Single-Trial Analyses of Novel Cognitive Tasks in an EEG-based Brain-Computer Interface Framework

Nathan Evans

EPFL :: Laboratory of Cognitive Neuroscience
Advisers: Dr. Lars Schwabe, Dr. Olaf Blanke

June 20, 2008

Contents

Chapter 1

Introduction

Over the past few decades parallel research in the domains of neuroimaging, neurobiology, signal processing, machine learning, psychology, and cognitive neuroscience have led to an exciting and promising practical nexus: the *brain-computer interface* (BCI). BCIs use advanced signal processing and machine learning techniques combined with a contemporary neuroanatomical and temporal understanding of brain processing in order to extract user intent from neurological data. More specifically, BCIs form a bi-directional communication channel between humans and machines in which users are able to relay commands in the form of brain activity patterns to software that recognizes and translates their intent into a corresponding action. In turn, humans can recognize the outcome of the translation of their intentions in the form of feedback and can hence modulate future efforts. This cyclical process is referred to as a *closed-loop BCI* (see Figure 1.1).

Since BCIs can offer the expression of ones intentions through control that does not depend on the normal outputs of the neuromuscular system, one of the primary motivations for BCI development has been to aid those with severe motor disabilities. With this common goal in mind a number of research groups have spent considerable time constructing a full-fledged BCI framework. While working toward this common end, it has become evident that BCI research is intrinsically multi-disciplinary and that each of the involved sub-disciplines can benefit from the collection of methodologies brought forth in BCI research. This thesis takes inspiration from BCI research and its methodologies by first implementing a typical BCI system and testing its effectiveness by confirming previous results on a standard mental motor imagery task. Next, using the same BCI framework and types of analyses, electroencephalogram (EEG) data from a novel set of cognitive imagery tasks are explored.

The higher-level organization of this thesis is as follows. We begin by introducing the goals, basic notions and overall context of the current work. Following the contextualization of this project, the basics of brain-computer interfaces are described and the current work is defined by the approaches selected in the different sub-components of the BCI taxonomy. We then move to the theoretical background underlying the methods implemented in the two experiments that were a part of this work. The theory of EEG signal preprocessing, feature extraction, and data classification is outlined. After building a theoretical foundation, the two experiments are described and their

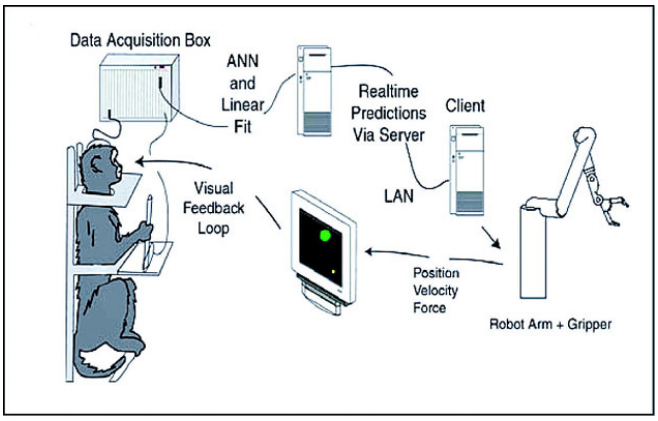

Figure 1.1: An example of the bi-directional communication established in a closed-loop BCI. Here, the monkey relays its intentions in the form of brain signals which are processed, interpreted, and translated into an action by a robotic arm. The monkey can then observe the robotic arm motion as a form of visual feedback to further aid future control. [**?**]

results are presented. Finally, a general and experiment-specific examination summarizes the results and future directions of this research.

This introductory chapter seeks to contextualize this thesis in light of past, related research in addition to motivating the need for the current work. First, the three primary goals of the project will be explicitly defined and spelled out. Next, a detailed literature review covering related research in the direction of each of these three primary goals will be given. This exploration will relationally highlight the specific location of the current work while providing further evidence to the need of such work.

1.1 Project Aim & Motivation

This project is largely motivated from prior success in motor imagery-based BCIs. Generally speaking, the primary goal in BCIs is to improve the speed and accuracy of the communication channel formed between ones brain and a computer. One interesting and alternative use of BCIs involves sidestepping this original goal of previous BCI systems by merely adopting their standards of experimental design, data evaluation, and interpretation of results. This alternative strategy moves beyond using BCIs in their original context and considers the BCI as a general research framework to be used to explore problems in other related domains. For instance, one can apply these methods from the standard BCI toolkit to other types of neuroscientific research. The advantages are bi-directional in that utilization of the BCI methodologies can help validate and discover neuroscientific findings while simultaneously advancing BCI research by augmenting its traditional paradigms with nonstandard techniques. Such an approach has been employed in this thesis, whose primary goals are two-fold: first, to explore the use of novel cognitive tasks as controls for an electroencephalogram (EEG)-based BCI and second, to apply the standard BCI

methodologies to explore and verify observations in the domain of cognitive neuroscience. The specific goals and steps taken to realize this work are outlined below.

Project Goals

1. **Develop a BCI framework**
   At the core of every BCI program is an elected set of particular technologies, experimental designs, algorithms, and data analysis techniques. As mentioned above, the selection of these widely-varied tools depends on the particular implementation and varies amongst BCI groups. Upon consideration of the form of our laboratory's previous experimental data as well as the future plans to extend our toolkit to new experimental paradigms, it was decided to develop an in-house BCI framework (see Section B).

2. **Verify the BCI framework by applying a standard BCI paradigm**
   Prior to performing novel data analyses within a framework one must confirm that the system works on pre-established paradigms. The most obvious way to validate a BCI system is to perform a typical BCI experiment and to ensure that the results confirm those found in the literature. Accordingly, a typical motor imagery-based BCI experiment (see Section 5) was carried out and results were evaluated against previous findings.

3. **Explore the use of novel cognitive tasks within an established BCI framework**
   Drawing on inspiration from findings in cognitive neuroscience and the standard motor imagery BCI paradigms, a new experimental design was created and tested (see Section 6). The usage of new cognitive tasks in BCI control serves as a useful research direction for BCIs and furthermore, the BCI methodology of data analysis can be used to further validate findings and provide new insights in neuroscience.

1.2 Related Research: A Literature Review

It has been nearly 20 years since such BCI systems first successfully managed to decode specific user intentions by exploiting certain characteristics of signals generated by a user's initiation of either overt (explicit) or covert (imagined) movements [39, 24]. These initial and promising results have since inflated the BCI research community from a couple of programs to a couple dozen programs. Several succinct topical reviews highlight the developments of BCI systems and the current state-of-art systems [36, 37, 38]. These reviews and individual publications indicate that though most research groups share the same end goal of improving the accuracy and speed of this brain-based communication channel, each differs in both research orientation and in the applied methods for their in-home BCI systems. For instance, some programs focus on benefits for the disabled [15], others on the signal processing or machine learning algorithms useful in such a context [18, 21], some on alternative BCI approaches [20], and a number of programs concentrate on end-user applications of BCI [3, 7]. The interdisciplinary nature of BCIs has led many signal processing and representation techniques, machine learning algorithms, and experimental setups

to draw inspiration from neuroscientific knowledge. On the other hand, neuroscience has done little to adopt the standard tools used in BCI to further strengthen the evidence of brain function.

1.2.1 BCI System Implementation

Due to inter-lab differences and the propensity to guard ones own implementation, there is a noteworthy lack of publicly-available, non-commercial BCI systems. Some open BCI frameworks and EEG analysis packages do exist [10, 33], but they are neither universally used nor readily adaptable to the interests and needs of particular research groups. Moreover, one must always balance the amount of time it would take to use and adapt a pre-existing system to tailored needs against the time it takes to build a new system from scratch. Unfortunately the latter option often wins, leading to further software disarray between research groups and a need to "re-invent the wheel" with respect to certain core BCI components. On the other hand, the techniques employed in different groups' research is openly published and can thus be recreated [6, 15, 20, 26, 34]. While choosing a specific approach to follow one must consider the technical constraints (*e.g.* choice of neuroimaging device and experimental environment), the end goal of the data analyses (*e.g.* artificial prosthetic control vs. neuroscientific research), and the target population (*e.g.* clinical patients or healthy subjects). After reviewing the different approaches from the most prominent BCI groups, it was decided to start with a fresh implementation closely mirroring the sensorimotor rhythm, EEG-based approach used by Graz University of Technology [29]. This choice was primarily motivated by the Graz system's many years of development, its traditional, well-published methods and its marked success.

1.2.2 Sensorimotor rhythm-based BCIs

The selected BCI approach includes an implicit decision of the experimental design, the types of signals, and signal features one extracts for BCI control (see Section 2.2 and Section 3.2). The Graz BCI is one of many groups that use sensorimotor rhythms (SMR) to control their BCI [2, 5, 14]. There are several reasons that SMRs are widely used as the control signal in a BCI, the first of which is that these signals are available at the non-invasive level of EEG. Next, SMRs have been shown to desynchronize with imagined movement or preparation for a movement making them useful in the context of mental imagery both for healthy subjects and those unable to perform overt motor movements [25]. Moreover, SMRs are rather quickly learned by the subject for control of a BCI whereas other control signals often take much more subject training [16]. Said differently, the overall time until seeing successful results with a naïve subject is reduced. Finally, with the advent of more advanced signal processing techniques, SMRs have been shown to be ubiquitously present in human adults [22]. The success and evidence presented in these studies influenced the design and construction of the base BCI system implemented in thesis as well as the first experiment (see Appendix Band Section 5).

1.2.3 "Cognitive" Tasks in EEG and BCI research

The vast majority of SMR-based BCIs use motor imagery-based tasks, such as imagining to move ones hand, foot, or tongue. This precedent is mainly due to the well-documented fact that imag-

ined motor movements activate similar and overlapping supplementary motor areas as their overt counterparts [27]. Said differently, if one imagines movement of a hand, similar and overlapping neural ensembles become activated as if one were physically initiating hand movement. Accordingly, with motor imagery one can expect certain predictable neural behavior which, in turn, can ideally yield a cleanly localized representation of brain state activity during a given task.

