Analyzing Brain Signals for Control

Henrik Eijsink

Master Thesis
Informatics: Robotics and Intelligent Systems
60 credits

Department of Informatics
The Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

May 2021

Analyzing Brain Signals for Control

Henrik Eijsink

Analyzing Brain Signals for Control

Henrik Eijsink

http://www.duo.no/

Print: Reprosentralen, University of Oslo

Abstract

The idea of direct communication between the brain and a computer emerged over 40 years ago and has been developing since. Despite progress, the potential of this idea to influence research in several fields and solve concrete problems is not yet fulfilled. In healthcare, brain computer interfaces (BCIs) may be used as a communication device that allows patients with a clear mind but lack of body control to control for example a robot hand, or a communication device. In principle, BCIs could help to restore neural function for patients who have suffered disease or trauma. One particular challenge is that some subjects seem unable to achieve high performance while trying to control a system by imagining a motoric movement, a phenomenon that is referred to as BCI illiteracy in the literature. The study described in this thesis was aimed at developing classifiers of brain signals (electroencephalograms). More specifically, the goal was to assess whether the use of brain signals from subjects performing actual motoric movements (motor execution) as training data affects the performance of a model when classifying brain signals produced during imagined movement (motor imagery), particularly for illiterate subjects.

A dataset was collected by having subjects perform equal amounts of actual movements and imaginary movements. Several classifiers were explored, and the random forest classifier, which showed the best results, was employed together with periodogram-based features. The classifier trained with motor imagery (MI) data, motor execution data (ME), and a combination of the two. The results of varying the training data indicated that for this particular classifier, brain signals from ME could be included in the training data to improve the

classification of MI-related brain signals, and there were weak indications that this also would apply to BCI illiterate subjects.

The fact that similar results could be obtained upon replacing MI data with ME data when developing an MI classifier, is encouraging. Generally, this could reduce the dependency on MI data, which are hard to collect reliably, in BCI development. Furthermore, the possibility of using ME data to train MI-based classifiers opens up the possibility to use ME data generated by a motorically competent person to be used in developing BCIs for disabled persons and BCI-illiterate persons. Of note, the present study was conducted on a limited number of subjects and several alternative approaches have remained largely unexplored, more work is needed to improve the statistical significance and explore alternative strategies.

Acknowledgments

I am very thankful to my supervisor Jim Torresen for his guidance throughout the last year. Jim provided valuable input and help during the research work and the writing of this thesis, for which I am very grateful. It has been an inspiring journey.

I also want to thank Tor Endestad from the Department of Psychology at the University of Oslo, for providing access to equipment and laboratory space that made the data collection possible, as well as for his encouraging input and expertise regarding the neurological aspects of this work.

The data collection would not have been possible if it were not for friends who volunteered to be subjects in response to my outreach on social media. GDPR rules prevent me from mentioning your names, but you know who you are, and I thank you all.

I would also like to thank my fellow master students Katrine Nergård, Knut Joar Strømmen, and Jakuba Kozlowski for the many long hours in the study hall, and countless interesting discussions, both professional and otherwise.

A special thanks goes to my parents, May Bente Brurberg and Vincent Eijsink, my sister, Malin Eijsink, and my friends (none mentioned, none forgotten), for continuous support and encouragement along the way.

Oslo, June 1, 2021

Henrik Eijsink

List of Abbreviations

BCI	Brain Computer Interface
EEG	Electroencephalogram
MI	Motor Imagery
ME	Motor Execution
SSVEP	Steady State Visually Evoked Potential
SCP	Slow Cortical Potential
ERD	Event Related Desynchronization
ERS	Event Related Synchronization
ERP	Event Related Potential
FT	Fourier Transform
FFT	Fast Fourier Transform
ANN	Artificial Neural Network
MLP	Multilayer Perceptron
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
RF	Random Forest
RL	Reinforcement Learning
kNN	k Nearest Neighbor
SVM	Support Vector Machine
GUI	Graphic User Interface

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The use of brain signals as input to intelligent systems is a fast-growing research area, with the healthcare and gaming industries as the main market drivers. Providing an alternative communication medium between brain and environment or the ability to control external devices with brain signals may help to restore independence for disabled individuals. Systems using brain signals for control are collectively referred to as Brain Computer Interfaces (BCIs). In healthcare, BCIs may revolutionize a wide variety of tasks, from diagnostics to rehabilitation after injury or disease. The gaming industry sees big potential in combining BCIs with Virtual Reality to provide a more authentic gaming experience or even a whole new genre of games. Such promising possibilities have prompted a steady growth in research of this area in the last two decades, which is easily visualized by the number of yearly BCI-related publications in Figure 1.

The simple goal for BCIs is to be able to control a computer system by using brain signals as input and this is already possible to some extent. However, despite numerous start-ups and large tech firms working on making safer, cheaper, and more accurate BCIs, penetration of this technology into commercial markets is still slow. There are some commercialized devices to record brain activity [1], but the majority of published research and achievements are still primarily based on using research-grade equipment. One of the challenges is to make a system that is easily set up by a doctor, caregiver, or layperson while maintaining the high accuracy of the state-of-the-art systems. For some uses, it is also essential that the

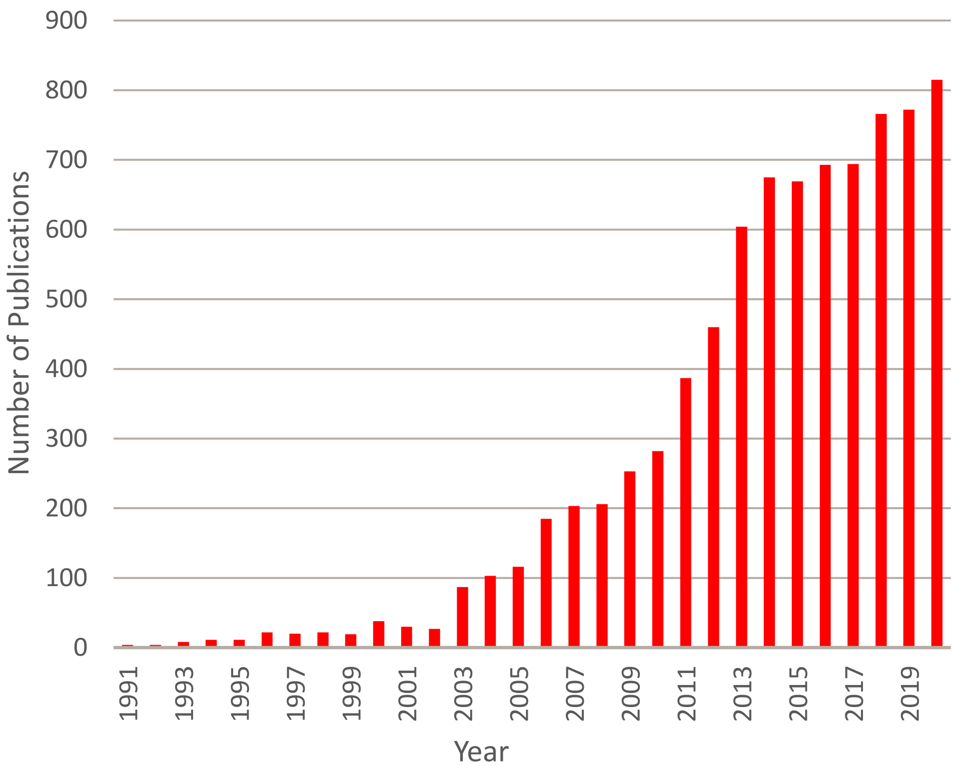


Figure 1. Number of BCI publications per year from 1991 to 2020. Obtained by the PubMed query "brain computer interface" on April 19, 2021.

device is wireless, which may add additional challenges related to latency or robustness.

A critical, and still largely untackled obstacle in BCI development for medical applications is the large variability in performance between, and even within, subjects. Some subjects are not able to achieve satisfactory performance levels even after an extensive training regimen. These subjects are referred to as BCI-illiterate and estimates, suggest that they account for up to 30 % of users [2]. A large fraction of BCI-illiterate subjects may prevent BCIs from being widely applied. While BCIs could be the next big step in the rehabilitation of patients with neurological injuries, BCI illiteracy and the need for extensive training programs, if such training is possible at all, may pose severe limitations.

For a subject to produce brain signals corresponding to physical motoric action, the subject can either carry out the action or imagine the action. These two scenarios are referred to as motor execution (ME), which is the actual performance of the movement, and motor imagery (MI), which is imagining performing the movement. MI can be seen as the planning of movement, with the overt execution being inhibited, and although these two abilities are evolved or impaired independently in the subject, both make the subject exhibit very similar brain signals from the same brain areas [3].

Healthcare-related BCI research is naturally focused on MI as most patients for which BCI could be a suitable enabling tool, have disabilities that make them unable to perform ME. For example, one could assume that an amputee has no way of performing ME and can only provide MI data to train the classifier. As an alternative, one could explore the possibility of training a classifier for use by an amputee on ME data from different subjects. It should be noted that Raffin et al. managed to distinguish between executing a movement of an amputees phantom limb and imagining moving the missing limb [4].

While the potential of BCIs is evident and while health care applications of BCIs will only increase with an aging population, there is a need to develop better BCIs that are less susceptible to failure due to BCI-illiteracy. This need was the main motivation for the work described in this thesis.

1.2 Research Goals

This project is a collaboration between the Department of Informatics and neuroscience experts at the Department of Psychology at the University of Oslo, who are working on a project to build a complete, complex BCI system. They plan on using multiple inputs, namely eye-tracking, electromyography (i.e., measuring electrical currents in skeletal muscles), in addition to brain signals obtained by functional near-infrared spectroscopy and EEGs. The first step will be to use EEGs to enable non-disabled volunteers to make a robot hand perform simple hand movements by either imagining or executing the movements with their own hands.

This work addresses the challenge of developing good BCIs and the problem of BCI illiteracy and aims at contributing to solving problems that limit BCI applications within the health care industry. Next to aiming at tackling the problem of BCI illiteracy, another goal was to provide a fresh perspective on the use of motor execution data versus motor imagery data and to identify possible approaches that could help to develop BCIs that are less susceptible to BCI illiteracy. This work was based on analyzing brain signals in the form of electroencephalogram (EEG) related to hand and foot control by human subjects.

1.3 Main Outputs

The main contributions of the work described in this thesis are as follows:

- A new dataset was collected consisting of both motor imagery and motor execution data, for developing and training BCIs based on hand gestures.
- A simple random forest model with periodogram-based features was implemented and used as a basis for the research.
- The impact of exchanging training data from motor imagery to motor execution on model performance was analyzed, with extra attention being given to possibly BCI illiterate subjects.

1.4 Thesis Structure

This thesis is structured in a way that chronologically represents the steps taken to produce this research. The chapters are structured as follows:

- Chapter 2: Background
  An introduction to the theoretical background of the research, addressing Electroencephalograms, BCIs, signal processing approaches, and classifiers.

- Chapter 3: Datasets
  A brief overview of publicly available datasets and the institutions that provide these, as well as a detailed summary of the collection process for, and the characteristics of, the dataset that was produced as part of this work.
- Chapter 4: Research Methods
  A detailed description of the process of handling a raw dataset and implementing a working classifier.
- Chapter 5: Results and Discussion
  An evaluation of the implemented models, and discussion of the performance of models trained with ME data versus purely MI-based models.
- Chapter 6: Summary and Conclusion
  A summary of the results, and conclusions with regards to the research goals presented in section 1.2.
- Chapter 7: Future Work
  A discussion of some uninvestigated aspects and possible topics for future research.

Chapter 2

Background

2.1 Brain Signals

There are several different approaches to acquire signals from brain activity which are divided into invasive and non-invasive methods. Invasive procedures are characterized by the American National Cancer Institute as “A medical procedure that invades (enters) the body, usually by cutting or puncturing the skin or by inserting instruments into the body.” [5]. In a BCI setting, invasive methods typically correspond to inserting electrodes into the brain. For example, Hochberg et al. used such methods to enable patients with long-standing paralysis to perform simple reaching and grasping actions with a BCI-controlled robot arm [6].

Invasive methods are resource-demanding and can easily cause immune reactions, infections, or other negative side effects. This has drawn the attention of the field to non-invasive methods that record brain activity from the scalp and have significantly better portability and lower cost while lowering damage to human bodies and being easier to use. One such non-invasive method, which is widely favored in BCIs because of its ease of use is electroencephalogram (EEG). When recording EEGs, cortical electrical activity is recorded through electrodes attached directly onto the scalp. As an example of the potential of this approach, Pfurtscheller et al. used EEG signals from partially paralyzed patients, together with functional electrical stimulation, which is a way to make the muscle of a patient move by electrical stimulation via electrodes, they enabled patients to reach and grasp with their own paralyzed hand [7].

2.1.1 The Origin of the Electroencephalogram (EEG)

As early as 1875, Richard Caton reported having measured electrical activity in the brains of rabbits and monkeys using a simple galvanometer, both from the scalp and directly from the cortex. Caton even suggested that the electrical currents measured were afflicted by functional activity and outside stimuli [8]. This quite groundbreaking discovery went largely unnoticed, and it took a surprisingly long time before Hans Berger took the next step and discovered similar electrical activity in the human brain in 1929. Berger was the first to coin the term Electroencephalogram (EEG) and his work touched upon different EEG phenomena, like EEG changes correlated with cognitive effort or attention [9]. The discoveries by Berger were disbelieved at first, until Adrian and Matthews reproduced the results in 1934 [10]. This marked the emergence of the EEG research field, and many important breakthroughs were made in the following years.

2.1.2 Brain Waves

Berger already coined the terms “alpha” and “beta” for denoting frequency ranges of brain waves, that is, much in the same way as

Frequancy Band	Frequency
Delta (δ)	< 3 Hz
Theta (ϑ)	3 - 6 Hz
Alpha (α)	6 - 14 Hz
Beta (β)	14 - 30 Hz
Gamma (γ)	> 30 Hz

Table 1. The frequency bands of brain waves

they are defined today [11]. Today, there are five well-established frequency bands, referred to as delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ), as shown in Table 1.

