used for EEG recordings. It is small and portable and has a sampling rate of 5 kHz. The 16 bit TTL trigger input enables the use of a large number of markers used to label important events during the EEG recording. Each amplifier has the capacity to record from 32 channels, furthermore, up to 8 amplifiers can be stacked together in order to record from up to 256 channels. Both active and passive electrodes can be used with this amplifier [20].

For this study the ActiCap from brain products was used. The ActiCAP uses

Figure 6.1: BrainAmp standard amplifier [20].

active electrodes, which have an integrated noise subtraction circuit [19]. Gel is used to minimize the impedance between electrode and scalp, which is displayed on the electrode using an LED (shown in Figure 6.2). The impedance range for each color is set before the cap is placed on the head. A total of 66 electrodes were plugged into the cap, according to Figure 6.3, 64 were used as channels and the remainder were used for the ground and reference point. The electrodes were evenly distributed over the scalps surface, including the locations present in the 10-20 system.

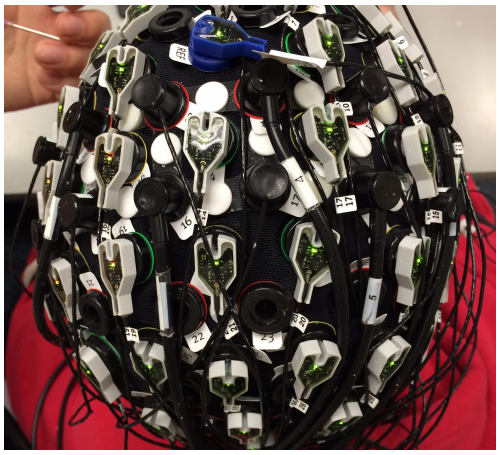

Figure 6.2: ActiCAP on subject showing impedances.Red: high impedance, yellow: medium impedence, green: low impedance

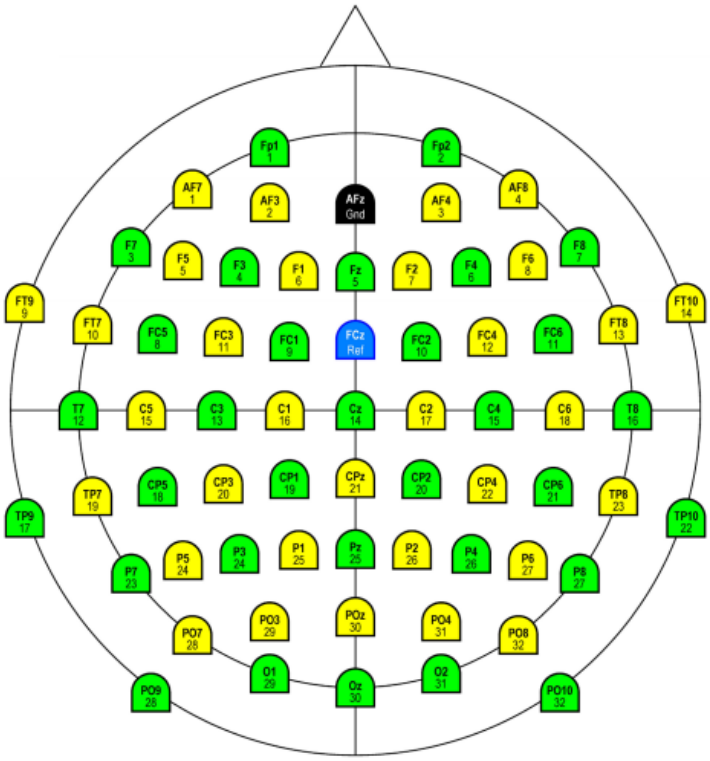

Figure 6.3: ActiCAP of electrodes on scalp [18].

6.3 Paradigm

Four conditions were implemented in this study: rest, motor only, imagery only, and mixed motor and imagery. For the rest condition subjects were seated in front of a gray screen, and instructed to relax and try not to move or talk. The rest condition was designed as a 6 to 8 minutes block. During the motor only condition the subject was instructed to click the left mouse button when a stimulus was presented on the screen. For the imagery only paradigm subjects were asked to think about tapping their finger when the stimulus was presented. Two types of stimuli were displayed during the mixed motor and imagery task. Depending on the type of stimulus, the subject had to either perform the motor action or think about tapping his or her index finger. The stimuli are randomly selected in order to reduce expectation from the subject.