It should be noted that motor imagery is technically a type of cognitive task, and the relationship between cognitive tasks their effect on EEG is still scientifically tenuous. It is currently unknown exactly how much one can rely on the correlation between the changes in an EEG signal and the individual cognitive process that supposedly cause them. Moreover, factors such as attention, memory, concentration, cognitive load, emotional state, and gender can dramatically affect the EEG signals in cognitive task-based experiments[8]. A further difficulty is unraveled with evidence pointing toward a correlation between particular features of unique cognitive tasks [1]. The brain seems to perform cognitive tasks in a highly distributed, and overlapping fashion. One key finding is the observed coherency in γ bands during performance of cognitive tasks [11]. This finding helps pinpoint potentially important frequency features of the EEG signal while performing cognitive tasks.

The implications of these studies provide strong motivation for further in-depth study of cognitive tasks in EEG experiments and in particular, for use in driving a BCI. To date, however, there is a notable lack of exploration into alternative cognitive tasks extending beyond imagination of motor movement. Exploration of alternative cognitive tasks may bear critical neuroscientific insight and perhaps prove to better drive BCIs as tasks can be tailored to a subject's imaginative strengths and weaknesses. There are, however, a couple of studies who move beyond motor imagery and consider other more cognitively-oriented tasks in the context of BCI. One such task has BCI users internally generating words beginning with the same letter [13]. Another compared the two tasks of imagined spatial navigation in a familiar setting and the internalized replaying of a familiar auditory tune against traditional motor imagery tasks [9]. Finally, a relaxation task (used as a baseline) and an arithmetic task where the subject successively subtracts seven from three-digit numbers have been used to control an online BCI [23].

These cognitively influenced studies are borne in same spirit as the present work, but differ in their scope and content. The general idea behind previous work has been to explore new cognitive tasks well-suited for distinguishing between intentional states, with the underlying intention to control a BCI in the most simple manner possible. This common research theme can be summarized as such:

> Identifying pairs of cognitive tasks which produced distinct EEG characteristics at the same recording site would have practical benefits. A suitable pair could then be selected for a subject and only single channel recordings would be necessary [8].

This principle captures only half of the motivation for investigating the two novel cognitive tasks of inner speech and own-body transformations found in the present work. The present work additionally concerns itself with the task of verification of neuroscientific findings. That is, finding an optimal set of cognitive tasks is not the sole goal. Additional data analyses will be performed across 64 channels to show that the underlying cognitive processes as seen from the analysis tools found in a BCI framework align with previous neuroscientific evidence.

1.2.4 Neuroscientific Verification of Cognitive Phenomena

As mentioned above, one of the principal intentions of this thesis is to explore a novel set of cognitive tasks which are selected for their neuroscientific evidence and feasibility (see Chapter 6 for associated experiment). The two tasks explored in this work, namely that of inner speech and own-body transformations have been localized in previous studies and could benefit from confirmation coming from other analytical angles. Inner speech has been shown to activate left-hemispheric regions associated to speech perception and speech motor control (*i.e.* the supplementary motor area) [32]. The act of imagining ones own body in a physical space external to the observer, also known as an own-body transformation, have shown to activate temporal-parietal junction [4].

The usage of BCIs in neuroscientific research is of strong potential merit as its methods can complement traditional neuroscientific research methods. This hybrid tactic is investigated using a fMRI-based BCI [35]. The authors point to the benefits of the unification of two analytical approaches: the manipulation of a neural substrate with behavior as the dependent variable and manipulation of behavior with the neural function as the dependent variable.

Recapitulation

This chapter developed the goals and context of the current project. More specifically, three main goals were laid out. First, a brain-computer interface framework will be developed. Second, a standard motor imagery-based BCI experiment will be conducted to verify the correctness of the BCI framework. Finally, using an already-verified BCI framework, an experiment using novel cognitive tasks will be explored in order to judge their use for BCI control and to further investigate these cognitive phenomena from a neuroscientific perspective. After providing the three goals for this project, a detailed literature review discussed the previous research in all of the related domains. This previous research helps contextualize the current work while providing crucial *apriori* knowledge that will be implicitly used throughout the remainder of this thesis.

Chapter 2

The Fundamentals of Brain-Computer Interfaces

Before delving into the details of the theory behind the individual components and tools used in the present BCI framework, it is essential to speak about BCIs at a higher, more abstract level. In this brief chapter we will explore the different constituent subsystems that all BCIs are made of and describe the overall flow of data in a BCI framework. Furthermore, we will categorize different BCI systems by distinguishing amongst the typical ways the BCI subcomponents are realized. Having established a BCI taxonomy, we can move forth in describing the current platform and experiments in the larger context of BCIs and by using the terms introduced in this brief overview of BCIs.

2.1 BCI Components

All brain-computer interfaces can be abstracted to reveal a set of subcomponent processes and common tools which implement these subcomponents. A BCI framework is also an integrated, abstract paradigm in which experimental design, certain analytical practices and technological systems work together. Below the common subsystems found in a typical BCI are provided with a short descriptor of their functional role. A birds-eye view in terms of the flow of information in a BCI is provided in Figure 2.1.

- **Acquisition System (Neuroimaging)**
  Brain data is captured using some neuroimaging modality. The type of hardware that one uses to collect data affects the experimental design, types of control signals that can be extracted, and often times dictates who the end users of the system are (healthy or a targeted clinical group).
- **Signal Preprocessing**
  Every neuroimaging modality has its technological and practical flaws. The data coming

from the brain is often times influenced by external noise and artifacts that must be first cleaned before continuing with analysis. The techniques used in the signal preprocessing stage are also dependent on the eventual type of feature classification one wishes to do, as certain classification algorithms work better with data preprocessed in specific ways.

- **Feature Extraction**
  Moving from a raw, preprocessed signal into a compact representation to describe the signal is one of the most critical steps in the information flow of a BCI. The features to be extracted from the signal depend on the acquisition system, experimental paradigm, and preprocessing techniques. The success of the eventual classification of the signals depends on the this step.

- **Feature Classification**
  Classification in a BCI involves the attempt to categorize the incoming feature representation of a signal into one of the pre-established types of signals. That is to say, after training itself on a number of examples where it can view the relationship between signal "type" (class) and the feature representation, the classification module will predict the label of a new, previously unseen feature representations of the signal. Classification success hinges on how separable the examples are in each class (output) and hence on the feature extraction method.

- **Application Output**
  After producing a best-guess for the class of what the incoming signal is, the output of the classifier is translated into output to some external application. This application can be a piece of software (e.g. a spelling application) or hardware (e.g. a robotic prosthesis).

- **Feedback**
  The final stage of information flow involves feedback to the BCI user of the application output. By observing if the system correctly translated his or her intentions, the BCI user can learn to regulate future signals accordingly.

2.2 BCI Taxonomy

Each BCI is defined by the way in which its subcomponents are implemented. To facilitate the use of terminology throughout this thesis, this section introduces the basic categories of BCIs while explicitly noting where the current work fits.

- Recording Technique (Invasive / Non-invasive)
  Neuroimaging techniques can be invasive or non-invasive. Common non-invasive recording techniques are electroencephalograms (EEG), magnetoencephalography (MEG), blood-oxygen level dependent response (BOLD), and near infrared spectroscopy (NIRS). Invasive techniques involve abrasion of the scalp, skull and dura and commonly include electrocorticograms (ECoG) and intercranial electrodes. The present work uses a non-invasive EEG-based acquisition system.

- General Approach (Biofeedback / Machine Learning)
  The general BCI approach refers to whether the burden of learning and adaptation lies on

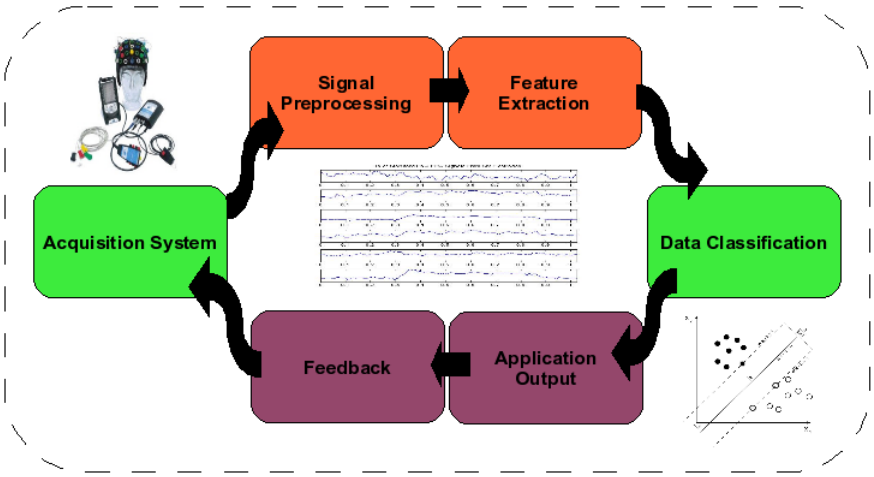

Figure 2.1: The basic information flow in a brain-computer interface as viewed from its various high-level constituents. [BCI acquisition system picture from Guger Technologies OEG]

the subject or on the machine learning algorithm. Using sophisticated machine learning techniques, this project follows the machine learning approach.

- Experimental Strategy (Focused Attention / Motor Imagery / Operant Conditioning / Cognitive)
  The experimental strategy is directly integrated with the type of control signal one wishes to obtain (see below). Successful BCI experimental designs include experiments in which the subject either focuses on an object, focuses on imagining the initiation of a motor task or a cognitive task. The current work focuses uses two experimental strategies, namely motor imagery and a set of cognitive tasks.

- Control Signal From Brain (Slow cortical potentials (SCP) / Sensorimotor Rhythms (SMR) / Evoked Potentials / Error Potentials / Etc)
  The signal features that one uses of the acquired data to control the BCI depend heavily on the experimental strategy and on the operative mode (see below). Several control signals have been used for control in BCIs and their details are outside of the scope of this project. One of the most common is the sensorimotor rhythms which are ubiquitously found in human adults and are exploited in the present work.