Alpha frequencies in the visual cortex are the strongest known EEG signals, and they are associated with visual processing. Markand has given an extensive summary of the numerous discoveries regarding alpha rhythms and their correlation to cognitive function, primarily visual processing and memory function [12]. A very interesting characteristic of alpha frequencies in the visual cortex is that they tend to be attenuated or blocked when the subject is presented with visual stimuli [13]. In the motor cortex (i.e., the part of the brain mainly responsible for planning and execution of movements), the frequencies in the alpha band are often referred to as mu frequencies(μ). With some similarity to normal alpha frequencies, mu frequencies are attenuated or blocked if the subject performs or visualizes a motor action [13].

Beta frequencies are found in the frontal and central regions of the brain. Much like alpha frequencies, brain waves in the lower part of the beta band have been shown to attenuate in association with certain types of brain activity, in particular in association with imagining or executing motor activity [14].

Gamma frequencies are the highest frequency oscillations in the brain and are almost continuously present. Increases in the amplitude of gamma waves have been suggested to be associated with both focused attention and powerful muscle contractions [15]. Furthermore, it has been shown that motor actions that attenuate the amplitude of alpha frequencies simultaneously increase gamma amplitudes [16].

In normal awake adults, both delta and theta frequencies are a minor part of the EEG frequency spectrum. Larger amounts of delta and theta frequencies are seen in children, as well as in adults who are in a drowsy, meditative, or sleepy state. Additionally, increased occurrence of delta and theta frequencies in awake adults is associated with brain damage and certain neurological diseases, such as a meningeal tumor or cerebral infarction [17]–[19].

Taken together, current knowledge on brain waves indicates that brain signals with frequencies in the alpha and lower parts of the beta band are dominant with regard to motor imagery and motor execution.

2.2 Brain Computer Interfaces

BCI systems in themselves consist of signal acquisition, signal processing, feature extraction, and classification (see Figure 2). On top of that comes the translation of classification results into actuation of the application system (e.g., making the robot hand move), which is outside the scope of this thesis. The different parts of such BCI systems may vary a lot, depending, for example on the type of brain signals that are used as input.

The possible ways in which EEG-based BCI systems translate the EEG signals from the brain fall into five different categories, which are based on steady-state visual evoked potential (SSVEP), slow cortical potential (SCP), event-related potential (ERP), event-related

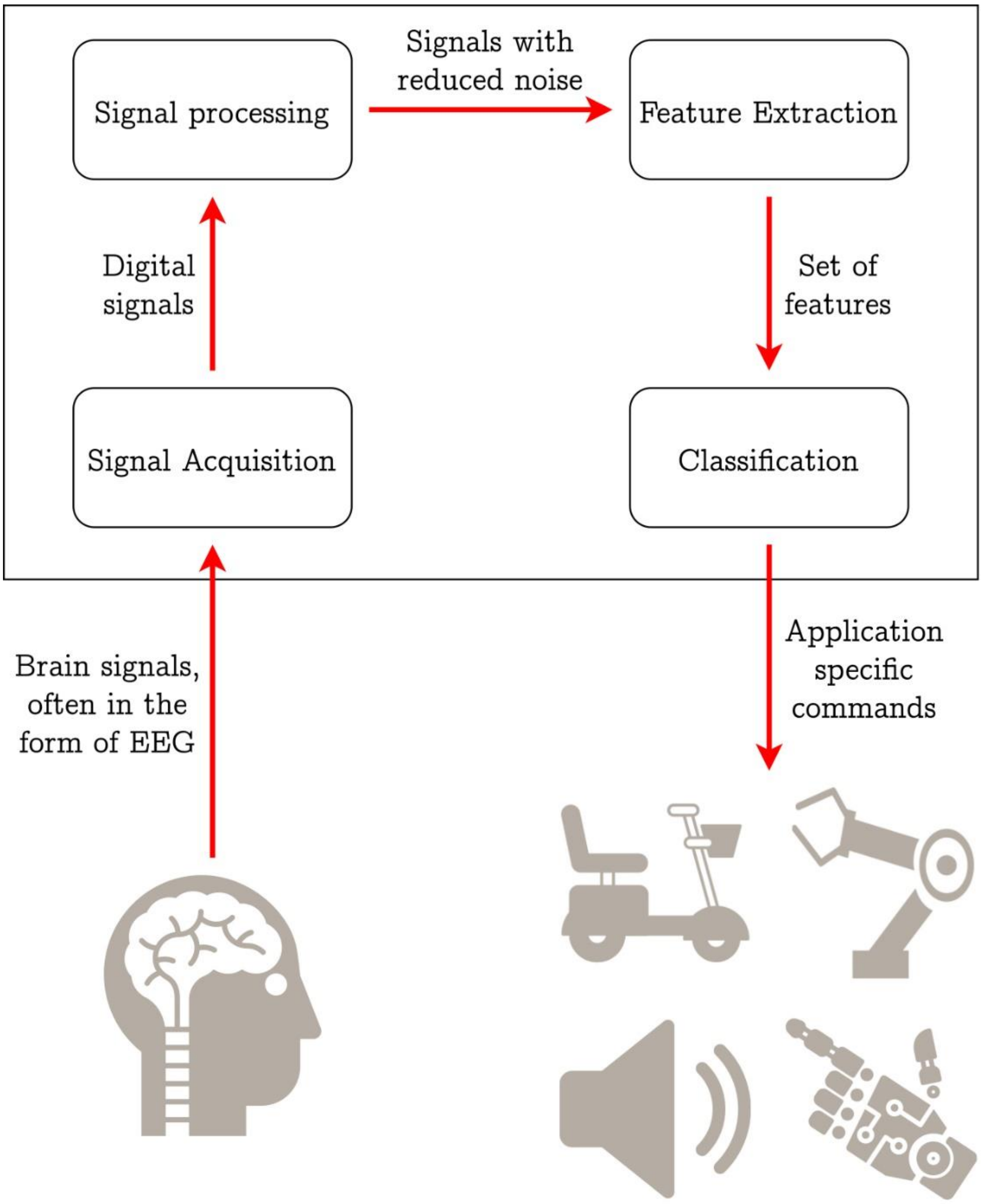


Figure 2. *A diagrammatic overview of a general BCI System.*

desynchronization/synchronization (ERD/ERS), and motor imagery/execution (MI/ME). These categories differ in terms of what kind of patterns in brain activity are being looked for and used. Thus, and more eminently, the choice translation system category also determines the type of input that is required from a subject to be able to use the BCI satisfactorily.

2.2.1 Steady State Visually Evoked Potential (SSVEP)

SSVEP is a response generated at the primary visual cortex by visual periodic stimuli. When the retina is excited by periodic stimuli between 3.5 Hz and 75 Hz the brain responds with electrical activity at the same frequencies in the area of the brain responsible for visual processing [20]. By equipping physical or virtual objects with different repetitive visual stimuli like checkerboards, or light-emitting diodes (LEDs) with appropriate frequencies, a BCI can determine which object the subject is looking at.

While SSVEP-based BCIs require little to no preparations or training time, because all the subject needs to do is look at what it wants to choose, the environment is required to have frequency-coded features/objects to work. This fits perfectly in for instance virtual reality applications, where the environment is highly customizable, however, it is not a great approach for most health care applications where the desired environment is simply the daily life of the patient, and thus SSVEP will not be focused on in this thesis.

2.2.2 Slow Cortical Potential (SCP)

Voltage changes in the brain with a frequency below 2Hz (in the lower parts of the delta band, see Table 1) are the lowest frequency EEG features, and these are called slow cortical potential (SCP).

Negative shifts in SCPs are typically associated with movement and other functions involving cortical activation, while positive shifts in SCPs are usually associated with reduced cortical activation, for example during sleep [21]. It has been shown that people can learn to control their SCPs and thereby control the movement of an object on a screen in one dimension [22]. Kübler et al. showed that two patients with late-stage Amyotrophic Lateral Sclerosis were able to learn to self-regulate their SCPs and use this to select letters or words in a language support program [23].

While good results have been achieved with SCP, it is only ever possible to control the SCP in two ways, positive and negative shifts. With only two actions, a working BCI would have to perform multiple binary decisions to choose from a large set of options. Combined with the extensive training period required to perform well, this leaves SCP unsuitable for BCIs aimed at motor control, and thus SCP will not be focused on in this thesis

2.2.3 Event Related Potential (ERP)

ERPs are voltage changes generated in the brain in response to specific events or stimuli and ERP waveforms are found by averaging EEG signals from multiple neighboring electrodes [24]. The idea behind this approach is to ignore the parts of the EEG signals that not a response to the event at hand. Because of this approach, ERPs are time-correlated to cognitive, sensory, or motor events. ERPs in humans can be divided into two categories. The early waves, peaking roughly within the first 100 milliseconds after the stimulus, are termed ‘sensory’, as opposed to the later, ‘cognitive’, waves which depend on how, and in which part of the brain, the subject processes the stimulus [25].

A lot of research has been done aimed at discovering and categorizing different ERP waveforms. They are described by their latency and amplitude and most of the identified waveforms have been linked to some type of cognitive or sensory event. In 1965 Sutton et al. [26] discovered a waveform termed P300, which has become a major component of research in the field of EEG research. A wide collection of strategies has been used to elicit the P300, of which the 'oddball' paradigm is the most used. The subject is presented with a series of similar stimuli, and occasionally a different one, i.e., the oddball. The oddball provokes the brain of the subject, and the P300 wave will be elicited [25].

ERPs, and the P300 waveform, is particularly used in language support programs, where the letters are highlighted on a screen one at a time, and the subject elicits the P300 when their desired letter is highlighted. ERPs are mostly suitable for BCIs where options can be shown on a screen, which is not optimal for the applications aimed at in this thesis.

2.2.4 Event Related Desynchronization/Synchronization (ERD/ERS)

Event Related Desynchronization (ERD) describes the regional amplitude attenuation of oscillations in the alpha and beta bands that occur in relation to an event [27]. Event Related Synchronization (ERS) describes the opposite effect, namely the regional increase of amplitude in the alpha and beta band. Both can be elicited either internally (i.e. by voluntary movement), or externally (i.e., in response to outside stimuli as with ERP) [28]. Early observations by Berger (1929), Jasper and Penfield (1949), and Chatrian et al. (1959) linked the attenuation or desynchronization of rhythms within the alpha and beta bands to the preparation and execution of movement [9], [29], [30].

ERD and ERS have been shown to be highly subject-specific, stable, and consistent over time, and suggested for further investigation as a biometric measure [31]. It can be seen as an underlying feature of brain signals, that to a slight extent explains motor imagery and motor execution (subsection 2.2.5), however, ERD and ERS, will not specifically be focused on in this thesis.

2.2.5 Motor Imagery / Motor Execution (MI/ME)

Motor Execution (ME) is simply performing a movement, while Motor Imagery (MI) can be seen as a mental rehearsal of a movement without any overt motor output. Brain signals elicited by MI and ME are somewhat dependent on the movement performed but generally exist in the alpha and beta band (Table 1).

MI/ME signals, which tend to be similar for the executed and the corresponding imagined movement [3], are important for the development of BCI systems for motor control. Subjects/patients can be trained to induce activity in the motor cortex by imagining motor movements without any limb movement or external stimulus [32]. Various machine learning methods can be used to recognize and classify distinct motor cortex activity and their corresponding physical movements. ME signals are not obtainable by paralyzed patients or amputees. However, since MI and ME produces very similar brain signals, it could be beneficial to consider them together, which is an important part of this study.

One disadvantage of MI is the large variation between subjects. Some people can perform MI reasonably well with little training, while others (10-50% of the general population) are not able to achieve satisfactory performance even after extensive training [33]. Such subjects are referred to as BCI-illiterate and this imposes a big challenge on the further development of BCIs that are usable by anyone. The problem of BCI illiteracy is an important challenge to be tackled in the upcoming years.

MI and ME is the natural choice for motor control based BCIs, and this work is solely focused on MI/ME as opposed to the other paradigms presented in this section.

2.3 Signal Processing

One of the biggest drawbacks of using EEGs as input to a BCI is the high occurrence of noise. The amplitude of EEG signals is in the order of microvolts, which makes the signal very sensitive to noise. The electrodes used to measure the signals can pick up on external electrical noise from various types of electrical equipment that are commonly found in close proximity to the subjects, such as computers, and electrical wiring. Additionally, the signals are susceptible to internal noise like flicker noise, i.e. noise from imperfect contact between two conducting materials, and thermal noise, where thermal energy causes free electrons to move randomly in a material [34]. Thus, major efforts, such as the use of robust electrode systems, are needed to minimize the noise, and preprocessing of the EEG signals is required to minimize the disadvantageous effects of noise.

2.3.1 Filtering

Preprocessing methods are aimed at maximizing the amount of useful information in the signal and minimizing the amount of irrelevant information. The first step is to increase the signal-to-noise ratio, i.e., removing noise. Methods to do so are mainly based on frequency-domain filtering, which aims at removing/reducing specific frequencies or frequency bands.

Notch filters can be used to remove stationary interference from for instance the electrical wiring in the room, by singling out the specific frequency of the alternate current in the electrical wiring and filtering out that frequency.

Band-pass filters are designed to reduce or remove all frequencies outside of a specified frequency range, for example, a band-pass filter between 8 and 13 Hz will extract the alpha rhythms of the signal. Variations of band-pass filters include high-pass and low-pass which reduces all frequencies below or above a certain threshold.

2.3.2 Feature Extraction

Recognizing and extracting features from the EEG signal is a central part of the signal processing required for this study, and vital for the to-be-developed BCI system. Most of the commonly used approaches involve some way of translating signals from the time domain to the frequency domain, to quantify the frequency information in the data, which has been shown to be valuable features for MI-based EEG classification [35]. Other methods aim at using mathematical procedures to extract the features that are the most distinguishable from each other. A brief introduction to some of the commonly used methods follows below.

Fourier-based methods are tools that use a mathematical transform to decomposes a function depending on time/space into a function depending on temporal/spatial frequency. The origin of Fourier-based methods is the Fourier transform (FT), discovered by Joseph Fourier in 1822 [36]. The FT produces the power spectral density of a signal, and other Fourier-based variants aim at producing an estimate of this.