The non-rest paradigms were divided into short inter-trial interval (ITI) and long ITI. The short inter-trial task displays the stimulus for 0.3 s and has an inter-trial

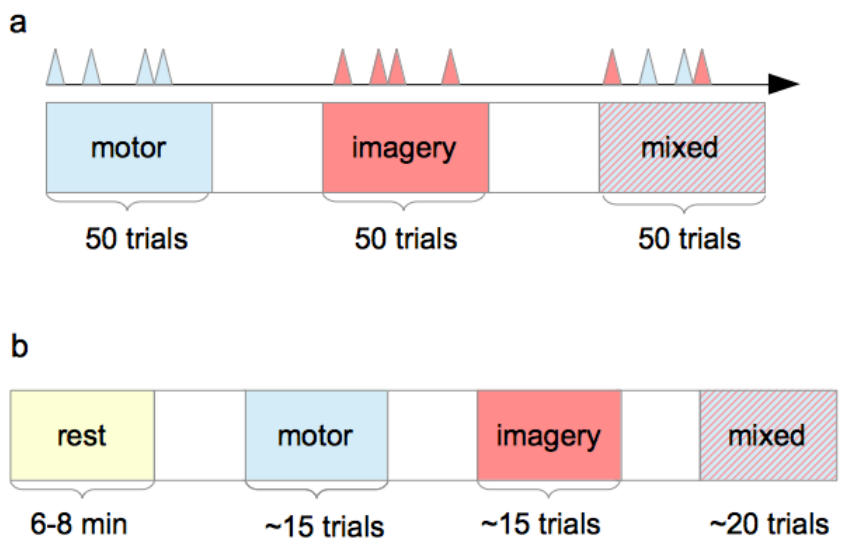

Figure 6.4: Paradigm: (a) short ITI experiment, (b) long ITI experiment.

interval of 0.5 to 1.2s. The long ITI task has a stimulus duration of 1s and an inter-trial interval of 13 to 15 s. Subjects either participated in the rest and long ITI conditions or short ITI conditions. Short ITI conditions each had 50 trials. The long ITI conditions varied in number of trials between 15 and 20.

6.4 Data Collection

Data was collected from 8 healthy volunteers using a BrainProducts BrainAmp standard 64 channel amplifier. Six subjects participated in the long ITI experiment and two participated in the short ITI experiment. All subjects signed a consent form approved by the Rutgers Institutional Review Board.

A total of 295 segments were recorded from all subjects for the imagery motor task, including the data recorded from the imagery paradigm and the mixed paradigm. For the motor task, 299 segments were recorded and for the rest condition 90 segments, with a duration of 0.4s, were obtained from the beginning of the rest period.

6.5 Data Analysis

After the EEG data was collected a common average reference (CAR) filter was applied. This has been shown to act as a high-pass spatial filter [46] accentuating components that are present in a large number of the electrodes. For further noise reduction ICA was applied.

There were a total of 295 vectors for the imagery condition, 299 vectors for the finger tapping condition and 90 vectors for the rest condition. Half of these vectors were used for training and the other half for testing. The vectors were chosen randomly; this was repeated 50 times for cross validation. The features were taken as the average value across the frequency band of $10Hz - 14Hz$ for each channel.

Classification between finger tap and rest condition was done using LS-SVM with a linear kernel, polynomial kernel and radial basis function (RBF) kernel. The number of features were reduced, using Fisher discriminant ratio (FDR), to create data sets of 56, 48, 32 and 16 features. Classification using a linear kernel was performed on each data set. The reduction of features using FDR was repeated to perform classification between motor imagery and rest condition.

A new method for feature reduction was designed by combining FDR with K-means. FDR was calculated for each feature in order to determine separability and sixteen features with the lowest score were removed. The remaining 48 features were divided into 32 clusters. Out of each cluster the feature with the highest FDR score was selected for classification. The same was repeated to select 16 features. This feature reduction method was applied to the classification of motor and rest condition as well as imagery and rest condition.

6.6 Results

6.6.1 Detecting Brain States of Motor and Rest Condition

The average classification accuracy using all 64 features was 100% for all kernels. Figure 6.5 shows a map of the training data using the two most separable features

as measured by FDR. It can be seen that the two features have a great separability. Classification using a linear kernel was repeated with a reduced number of features. The features with the highest FDR were selected. Figure 6.6 show how accuracy changes dependent on the number of features used for classification. As features are reduced accuracy decreases nonlinearly. A map of electrode location and their corresponding feature separability is shown in Figure 6.7. Channels with the highest separability could indicate activated brain areas during the motor task. Features corresponding to adjacent channels might not provide additional information about the brain state.