- Operative Mode (Synchronous / Asynchronous)
  The manner in which the BCI operates is also closely tied with the experimental strategy. Asynchronous BCIs require the computer to prompt the user with a cue signal periods where commands can be interpreted. Synchronous BCIs are always waiting to interpret commands and do so by first attempting to understand when a user wishes to send a command. The BCI implemented in this thesis is asynchronous.

- Feedback (Continuous / Discrete / Modality)

All online BCIs must give feedback to the user so that they can regulate their signals. The present work is an *offline* BCI, that is, feedback is disabled. This essentially means that the communication channel is one-directional.

- End Users (Healthy Users / Clinical Target Groups)
  The end user is closely tied with the recording modality (invasive vs noninvasive) as well as the applications. The present study wishes to look at healthy EEG users and uses the BCI framework as a neuroscientific research tool.

Recapitulation

We have considered the BCI as a framework under which information flows in a cyclic manner. A BCI is defined by the summation of the types of approaches used in its subcomponents. The BCI framework implemented in the present work is explicitly defined and placed into context with respect to other BCI systems. After reading this chapter the reader is prepared to better understand the upcoming chapters as their relative importance is now put into the frame of reference of an entire BCI framework. Furthermore, this chapter has introduced the key terminology used throughout this paper.

Chapter 3

EEG Signal Preprocessing and Feature Representation

The data in the present work are recorded noninvasively by measuring the surface electrical activity of the brain via extra-cranial electroencephalography (EEG) (see Figure 3). It is worth briefly mentioning the characteristics of an EEG signal, as most data processing algorithms are inspired by the biophysical features of the signal. The flow of ions in and out of neuronal dendrites creates a compensatory extracellular current captured as a potential difference (voltage) in EEG recordings. Since this electrical activity is constantly changing and modern EEG equipment allows one to sample at a very high rate, EEG has an excellent temporal resolution.

Despite understanding the source of the signals, one of the strongest disadvantages to using EEG as a neuroimaging technique is that extra-cranial EEG signals portray neither the precise the origin of the signal nor the individual neuronal activity at this source. Signals instead represent the aggregate synchronous activity of tens of thousands of underlying neurons sharing the same radial spatial orientation. Additionally, surface EEG electrodes sense voltages that tend to originate from near the surface of the cortex, as voltage fields dissipate from deeper brain sources by the time they reach the skull. In addition to the previous shortcomings, anatomical factors such as the intervening tissue and the skull act as a signal attenuators further muddying the spatial resolution of the oscillations. The poor spatial resolution of EEG calls for data preprocessing methods which can spatially filter the data and enhance the resolution and increase the signal-to-noise ratio.

After applying signal preprocessing operations to the raw EEG signal, one can take one step further by transforming the signal into an alternate representation. With the end goal being to classify the EEG signal, the manner in which the signal is represented to the classifier is critical. The process by which one represents the signal as a conglomeration of certain features is herein referred to as *feature extraction*. Feature extraction can be performed manually by taking hints from the underlying characteristics found in EEG signals while performing a given task. The subsequent section on feature extraction discusses the various manual signal representations used in the present work. Alternatively, feature extraction can be performed automatically with *automated feature selection*, a process that naïvely searches for "important" features as defined by some criteria.

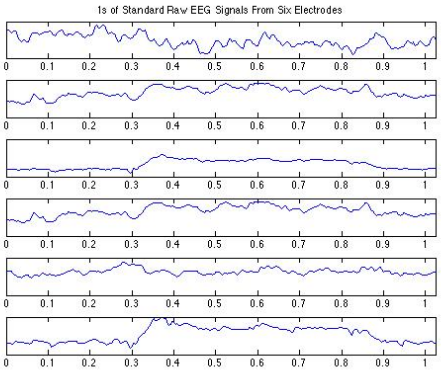

Figure 3.1: *Left:* A musician dramatically demonstrates a surface EEG cap of 30-active electrodes for use in a brain-controlled music performance. [*Quintephone brainwave regen poster colour* by Quintist Chris Aimone - 2007]. *Right:* Six typical EEG signals for a 1 second period recorded at 256Hz.

3.1 Raw Data Cleaning and Preprocessing

EEG data is additionally susceptible to external electrical noise coming from the recording environment and the body. The key sources of EEG artifacts are electrooculographic (EOG) signals coming from the blinking of the eyes, electromyographic (EMG) signals coming from the flexion of muscle tissue such as the neck, and electrocardiographic (ECG) signals from the heart. Electrical outlets and appliances (such as lights) are the primary cause for the environmental electrical noise found in EEG signals. It is thus imperative to first "clean" the data of artifacts via manual or automated artifact detection. Though the description of automated artifact detection is beyond the scope of this work, the data cleaning methods used in the present experiments can be found in subsequent chapters (see Chapter 5and 6). Once the data is cleaned of artifacts, both temporal and spatial EEG preprocessing operations can be performed on the raw signal.

For notational purposes we can consider the raw EEG matrix to be a two-dimensional raw EEG matrix of M channels (*i.e.* electrodes) and N features (*i.e.* time points). Let S_C be the vector representing a single channel of length N in the raw EEG matrix. Conversely, let S_F be the vector of length M representing a single feature (time point) across all channels. It is often useful in stimulus-locked experiments to make a clear distinction between the pre-stimulus S_C^B and post-stimulus S_C^A period for each channel C.

3.1.1 Temporal Preprocessing Techniques

Temporal preprocessing methods operate across the time (feature) dimension of the raw EEG matrix.

- **Baseline normalization**

The idea behind baseline normalization is to take the average voltage for each channel during a baseline (S_B) period, and to subtract this per-channel average from each time point in the corresponding channel. This has the effect of normalizing each channel in the raw signal to the mean value during a controlled period. Thus, for each channel C :

$$S_C^{BASE} = S_C - \mu(S_C^B) \tag{3.1}$$

- **Zero-mean normalization**
  A zero-mean normalization subtracts the mean of the signal across all time points for each channel from each corresponding channel and then divides the result by the standard deviation across the time points. This has the effect of changing each feature in the signal to have a mean of 0 and unit representing the number of standard deviations away from the mean. That is, for each channel C:
  $$S_C^{ZERO} = \frac{S_C - \mu(S_C)}{\sigma(S_C)} \tag{3.2}$$

- **Unit variance normalization**
  This normalization scheme rescales each feature for each channel to unit variance by dividing each feature by the square root of the variance across all features. This type of normalization is useful if the data is eventually submitted to a neural network classification, but can be deleterious for other learning algorithms. Mathematically speaking, for each channel C one computes:
  $$S_C^{UNITVAR} = \frac{S_C}{\sqrt{\sigma(S_C)}} \tag{3.3}$$

3.1.2 Spatial Preprocessing Techniques

Since the channels (electrodes) have a particular spatial relationship with one another on the scalp, spatial preprocessing involves operations performing operations across the channel dimension of the raw EEG matrix. The following four methods are the most common spatial preprocessing techniques have been compared in the context of BCI use [19]. It should also be noted that the method of Common Spatial Patterns (CSP) is often used but left out of the present discussion [12]. A graphical illustration of the involved electrodes for each of these spatial filters can be seen in Figure 3.1.2.

- **Ear reference**
  The most traditional way to remove noise from electrodes is to attach an electrode to some neutral electrical ground (such as the ear) and refer all other channels to this neutral signal. The main principle is that if there is environmental or body-generated electrical noise captured at the ears, this noise will reach the EEG electrodes. Subtracting each channel by this reference would then ideally remove the environmental noise. Let S_F^{REF}be the reference signal for feature F. The ear reference is computed for each feature F as follows:
  $$S_F^{EAR} = S_F - S_F^{REF} \tag{3.4}$$

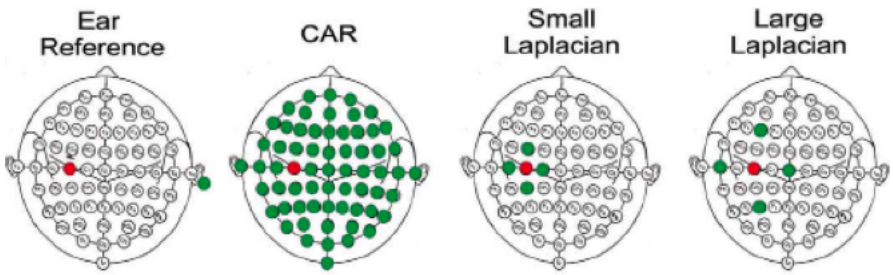

Figure 3.2: The involved electrodes in four different spatial filters. The red electrode represents the electrode at which the filter is applied and the set of green electrodes are used as references according to the methods described in Section 3.1.2 [19].

- **Common Average Reference (CAR)**
  The CAR functions according to a similar principle as ear referencing. In order to subtract common environmental noise, the mean value for a given feature across all channels in the montage computed and subtracted from the feature values of each channel. More formally, for each feature F we compute:

$$S_F^{CAR} = S_F - 1/C \sum_{j=1}^{C} S_F^j \tag{3.5}$$

- **Small/Large Laplacian**
  As mentioned above, electrical signals captured by EEG flow radially from their source. The principle behind a surface Laplacian filter is that it would be ideal to capture the electrical activity directly beneath a given electrode and to filter out activity under adjacent electrodes. Computation of the standard surface Laplacian is expensive as it involves interpolation of splines and a second-order spatial derivative computation at each electrode. In the present work a spatial Laplacian filter is instead approximated by the discrete *finite-difference method*, an often-used approximation that involves simply computes and subtracts the mean of a defined neighborhood of electrodes from each electrode of interest. The neighborhood function for each electrode can be defined in any arbitrary way, though this work follows the literature by defining a *Small Laplacian* as the four nearest-surrounding neighbors (about 3cm from the electrode in question). The *Large Laplacian* requires the same computations only differing in neighborhood definition: the next-nearest neighbors (about 6cm from the electrode in question). Note that the neighbors of each electrode are unique. Taking a neighborhood of four electrodes for each channel D_C and individual neighbors j , a Laplacian filter is applied as:

$$S_C^{LAP} = S_C - 1/4 \sum_{j \in D_C}^{4} S_C^j \tag{3.6}$$

3.2 Feature Extraction

With the overall goal of classifying EEG data in mind, we wish to construct a representation of the EEG signal that maximizes the discriminability of each type of task in the experimental design. Ideally this means molding the set of clean, preprocessed, raw EEG signals into some unique representation with respect to a certain elicited brain state. Though EEG is recorded continuously during an experiment, there exist short periods of interest known as a *trial* during which the experimenter asserts that the subject is in a particular brain state. Trials are generally a collection of short EEG signals that, in the current work, are one second in length (see Chapters 5and6). It is worth noting the amount of flexibility built into a representational system; these trials can be represented in any arbitrary fashion. For instance, the timeseries information can instead be seen as the combination of a number of sub-features computed by one or many methods with different parameters such as particular EEG electrodes or selected frequency bands.