One variant of the FT is the aptly named Fast Fourier Transform (FFT). The FFT divides the signal into multiple smaller parts, to compute the FT of each part and put them together to form an estimated power spectral density. This reduces computation time drastically.

The wavelet transform (WT) is an algorithm that uses a basis function that introduces additional special properties in the transform. It produces a time-frequency plot with variable resolution in both time and frequency. It provides good time resolution and poor frequency resolution at high frequencies while the opposite is the case for low frequencies.

The periodogram is an estimate of the power spectral density, that is computed by performing the FT of the signals estimated autocorrelation sequence. This method exploits the convolution theorem, which states that the FT of a convolution of two signals is the pointwise product of their FTs. Since the estimated autocorrelation is the signal convolved with a copy of itself, the calculation of the periodogram becomes quite simple. To reduce bias, the signal can be split into overlapping segments, for which the periodogram is calculated for each and array averaged, to result in an averaged periodogram. This method is commonly known as the Welch method, and is the method that is implemented in this work.

Common Spatial Pattern (CSP) is a technique that has been shown to produce strong results on MI-based BCI, exemplified by Ang et al. [37], who won a BCI competition which was very relevant to this work and will be covered in subsection 3.1.1. CSP weighs the electrodes according to their significance in the classification task and suppresses noise in individual channels by using correlations with neighboring electrodes [38]. CSP has the disadvantage of needing an extensive number of electrodes, which increases the electrode application time and cost, while reducing the portability of the system.

2.4 Feature Classification

BCIs are based on using the brain as the medium to communicate commands to a system. This requires the system to recognize which commands the subject is trying to convey by classifying the brain signals into predefined actions or decisions of for example a robot hand or computer simulation. Some of the most established methods are described in the following subsections

The methods used for feature classification are based on machine learning and classical statistics and these methods are often categorized into supervised and unsupervised methods. Supervised methods use data that are labeled, which makes it easy to evaluate the result, while unsupervised methods use unlabeled data and instead of classifying input into predefined classes, groups input that are similar into unnamed classes. Some interesting work has been done with unsupervised learning, as a means for making BCIs that are able to self-adapt while in use [39], however, it seems to be the common conception is that supervised learning is the way to go for BCIs.

2.4.1 Deep Learning

Deep learning is a group of modern and powerful classification techniques, based on Artificial Neural Networks (ANNs). ANNs are inspired by the biological neural networks in human brains. The most basic ANN called the Multilayer Perception (MLP) is based on a network of neurons organized in layers with weighted connections between them. Input signals are fed into the first layer, and processed through all the layers, to produce results in the last layer. Each neuron gets input from the neurons in the previous layer, processes it, and signals to the neurons in the next layer based on the outcome. The output from the last layer decides the classification, and the result is checked against the true value for the input that was provided. If the classification was correct, the weights of all inter-neuron connections are adjusted in a way that will make sure that other inputs that are very similar will be classified the same. If the classification was incorrect, the weights are adjusted in a way that will make sure that similar inputs will be classified differently.

Deep learning techniques require a lot of data, however, if enough data is available, deep learning techniques are often able to find patterns in the data that are impossible to find with traditional algorithms (Figure 3). The connections between neurons and the function for updating weights can be implemented in various ways to make variations of ANNs. These include Convolutional Neural Networks (CNNs) which take advantage of the spatial and temporal structure of the input data and Recurrent Neural Networks (RNNs),

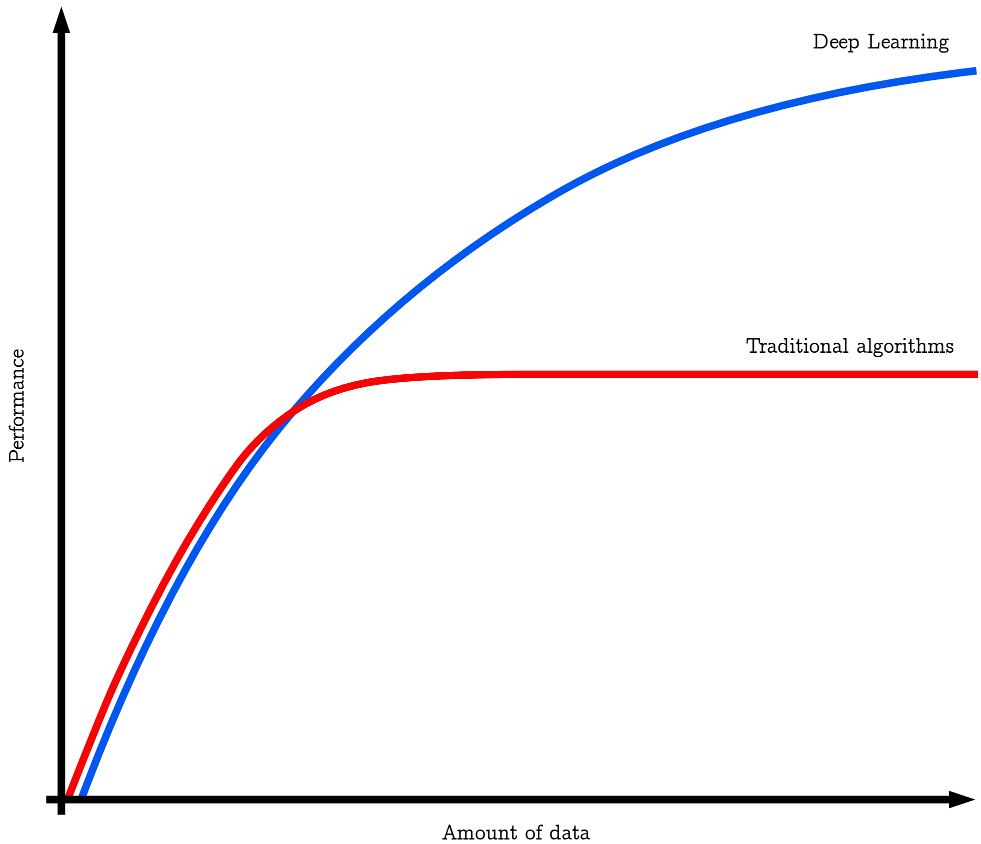


Figure 3. The relationship between the amount of available data and classification performance in deep learning and traditional algorithms.

which give the neurons memory to use the previous output as additional input.

2.4.2 Support Vector Machine (SVM)

Support vector machines (SVMs) are increasingly popular in various biological applications [40]. An SVM is a supervised learning algorithm that induces a linear optimal hyperplane to discriminate between classes based on the maximum margin principle. If it is not possible to classify the data linearly, SVMs can use a kernel function to map the data into a higher-dimensional space in which linear classification is possible. Since SVMs use a linear hyperplane they are typically used for binary classification tasks, but they can also be implemented as multiclass classifiers by breaking the problem down into multiple binary classification cases.

2.4.3 Nearest Neighbor Classifiers

A simpler group of classifiers are called Nearest Neighbor Classifiers, where a feature vector is assigned to a class based on the class of its nearest neighbor (s). The majority vote of the k nearest neighbors is called k Nearest Neighbor (kNN) and is the most widely used classifier in this group. The kNN approach requires lower training and testing time than the more complex SVM and MLP, but with comparable classification accuracy [41]. Nearest neighbor classifiers are inherently unsupervised, but can be modified to include labels, which makes them more suited for BCI classification.

2.4.4 Random Forest (RF)

Random Forest (RF) is a classification algorithm that uses standard decision trees, which are simple classifiers that predict the value of an output based on binary decisions on each input value. Such a simple classifier could be effective in some cases, but it is prone to fitting too closely to the training set, causing it to perform badly on new independent data. This is referred to as overfitting and RF aims to solve this by fitting many different trees and setting the final classification based on a vote by all trees. This reduces the variance and creates a more robust classifier. RF has performed surprisingly well considering its simplicity and is believed to have the potential for BCIs, thus it will be the main classification algorithm in this work.

2.4.5 Reinforcement Learning (RL)

Reinforcement Learning (RL) is a machine learning algorithm with similarities to the way a human child learns to operate in the world, by trial and error. Initially, the system tries to process input (often simply the state of the environment) and executes an action. The environment responds with a negative or positive reward (i.e., reinforcement), and the system learns based on these rewards. If the system receives a strong negative reward, the system will learn that it was a bad action, in response to the input received, and thus be less likely to repeat that action if presented with similar input. Reinforcement Learning is very powerful but is dependent on a good way to set the strength of the reinforcements, which can be difficult for many classification problems, including BCI-related classification. Although RL is not explored any further in this thesis, a simple idea of how to implement Reinforcement Learning in BCI is proposed in section 7.4

2.4.6 Transfer Learning

Transfer learning is not a classification algorithm in itself but is a technique that aims to exploit patterns found in a dataset from one problem to help find patterns in another dataset, potentially from another problem. This is done by training a model on one dataset, and then continuing the training with another dataset, with little modifications in between trainings, transferring as much trained information from the first to the second dataset. For the problem described in this thesis, transfer learning can be to train a model on data from one subject, and retraining on another model, or to train a model on ME data, and retraining on MI data.

Chapter 3

Datasets

A key aspect of machine learning algorithms is the need for large data sets to be able to train a classifier. It is also of great importance that the data used for training is of high quality, with little noise, and has a high resemblance to the data that would be used in a real-world application. The general conception is that the more available data, the better performance can be achieved, with eventual convergence towards the upper boundaries that are set by the algorithm and/or the classification problem.

In many research areas, deep learning has made great positive impacts over the last decade. This is mainly because deep learning methods have a higher upper boundary of performance compared to traditional methods. Since the amount of available data is growing exponentially in many fields, reaching these higher upper boundaries of performance becomes increasingly realistic. With a large enough dataset, deep learning methods have the potential to achieve better performance than traditional methods (see Figure 3).

In the field of EEG-based BCI, data collection is considerably more resource-demanding than many other fields, which makes the growth of available data lower compared to for example various image-based classification problems.

3.1 Benchmarking and Coordination

When research in BCIs started to take off around the year 2000 (Figure 1), algorithms were reported to give impressive results quite quickly [42]. However, due to the naturally large variability in datasets and performance metrics in BCI systems, there was no good way to compare research results to assess their true quality and impact.

3.1.1 BCI Competitions

In 2003 a group of researchers from multiple research groups started a collaboration on organizing BCI competitions with publicly available datasets that could serve as benchmarks for various BCI applications [43]. The collaboration included researchers from Germany, Austria, and the USA, including the eminent Gert Pfurtscheller and Niels Birbaumer, who are two of the most influential researchers in the field of BCI.

By encouraging researchers to implement their methods on the same datasets, within the framework of a small set of simple guidelines, the competitions served as benchmarks for BCI research for some time. Although the last competition was in 2008 and other datasets are taking over as benchmarks today, research is still being published using the original BCI competition datasets.

3.1.2 BNCI Horizon 2020

In 2013, an EU coordination and support action was funded, named the Brain/Neuronal Computer Interaction Horizon 2020 (BNCI Horizon 2020). The initiative aimed at providing a roadmap for research efforts in the BCI field until 2020 and beyond [44]. Some of the researchers involved in the organization of the BCI competitions were also involved in the BNCI Horizon project.

BNCI Horizon has provided the research community with more than 25 different publicly available data sets that, in the same way as the earlier BCI competitions, contribute to better benchmarking of research in the field, provided that these data sets are being used.

3.2 Producing a Dataset

Despite the possibility of using publicly available datasets, this research included the production of a unique dataset that was tailored for addressing the research questions that we asked. The motivation behind producing a dataset specifically for this master's thesis work was to be able to simplify the BCI problem and separate the problem into multiple levels of complexion (as presented in a list in section 3.2.1, below). This approach was chosen upon recommendation by experts in neuroscience at the University of Oslo with strong experience in EEG research, to increase the chance of producing useful results.

Another motivating factor for producing a dataset specific for this work was to have a dataset with a combination of MI tasks and ME tasks, with data for both tasks being generated with otherwise identical parameters. The technical challenges of producing the dataset and the considerations made are described in the following subsections.