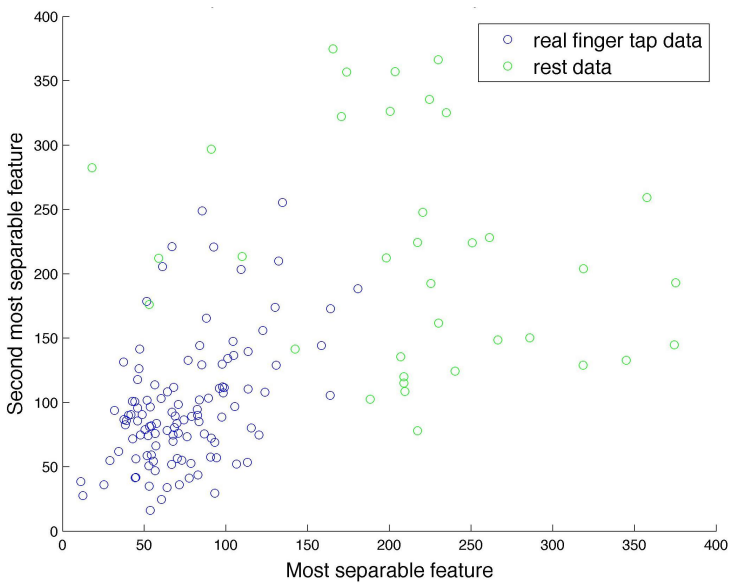

Figure 6.5: 2D Representation of the 2 Most Separable Features.

The combined feature reduction method was used in order to maximize separability and minimize redundancy between features. Table 6.1 compares the classification using the two feature reduction methods. The combined method allows for a lower number of features with little loss in accuracy.

Table 6.1: Comparison of accuracies using two different feature reduction methods

Number of Features	**Accuracy using FDR only** %	**Accuracy using combined** %
32	99.03	99.70
16	97.90	99.05

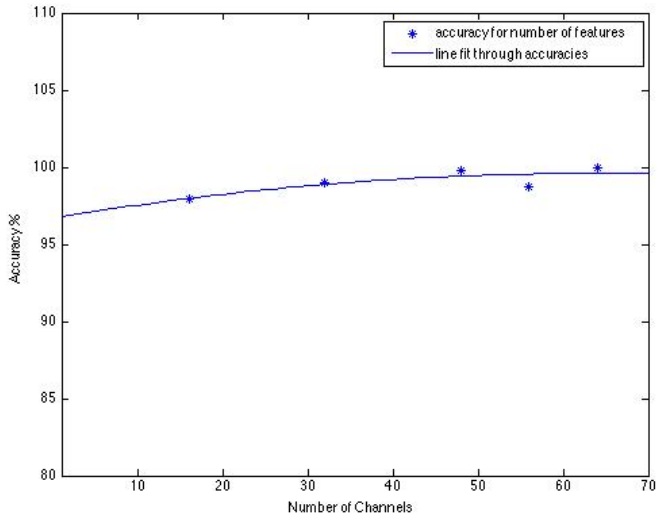

Figure 6.6: Accuracy of Linear Classification as a Function of Features for Motor and Rest.

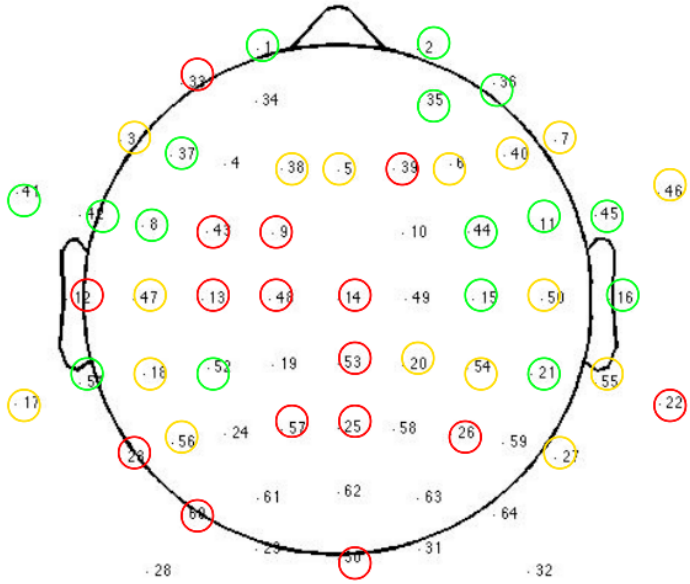

Figure 6.7: Map of Electrode Location and Feature Separability. The green features have the highest FDR, followed by yellow and red. The white features have the lowest FDR. There are 16 features of each color.