In general, BCI has seen greater success with signals represented in the frequency domain, as many tasks have been shown to have distinctively associated spectral features. All of the feature extraction methods presented below are based on first transforming the EEG signals from the time domain to the frequency domain. Since frequency domain analysis and its applications is a large field, only those techniques particularly pertinent to the present work are discussed below.

- **Spectrograms and Spectral Power Estimation**
  One first simple alternative signal representation is to transform the timeseries data for a trial into the frequency domain via a *short-term Fourier transform (STFT)* across various frequency bands using a sliding window and then computing the power spectral density (PSD) under the sliding window. The result is referred to a *spectrogram*, where one information for the intensity (power) for each frequency in the signal across time. The spectrogram is computed as the absolute magnitude of the result of the STFT:

$$spectrogram(t, \omega) = |STFT(t, \omega)|^2 \tag{3.7}$$

  There are several major disadvantages to performing a discrete frequency domain transformation using a STFT within the context of BCIs who are generally characterized by short trials. First, the STFT generates certain edge effects as the first result comes from the middle of the first sliding window, thus intensity data for frequencies beginning in time periods before the middle of the first sliding window are lost. If the intensities of importance lie within the beginning of the signal, this is unacceptable. One potential workaround is to take extra data around the trial. However, since trials tend to begin directly following stimulus presentation, one would be computing the spectrogram for a mix of pre and post-stimulus signals, which is often undesirable. Second, to avoid aliasing problems and to capture higher frequencies, the window size of the sliding window in the STFT must be sufficiently long. This requirement is difficult to meet as EEG trials can be short. Finally, the computation of the power in a series of sliding windows can be computationally expensive. This additional computation is often unacceptable for real-time BCI systems. For these reasons,

  BCI systems often estimate the spectral power in trials via power density estimation. One estimation scheme is the *maximum entropy (all-poles) method* that uses an auto-regressive

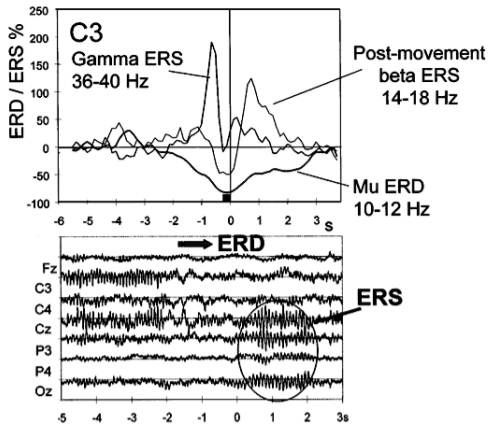

Figure 3.3: Top: Band power as a function of time for three frequency ranges (10-12Hz, 14-18Hz, and 36-40Hz) displaying an electrode (C3) contralateral to right finger movement. Note that the μ ERD peaks around stimulus onset of *t=0*.
Bottom: ERD can be clearly seen in the muting of the signals Fz, C3, C4 and Cz on and around the initiation of motor imagery. Moreover, the post-movement β synchronization following imagined motor movement is clear in parietal electrodes.

model (AR) to compute coefficients of the spectral power [31]. The advantage to spectral estimation is that an estimate can be provided even for short sequences with no data loss as in the sliding window STFT. On the other hand, the underlying auto-regressive model requires finding a proper value for the model order hyperparameter in order to provide a more precise estimate. This hyperparameter is eventually selected automatically in the SINGLETRIAL MATLAB TOOLBOX (see Section 4.1).

- **Event-related Synchronization/Desynchronization (ERS/ERD)**
  Sensorimotor rhythms manifest themselves over somatosensory cortices and generally consist of an α-rhythm centered around 8-12Hz (referred to as the μ-rhythm), a secondary β-rhythm component around 16-20Hz, and a γ-component around 36-40Hz. When preparing or performing a motor movement, an increasing number of neural populations synchronize causing the slowing of oscillations in particular neural ensembles. This phenomenon is referred to as an *event-related desynchronization* (ERD) and is most noticeable in the μ-rhythms 3.2. ERDs can be exploited as a feature representation for BCI use by first computing the PSD for pre-and post-stimulus data. Then, one defines a pre-stimulus baseline period and computes the percent change in the PSD across the stimulus. More formally, consider that one already has the power for a pre-stimulus reference (baseline) period $P_R$and the power for the after-stimulus period P_A. The percent in ERD change is then computed as:

$$ERD\% = 100\left(\frac{P_A - P_R}{P_R}\right) \tag{3.8}$$

- **Event-related Spectral Perturbation (ERSP)**
  The downside to ERD analysis is that it only captures very narrow selective frequency bands. The *event-related spectral perturbation* (ERSP) generalizes this by showing the mean frequency-by-latency across an experimental event. ERSPs are commonly used as an EEG data analysis tool to observe how strongly event-related power increases or decreases with respect to a baseline power. It can be computed in a straightforward manner from the timeseries by taking the mean of the power during a reference period P_R and the mean power during the after-stimulus period P_A. Then, these mean PSDs are converted into the decibel domain (below noted as the $db()$ operator) to regularize the eventual logarithmic computation:
  $$ERSP = log\left(\frac{db(\mu_{P_A})}{db(\mu_{P_R})}\right) \tag{3.9}$$

- **Other BCI Features**
  The above-discussed features are all possible feature representations based on the experimental paradigm used in this project. Namely, having both well-defined stimuli (from which pre and post-stimulus periods can be defined), and motor-imagery based task make these feature representations natural candidates. For clarity, it should be mentioned that BCI groups control their BCIs and select their experimental design according to a number of other paradigms such as *slow cortical potentials* (SCP), *readiness potentials* (Bereitschaftspotential), *steady-state visual evoked potentials* (SSVEP), and the so-called *P300* "abnormal-event" signal. The explanation of these signal features is outside of the scope of the present work but can be found in any basic review of BCIs.

3.3 Automated Feature Selection

Feature extraction as described above requires one to embed outside knowledge of the underlying neurophysiological characteristics (e.g. for deciding particular EEG electrodes) or task-related signal characteristics (e.g. for selection of frequency bands or deciding to use ERD). An alternative approach is to assume nothing about the signal and to naïve way to select especially distinctive features. Automated feature selection is used to this end and in the context of BCIs is used on two types of features: EEG electrode selection and frequency band selection. Using automated algorithms that highlight those channels and important frequencies in classifying the data is beneficial in two key ways. First, it provides one with a way to reduce the dimensionality of the data by electing to classify the data only on the most important features. Second, one can verify that the selected electrodes spatially correspond with the expected neurophysiological areas for a given task and that the important frequencies align with prior frequency-domain EEG research. This second advantage is a way in which neuroscientific research can be doubly verified using a common BCI tool.

- **R^2 Feature Selection**
  Feature selection can be seen as the task of giving each feature a score based on a scoring metric and then to select features whose scores are highest. One such statistical, correlative scoring metric is the R^2 criterion, whose value represents the proportion of variance of the

EEG voltages that is accounted for by each task (*e.g.* left-hand or right-hand motor imagery). The R^2 value is a useful metric for feature selection as it reflects the signal-to-noise ratio. The present work, however, favors more sophisticated feature selection techniques described below.

- **Distinction Sensitive Learning Vector Quantization (DSLVQ)**
  This feature selection method has been used to select both optimal channel and frequency bands in a set of motor imagery-based BCI experiments [30]. In short, DSLVQ operates by using a modified version of the LVQ algorithm in which a weight vector assigns a value to each dimension while the learning of the network takes place. Those dimensions aiding in the correct classification of new samples are assigned higher weights whereas those contributing to misclassification are penalized with a lower weight. After learning takes place one has a weight vector for each input dimension (feature) and can use these to grade how important each feature was in the discrimination of two tasks. Though The Singletrial Matlab Toolbox contains tools to compute and view features selected by DSLVQ, this type of analysis is not used in the experimental results of this thesis.

- **Joint Classification & Feature Selection Methods**
  There are a number of algorithms which have been shown to do joint EEG classification and automated feature selection [21]. While classifying EEG data these techniques maintain a high-dimensional weight vector representing a decision hyperplane in multidimensional space. The weights in this weight vector are directly correlated with the importance of their corresponding dimension, as in DSLVQ. That is, by taking the absolute value of the feature's respective weight in this weight vector, one can obtain a direct measure of the importance of that feature. This can again be done for both electrode selection and frequency band selection (see Section 4.3 for the theory and Chapter 5for experimental results showing how these can be used). This is the method of automated feature selection used in the current work for all experimental analyses.

Recapitulation

This chapter provides a concise summary of the ways of manipulating raw EEG signals in order to prep them for classification algorithms. We began by going through the raw preprocessing techniques which involve processing the EEG data over the time dimension or the spatial dimension (across the channels). We next described how to extract sensorimotor and frequency features out of the timeseries data in order to more precisely represent a period of EEG data for classification. Finally, we contrasted the differences between feature extraction and automated feature selection by providing a number of methods which are used to find channel and frequency features automatically out of EEG data. The theory presented in this chapter forms a backbone for all of the data analysis techniques used in the two conducted experiments.