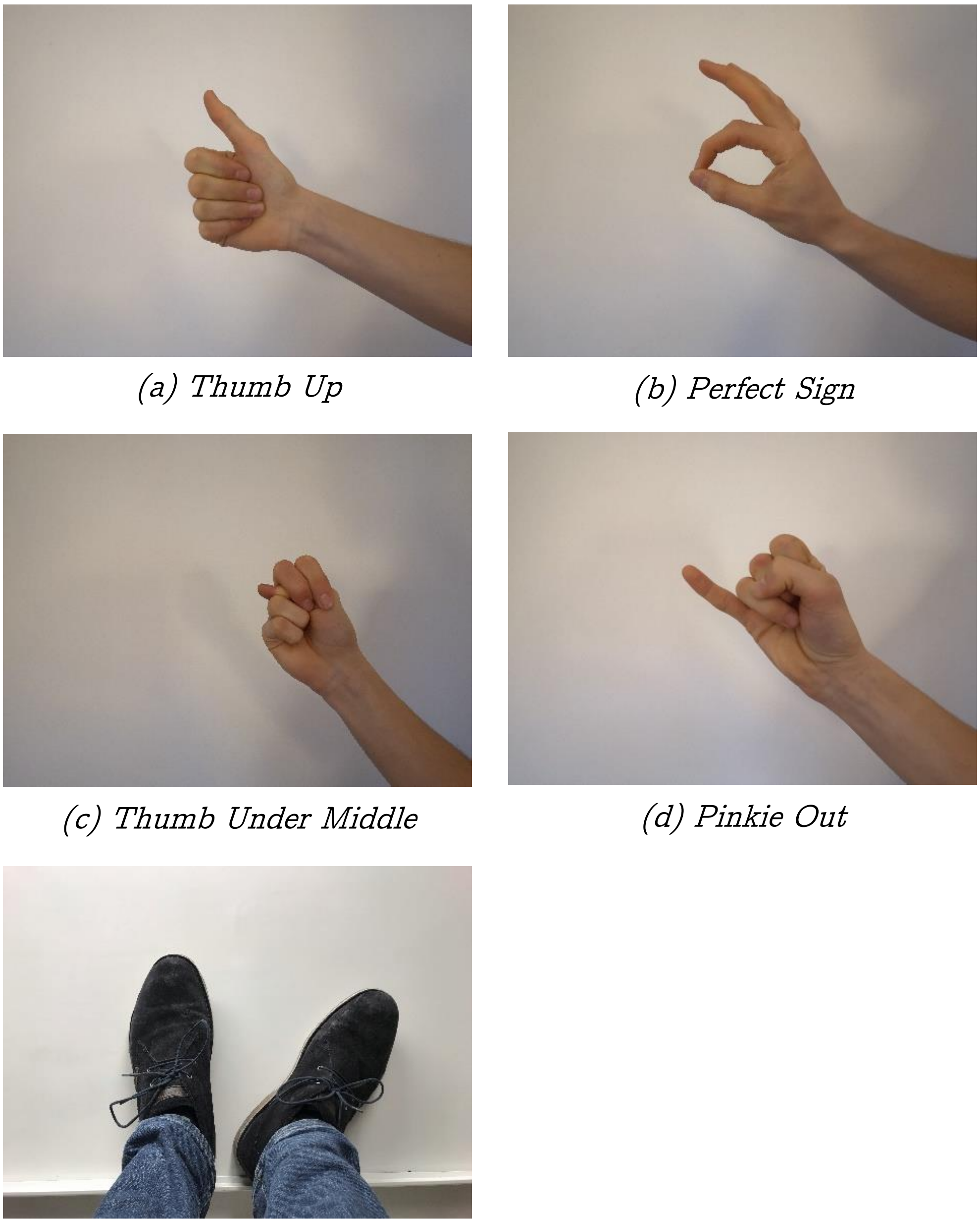
3.2.1 Gestures

The nature of the dataset used for BCI development in studies like the present needs attention. Most of the available datasets on hand gestures either is a binary right hand / left hand problem or include many similar gestures, which poses interesting challenges, because discriminating between gestures is difficult, making it difficult to obtain high classification accuracy between those gestures. Therefore, in the present study, a mixture of both quite similar and quite different gestures was chosen, to allow working at different levels of complexity. Five gestures were chosen, including a foot gesture (slight outwards rotation of foot), to be used as a reference, and four hand gestures (Figure 4). Subjects were explicitly instructed to do the foot gesture and the hand gestures with the hand/foot at the same side.

Among the four selected hand gestures, two gestures are very simple and are expected to be natural for most subjects (Thumbs Up and Perfect Sign). The other two gestures are more difficult and are expected to feel somewhat odd for the subjects and to demand a higher level of focus (Thumb Under Middle and Pinkie Out).

The different levels of the classification can be simplified as follows:

- Differentiate between hand gestures and the foot gesture
- Differentiate between “complex” and “simple” hand gesture
- Differentiate between all gestures



(a) Thumb Up

(b) Perfect Sign

(c) Thumb Under Middle

(d) Pinkie Out

(e) Foot Movement

Figure 4. Pictures of gestures that were presented to the subjects. Picture (e) refers to an outwards rotation of the right foot, this was clarified prior to the first run.

Before the experiment, each subject went through the gestures one by one, to make sure that the subject was able to execute the gestures. It was explained to the subject that specific angles or alignment of their fingers was not important, as long as the subject had their way of doing the gestures consistently.

3.2.2 Timing Scheme

Each session comprised up to 16 runs of three minutes each, separated by short breaks, of anywhere from 30 seconds to five minutes. Eight of the 16 runs were used to generate Motor Imagery data, and the other eight to generate Motor Execution data. Each run consists of 35 trials (seven trials for each of the five gestures), yielding a total of up to 280+280 trials (MI+ME) per subject.

Every run was initialized by presenting a grey screen to the subject. After an initial 5 seconds, the first trial started by presenting a gesture image in the middle of the screen. The images were presented in a randomized order, such that the subject did not know which image to expect next, at any given time during a run. The gesture image was shown for 3 seconds, followed by a 2-second break, before the next trial (Figure 5). The subjects were instructed to

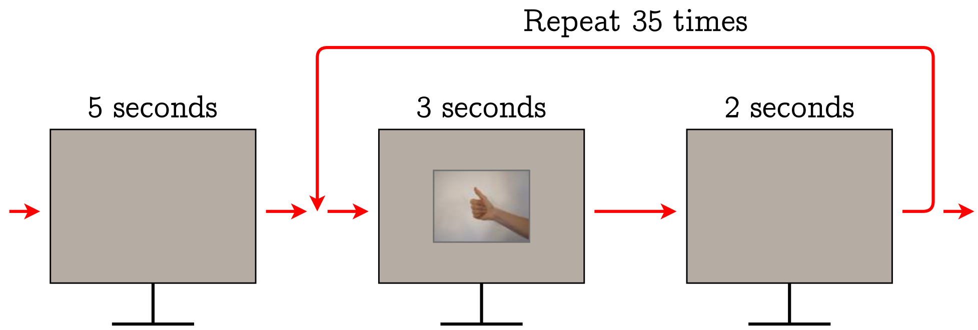


Figure 5. Timing scheme for one run.

execute/imagine the gestures for the full 3 seconds that the image was shown.

After each run, the subjects had the opportunity to drink some water, and decide the duration of the break before the next run. The subjects were asked to share how they felt they performed. If a subject expressed signs of severe fatigue and/or drop in focus, or upon request by the subject, the session was cut. For one of the subjects in the final dataset, this was the case, and as such, only one run of MI data is available in addition to ME data is available for that subject (later referred to as subject 1).

3.2.3 Data Recording

The experiment was performed in a laboratory at the Department of Psychology at The University of Oslo using equipment for recording

Figure 6. The author of this thesis wearing a head cap with attached electrodes. Picture by Katrine Nergård.

EEG from BioSemi [45]. The EEG signals were recorded with sixty-four Ag-AgCl active electrodes with a sampling frequency of 2048 Hz. The electrodes were arranged on the scalp according to the international 10-20 system [46], illustrated by a diagram provided by BioSemi shown in (Figure 7). This was achieved by using BioSemi head caps which are available in multiple sizes and thus can be adapted to the size of the subject's head (Figure 6). The head caps have multiple holes in them, corresponding to the correct sites for the electrodes to be attached, given that the head cap is correctly fitted to the subject. Before attaching the electrodes, these holes were filled with saline gel to maximize contact between the scalp and the active electrodes. Fitting the head cap to consequently achieve the intended electrode placement (Figure 7) was performed as accurately as possible to ensure comparability between subjects.

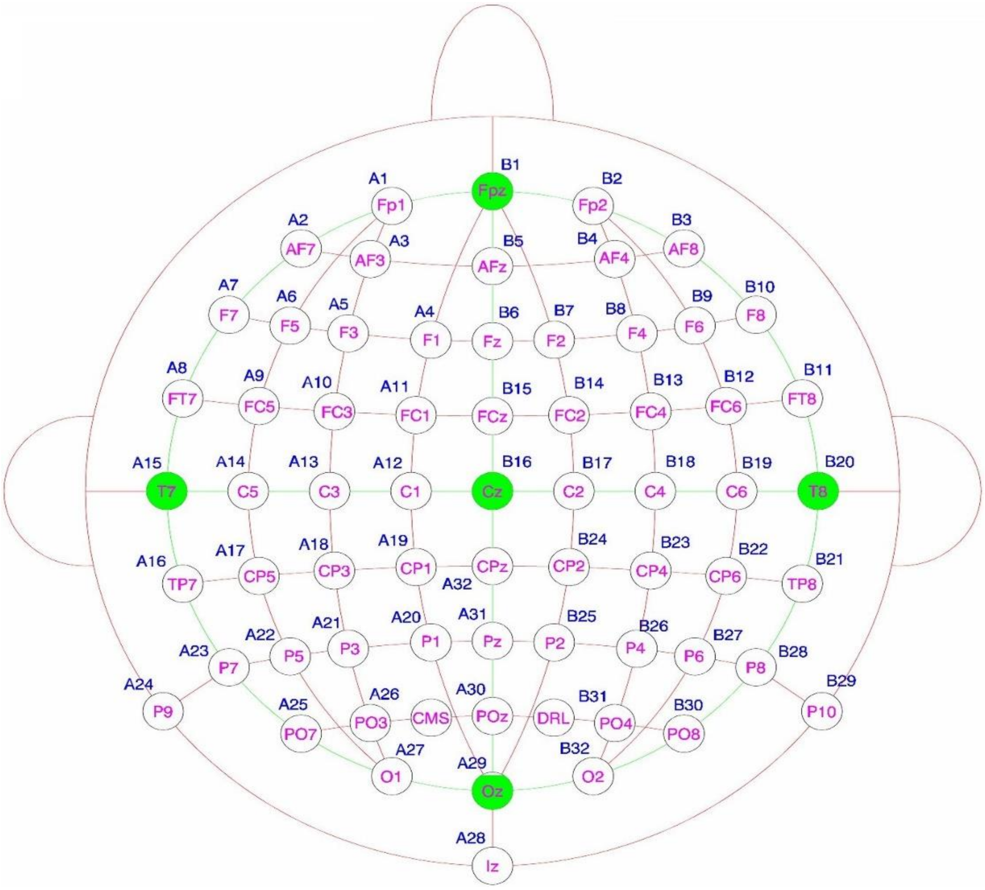


Figure 7. The layout of the BioSemi head cap. The labels in pink font are the electrode names according to the international 10-20 system. The labels in blue font are the electrode labels as they are used in the datasets. The cap is adjusted so that Fpz and Oz are placed on the line from Nasion to Inion at intervals 10%, 80%, and 10%. The Nasion is also known as the bridge of the nose, and the Inion is can be located by finding the small bump on the lower backside of the skull. Similarly, T7 and T8 are placed on the line from preauricular to preauricular at intervals 10%, 80%, and 10%. The preauriculars are the small holes in front of the upper ear on each side of the skull. Image Source: BioSemi [38]

3.2.4 Subject Preparations and Environment

The subjects were asked to make sure that they were well-rested and to eat a proper meal before the experiment. Upon arrival at the experiment site, the subject was presented with a brief explanation of the project and details about the methodology, followed by familiarization with the task at hand (i.e., executing and imagining the gestures). When the subject had no more questions and seemed to understand the experiment, the subject was given the opportunity to go to the toilet, drink some water, and make any other last-minute preparations of choice.

Subsequently, the subject was seated in a comfortable, adjustable armchair in front of a computer screen in a separate room, with their face approximately 1 meter from the screen (see Figure 8).

3.2.5 Synchronization and Data Format

When producing a dataset with brain signals and markers that represents the visual stimuli presented, it is important to achieve highly accurate synchronization between the brain signals and the markers. Synchronization is crucial for obtaining accurate information about exactly which brain signals were produced during the stimuli periods (i.e., when the picture was present on the screen). Synchronization of the data collected in this study was optimized

Figure 8. The setup of the experiment; the motor execution task "Thumb Under Middle" is being performed by the author of this thesis. Picture by Katrine Nergård.

using *Labstreaminglayer*, which is known for its sub-millisecond accuracy [47]. *Labstreaminglayer* is an open-source and cross-platform software package that was developed to streamline the data acquisition process and supports data collection from a large range of hardware, including BioSemi, and the python package *psychopy* [48].

Psychopy is a simple tool for presenting stimuli to subjects. By creating a simple loop in python, stimuli were presented, and accurate time stamps were sent to the *Labstreaminglayer*, which saved the data in the xdf-file format. The xdf-file format is a compressed binary file, that can easily be opened in python with the *pyxdf* package [49].

3.2.6 Problems Encountered

After the initial data collection (9 subjects), analysis of the data revealed signals that caught attention because they looked like a type of artifact that was not expected. The raw signal showed short spikes in voltage with incomparably higher amplitude than all other rhythms seen in the data. While these spikes could just represent noise, their magnitude rendered them impossible to remove, even with a thorough assessment of various noise removal procedures. After investigating the BioSemi system, it was concluded that the observed spikes were due to an erroneous electrode set-up, which had caused the active electrodes to effectively only record strong signals derived from the blinking of an error light that overshadowed the weaker brain signals. Upon instruction from the developers of BioSemi, the problem was fixed by attaching BioSemi's Common Mode Sense electrode, and Driven Right Leg electrode, a set-up that essentially replaces the ground electrodes in traditional systems, and should always be done when using BioSemi's active electrode system.

Unfortunately, the originally collected data from nine subjects had to be discarded, which was a big setback, as multiple weeks had passed when working on data collection, and fixing the problem.

3.2.7 Final Dataset

At the time of the second round of data collection, the infection rate of the Covid-19 pandemic had reached new heights in Oslo. The university naturally had to restrict access to the EEG lab and decided to forbid access for non-University employees, non-students, as well as employees and students that did not already have access to the lab for their own projects. This made data collection difficult, as access to subjects was severely limited. Considering the reduced number of possible subjects, the choice was made to only collect data from three well-instructed subjects, assuming this would provide sufficient data to critically and reliably assess traditional algorithms and yield results that could identify limitations as well as identify opportunities for subsequent studies with larger datasets. The view was that if the situation would improve, more data could be collected at a later stage.

Chapter 4

Research Methods

4.1 Preprocessing

When presented with EEG data, either in a live BCI system or after collecting data in an experiment, as has been done here, the raw signals contain a considerable amount of noise. Naturally, the next part of the successful development and application of a BCI is to process these raw signals into filtered signals that contain as much useful information as possible, with as little noise as possible. This section provides a step-by-step description of the preprocessing methods that were used in this project.

4.1.1 Selecting and discarding bad channels

The equipment and procedure used to capture EEG signals, with its 64 electrodes, has elements that are vulnerable to both user errors, and technical errors. Other than severe errors, as described in subsection 3.2.6, which need to be corrected, it is quite common to experience "bad channels", which are either "dead" (exclusively white noise) or significantly noisier than the other channels. This could be due to an electrode having a twisted or even broken cable, improper placement of electrodes, or bridging of two electrodes by the saline gel.