6.6.2 Detecting Brain States of Imagery and Rest Condition

In the classification of imagery finger tapping and rest condition a linear kernel was used and features were initially reduced just using FDR. It can be seen in Figure 6.8 that using FDR alone can produce good results. However, accuracy can be improved over the FDR method when the combined FDR and K-means method is used (see Table 6.2).

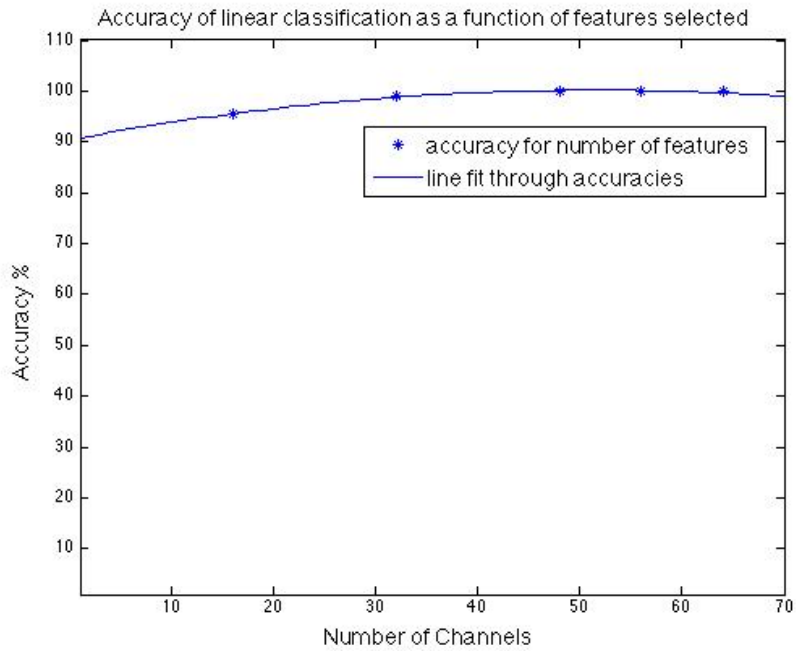

Figure 6.8: Accuracy of Linear Classification as a Function of Features for Motor Imagery and Rest.

Table 6.2: Comparison of accuracies using two different feature reduction methods

Number of Features	**Accuracy using FDR only** %	**Accuracy using combined** %
32	99.16	99.6
16	95.39	97.65

The results from this study were compared to literature (Table 6.3). This study achieves better classification accuracies than similar studies.

Table 6.3: Comparison of Results with Previous Classification Methods using EEG.

Source	**Method**	**Feature Space**	**Task**	**Average Accuracy** %
this study	linear SVM	average power of bandwidth (10 Hz - 14 Hz)(combined method to reduce features to 32)	motor and rest	99.70
this study	linear SVM	average power of bandwidth (10 Hz - 14 Hz)(combined method to reduce features to 16)	motor and rest	99.05
this study	linear SVM	average power of bandwidth (10 Hz - 14 Hz)(combined method to reduce features to 32)	motor imagery and rest	99.6
this study	linear SVM	average power of bandwidth (10 Hz - 14 Hz)(combined method to reduce features to 16)	motor imagery and rest	97.65
[75]	weighted mean of correlations	ERD on each time-frequency grid point (12 channels)	left and right motor imagery	79
[75]	weighted mean of correlations	ERD on each time-frequency grid point (20 channels)	left and right motor imagery	80
[63]	LDA	ERD of the μ and β rhythms (6 electrodes)	left vs. right motor imagery	86.8
[63]	LDA	ERD of the μ and β rhythms (64 electrodes)	left vs. right motor imagery	91.7
[41]	linear SVM	DSP of MRP (64 electrodes)	left and right finger tap	79.93
[41]	linear SVM	CSP of ERD (64 electrodes)	left and right finger tap	79.5
[41]	linear SVM	DSP of MRP and CSP of ERD (64 electrodes)	left and right finger tap	86.26

6.7 Conclusion

Motor and imagery motor tasks can be differentiated very well from rest using EEG recordings. Feature reduction helps decrease classification complexity and can significantly improve classification speed. Applied to brain imaging, it can help identify specific areas of the brain activated during an action. However, the features and the task need to be carefully selected in order to properly reflect the intended action. In the motor task for this study the subject performed multiple actions: identification of type of stimulus and decision of action, which can be seen as high separability in the prefrontal cortex, and of course the tapping which should be seen in the motor cortex. The fact that a task cannot isolate a single action combined with the poor spacial resolution of EEG signal, make it harder to detect activation through feature selection.