Chapter 4

EEG Data Classification

One of the primary aims of a BCI is to translate the EEG signals and assign them to a state; in other words, to classify the EEG data. This amounts to finding underlying patterns within the brain activity as represented by a preprocessed, feature-extracted representation of the original electrical signals. This chapter is dedicated to the process of data classification and represents the subcomponent of the BCI that is a pattern classification machine. We begin by briefly describing the staple methods used in data classification in the BCI context and catalog the fundamental operations taking place when classifying data during the two experiments found in this work. After conveying the basic principles, the theory is lain for the two classification methods analyzed and compared throughout the rest of this thesis. This chapter is not a primer on machine learning and pattern classification, as these wide-ranging topics exceed the boundaries of the present work. Furthermore, only a fraction of the subdomain of BCI machine learning is touched as only those techniques used in the present work are discussed (see implementation information in Appendix B). Comprehensive reviews exist which cover the ideas from this chapter in much greater depth and breadth [17, 21].

4.1 The General Principle and Methods

We consider the problem of two-class, supervised classification in BCIs. More specifically, every data point has an associated label out of a set of two labels. In the greater context of the BCI process, a data sample will at this point be a feature-extracted representation of the raw, preprocessed EEG signals coming from a time period where the state of the brain (*i.e.* the label) was controlled. At the classification level the data has been abstracted from its source; as far as the classifier is concerned, the data can represent anything – it is just a point in a high-dimensional space. The task of the classifier is to build a model by training itself on training examples. By observing the label of a new training example, the classifier assimilates into its model the fact that data of this label can "look" like the new training example. With more and more training examples the model becomes intricate and progressively more robust. The model begins to capture the underlying patterns in like-labeled data (a class) and finds a way to separate data from this class from those

belonging to another class. Each particular classifier has a unique criterion by which it chooses how to make its class-assigning and model-updating actions based on the presentation of a new sample.

Once the model is trained on a set of training examples, the classifier is ready to be put to the test. To test for classification performance a new data point is given to the classifier who, without looking at its associated class label, determines a class label for the sample given its internal model and its decision criterion. This proposed label is then compared against the true label. Classification performance is then gaged as the number of correct assignments out of the total number of assignments the classifier makes.

There are a number of practices which are critical while doing machine learning to avoid "cheating". The first common practice is to randomly sample the data which is fed to a classifier. If the classifier were to always see the same data, it could learn to perfectly classify without properly learning the underlying patterns. A similar problem happens if a classifier's model is trained with too few data. This problem is referred to as *overfitting* as the classifier learns to perfectly classify the training data but cannot generalize well to new, unknown data. Other than having enough data and randomly sampling, one should designate and set aside a training and a test set. While training the model for a classifier, the classifier may never assimilate data into its model from the test set – these are forbidden for training. There are a number of other simple mistakes that one can make to "cheat" in machine learning but listing all of these is out of the scope of this thesis.

Particular Concepts

It is particularly instructive to define a few brief classification algorithm concepts in order to later define the classification algorithms used in this thesis.

- *Generative vs. Discriminative*
  The classification approach can be either to learn how to discriminate the class membership for direct classification of a new feature vector (discriminative), or to learn how to compute the likelihood that a new feature vector belongs to each class (generative). Both the linear fisher discriminant (LFD) and sparse linear programming machine (SLPM) are discriminative classifiers. This means that these two learning algorithms use training samples to build a multidimensional, separating hyperplane which, when given a new feature vector, class membership can be immediately computed.

- *Sparse classifiers*
  Most discriminative classifiers use a high-dimensional vector known as the *weight vector* to describe the separation between classes. Sparse classifiers are variants upon general classification algorithms which favor sparse weight vectors, namely they return weight vectors where most entries are zero. This can be useful in determining the important directions (dimensions) for classification and for automated feature selection, in the case of BCIs.

- *Cross-validation and Nested Cross Validation*
  The ideas presented above about using a separated training and test set provides a classification performance for a given random split of the data. Since this random split could,

by chance, be "lucky" or "unlucky" in the sense that the data used for training the classifier may be unrepresentative of the population of the data as a whole. To help further generalize classification results a technique known as *cross validation* is used. Cross validation comes in a number of flavors, all using the same principle that the data is to be resampled a number of times and the average classification performance across these different samplings should be instead used. This creates a more generalized estimate of the classification performance on unseen data.

Three common variants of cross validation exist and are used in the current work. The first is *k-fold cross validation*. The idea is that one divides the data into k clusters. Training of the classifier is performed for data from k-1 of the clusters and the final cluster is used for testing. This process represents a single fold. Then, the cluster which was used as testing for the first fold becomes part of the training set for the second fold and in its place a different cluster from the k-1 clusters becomes the test data. Each fold thus builds a model on k-1 data clusters and tests the model on one data cluster. This is repeated k times and the average is taken to be the generalized classification performance. A similar, more extreme version of k-fold cross validation is *leave-one-out (LOO) cross validation*. LOO splits the entire data set into one cluster per data point. At this extreme a model is trained for all data except for one point. Then the classifier is tested on this one data point. The data point left out for testing rotates such that all data points are at some point both a part of the training set and are the test point. The average over the number of classifications performed gives the generalized classification error.

One final variant of cross validation is the so-called *nested cross validation*. Nested cross validation is used when there are hyperparameters that must be estimated for the data set. It is unfair to learn hyperparameters based on all data and then to split the data into a training and test set. Instead, the idea behind nested cross validation is that you embed a cross validation for model selection inside of the outer cross validation scheme. More specifically, the inner cross validation performs cross validation on just the training set for the outer cross validation. Finding the best parameters given the outcome of the inner cross validation on the training set allows one to set the parameters at the outer level and compute the classification performance. This is a "fair practice" algorithm that is often times overlooked when performing parameter estimation in cross validation schemes.

4.2 Linear Fisher Discriminant

The linear fisher discriminant (LFD) is the first of the two classifiers used in this work. The idea behind it is to maximize the difference between the means of the two classes while simultaneously minimizing the variance within each class. Let s_1^2be the variance for class one and s_2^2 be the variance for class two. Moreover, we define μ_1 as the average for class one and μ_2 as the average for class two. The linear fisher discriminant seeks to find the maximum value of the Fisher criterion for all linear projections of w where the Fisher criterion is defined as:

$$J(w) = \frac{|\mu_1 - \mu_2|^2}{s_1^2 + s_2^2} \tag{4.1}$$

The result is a linear projection w who defines a hyperplane in multi-dimensional space dividing the two classes. To obtain a class label one simply projects the data onto the hyperplane and takes the $sign()$of the output as the class label.

4.3 Sparse Linear Programming Machine (SLPM)

The Sparse Linear Programming machine is a sparse variant of a Support Vector Machine (SVM). SVMs operate on the principle that a high-dimensional hyperplane can exist that perfectly separate the data. The goal of a SVM then is to maximize the margin, *i.e.* the distance from the separating hyperplane to the nearest point in each class. The precise theory of SVMs is outside of the scope of this work and can be found in any pattern classification textbook. It suffices for our purposes to say that SVMs can be formed as a mathematical optimization problem where one is looking for a projection w such that:

$$\min \frac{1}{2}||w||^2 \quad s.t. \quad y_i(w^T x_i + b) \geq 1$$

Where y_i is the class label of [-1, 1] for sample x_i and b is a bias (or a simple offset for the hyperplane). The Sparse variant of the SVM simply uses the 1-Norm instead of the 2-Norm in this mathematical optimization. By switching to a 1-Norm the SLPM favors sparsely populated weight vectors. SLPM is implemented by using a mathematical programming approach, i.e. setting up a large system of equations and solving for the optimized variable values. More details about SLPM and its relationship to SVMs can be found in [21].

Recapitulation

This chapter has briefly covered the basics needed to understand machine learning in the context of BCIs. The basic data representation as well as the goals of data classification in BCI have been lain out. Furthermore, the theoretical background for the basic tools used throughout the analysis of the experiments have been presented. Finally, a quick introduction to the theory behind the two classifiers compared in the results was provided. Understanding the place of data classification in a BCI will help one understand and interpret what the results of the experiments represent.

Chapter 5

Experiment 1: Motor Imagery

In order to achieve the goals set forth for the current work, we begin with an experiment that directly follows a well-established line of research using motor imagery in BCIs. More specifically, the protocol used in the Graz BCI is recycled with particular dedication toward reproducing similar results in the SINGLETRIAL MATLAB TOOLBOX as are found in the literature [28]. By closely following a standard BCI protocol and using the SINGLETRIAL MATLAB TOOLBOX to decompose and analyze the data, we are able to simultaneously achieve the first two goals of the present work (see Section 1.1).

The chapter is organized by linearly running through the various stages of experiment. We begin by describing the manner in which the experiment was performed, namely how many subjects participated, the neuroimaging hardware used, the experimental environment, and experimental design. Next, we provide an in-depth look at the EEG data by describing the shape and preprocessing of the captured data. The different methods and techniques used to extract features out of the preprocessed data are then chronicled, followed by a description of the data classification methodologies. Finally, the experimental data are analyzed and detailed results are provided. Throughout the results section data analysis and qualification is performed but further discussion of the details and implications of the results is left for Section 7.2.

5.1 Methods

Subjects
A motor imagery-based EEG experiment with nine healthy, adult subjects (seven males, two females) was performed. None of the subjects (with the exception of one) had previously participated in a BCI experiment. Each subject participated in one EEG session lasting between 25-40 minutes, with interjected breaks every 5-7 minutes. All subjects were volunteers and provided written, informed consent. Data for eight subjects is included in the following analyses, excluding one subject (subject 3) whose recordings were discarded due to noise and lack of experimental attention.

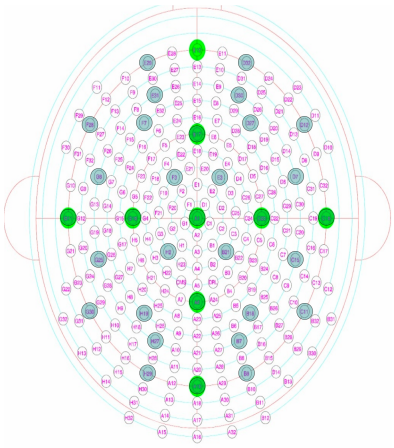

Figure 5.1: The electrode placement for 32 channels of the 256-channel Biosemi cap used for eight of nine subjects in the motor imagery study. The subset of electrodes used are highlighted in blue.