The most common way of detecting bad channels is by visual inspection and tools to do so are a standard feature of most established EEG-handling software. Since this work relied on implementing a custom EEG-handling software, a tool was implemented that looped through every raw EEG file and provided sufficient visual information to be able to detect the bad channels. For each raw file in the selected folder, the function presented two windows, a simple graphic user interface (GUI), and an interactive plot of the raw EEG data (Figure 9 and Figure 10). The plot allows the user to hide and show individual channels as well as to zoom in and move the plot around if needed. Furthermore, the GUI shows basic metrics to help the user identifying potentially bad channels and provides a check box per channel that is to be ticked if the channel is deemed bad. Selected channels were removed from the data, and this was done by setting their values to 1.0, thus ensuring that the dimensions of the data were kept constant.

The metrics that were used as a guideline to identify potential bad channels, were selected by trial and error. This means that an initial visual inspection was performed, and the metrics were estimated based on the channels deemed bad by the inspection. Thus, the selection of these metrics is not thoroughly justified but proved to help speed up the visual inspection process when inspecting processing the entire dataset. The metrics considered were:

- win_stdev_max – The maximum value for the windowed standard deviation with window size = 1024 samples (0.5 seconds).

- win_stdev_mean – The average value for the windowed standard deviation with window size = 1024 samples (0.5 seconds). A low value could indicate a "dead" channel
- stdev_ratio – The ratio win_stdev_max / win_stdev_mean. A large value indicates large variations in the signal's behavior over time. A high value indicates sudden spikes/drops in voltage that are considerably more powerful than the general fluctuations in the signal, which could be a sign of a loose contact or damaged wire.
- mean – The average of the entire channel values. A very high value indicated a channel with a bad connection.

As examples, the flat grey line at about 250 000 μV in Figure 9 has a very high mean, a very low win_stdev_mean. The pink line at about 100 000 μV in Figure 9 also has a quite high mean, and has an extremely high win_stdev_mean, visualized by the large oscillations compared to the rest of the signals. By just visual inspection, both the mentioned channels were almost immediately identified as "dead" electrodes (Figure 9).

4.1.2 Removal of high-frequency components

There is some disagreement on how far the gamma frequency band of the brain waves stretches (Table 1), where some studies indicate the presence and relevance of frequencies well above 100 Hz [50]. However, it seems to be the general conception that for motor tasks, both executed and imagery, only frequencies from the theta band up to and including the lower part of the gamma band (≈ 45 Hz) are essential [51]. With a sampling frequency of 2048 Hz and, consequently, a Nyquist frequency of 1024 Hz, the data collected in this study included frequency information from 0 to 1024 Hz. The

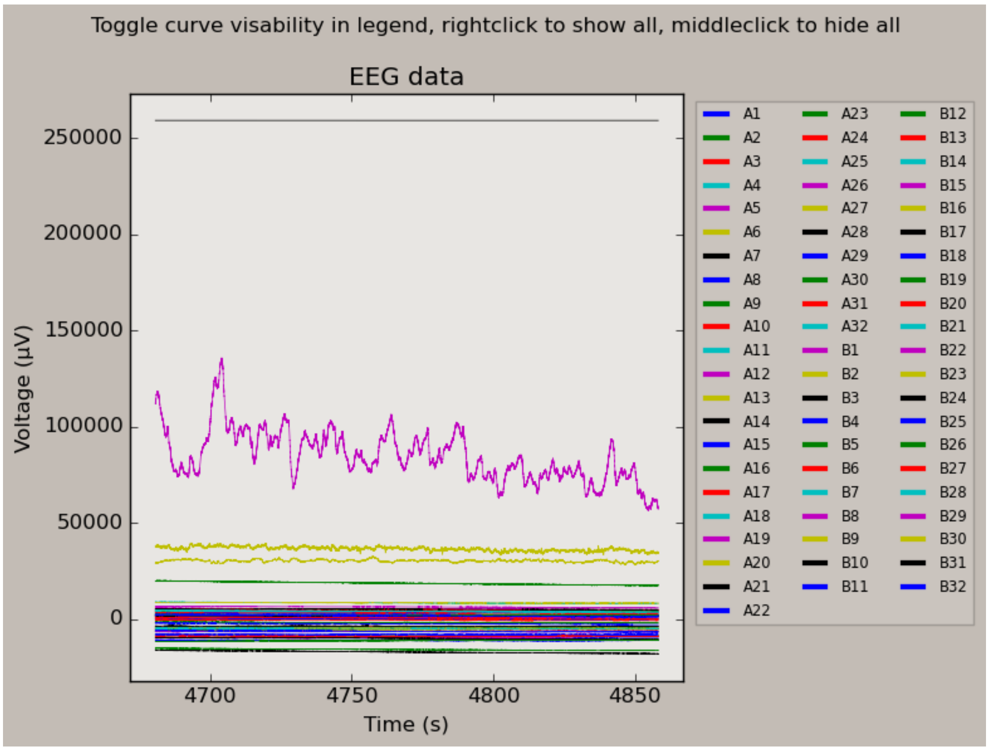
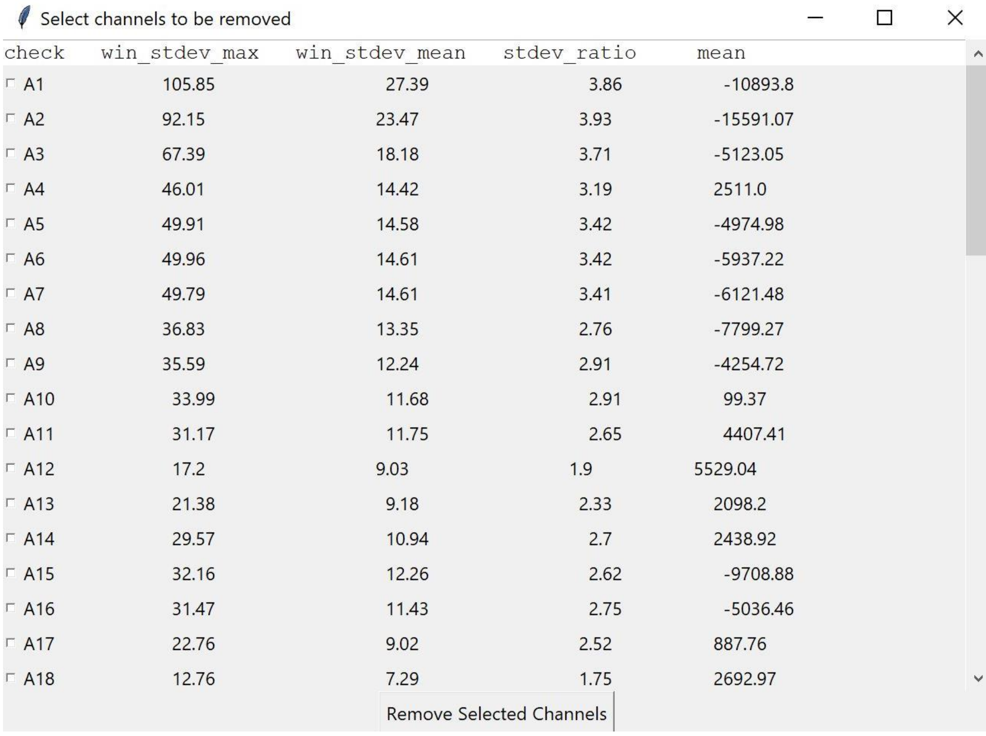


Figure 9. Interactive plot for visual inspection of potential bad channels.

overshooting frequency information can be filtered out using a bandpass or lowpass filter.

The straightforward thing to do is to filter the data with a bandpass filter ranging from 3 Hz to 45 Hz. However, from a machine learning perspective, where more is better in terms of data amount, it is interesting to include some margin in the hope that additional information can be extracted from outside the known essential frequency range. Two separate Butterworth filters were created using the python package *Scipy* [52], namely, one lowpass filter with cutoff at 80 Hz, and one bandpass filter with passband 3 Hz – 45 Hz (Figure 11). Hereinafter, the two variants of the datasets will be referred to as dataset A (lowpass at 80 Hz) and dataset B (bandpass at 3 Hz - 45 Hz). Both filter variants, were implemented as



check	win_stdev_max	win_stdev_mean	stdev_ratio	mean
A1	105.85	27.39	3.86	-10893.8
A2	92.15	23.47	3.93	-15591.07
A3	67.39	18.18	3.71	-5123.05
A4	46.01	14.42	3.19	2511.0
A5	49.91	14.58	3.42	-4974.98
A6	49.96	14.61	3.42	-5937.22
A7	49.79	14.61	3.41	-6121.48
A8	36.83	13.35	2.76	-7799.27
A9	35.59	12.24	2.91	-4254.72
A10	33.99	11.68	2.91	99.37
A11	31.17	11.75	2.65	4407.41
A12	17.2	9.03	1.9	5529.04
A13	21.38	9.18	2.33	2098.2
A14	29.57	10.94	2.7	2438.92
A15	32.16	12.26	2.62	-9708.88
A16	31.47	11.43	2.75	-5036.46
A17	22.76	9.02	2.52	887.76
A18	12.76	7.29	1.75	2692.97

Figure 10. Simple GUI to select bad channels for removal. The metrics are listed on each row are listed and explained in subsection 4.1.1

Butterworth filters, with the cutoffs defined as the point where the gain was reduced to ${}^{1}/_{\sqrt{2}} \approx 0.71$.

4.1.3 Downsampling

After removing/reducing a large portion of the frequencies using the filters described above, the size of the data is still approximately the same. The Nyquist frequency is still 1024 Hz, which theoretically still enables us to identify the magnitude of frequencies that are outside the desired ranges, even if they are insignificant. We can downsample the data by simply keeping only the n_{th} sample, without losing frequency information, as long as $f_s / 2n > f_{max}$, where f_s is the sampling frequency, f_{max} is the largest frequency component in the

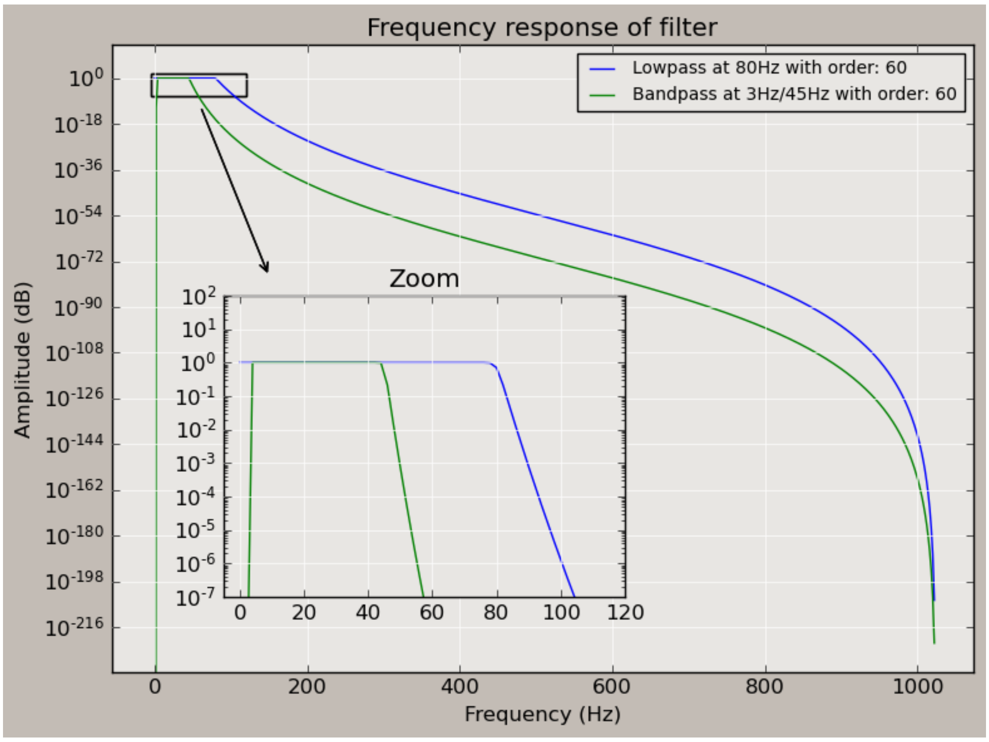


Figure 11. Frequency response of the lowpass filter (blue) and the bandpass filter (green).

signal (after filtering). This process is referred to as decimating by a factor of n and essentially lowers the Nyquist frequency to be just above the maximum frequency that is of interest. Dataset A was decimated by a factor of 12 and dataset B was decimated by a factor of 22.

Order of downsampling and high-frequency removal

It is important to note the impact of the order of the two operations described above, downsampling and filtering out high-frequency components. If a signal that is downsampled has frequency components larger than half the new sampling frequency aliasing noise will be introduced into the data. Aliasing noise is when voltage changes from frequencies higher than the Nyquist frequency (i.e., frequencies which cannot be observed because the sampling frequency is too low), are projected into lower frequencies. Bandpass or lowpass filters that are used before downsampling are often referred to as anti-aliasing filters, as their main goal can be considered to be avoidance of introducing aliasing noise during downsampling.

Reusing discarded samples when downsampling - Splitsampling

To fully utilize the data available, a method was employed that reuses the samples that would otherwise be discarded when downsampling. This method, introduced by Frydenlund et al. [53], involves making n new dataset, taking the n_{th} sample starting at sample 0,1,...,n-1. This gives n "parallel" datasets and, that can be seen similar versions of the same data, but with varying white noise. This data augmentation method will be referred to as splitsampling.

The method is similar to a data augmentation method widely used in image recognition problems where multiple rotated and/or scaled variants of an image are included in the training set to increase scale- or rotation-invariance. In the case of time series like EEG data, the method is adding slight temporal invariance [53].

4.1.4 Removal of Power Line Interference

The power line frequency is the theoretical frequency of the alternating current available in normal electrical outlets. This frequency is different around the world, and the most common is either 60 Hz or 50 Hz. In Norway, where this data was collected, was expected, and found, to be 50 Hz. As the voltage in the power lines is in the range of 10^6 - 10^7 as high as the voltage of the EEG signals, the power lines are very likely to contaminate the EEG signals.