Therefore, feature selection cannot be done simply by using previous knowledge about the area of activation. Combining FDR and K-means for feature reduction produces better classification accuracies than the FDR feature reduction method alone. FDR only detects separability between classes of a feature but does not compare features. K-means groups similar features together without regard to their separability. By combining the two algorithms both inter-class and intra-class similarities are reduced. This method for feature reduction works well on EEG data because adjacent channels can have similar signals.

Chapter 7

Conclusion

Detecting brain states is a complex problem that has many influencing factors including imaging technique, experimental task, feature extraction and selection and classification method. To optimize performance, these factors should not be chosen individually but rather the strength of each should be considered such that the choices for each factor complement each other.

NIRS has good spatial resolution and should be used for classification of tasks that have different neuronal activation areas in order to maximize performance. EEG has good temporal resolution and can detect rhythms of high frequency. By examining the frequency band of the brain rhythms the difficulty or intensity of the task can be detected. Each channel picks up a combination of the neuronal activity in the brain, therefore, channels that are close to each other have similar signals, which creates redundancy in features.

The task used in this study simulated three conditions: rest, motor and motor imagery. The motor and motor imagery conditions share some neuronal activation areas [64]. However, studies suggest that the motor conditions also has neuronal activations which are not shared with the motor imagery condition [64]. This is supported by this study since classification between motor and motor imagery using NIRS achieved an accuracy of 77.78%. The similarity in brain states of the two conditions can be seen in the EEG study. The performance of the classification of motor and rest and that of motor imagery and rest are similar with the former being slightly higher most likely due to increased brain activity. As discussed in Chapter 2 activities that are more difficult or intense will cause brain rhythms with higher frequency. The motor task is considered slightly more difficult than the motor imagery task because it requires both intent and

motor performance.

In this study the choice of features was based on the literature. For the NIRS study the features chosen were the changes in concentration of oxygenated hemoglobin and deoxygenated hemoglobin over a period of time. Several window sizes were examined. The best performance was achieved with a time segment that included only one finger tap. For the EEG study the features were extracted from the frequency domain. This study did not compare different features but focused on feature selection.

Brain imaging data often has many features but only few samples, which makes feature selection essential for classification. Two methods were compared for feature selection FDR and the combined method which uses FDR and K-means. FDR focuses on features that are highly separable but does not detect features that are very similar. To remove the redundancy added by features that are similar K-means was added to the feature selection algorithm. In the NIRS study, using FDR for feature selection produced a better classification accuracy than the combined method. This implies that there are not many similar features in the data. Separability of each individual feature is the better criteria for feature selection. In the EEG study it was observed that the combined method resulted in better classification accuracies than the FDR only method. This is due to the fact that channels that are placed in proximity of each other usually have a similar signal. Removing redundancy due to similarity, therefore, is very important when using EEG data.

7.1 Future Work

In this study, the detection of brain states, using two brain imaging techniques individually, was performed. The two techniques can be combined into one classification problem. The advantages of each technique can be used to improve performance. Additionally, the classification problem can be extended to more classes. Mental tasks could be included in addition to motor tasks. These two tasks could highly separable which is more important in a multi-class detection.

The combined feature selection algorithm accounts for both usefulness of features

and redundancy due to similarity. However, there can be improvements made to the algorithm to maximize performance. The original method had a fixed number of clusters that were formed after a number of features were already discarded. The number of features discarded was decided by the user, however, a better way would be to automate that decision using a threshold based on the highest score. Additionally, the algorithm can be improved by not having a fixed number of clusters in order to avoid features grouped together without being similar. This can be achieved by setting a maximum distance from the centroid of the clusters that is allowed in order to belong to that cluster. Finally, adding the FDR score as a feature in the clustering algorithm could ensure that features that are clustered together have similar FDR scores.

References

[1] Franck Amyot, Trelawny Zimmermann, Jason Riley, Jana M Kainerstorfer, Victor Chernomordik, Eric Mooshagian, Laleh Najafizadeh, Frank Krueger, Amir H Gandjbakhche, and Eric M Wassermann. Normative database of judgment of complexity task with functional near infrared spectroscopyapplication for tbi. *Neuroimage*, 60(2):879–883, 2012.

[2] Kai Keng Ang, Cuntai Guan, Kerry Lee, Jie Qi Lee, Shoko Nioka, and Britton Chance. A brain-computer interface for mental arithmetic task from single-trial near-infrared spectroscopy brain signals. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3764–3767. IEEE, 2010.