Data Collection

Data was captured using 32 active Ag/AgCl electrodes and a Biosemi ActiveTwo AD amplifier and converter. Electrodes were placed atop a layer of electroconductive gel according to the closest-fit of the 32-channel, 10/20 Biosemi cap inlayed on the 256-channel, ABC layout Biosemi cap. For eight subjects these 32 of 256 channels were used for recordings and for one subject the closest-fit of 32 of the 192-channel, 10/20 Biosemi cap were used (see Figure 5.1). To interpolate electrooculargrams (EOG) for later artifact detection, four flat electrodes were placed on the face: two above and below the left eye, one to the left of left eye, and one to the right of the right eye. Impedance in all electrodes was verified to be at an acceptably low level before beginning the experiment. EEG data for 32 scalp and 4 EOG electrodes were digitized at 256Hz and stored for later analysis.

Experimental Design

Subjects were seated one meter in front of a stimulus monitor in a dark, electric-shielded, sound-proof room. Each subject was led through a series of practice trials and demonstrated the time periods when they were able to relax and were otherwise asked to sit as still as possible and to avoid blinking of the eyes. Following the standard Graz BCI paradigm, the goal was to discriminate between two mental tasks. Upon presentation of one of two visual cue stimuli, subjects were asked to mentally envision left-hand or right-hand movement while avoiding physical initiation of the movement. A block consisted of forty trials, each trial lasting eight seconds on average and subjects performed between 160-240 (half left-hand, half right-hand) trials in a session.

A single trial began with a two-second period where subjects were asked to relax and fixate on a point in the middle of the screen. Following this, a green circle overlays the fixation point as a one-second warning of the upcoming stimulus. The stimulus is then presented inside of the green circle as a randomly-decided left or right arrow for one second. Following this is a three second rest period and then a randomized inter-trial interval lasting 0.5-1.5 seconds. See Figure 5.2 for a graphical representation of the experimental design of a single trial and the accompanying visual

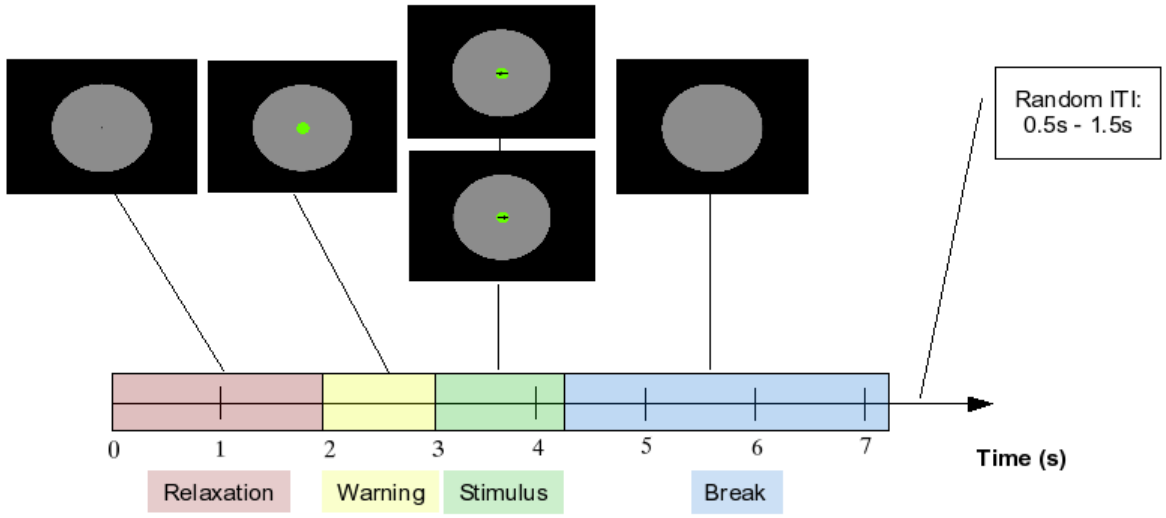

Figure 5.2: A single trial of the motor imagery study with accompanying visual stimuli. Subjects are asked to relax, take note of the stimulus warning, and then perform the mental task of imagining left-hand or right-hand movement according to the direction of the arrow. Afterward, the subject is given a few seconds of rest and another trial begins.

stimuli.

5.2 Data Preprocessing

The selection of preprocessing techniques has great effect on the eventual classification results and highlights the trade-off between not having enough data to classify high-dimensional data and the danger of overfitting. Because of the experimental setup and ample relaxation periods, it was elected to keep all trials for all subjects in the analysis. This differs dramatically from previous approaches as they perform manual selection of artifact-free trials often leading to trial rejection rates as high as 60%. Though these trial rejections can help eventual classification results, it was determined that keeping all data is better suited for the eventual move to an online BCI using robust classification algorithms. This idea is the general BCI approach popularized by the Berlin BCI group with the slogan “let the machines do the learning”.

For a few subjects data was captured at 1028Hz and was hence downsampled to 256Hz for consistency amongst subjects. Though all trials were kept in the data analyses, any significantly noisy channels were discarded. On average, less than one channel was discarded per subject. The data was first spliced into a series of labeled trials by using timestamps sent by the synchronized experimental presentation software EPrime. The period of interest for each trial spanned from 2500 ms before the stimulus presentation until 1250 ms after stimulus presentation. A baseline

period of one second was defined ranging from -2500 ms to -1500 ms (with respect to stimulus onset) and the task period of one second defined between 250 ms and 1250ms. The baseline was selected to be one second lying in the middle of the relaxation period, prior to the green circle stimulus warning. The task period was offset 250 ms from stimulus onset to avoid capturing the visual-evoked potentials from the presentation of a visual stimulus.

The split, raw data was first temporally preprocessed by baseline normalization (see Section 3.1.1) and then spatially preprocessed in one of two ways, depending on the eventual destination of the data:

1. **Manual Channel Reduction and Interpolation**
   The Graz BCI uses only two EEG channels recorded over sensorimotor cortex (one over the left hemisphere, one over the right hemisphere) and a third EEG channel over central parietal lobe. From these three channels they spatially interpolate values at the two sensorimotor channels using a dipole method. The net effect is that features are eventually computed only for two channels. Since this work aimed to first replicate results found in the previous literature, the 32-channel recordings had to be manipulated into two channels. To do so, first a CAR spatial filter was applied to all channels (see Section 3.1.2). Following this, a *small, reduction Laplacian* spatial filter was applied. This filter is a combination of manual selection of the two channels over sensorimotor cortex lying directly over "hand areas" (C3 and C4) and a simultaneous local, small Laplacian. More specifically, the average value of the four nearest neighbors around each of these two channels is taken and subtracted from the corresponding channel. The result is a reduction of 32 channels to two, interpolated channels.

2. **Feature Selection: Channel and Frequency Approach**
   For those data destined to feature selection algorithms there are two perspectives from which to view feature selection: the channel perspective or the frequency perspective. The difference lies in whether one wishes to eventually have a feature vector composed of elements which represent channels or frequencies. When performing channel selection, the spatial preprocessing was done by keeping all channels and applying a spatial CAR filter to all channels for all trials. When performing frequency selection, a spatial CAR filter is applied to all channels and then the data is reduced to a single, hand-picked channel which, after feature extraction, is represented by various frequency components.

5.3 Feature Extraction and Data Classification

Each trial is represented by two seconds of preprocessed EEG data where the first half represents baseline activity and the latter half represents task-based activity. Feature extraction was executed in one of three ways, again depending on the type of final analysis to be done. The feature extraction process is graphically depicted in Figure 5.3 for several of the three tactics.

1. **Graz BCI Features**
   Following the Graz BCI protocol, the two seconds of data for each trial was split into eight

	Sub 1	*Sub 2*	*Sub 4*	*Sub 5*	*Sub 6*	*Sub 7*	*Sub 8*	*Sub 9*
Low Band (μ)	9-13	8-12	9-13	8-12	9-13	8-11	8-11	8-12
Middle Band (β)	20-25	19-25	19-24	18-26	19-26	18-26	18-26	18-25

Table 5.1: The two hand-picked frequency bands (in Hz) for each subject used in computation of the Graz BCI features for classification of left-hand vs. right-hand motor imagery.

windows of 250ms. Then, using Maximum Entropy Method spectral estimation (see Section 3.2), an estimation was made of the spectral power in each 250ms chunk across two subject-dependent frequency bands. It is has been well documented that inter-subject differences exist in terms of important frequencies for classification [28]. Subject-dependent frequency bands were selected by hand based on observation of their respective effectiveness. Each subject's two frequency bands were chosen from μ and β rhythms and were centered roughly around 8-13Hz and 16-24Hz. Table 5.1 provides the subject-dependent frequency bands. A spectral estimation was made every 1Hz across each of the two frequency bands and the average spectral estimation across each band was taken. For example, the spectral estimation for 8-13Hz is the mean spectral estimation for the estimations at 8Hz, 9Hz, 10Hz, etc.

Next, the ERD was computed for each 250ms across two frequency bands. This is a paired operation in which the estimation for each band from the first 250ms of the baseline period is subtracted and normalized from the first 250ms of task-related data according to the ERD computation found in Section 3.2. Then, the ERD for each band from the second 250ms period is computed and so forth. After feature extraction a trial is represented by four ERD values for each of two channels and each of the two frequency bands.

2. **Joint Feature Selection and Classification: The All-Channel Approach**
Since we wish to perform feature selection across the channels, it is necessary that the feature vector contain information for each channel. To achieve this, we begin with the same process as used in the Graz BCI features. Namely, four ERD values for each trial and for each channel across a single frequency band are computed as above. The frequency band can be arbitrarily defined and in the following analyses is a narrow band of 1Hz. The average of these four ERD values is taken as the feature for each channel. Thus, after feature extraction each trial is represented by one ERD value for each of the 32 channels.