To remove the noise from the power line, a notch filter was implemented using the python package *Scipy* [52]. The center frequency was set to 50 Hz, and the width of the filter was set to 0.25 Hz. Figure 12 shows the power spectrum of one channel from dataset A before and after the notch filter was applied, as well as the frequency response of the notch filter itself.

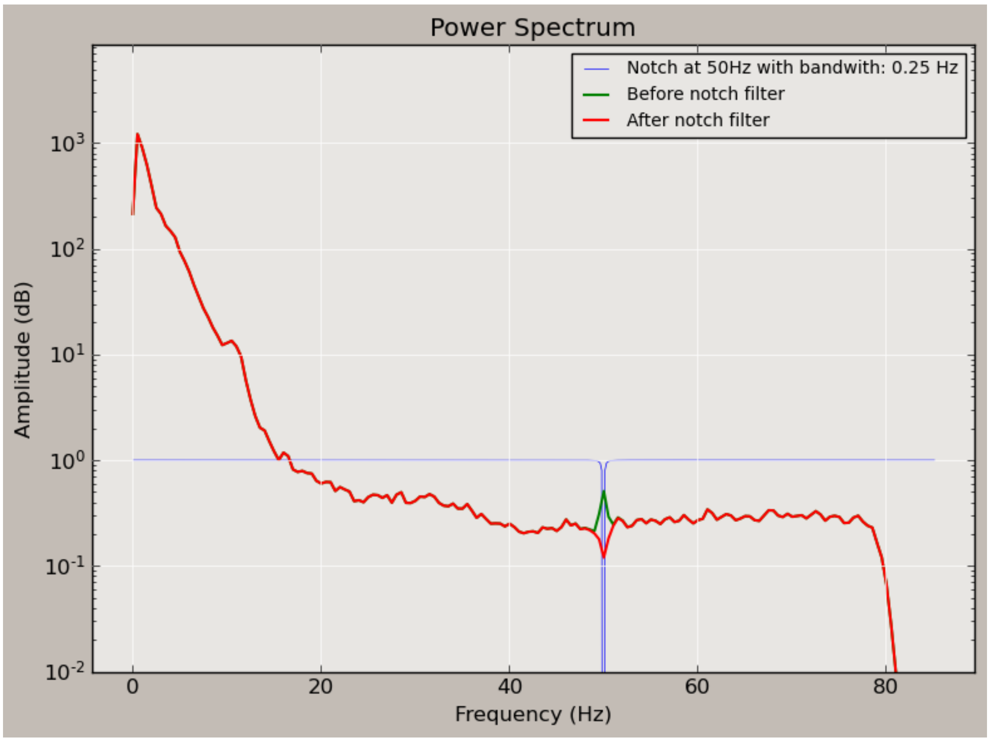


Figure 12. Application of the notch filter. The green and red lines are the estimated power spectral density of electrode A1 (Fp1) from one run in dataset A. The green is before, and the red is after the notch filter was applied. The blue line is the frequency response of the notch filter.

4.2 Feature extraction

4.2.1 Data Chunking

An important aspect of time series classification is deciding how big chunks of the data (defined by a time range, or $t_{chunksize}$) that are to be handled at a time. Generally, the goal is to make a system that classifies the incoming signal live, which equates to reducing $t_{chunksize}$ to near-zero. However, the more $t_{chunksize}$ is reduced, the less low-frequency information remains in the chunk. This is because the relationship between frequency f and the duration T of a full oscillation: $f = {}^{1}/_{T}$. For example, in a chunk with a length of 0.05 seconds, frequencies under 20 Hz will not be recognizable, as the chunk is not long enough for a full oscillation.

The total lag is equal to $t_{tot} = t_{chunksize} + t_{computing}$, where $t_{computing}$ is the time that is needed to do the classification. While this is very important when eventually implementing a real-life system, optimization of this aspect was not part of the study described in this thesis, and therefore, optimization of these metrics is not discussed any further. The data were split into 3 seconds (i.e., the time period during which the stimulus was present on the screen (see subsection 3.2.2), and each 3-second chunk was processed as one trial, with a corresponding label (i.e., the name of the gesture).

4.2.2 From Temporal to Frequency Information

Because of the large amounts of noise in EEG signals, the raw signals in the time domain is are not appropriate for use as classifier input. As discussed in subsection 2.3.2, there are many ways to convert the signals into features based on their characteristics, with most

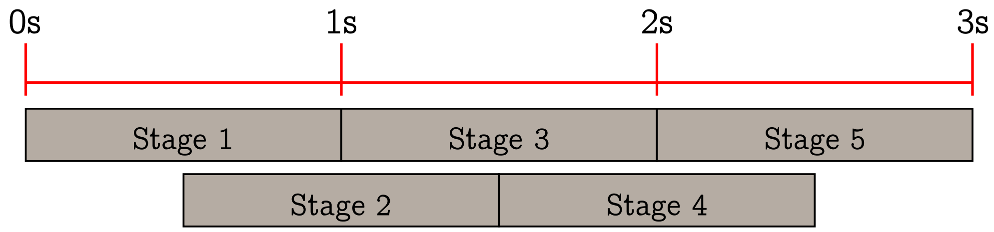


Figure 13. Diagrammatic overview of a trial. The trial, with a total length of 3 seconds, was split into 5 1-second stages with 0.5 second overlap.

methods including a transformation from the time domain to the frequency domain. The method that was used in this project is based on generating periodograms with the Welch method as described in subsection 2.3.2. This is one of the most commonly used methods to estimate the spectral density of a signal.

To retain some temporal information, each trial was split into five 1-second parts with a 0.5-second overlap, and these 1-second parts will be referred to as stages (Figure 13). The Welch periodogram for each stage was computed using the *Scipy* package in python[52]. Since we expect the most influential frequencies to be in and around the alpha band and lower parts of the beta band (Table 1), these frequencies should be weighted heaviest. The Welch periodogram produced an estimate of the power spectral density with a frequency resolution of 1 Hz, which equals frequency bands of width 1 Hz. In the range 5 – 16 Hz, these bands were taken directly as features, while outside that range, the bands were averaged to widths of 2, 4, or 8, as visualized in Figure 14.

This leads to a feature vector that, for dataset A, consists of:

$$5\ stages\ \times 25\ frequency\ bands\ \times 64\ channels = 8000\ features.$$

For dataset B, these numbers are:

$$5\ stages\ \times 19\ frequency\ bands\ \times 64\ channels = 6080\ features.$$

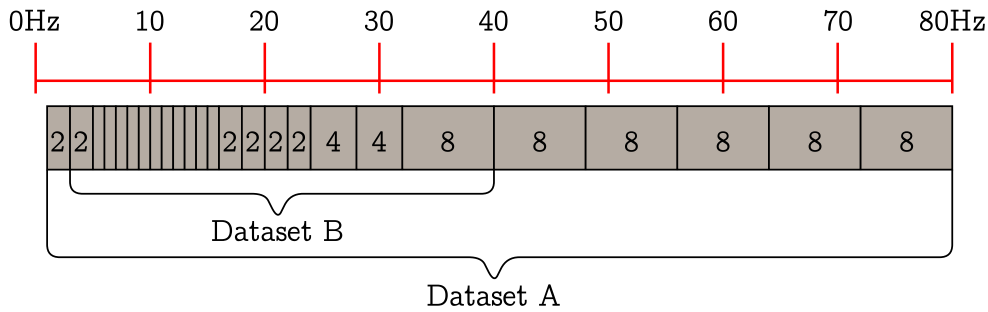


Figure 14. Diagrammatic overview of the frequency bands. The grey boxes correspond to the frequency bands that were used as features, and the number within them corresponds to the width of the frequency band. Non numbered bands have a width of 1 Hz

4.3 Classification

For classification, Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) algorithms were implemented. During the initial exploration of models, the RF algorithm outperformed the other models and was therefore chosen as the method for further development within the boundaries of this study.

As described in subsection 2.4.4, RF has several advantages, including low bias and variance, reasonably fast computation time, and good performance on small datasets. The RF was implemented using the *scikit-learn* package in python [54], which includes versatile tools for machine learning algorithms, including RF, and performance analytics.

One of the main goals of this research was to look at the possibilities to utilize ME data when training an MI-based BCI. Apart from using ME and MI data separately as training data, in one case, the

Training Data	Test Data
ME	MI
MI	MI
ME + MI	MI
ME -> MI	MI
ME	ME

Table 2. Training and test data configurations. ME + MI denotes training on a dataset that consists of all available ME and MI data together. ME -> MI denotes training on ME data first, then retraining on MI data. The set-up depicted in the lowest row was included for benchmarking as this set-up is expected to yield high classification performance.

classifier was trained on ME and MI data together, while in another case the classifier was trained on ME data first and then retrained on MI data. Table 2 provides an overview of the various configurations of training and test data that were used.

Because of the splitsampling technique that was employed to increase dataset size, extra measures had to be taken to avoid ensure proper generalization. Typically, in classification problems similar to this, a test set is drawn at random from the dataset. In the case of splitsampling, this would imply a probability for testing on brain signals related to the same physical movement as for brain signals in the training set, however, sampled in parallel. Instead, the test data was chosen as all the trials from one specified run, this ensured that no trial would have splitsampled versions in both the training and test set.

4.4 Model Testing

To assess the different configurations of which data is used for training and testing (MI, MI, or a combination) and find the parameters that would produce the most optimal classifier, a search was conducted, training classifiers using a multitude of combinations of configurations and parameters. In the following list, the variations are described:

- Training data – Either ME, MI, ME + MI, or ME -> MI, as shown in the first column of Table 2
- Test data – Either ME or MI, as shown in the second column of Table 2
- Number of classes – Classify all 5 classes or classify only foot movement versus hand gesture (Figure 4).
- Which subjects – All 3 subjects, one individual subject, the two best performing subjects, train on the best or the two best subjects and test on the worst subject.
- Test run – Which run to use for testing, ranging from the first run to the eighth and last run that was performed (for either MI or ME). This run was not included in the training data.
- Number of trees – How many trees were trained in the random forest. Ranging from 5 to 85 in steps of 5
- Number of additional trees for retraining – In the case of ME -> MI training data, 10, 20, 30, or 40 trees were added for retraining with MI, after the initial training with ME

4.4.1 Performance Metrics

To properly assess the performance of the classifier, different performance metrics were used. The first and most basic one was accuracy, which is simply the number of instances that were classified correctly divided by the total number of instances. The accuracy is not a very good performance metric as it will not consider the expected accuracy by a random classifier, and it will suffer when the number of instances differs a lot between the various classes. However, accuracy is far more intuitive than other, more informative performance metrics. Therefore, accuracy is sometimes included in the presented results to help understand what the performance would equate to in a live system.

Of the more informative, and robust performance metrics, the Cohen's Kappa metric was used in this study [55]. Although this metric has been criticized by some [56], Cohen's kappa has been widely used as a performance metric in similar classification problems, for example in BCI Competition IV, dataset 2a and 2b [43]. Kappa is a measure of accuracy with random classification excluded, and it is calculated like this:

$$Kappa = \frac{(observed\ accuracy - expected\ accuracy)}{(1 - expected\ accuracy)}$$

Here, the expected accuracy is the accuracy one would expect if using a random classifier.

4.4.2 Cross-Validation

To properly assess the performance of the models, leave-one-out cross-validation was employed. This means that for every model, eight separate trainings were performed, each leaving one run out of the training data, which was used as testing data. This is the same parameter listed as test run in section 4.4. The mean performance for these eight trainings is the cross-validated kappa, later also referred to as mean kappa. A large surprisingly variation in kappa within these eight trainings were discovered, therefore, the maximum kappa was also reported (this is thoroughly motivated in section 5.3).

Chapter 5

Results and Discussion

5.1 Summary of Data Structure

Data collection is described in detail in Chapter 3 and was done with three subjects. Subjects 2 and 3 generated eight ME and eight MI runs, while subject 1 generated eight ME runs and one MI run (i.e., the data collection was terminated early by request from the subject). Each run comprises 35 trials, corresponding to 3 seconds of active execution/imagining of a gesture, yielding a total of 840 ME trials and 596 MI trials to be used for training and testing. Sixty-four electrodes were used to record the data, of which, between one and eight had to be discarded due to the technical issues described in subsection 4.1.1.

Further processing of the data, namely noise removal and downsampling, is described in detail in section 4.1 and resulted in the creation of two variants of the dataset, referred to as dataset A (lowpass filter at 80 Hz) and dataset B (bandpass filter at 3 Hz/45 Hz). The splitsampling method described in subsection 4.1.3, effectively multiplied the amount of data available for training and testing, by a 12 or 22 for dataset A and dataset B, respectively.

During the initial assessment of the results, no obvious performance variation was found between models trained on the two different datasets. Therefore, the difference between using dataset A and dataset B was not further investigated, and all results presented, are from models trained with dataset B, as dataset B is in compliance with the accepted best practice [51].

Feature extraction (section 4.2) was set to extract 3-second-long chunks of data, corresponding to one trial, and convert the temporal information to frequency information that was then to be used for classification. As explained in subsection 4.2.2, we ended up with 8000 (dataset A) or 6080 (dataset B) features for each trial (5 stages * 64 channels * 25 or 19 frequency bands).

5.2 Summary of Model Development and Testing

Several thousands of models were set up, trained, and tested, using every possible combination of the parameters listed in section 4.4. The type of training and testing data (MI, ME, or a combination of the two), full classification of all classes or only foot versus hand, all subjects, or individual subjects, varying the run used for testing, the number of trees that were trained in the random forest, and the number of additional trees for retraining (in the case for ME->MI training data). In all cases, the run, which was used as testing data, was explicitly kept separate from the training data.