[3] Bernard J Baars and Nicole M Gage. *Cognition, brain, and consciousness: Introduction to cognitive neuroscience*. Academic Press, 2010.

[4] John Bibby, Helge Toutenburg, and Helge Toutenburg. *Prediction and improved estimation in linear models*. Wiley, 1977.

[5] Benjamin Blankertz, Motoaki Kawanabe, Ryota Tomioka, Friederike Hohlefeld, Klaus-robert Müller, and Vadim V Nikulin. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In *Advances in neural information processing systems*, pages 113–120, 2007.

[6] David A Boas, Anders M Dale, and Maria Angela Franceschini. Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *Neuroimage*, 23:S275–S288, 2004.

[7] Alper Bozkurt, Arye Rosen, Harel Rosen, and Banu Onaral. A portable near infrared spectroscopy system for bedside monitoring of newborn brain. *Biomedical engineering online*, 4(1):29, 2005.

[8] Richard B Buxton, Kâmil Uludağ, David J Dubowitz, and Thomas T Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23:S220–S233, 2004.

[9] Shirley M Coyle, Tomás E Ward, and Charles M Markham. Brain–computer interface using a simplified functional near-infrared spectroscopy system. *Journal of neural engineering*, 4(3):219, 2007.

[10] Xu Cui, Signe Bray, and Allan L Reiss. Speeded near infrared spectroscopy (nirs) response detection. *PLoS One*, 5(11):e15474, 2010.

[11] K De Brabanter, P Karsmakers, F Ojeda, C Alzate, J De Brabanter, K Pelckmans, B De Moor, J Vandewalle, and JAK Suykens. Ls-svmlab toolbox users guide. *ESAT-SISTA Technical Report*, pages 10–146, 2011.

[12] Stefan Debener, Falk Minow, Reiner Emkes, Katharina Gandras, and Maarten Vos. How about taking a low-cost, small, and wireless eeg for a walk? *Psychophysiology*, 49(11):1617–1621, 2012.

[13] Arlene Duncan, Judith H Meek, Matthew Clemence, Clare E Elwell, Lidia Tyszczuk, Mark Cope, and D Delpy. Optical pathlength measurements on adult head, calf and forearm and the head of the newborn infant using phase resolved optical spectroscopy. *Physics in medicine and biology*, 40(2):295, 1995.

[14] Monica Fabiani, Gabriele Gratton, and Michael GH Coles. Event-related brain potentials. *Handbook of psychophysiology*, 2:53–84, 2000.

[15] Jonathan R Folstein and Cyma Van Petten. Influence of cognitive control and mismatch on the n2 component of the erp: a review. *Psychophysiology*, 45(1):152–170, 2008.

[16] Terrence S Furey, Nello Cristianini, Nigel Duffy, David W Bednarski, Michel Schummer, and David Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906–914, 2000.

[17] Deon Garrett, David A Peterson, Charles W Anderson, and Michael H Thaut. Comparison of linear, nonlinear, and feature selection methods for eeg signal classification. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):141–144, 2003.

[18] Brain Products GmbH. acticap 64ch standard-2, 2012.

[19] Brain Products GmbH. Acticap, 2014.

[20] Brain Products GmbH. Brainamp standard, 2014.

[21] Brain Products GmbH. Eeg, 2014.

[22] Todd R Golub, Donna K Slonim, Pablo Tamayo, Christine Huard, Michelle Gaasenbeek, Jill P Mesirov, Hilary Coller, Mignon L Loh, James R Downing, Mark A Caligiuri, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science*, 286(5439):531–537, 1999.

[23] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.

[24] Matti S Hämäläinen and Risto J Ilmoniemi. *Interpreting measured magnetic fields of the brain: estimates of current distributions*. Helsinki University of Technology, Department of Technical Physics, 1984.

[25] Moritz Helias, Susanne Kunkel, Gen Masumoto, Jun Igarashi, Jochen Martin Eppler, Shin Ishii, Tomoki Fukai, Abigail Morrison, and Markus Diesmann. Supercomputers ready for use as discovery machines for neuroscience. *Frontiers in neuroinformatics*, 6, 2012.

[26] Uwe Herwig, Peyman Satrapi, and Carlos Schönfeldt-Lecuona. Using the international 10-20 eeg system for positioning of transcranial magnetic stimulation. *Brain topography*, 16(2):95–99, 2003.

[27] Thomas Hofmann, Bernhard Schölkopf, and Alexander J Smola. Kernel methods in machine learning. *The annals of statistics*, pages 1171–1220, 2008.

[28] Anil K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651 – 666, 2010. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR) 19th International Conference in Pattern Recognition (ICPR).