Regardless of the manner in which features are extracted from the signal, the representation must be standardized for the classification algorithms. This is performed in a linear unfolding of feature matrices into a single feature vector. For example, the Graz BCI representation results in a 2 channel x 4 ERD-values x 2 frequency band feature matrix which is flattened into a single 16x1 feature vector. The data are now packaged in a generalized matrix ready for classification, namely each trial is represented a single feature vector. Class labels (*i.e.* left or right) are assembled for each of the trials.

A 10-fold cross validation classification scheme was performed with classification by two types of classifiers: a Linear Fisher Discriminant (LFD) and a Sparse Linear Programming Machine (SLPM) (as described in Chapter 4). Nested cross validation was also implemented to be used in this context but results are not given (see Section 7.4). It was necessary to manually select values

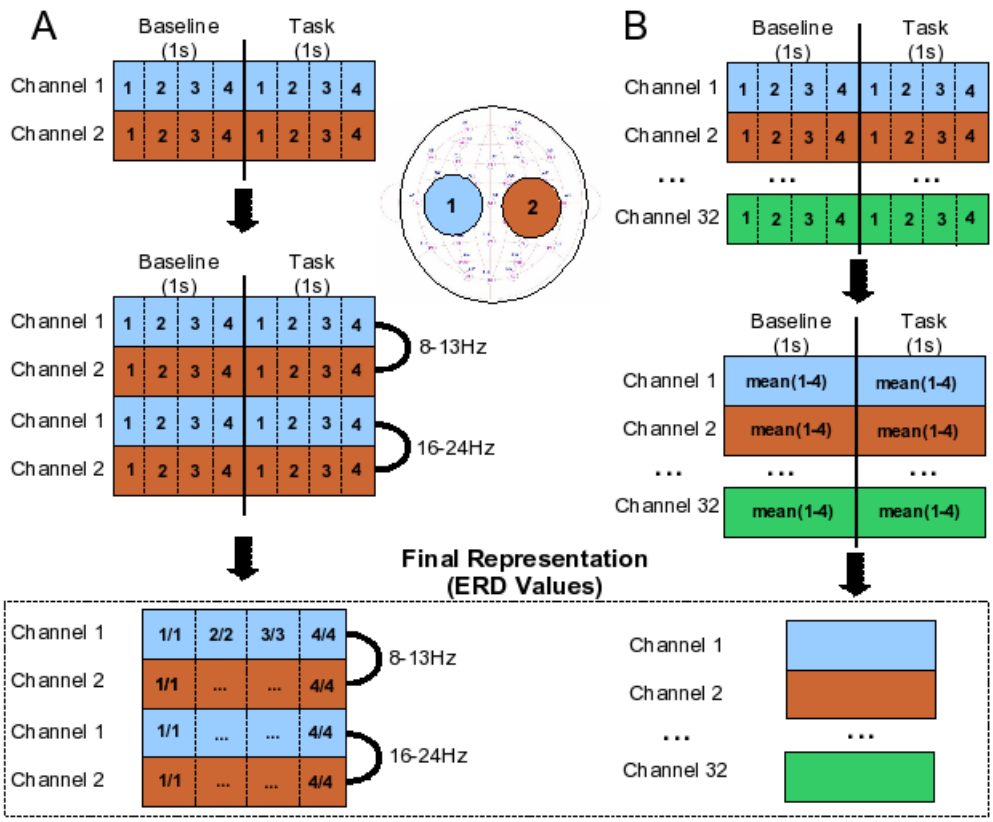

Figure 5.3: *Left (A).* Graz BCI-like features extracted from preprocessed EEG signals. First, each trial is broken into eight windows of 250ms for the two channels. Spectral estimations are then made across two frequency bands. Finally, ERD is computed by matching the estimates from first 250ms period of task-related data to the first 250ms period of the baseline period, the second to the second, etc. This leaves one ERD per 250ms for two channels and two frequency bands. *Right (B).* Features extracted for eventual automated channel feature selection. Each trial is broken into eight windows of 250ms for all 32 channels. Spectral power estimations are made for each 250ms and the average is taken for the baseline period and the task-related period. Finally, a single ERD value is computed using the averaged spectral powers leaving one ERD per channel.

	Sub 1	*Sub 2*	*Sub 4**	*Sub 5*	*Sub 6*	*Sub 7*	*Sub 8*	*Sub 9*	*Subject Average*
LFD	52	**69**	**63**	**54**	**54**	**64**	**61**	**57**	59
SLPM	**53**	68	63	52	52	62	59	57	58

Table 5.2: Classification performance (in %) for two classifiers (LFD and SLPM) using the Graz BCI feature vectors. The best per-subject performance is in bold. It is worth noting that there is little difference in classification performance between classifiers with this low dimensional feature vector.*(*)* Subject 4's results were consistently inversing classification decisions (see Section 7.2).

for both the model order hyperparameter in MEM estimation and the regularization (sparsity) hyperparameter used in the SLPM. The selected parameter values are listed in Appendix A.2 in Table A.1.

5.4 Results

Analysis was performed for the three different feature representations across two dimensions: classification performance and automated feature selection. In this section the results are accompanied by minimal analysis and observations. See Section 7.2 for further, in-depth discussion of the results.

5.4.1 Classification Results

The overall goal for the classifier is to correctly classify a new, unseen trial as being either right-hand or left-hand motor imagery. Since discrimination is amongst two classes, a classifier performance of 50% indicates no significant distinguishability exists between the chosen representation of the two mental tasks and 100% indicates perfect distinguishability. Classification performance was computed using a cross validation scheme for two types of classifiers a LFD and a SLPM for the following two feature representations.

1. **Graz BCI Features**
   Table 5.2 displays the results obtained by classification using the Graz BCI feature representation. It is clear from the beginning of our analysis that certain subjects stick out as having better performance, namely subjects 2, 7, and 8. Furthermore, note that there is very little difference in performance between the two classifiers with this low-dimensional feature vector representation. Knowing that classification is working well already for certain subjects, we move on to more sophisticated techniques and analyse the data using different feature vector representations.

2. **All-Channel Approach: All Channels, Individual Frequencies**
   Using a feature vector representing one ERD value per channel, several analyses were made. First, raw classification performance for the two classifiers across the same two subject-dependent low/high frequency bands as above were computed (Table 5.3). These data show us what direct effects the feature representation can have. More specifically, these results exhibit the effect of classifying using data from all channels whereas the Graz features reduce

	Sub 1	*Sub 2*	*Sub 4*	*Sub 5*	*Sub 6*	*Sub 7*	*Sub 8*	*Sub 9*	*Subject Average*
LFD (*low*)	**60**	65	**58**	56	**56**	64	56	51	58
LFD (*high*)	55	73	55	53	52	61	55	51	57
SLPM (*low*)	56	81	58	**60**	51	**73**	60	51	61
SLPM (*high*)	56	**87**	55	56	53	66	**68**	**55**	62

Table 5.3: Summary of classification performance (in %) for two classifiers (LFD and SLPM) in low and high frequency bands using the all-channel feature vectors in the motor imagery experiment. The low and high frequency bands were selected according to Table 5.1. The best per-subject performance is in bold. SLPM shows better overall performance with the all-channel feature vectors.

the dimensionality by hand-picking features (channels and frequencies). If using Graz-like BCI features were a better-distinguishing representation, the respective results for LFD and SLPM in the low and high frequency bands should be particularly different. This is not the case and in certain instances classification performance is better for the all-channel feature representation. When comparing LFD and SLPM for the all-channel feature representation, it becomes clear that SLPM seems to better handle the higher dimensionality of the discrimination problem (also see Section 7.2).

We now look at the results from SLPM classification at a finer-grained scale. Particularly, Figure 5.4 shows the SLPM classification performance at particular frequencies between 5-100Hz across all subjects. This whisker plot nicely highlights the mean classification error at each frequency, the variance amongst subjects and the lowest and highest error. Note that these results are represented in classification error and not classification performance, where classification error is defined as $error = 1 - performance$. Thus, an error of zero is most desired. This figure shows us that frequencies less than 30Hz are best at distinguishing between the two tasks for the majority of all users. After about 30Hz the mean classification error hovers around 50%, indicating little distinguishability.

A tabular summary of these classification results averaged over 4Hz bins for each subject can be found in Table 2. This tabular view gives credit to those subject's results whose classification performance is otherwise anonymized in the whisker plot. It further shows us which frequency bins yielded best classification performance and cleanly shows more success for certain subjects (as high as 87% performance for subject 2, for instance).

5.4.2 Automated Feature Selection with SLPM

In addition to obtaining classification results with SLPM, we can perform automated feature selection using the high-dimensional weights for the decision hyperplane as described in Section 3.3. In this section we observe those channels and frequencies which were naïvely selected during the classification process.