5.3 Initial Assessment of the Results

Upon a first look at the results, one significant and unexpected feature was noted: The performance of otherwise equal classifiers depends heavily on which run that was used for testing the trained model. For example, models achieving a cross-validated kappa score of 0.6 could have maximum and minimum kappa scores of 0.9 and 0.0, respectively. It should be pointed out that the runs used for testing were not included in the training data, as outlined in subsection 4.4.2. Generally, the standard deviations for kappa obtained when testing the same model on different runs were between 0.15 and 0.30, which is considered rather high. Effects as large as those observed were not expected and may suggest that the models do not perform reliably.

There are possible explanations for the relatively high sensitivity of kappa for the choice of testing data. The data collection (Chapter 3) is a wearing procedure for the subjects. Both MI and ME were generally perceived as quite tedious, and especially MI was expressed by the subjects to be very mentally demanding. In both situations, intense concentration is demanded from the subjects. The total time of concentrated attention equated to a maximum of $2 \times 23{:}20\, minutes = 46{:}40\, minutes$, which might have been too much to ensure adequate focus for the full duration of the procedure.

When recording data, comments given by the subjects in between runs indicated that there could be seemingly random variations in how successful a run was. When asked about why a run felt unsuccessful, it was suggested by the subjects that a failed, clumsy, or uncomfortable trial at the beginning of a run made the subjects lose their concentration. Such variability corresponds well with the seemingly random distribution of performance obtained when varying which run is being used for testing of one classifier.

Improvements could likely be obtained by revising the data collection procedures. For example, while it is useful to have MI and ME data from the same subject, achieving this prolonged the sessions per subject, likely reducing the level of concentration maintained by the subject.

Considering the plausible explanations for the relatively large kappa value standard deviations, there is no reason to assume that the models were generally flawed. In the discussion of the outcomes below, if nothing else is specified, the kappa performance metrics will be reported as the cross-validated kappa, hereinafter referred to as mean kappa, with maximum kappa in parenthesis, which is the kappa value for the most successful choice of test run, for that specific model. The reason for including the maximum is that this could be perceived as the potential for the model. A live BCI system would require high levels of concentration, and this will equate to removing or reducing the number of runs that were deemed unsuccessful by the subjects, which would drive the average performance up towards the maximum.

5.4 Motor Execution (ME) Classification

The results for various combinations of the three subjects and the number of classes were found by varying the number of trees in the random forest algorithm, as described in section 4.4. The performance of the best models is shown in Table 3. The number of trees in the random forest for each of the models were between 10 or 15 for the single-subject models on the two-class problem, and between 45 and 85 for all other models, which results are presented in Table 3. This indicates that more complex datasets, both in the sense of multiple subjects and multiple, less differentiable classes, require more trees in the random forest, to converge to the best possible performance. This corresponds well to previous studies on the number of trees in random forests [57], as such, the number of trees will not be discussed any further.

	2 Classes Kappa		5 Classes Kappa	
Subject	Mean	Max	Mean	Max
1	0.57	1.00	0.23	0.40
2	0.21	0.67	0.22	0.30
3	0.74	1.00	0.24	0.31
1, 3	0.63	0.97	0.22	0.31
1, 2, 3	0.36	0.60	0.21	0.26

Table 3. Results for the best classifiers that were trained and tested on ME data. Subject numbers separated by comma refers to a model trained and tested on data from both/all subjects.

5.4.1 Inter-Subject Difference in Performance

For the two-class problem, the best performance scores were 0.57 (1.00) for subject 1, 0.21 (0.67) for subject 2, and 0.74 (1.00) for subject 3. Thus, there is considerable variation in performance between subjects. BCI illiteracy is mainly considered in MI-based applications, however, it has been shown that BCI illiteracy exists for other paradigms as well [58], [59]. The results indicate that subject 2 might be BCI illiterate. Of note, all subjects gave similar results when using five classes but in this case, the performance scores were generally low.

5.4.2 Two-Class Versus Five-Class

The results for the five-class problem including all hand gestures for the best subject are not impressive, with a kappa score of 0.24 (0.31), and an accuracy of 0.38. For the two-class problem (foot movement versus hand movement), the scores were considerably better (Table 3) and in that case, the highest kappa score was 0.74 (1.00) corresponding to an accuracy of 0.92. It is thus clear that either the models were not good enough, or the dataset not big enough to succeed in solving the five-class problem. Of note, for the putatively BCI-illiterate subject 2, there is no significant difference in the score for the two-class and five-class problems.

5.4.3 Multi-Subject Models

If the classifier would be invariant to subject-specific features, the performance of a multi-subject model would be expected to be equivalent to the mean performance of models based on individual subjects. For subjects 1 and 3, the average score is 0.66 (1.00), and the score for the combined model is 0.63 (0.97), which could be considered equivalent. However, if we include subject 2, the average score is 0.51 (0.89), and the score for the combined model is 0.36 (0.60), which implies a considerably greater difference.

Considering the deviating results for subject 2, leading to the suggestion that subject 2 is BCI illiterate, it is not surprising that the combined model based on all three subjects did not perform well. This underpins the importance of recognizing such subjects when dealing with datasets, not only because a model trained on the data from the illiterate subject performs poorly, but also because such data weakens the results of a multi-subject model.

5.5 Motor Imagery (MI) Classification

5.5.1 Models Trained on Motor Imagery (MI) data

The straightforward way to do MI classification is to train a model on MI data, and this will be the benchmark for comparisons with models incorporating ME data in training (subsection 5.5.2). The best-performing MI-based models are presented in Table 4. Note that

	2 Classes Kappa		5 Classes Kappa	
Subject	Mean	Max	Mean	Max
2	0.06	0.28	0.08	0.21
3	0.68	0.94	0.24	0.36

Table 4. Results for the best classifiers that were trained and tested on MI data. Subject 1 is not included because this subject was not able to perform a sufficient number of MI runs.

subject 1 is not included as it was not able to perform more than one run of MI. The best model for subject 3 has a kappa score of 0.68 (0.94) for solving the two-class problem, which is reasonably good. On the other hand, the best model for subject 2 has a score of 0.06 (0.28), which is very bad. Combining these observations with those described for ME classification (section 5.4) supports the idea that subject 2 is BCI illiterate.

5.5.2 Introducing Motor Execution (ME) Data to Improve Performance

To investigate how ME data would affect the training of a MI classifier, three variations were implemented. One model was trained on ME data, one was trained on ME and MI data together, and one was trained on ME data, and then retrained on MI data. The results are presented in Table 5, and to clarify, the rows with MI as training data correspond to the two rows of Table 4.

Subject	Train Data	2 Classes Kappa Mean	Max	5 Classes Kappa Mean	Max
2	MI	0.06	0.28	0.08	0.21
	ME	0.10	0.30	0.05	0.23
	ME + MI	0.07	0.15	0.11	0.23
	ME -> MI	0.11	0.23	0.10	0.19
3	MI	0.68	0.94	0.24	0.36
	ME	0.56	0.81	0.17	0.34
	ME + MI	0.62	0.91	0.30	0.36
	ME -> MI	0.65	0.91	0.24	0.39

Table 5. Results of the best MI classifiers, trained on varying combinations of data. MI and ME corresponds to training solely on MI or ME data, respectively. ME + MI corresponds to using a training data set that is a combination of ME and MI data. ME -> MI corresponds to initial training solely with ME, followed by continued training with MI. Subject 1 is not included because this subject was not able to perform a sufficient number of MI runs.

The total number of trees in the random forest that gave the best results for the models that are presented in Table 5 varied greatly, from 5 to 85 trees. Findings by Oshiro et al. [57] suggest that the convergence of the number of trees (i.e. the point at which the model will not improve significantly by increasing the number of trees) is expected to be between 64 and 128. This might indicate that the dataset is too small and that improved results could be achieved by increasing the size of the dataset.

When analyzing the results in Table 5 for subject 3, there does not seem to be any significant difference in performance between the models trained solely on MI data and models trained combinations of MI and ME. This could indicate a potential for including ME in the training data for an MI-based BCI, without negatively impacting performance. This could have a positive influence on data collection, because a portion of the mentally demanding MI process could be exchanged for less demanding ME, making subjects able to collect more data. It should be noted though that models that were trained on a combination of MI and ME data were trained on more data compared to the models trained on only MI or ME data since ME data were added to the MI data rather than replacing parts of these.

It is worth noting that training of the MI classifier with only ME data gave a reasonable result, with a kappa score of 0.56 (0.81). While this value is lower than the value obtained with all the other set-ups, all including MI data in the training set, it is still reasonable. This adds to the idea that indeed, it should be possible to utilize ME data to train MI classifiers

It should be noted that for subject 2, which is considered BCI illiterate, the ME-data-including variations of the classifier all performed marginally better, while for subject 3, they performed marginally worse. Although this observation is not statistically significant, it is tempting to interpret this observation as a slight indication in support of what was one of the goals of this research, namely the use of ME data in training, to improve MI classification for BCI illiterate subjects.

Chapter 6

Summary and Conclusion

This thesis presents research aimed at providing insight into BCI development, with a special focus on BCI illiteracy. One specifical goal was to explore the possibility of incorporating ME into the training data for a model that aims to classify MI data. A novel dataset is presented that consists of both ME and MI data for a combination of hand gestures and one foot movement. While the dataset is of limited size, for reasons described in subsection 3.2.6, it proved useful for exploring the issues that we set out to explore. This being said, clear limitations were also met, such as the low performance in solving the five-class problem. The quality and quantity of the dataset may have something to do with this, although further improvements may also come from improving methods and exploring other strategies, as outlined in Chapter 7.

The large variation in the number of trees that made up the best random forest classifiers for the different data set configurations, subjects, and the number of classes, further indicates that more data could increase performance towards a point of convergence.

ME data tend to be overlooked a bit in BCI development for health care applications because BCIs naturally are most needed to help individuals lacking motor abilities. Thus, a goal was to assess the value of ME data in developing MI-based classifiers, the premise being that ME data should not be overlooked in BCI, even in applications where the end goal is an MI classifier. The results show that models trained on a combination of ME and MI data generally performed at a similar level as purely MI-based models, with a strong subject achieving marginally worse performance and a BCI illiterate subject achieving marginally better performance.

The fact that similar results could be obtained upon replacing MI data with ME data when developing an MI classifier, is encouraging since this could reduce the dependency on MI data in BCI development. Collection of ME data is easier because it puts less strain on the subject compared to the collection of MI data. Furthermore, BCI illiterate subjects are usually better at producing useful ME data, compared to MI data, also suggested by observations in this study. Thus, increased use of ME data means that more subjects can contribute to BCI development. Most importantly, the possibility of using ME data to train MI-based classifiers opens up the possibility to use ME data generated by a motorically competent person to be used in developing BCIs for disabled persons. Inter-subject models are of great interest and should be the focus of future work (see section 7.1).

Importantly, the present study is too limited to reach firm conclusions and point decisively at future directions. For example, the very interesting observation that for a BCI illiterate subject, including ME in the training data could even improve the performance of an MI classifier is based on differences that are not statistically significant. Further work is needed to back up the observed trends, which may, when investigated with larger datasets, turn out not to be trends at all. Regardless, it is encouraging that the data collected in this study do not dismiss the possibility of increased use of ME data.

Chapter 7

Future Work

As summarized in Chapter 6, the results presented in this thesis provide slight indications as to if, and how, ME data can help tackle the BCI illiteracy problem. On the way towards developing successful BCIs, in particular, BCIs that work well for BCI illiterate subjects, much more work and larger datasets are needed. The sections below describe some lines of research that could be explored in future work

7.1 Further Investigation of ME-based MI classifications

Clearly, the results and preliminary conclusions presented in this thesis need to be backed up by additional work. By increasing the size of the dataset, it should be possible to establish the statistical foundation needed to either verify or reject the preliminary conclusions that were presented.

Another natural next step would be to investigate inter-subject models, that is, models that are tested and used on a subject that did not contribute with training data. This type of approach is crucial for increasing the possibility of successful development of BCIs for BCI-illiterate subjects, including motorically disabled subjects. Alternatively, one could produce a base model from multiple subjects, and employ transfer learning (subsection 2.4.6) to fine-tune the model for the proposed user of the system. This would make the implementation of a BCI system for healthcare purposes much less extensive, as the use of an already trained base model would reduce the training time needed for the patient in question.

7.2 Visual response vs motor response

Both Motor Imagery and Motor Execution rely on performing an imaginary or physical movement. Importantly, for the dataset presented in this work, there is no obvious way to separate between brain signals related to the subject seeing and understanding the visual stimuli, and the actual MI/ME. It would be interesting to investigate the effect this has on the performance of the model. One simple way to test this is to set up a similar data collection procedure as in this work, but with an added text as part of the visual stimuli, that for example prompts the user to "imagine" or "ignore". Results from such an experiment could potentially be used to exclude certain features related to the visual stimuli from the training data, to ensure that the BCI model uses only brain signals related to the actual movement or the imagining of the movement. Another possible approach to remove the effects of the visual stimuli on the brain signals is having the stimuli be removed and instruct the subjects to start the execution or imagining of the movement after the stimuli is no longer visible.

7.3 Executing movement versus maintained position

In the dataset presented in this work, subjects were to perform and maintain a gesture for 3 seconds. It would be interesting to look closer at which parts of these 3 seconds that are the most important for classification, and if it is even possible to classify a maintained position when the first 1-2 seconds is left out. For a live BCI system as proposed in this thesis, it would not be particularly interesting to use anything other than the first second of a movement, or preferably less. However, investigating this could generate insight into what kind of brain signals can be classified. It would also be interesting to look at the difference between MI and ME, as MI requires continuous concentration to maintain a position even for very simple gestures, while ME does not.

7.4 Reinforcement Learning

Established BCI techniques are implemented as “communication devices”, that communicate discrete actions to a computer, rather than interpreting the natural signals from the brain, translating these into a continuous space of fine-tuned motor control or communication. BCIs, as solved today, will never contain enough classes of hand movements to enable a guitarist or artist with an amputated hand be able to fully recover their original fine motor skills. In other words, it is currently not possible for a BCI user to invent a ‘new’ hand gesture, even with state-of-the-art hand control BCIs, the user is limited to a specified set of gestures.