[29] A Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, 14:841, 2002.

[30] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin J Mckeown, Vicente Iragui, and Terrence J Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(02):163–178, 2000.

[31] Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*, volume 4. McGraw-Hill New York, 2000.

[32] M Jawad Khan, Melissa Jiyoun Hong, and Keum-Shik Hong. Decoding of four movement directions using hybrid nirs-eeg brain-computer interface. *Frontiers in human neuroscience*, 8, 2014.

[33] Hyun Seok Kim, Min Hye Chang, Hong Ji Lee, and Kwang Suk Park. A comparison of classification performance among the various combinations of motor imagery tasks for brain-computer interface. In *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, pages 435–438. IEEE, 2013.

[34] L Kocsis, P Herman, and A Eke. The modified beer–lambert law revisited. *Physics in medicine and biology*, 51(5):N91, 2006.

[35] Ron Kohavi and George H John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324, 1997.

[36] Zoltan Joseph Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical eeg. *Electroencephalography and clinical Neurophysiology*, 79(6):440–447, 1991.

[37] Sun Yuan Kung. *Kernel Methods and Machine Learning*. Cambridge University Press, 2014.

[38] Yongjin Kwon, Myong K Jeong, and Olufemi A Omitaomu. Adaptive support vector regression analysis of closed-loop inspection accuracy. *International Journal of Machine Tools and Manufacture*, 46(6):603–610, 2006.

[39] Thomas Navin Lal, Michael Schroder, Thilo Hinterberger, Jason Weston, Martin Bogdan, Niels Birbaumer, and Bernhard Scholkopf. Support vector channel selection in bci. *Biomedical Engineering, IEEE Transactions on*, 51(6):1003–1010, 2004.

[40] Eric C Leuthardt, Kai J Miller, Gerwin Schalk, Rajesh PN Rao, and Jeffrey G Ojemann. Electrocorticography-based brain computer interface-the seattle experience. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 14(2):194–198, 2006.

[41] Xiang Liao, Dezhong Yao, Dan Wu, and Chaoyi Li. Combining spatial filters for the classification of single-trial eeg in a finger movement task. *Biomedical Engineering, IEEE Transactions on*, 54(5):821–831, 2007.

[42] Fabien Lotte, Marco Congedo, Anatole Lécuyer, and Fabrice Lamarche. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of neural engineering*, 4, 2007.

[43] Fabien Lotte and Cuntai Guan. Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms. *Biomedical Engineering, IEEE Transactions on*, 58(2):355–362, 2011.

[44] Man-Wai Mak and Sun-Yuan Kung. A solution to the curse of dimensionality problem in pairwise scoring techniques. In *Neural Information Processing*, pages 314–323. Springer, 2006.

[45] Jaakko Malmivuo and Robert Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, 1995.

[46] Dennis J McFarland, Lynn M McCane, Stephen V David, and Jonathan R Wolpaw. Spatial filter selection for eeg-based communication. *Electroencephalography and clinical Neurophysiology*, 103(3):386–394, 1997.

[47] Judith Meek. Basic principles of optical imaging and application to the study of infant development. *Developmental Science*, 5(3):371–380, 2002.

[48] Thomas Muehlemann, Daniel Haensse, and Martin Wolf. Wireless miniaturized in-vivo near infrared imaging. *Optics express*, 16(14):10323–10330, 2008.

[49] Noman Naseer and Keum-Shik Hong. Classification of functional near-infrared spectroscopy signals corresponding to the right-and left-wrist motor imagery for development of a brain–computer interface. *Neuroscience letters*, 553:84–89, 2013.

[50] Ernst Niedermeyer. 9. the normal eeg of the waking adult. *Electroencephalography: Basic principles, clinical applications, and related fields*, page 167, 2005.

[51] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[52] Takuya Ozawa, Takatsugu Aihara, Yusuke Fujiwara, Yohei Otaka, Isao Nambu, Rieko Osu, Jun Izawa, and Yasuhiro Wada. Detecting event-related motor activity using functional near-infrared spectroscopy. In *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, pages 1529–1532. IEEE, 2013.

[53] Roberto D Pascual-Marqui, Christoph M Michel, and Dietrich Lehmann. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of psychophysiology*, 18(1):49–65, 1994.

[54] Roberto Domingo Pascual-Marqui. Review of methods for solving the eeg inverse problem. *International journal of bioelectromagnetism*, 1(1):75–86, 1999.

[55] Paul Pavlidis, Jason Weston, Jinsong Cai, and William Noble Grundy. Gene functional classification from heterogeneous data. In *Proceedings of the fifth annual international conference on Computational biology*, pages 249–255. ACM, 2001.