So far regression models, which are inherently non-discrete, have been largely overlooked in BCI research. Regression models are not restricted by a discrete solution space of classes and have an infinite number of solutions in continuous space. This would equate to being able to make any combination of physically possible movements and rotations in all joints of a robot hand, with precision equal to that of the robot hand actuators. The implementation of regression models for BCIs has been limited, because brain signals related directly to the set of specific muscle contractions that comprise a movement are incredibly hard, if even possible, to discover.

One approach for moving in the direction of BCIs with a less discrete solution space (i.e., a system that can be trained, by the subject, to perform new movements) could be to implement a reinforcement learning model. In such an approach, the subject simultaneously trains the control of his/her brain waves and the way the BCI handles these brain waves. To obtain this, the subject should be asked to complete a series of tasks that entailing various movements of a robot hand. Equipped with a keyboard, the subject should give feedback (i.e., an integer value between 1 and 10) that corresponds to how well he/she felt the robot hand movement responded to their intentions for each task. With the feedback from the subject as reinforcement rather, the reinforcement learning algorithm can effectively learn any new movement while in use, much like a child learns by trial and reinforcement by parents.

References

[1] P. Sawangjai, S. Hompoonsup, P. Leelaarporn, S. Kongwudhikunakorn, and T. Wilaiprasitporn, "Consumer Grade EEG Measuring Sensors as Research Tools: A Review," *IEEE Sens. J.*, vol. 20, no. 8, pp. 3996–4024, Apr. 2020, doi: 10.1109/JSEN.2019.2962874.

[2] C. Vidaurre and B. Blankertz, "Towards a Cure for BCI Illiteracy," *Brain Topogr.*, vol. 23, no. 2, pp. 194–198, Jun. 2010, doi: 10.1007/s10548-009-0121-6.

[3] R. Bauer, M. Fels, M. Vukelić, U. Ziemann, and A. Gharabaghi, "Bridging the gap between motor imagery and motor execution with a brain–robot interface," *NeuroImage*, vol. 108, pp. 319–327, Mar. 2015, doi: 10.1016/j.neuroimage.2014.12.026.

[4] E. Raffin, P. Giraux, and K. T. Reilly, "The moving phantom: Motor execution or motor imagery?," *Cortex*, vol. 48, no. 6, pp. 746–757, Jun. 2012, doi: 10.1016/j.cortex.2011.02.003.

[5] "Definition of invasive procedure - NCI Dictionary of Cancer Terms - National Cancer Institute," Feb. 02, 2011. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/invasive-procedure (accessed Sep. 28, 2020).

[6] L. R. Hochberg *et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, May 2012, doi: 10.1038/nature11076.

[7] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "'Thought' – control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neurosci. Lett.*, vol. 351, no. 1, pp. 33–36, Nov. 2003, doi: 10.1016/S0304-3940(03)00947-9.

[8] R. Caton, "The Electric Currents of the Brain," *Am. J. EEG Technol.*, vol. 10, no. 1, pp. 12–14, Aug. 1875, doi: 10.1080/00029238.1970.11080764.

[9] H. Berger, "Über das Elektrenkephalogramm des Menschen," *Arch. Für Psychiatr. Nervenkrankh.*, vol. 87, no. 1, pp. 527–570, Dec. 1929, doi: 10.1007/BF01797193.

[10] E. D. Adrian and B. H. C. Matthews, "THE BERGER RHYTHM: POTENTIAL CHANGES FROM THE OCCIPITAL

LOBES IN MAN," *Brain*, vol. 57, no. 4, pp. 355–385, Dec. 1934, doi: 10.1093/brain/57.4.355.

[11] Thomas F Collura, "History and Evolution of Electroencephalographic Instruments and Techniques," 1993.

[12] O. N. Markand, "Alpha Rhythms," *J. Clin. Neurophysiol.*, vol. 7, no. 2, pp. 163–190, Apr. 1990.

[13] E Garcia-Rill, "Chapter 8 - Waking and the Reticular Activating System in Health and Disease," 2015. Accessed: Oct. 09, 2020. [Online]. Available: https://reader.elsevier.com/reader/sd/pii/B9780128013854000082?token=5F8747AD3A8AAEDCBB3968B9EA8EA07D2CFF6D0E435B8964BF55EFCB1F61FD2581CC67FD84B91E976EA2F745D6957C9B

[14] J. D. Kropotov, "Chapter 3 - Beta Rhythms," in *Quantitative EEG, Event-Related Potentials and Neurotherapy*, J. D. Kropotov, Ed. San Diego: Academic Press, 2009, pp. 59–76. Accessed: Oct. 13, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780123745125000037

[15] T. Mima, N. Simpkins, T. Oluwatimilehin, and M. Hallett, "Force level modulates human cortical oscillatory activities," *Neurosci. Lett.*, vol. 275, no. 2, pp. 77–80, Nov. 1999, doi: 10.1016/S0304-3940(99)00734-X.

[16] B. D. Mensh, J. Werfel, and H. S. Seung, "BCI competition 2003-data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1052–1056, Jun. 2004, doi: 10.1109/TBME.2004.827081.

[17] A. Kübler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw, and N. Birbaumer, "Brain–computer communication: Unlocking the locked in.," *Psychol. Bull.*, vol. 127, no. 3, p. 356, May 2001, doi: 10.1037/0033-2909.127.3.358.

[18] P. M. D. Gloor, G. Ball, and N. M. D. Schaul, "Brain lesions that produce delta waves in the EEG," *Neurology*, vol. 27, no. 4, pp. 326–333, Apr. 1977.

[19] Lara V Marcuse, Madeline C Fields, and Jiyeoun Yoo, "2 - The normal adult EEG | Elsevier Enhanced Reader," 2015. Accessed: Oct. 21, 2020. [Online]. Available:

https://reader.elsevier.com/reader/sd/pii/B9780323353878000020?token=A3816AE08319B86B23BEC708779AAE3D144D780C2E21329B789836FFC5CEE490E4A15A1C19377BE0A2ED9FBF99A9448C

[20] F. Beverina, G. Palmas, S. Silvoni, F. Piccione, S. Giove, and T. Agrate, "User adaptive BCIs:," Dec. 2003.

[21] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002, doi: 10.1016/S1388-2457(02)00057-3.

[22] N. Birbaumer, T. Elbert, A. G. Canavan, and B. Rockstroh, "Slow potentials of the cerebral cortex and behavior.," *Physiol. Rev.*, vol. 70, no. 1, pp. 1–41, Jan. 1990, doi: 10.1152/physrev.1990.70.1.1.

[23] A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. P. Birbaumer, "Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication," *Arch. Phys. Med. Rehabil.*, vol. 82, no. 11, pp. 1533–1539, Nov. 2001, doi: 10.1053/apmr.2001.26621.

[24] D. H. R. Blackwood and W. J. Muir, "Cognitive Brain Potentials and their Application," *Br. J. Psychiatry*, vol. 157, no. S9, pp. 96–101, Dec. 1990, doi: 10.1192/S0007125000291897.

[25] S. Sur and V. K. Sinha, "Event-related potential: An overview," *Ind. Psychiatry J.*, vol. 18, no. 1, pp. 70–73, 2009, doi: 10.4103/0972-6748.57865.

[26] S. Sutton, M. Braren, J. Zubin, and E. R. John, "Evoked-Potential Correlates of Stimulus Uncertainty," *Science*, vol. 150, no. 3700, pp. 1187–1188, Nov. 1965, doi: 10.1126/science.150.3700.1187.

[27] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 42, no. 6, pp. 817–826, Jun. 1977, doi: 10.1016/0013-4694(77)90235-8.

[28] G Pfurtscheller, "Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest," Dec. 1991, Accessed: Oct. 09, 2020. [Online]. Available: https://reader.elsevier.com/reader/sd/pii/0013469492901333?token=08463C22237384C564254C247843A355D07B3CB685000C1791FCEE9E65BD5B11BC036EAB4BFA5D273C577F24FEEA3F9B

[29] H. Jasper and W. Penfield, "Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus," *Arch. Für Psychiatr. Nervenkrankh.*, vol. 183, no. 1, pp. 163–174, Jan. 1949, doi: 10.1007/BF01062488.
[30] G. E. Chatrian, M. C. Petersen, and J. A. Lazarte, "The blocking of the rolandic wicket rhythm and some central changes related to movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 11, no. 3, pp. 497–510, Aug. 1959, doi: 10.1016/0013-4694(59)90048-3.
[31] G. Pfurtscheller and C. Neuper, "Future prospects of ERD/ERS in the context of brain–computer interface (BCI) developments," in *Progress in Brain Research*, vol. 159, C. Neuper and W. Klimesch, Eds. Elsevier, 2006, pp. 433–437. Accessed: May 29, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079612306590284
[32] "A Step-by-Step Tutorial for a Motor Imagery-Based BCI," *ResearchGate*, Jan. 2018, Accessed: Oct. 08, 2020. [Online]. Available: https://www.researchgate.net/publication/323780357_A_Step-by-Step_Tutorial_for_a_Motor_Imagery-Based_BCI
[33] R. Zhang, F. Li, T. Zhang, D. Yao, and P. Xu, "Subject inefficiency phenomenon of motor imagery brain-computer interface: Influence factors and potential solutions," *Brain Sci. Adv.*, vol. 6, no. 3, pp. 224–241, Sep. 2020, doi: 10.26599/BSA.2020.9050021.
[34] W. M. Leach, "Fundamentals of low-noise analog circuit design," *Proc. IEEE*, vol. 82, no. 10, pp. 1515–1538, Oct. 1994, doi: 10.1109/5.326411.
[35] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative Analysis of Spectral Approaches to Feature Extraction for EEG-Based Motor Imagery Classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 4, pp. 317–326, Aug. 2008, doi: 10.1109/TNSRE.2008.926694.
[36] J. B. J. baron Fourier, *The Analytical Theory of Heat*. The University Press, 1878.
[37] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter Bank Common Spatial Pattern Algorithm on BCI Competition IV Datasets 2a and 2b," *Front. Neurosci.*, vol. 6, 2012, doi: 10.3389/fnins.2012.00039.

[38] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI)," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 447–456, Dec. 2000, doi: 10.1109/86.895947.
[39] J. Gan, "Self-adapting BCI Based on Unsupervised Learning," Jan. 2006.
[40] W. S. Noble, "What is a support vector machine?," *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006, doi: 10.1038/nbt1206-1565.
[41] N. Isa, A. Amir, M. Ilyas, and M. Razalli, "The Performance Analysis of K-Nearest Neighbors (K-NN) Algorithm for Motor Imagery Classification Based on EEG Signal," *MATEC Web Conf.*, vol. 140, p. 01024, Jan. 2017, doi: 10.1051/matecconf/201714001024.
[42] P. Sajda, A. Gerson, K.- Muller, B. Blankertz, and L. Parra, "A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 184–185, Jun. 2003, doi: 10.1109/TNSRE.2003.814453.
[43] "BCI Competitions." http://www.bbci.de/competition/ (accessed Apr. 13, 2021).
[44] C. Brunner *et al.*, "BNCI Horizon 2020 – Towards a Roadmap for Brain/Neural Computer Interaction," in *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access*, Cham, 2014, pp. 475–486. doi: 10.1007/978-3-319-07437-5_45.
[45] "Biosemi EEG ECG EMG BSPM NEURO amplifier electrodes." https://www.biosemi.com/ (accessed Apr. 17, 2021).
[46] C. Elger G. Klem, H. Lüders, and H. Jasper, "The ten-twenty electrode system of the international federation," *Guidel. Int. Fed. Clin. Physiol.*, 1958.
[47] "LabStreamingLayer's Documentation — Labstreaminglayer 1.13 documentation." https://labstreaminglayer.readthedocs.io/index.html (accessed Apr. 19, 2021).
[48] "Home — PsychoPy v3.0." https://www.psychopy.org/ (accessed Apr. 19, 2021).

[49] C. Kothe, *pyxdf: Python library for importing XDF (Extensible Data Format)*. Accessed: Apr. 19, 2021. [Online]. Available: https://github.com/xdf-modules/xdf-Python

[50] M. Siegel and P. König, "A Functional Gamma-Band Defined by Stimulus-Dependent Synchronization in Area 18 of Awake Behaving Cats," *J. Neurosci.*, vol. 23, no. 10, pp. 4251–4260, May 2003, doi: 10.1523/JNEUROSCI.23-10-04251.2003.

[51] A. Ramos-Murguialday and N. Birbaumer, "Brain oscillatory signatures of motor tasks," *J. Neurophysiol.*, vol. 113, no. 10, pp. 3663–3682, Mar. 2015, doi: 10.1152/jn.00467.2013.

[52] "SciPy.org — SciPy.org." https://www.scipy.org/ (accessed May 09, 2021).

[53] A. Frydenlund and F. Rudzicz, "Emotional Affect Estimation Using Video and EEG Data in Deep Neural Networks," in *Advances in Artificial Intelligence*, vol. 9091, D. Barbosa and E. Milios, Eds. Cham: Springer International Publishing, 2015, pp. 273–280. doi: 10.1007/978-3-319-18356-5_24.

[54] "scikit-learn: machine learning in Python — scikit-learn 0.24.2 documentation." https://scikit-learn.org/stable/ (accessed May 15, 2021).

[55] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960, doi: 10.1177/001316446002000104.

[56] R. Delgado and X.-A. Tibau, "Why Cohen's Kappa should be avoided as performance measure in classification," *PLOS ONE*, vol. 14, no. 9, p. e0222916, Sep. 2019, doi: 10.1371/journal.pone.0222916.

[57] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How Many Trees in a Random Forest?," in *Machine Learning and Data Mining in Pattern Recognition*, Berlin, Heidelberg, 2012, pp. 154–168. doi: 10.1007/978-3-642-31537-4_13.

[58] B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser, "BCI Demographics: How Many (and What Kinds of) People Can Use an SSVEP BCI?," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 107–116, Apr. 2010, doi: 10.1109/TNSRE.2009.2039495.

[59] C. Guger *et al.*, "How many people are able to control a P300-based brain–computer interface (BCI)?," *Neurosci. Lett.*, vol. 462, no. 1, pp. 94–98, Sep. 2009, doi: 10.1016/j.neulet.2009.06.045.