[56] Maria Peifer, Li Zhu, and Laleh Najafizadeh. Real-time classification of actual vs imagery finger tapping using functional near-infrared spectroscopy. In *Biomedical Optics*, pages BM3A–34. Optical Society of America, 2014.

[57] Veronica B Perez, Edward K Vogel, SJ Luck, and ES Kappenman. *Oxford handbook of event-related potential components*. Oxford university press, 2012.

[58] G Pfurtscheller, C Brunner, A Schlögl, and FH Lopes Da Silva. Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks. *Neuroimage*, 31(1):153–159, 2006.

[59] Gert Pfurtscheller and Fernando H Lopes da Silva. Event-related eeg/meg synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11):1842–1857, 1999.

[60] Gert Pfurtscheller, Christa Neuper, Alois Schlogl, and Klaus Lugger. Separability of eeg signals recorded during right and left motor imagery using adaptive autoregressive parameters. *Rehabilitation Engineering, IEEE Transactions on*, 6(3):316–325, 1998.

[61] Christophe Phillips, Michael D Rugg, and Karl J Friston. Anatomically informed basis functions for eeg source localization: combining functional and anatomical constraints. *NeuroImage*, 16(3):678–695, 2002.

[62] John Polich. *Oxford handbook of event-related potential components*. Oxford university press, 2012.

[63] Florin Popescu, Siamac Fazli, Yakob Badower, Benjamin Blankertz, and Klaus-R Müller. Single trial classification of motor imagination using 6 dry eeg electrodes. *PloS one*, 2(7):e637, 2007.

[64] Carlo A Porro, Maria Pia Francescato, Valentina Cettolo, Mathew E Diamond, Patrizia Baraldi, Chiara Zuiani, Massimo Bazzocchi, and Pietro E Di Prampero. Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study. *The Journal of neuroscience*, 16(23):7688–7698, 1996.

[65] Hillel Pratt. *Oxford handbook of event-related potential components*. Oxford university press, 2012.

[66] Walter S Pritchard, Scott A Shappell, and Michael E Brandt. Psychophysiology of n200/n400: A review and classification scheme. *Advances in psychophysiology*, 4:43–106, 1991.

[67] Saeid Sanei and Jonathon A Chambers. *EEG signal processing*. John Wiley & Sons, 2013.

[68] Felix Scholkmann, Stefan Kleiser, Andreas Jaakko Metz, Raphael Zimmermann, Juan Mata Pavia, Ursula Wolf, and Martin Wolf. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage*, 85:6–27, 2014.

[69] John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[70] Ranganatha Sitaram, Haihong Zhang, Cuntai Guan, Manoj Thulasidas, Yoko Hoshi, Akihiro Ishikawa, Koji Shimizu, and Niels Birbaumer. Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain–computer interface. *NeuroImage*, 34(4):1416–1427, 2007.

[71] Fren TY Smulders and JO Miller. *Oxford handbook of event-related potential components*. Oxford university press, 2012.

[72] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[73] Arno Villringer and Britton Chance. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in neurosciences*, 20(10):435–442, 1997.

[74] Edward K Vogel and Steven J Luck. The visual n1 component as an index of a discrimination process. *Psychophysiology*, 37(02):190–203, 2000.

[75] Tao Wang, Jie Deng, and Bin He. Classifying eeg-based motor imagery tasks by means of time–frequency synthesized spatial patterns. *Clinical Neurophysiology*, 115(12):2744–2753, 2004.

[76] Jonathan R Wolpaw, Niels Birbaumer, William J Heetderks, Dennis J McFarland, P Hunter Peckham, Gerwin Schalk, Emanuel Donchin, Louis A Quatrano, Charles J Robinson, Theresa M Vaughan, et al. Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering*, 8(2):164–173, 2000.

[77] Geoffrey F Woodman. A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, 72(8):2031–2046, 2010.

[78] AG Yodh and DA Boas. Functional imaging with diffusing light. *Biomedical photonics handbook*, 21(1):21–45, 2003.

[79] Han Zhang, Yu-Jin Zhang, Chun-Ming Lu, Shuang-Ye Ma, Yu-Feng Zang, and Chao-Zhe Zhu. Functional connectivity as revealed by independent component analysis of resting-state fnirs measurements. *Neuroimage*, 51(3):1150–1161, 2010.

[80] Li Zhu, Maria Peifer, and Laleh Najafizadeh. Towards improving the detection. In *Biomedical Optics*, pages BM3A–29. Optical Society of America, 2014.