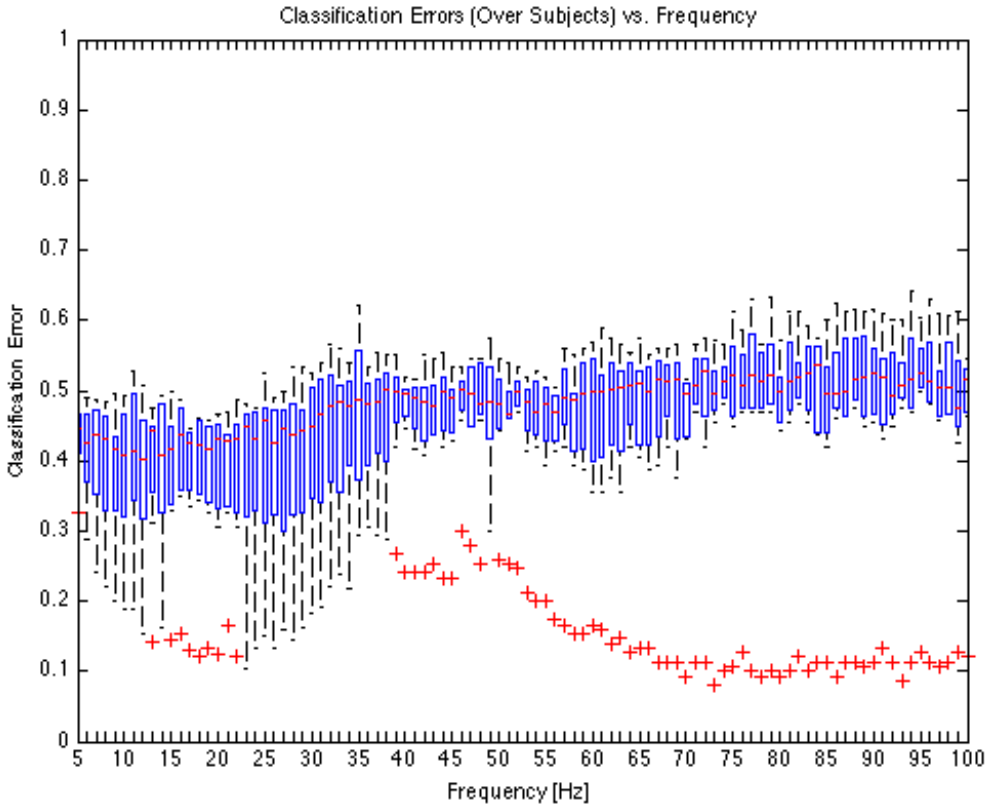

Figure 5.4: Whisker plot of classification error for different 1Hz-wide frequency bands for all subjects using the All-Channels feature representation. The box at each frequency shows error variance as well as the average classification error (red line). Dotted lines extend to classification errors within two standard deviations of the mean. Red crosses indicate statistical outliers. It is clear that classification performance for the average subject is best before 30Hz, and most notably between 10-27Hz.

Frequency Band	*Sub 1*	*Sub 2*	*Sub 4*	*Sub 5*	*Sub 6*	*Sub 7*	*Sub 8*	*Sub 9*	*Subject Average*
5-9Hz	56	75	57	58	51	69	57	53	59
9-13Hz	**57**	83	**58**	**59**	51	**72**	60	52	61
13-17Hz	53	85	56	58	53	67	63	55	61
17-21Hz	55	**87**	54	57	**56**	65	66	**57**	62
21-25Hz	56	**87**	53	55	53	66	69	53	62
25-29Hz	55	85	54	54	49	64	73	54	61
29-33Hz	54	80	50	51	46	60	70	51	58
33-37Hz	53	72	50	52	44	53	70	46	55
37-41Hz	52	60	50	49	48	49	**76**	50	54

Table 5.4: Per-subject SLPM classification results (in %) for the motor imagery experiment between 5-41Hz for 4Hz-wide frequency bands using the All-Channels feature represntation. Values in bold indicate the frequency band at which the highest classification performance was obtained. Frequency bands in which highest classification performance was achieved for some subject are additionally highlighted. This clearly shows (with the exception to subject 8) that highest discrimination between left-hand and right-hand motor imagery are found in μ and β rhythms.

Selected Channels

In order to verify that the selected channels lie over the expected neurophysiological regions it is useful to visually inspect the selected channels on a scalp plot. Figure 5.4.2 shows 360-degree scalp plots for two different frequency bands. Since it has been demonstrated in the classification results that μ and β rhythms are most important in classification performance, we display the weight vectors as projected on the scalp for the average subject for 8-13Hz and 18-26Hz. These scalp plots undeniably indicate that the selected channels are precisely where one would expect them to be, namely directly above the hand area of the motor homunculus in the left and right hemispheres. An additional scalp plot is given for 40-44Hz in Figure 5.4.2 to demonstrate that the tight channel localization found in the μ and β regions across the motor cortex, and more particularly directly above the hand area of the motor homunculus, is not so for the selected-channels outside of these two ranges.

Selected Frequencies

As established in the introductory chapters, the literature tells us that we should expect to high distinguishability in the frequencies associated with sensorimotor rhythms. We have so far seen two implicit pieces of evidence that selected frequency features do in fact correspond to to sensorimotor rhythms. First, the classification results in Table 2show us that classification performance was highest for all subjects (except one) in the frequencies corresponding to μ and β rhythms. The second piece of evidence comes from the channel selection scalp plots which clearly show that channel selection aligns with our expectations in the sensorimotor rhythm bands and does not align in the case of frequencies outside of this area (40-40Hz, for instance). Direct evidence for selected frequencies is provided in Figure 5.4.2 that displays the weights of different frequencies for the average subject. These are computed by summing all weights for each subject at each frequency.

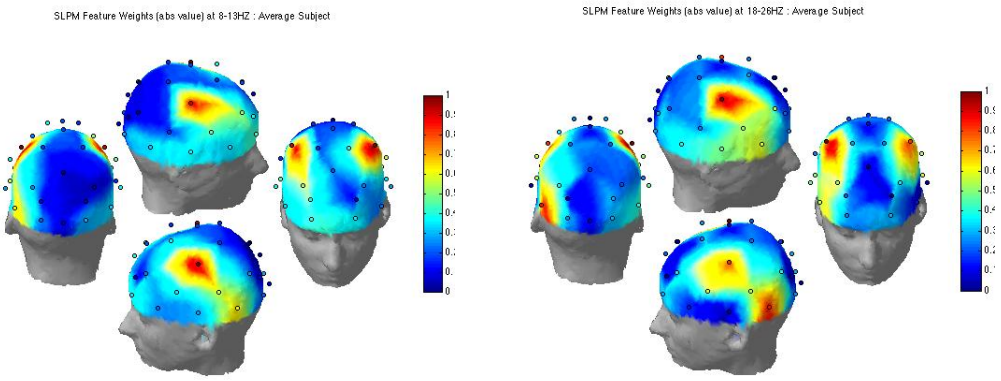

Figure 5.5: *Left:* Topographic scalp plots for the average subject inμ rhythms (8-13Hz) with superimposed feature weights for each channel as generated by the SLPM joint classification and feature selection algorithm. The most important channels are cleanly localized over left and right hand areas in the motor cortex. *Right:* The scalp plots for β rhythms (16-24Hz). Selected features also clearly lie in the expected neurophysiological areas.

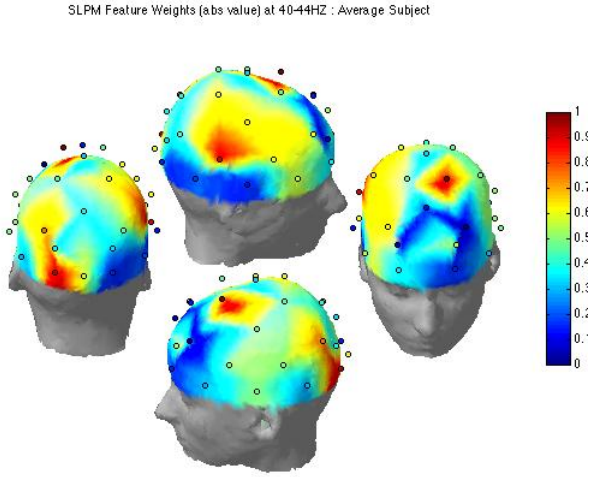

Figure 5.6: Topographic scalp plots for the average subject in higher EEG frequencies (40-44Hz) with superimposed feature weights for each channel as generated by the SLPM joint classification and feature selection algorithm. These frequencies lie outside of sensorimotor rhythms and do not show the same neurophysiological correspondence with respect to selected channels.

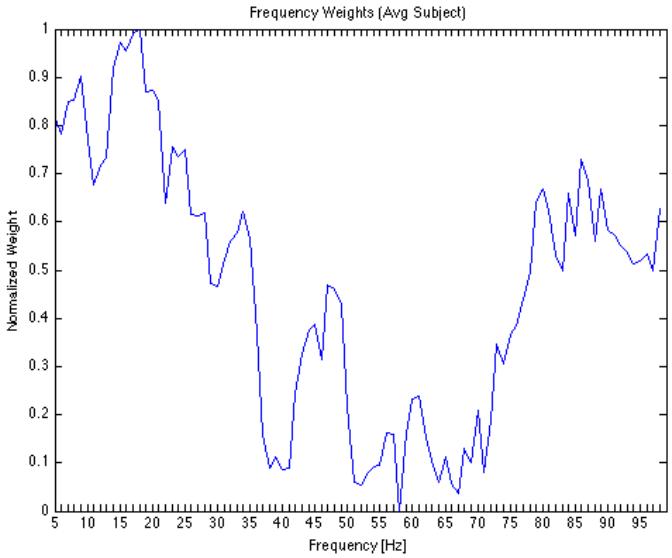

Figure 5.7: Min-max normalized weights for each frequency using joint automated feature selection and classification with SLPM. Note that the selected frequencies (strongest weights) lie between 8-11Hz and 15-22Hz and align with the expectation of sensorimotor rhythm importance.

Then, the aggregate weights for each frequency are normalized with a min-max normalization, namely the minimum value is subtracted from all weights and then all weights are divided by the maximum weight. This has the effect of scaling all weights between 0 and 1. This figure shows the selected frequencies to be clearly highest in the μ and β rhythms with a sharp fall off around 35Hz. The strongest frequencies are shown to in fact be in the 16-22Hz range which makes sense since the best classification performances were often found in this band.

Recapitulation

Throughout the course of this chapter we have described in detail the first of two experiments done in the present work. The success as marked by the results of the experiment indicates the successful completion of two goals of the current thesis (see Section 1.1). First, we have verified that the implemented BCI framework (The Singletrial Matlab Toolbox) properly works by putting its tools to the test. Second, a traditional motor-imagery based BCI experiment was performed for nine subjects and data were analyzed. The results mirror those found in the literature, namely that nontrivial classification rates can be achieved using sensorimotor-based features for imagined left-hand and right-hand motor movements. Classification performance rates of up to 87% were found. Furthermore, sophisticated joint classification and feature selection algorithms increased the classification performance for higher-dimensional feature vector representations while simultaneously providing naïvely selected features that agree with the current neurophysiological

and frequency-domain understanding of EEG signals based on imagined motor movement. More particularly, the automated feature selection algorithms chose EEG channels lying directly over the left and right hand areas in the motor cortex and favored frequencies originating from the well-studied sensorimotor rhythms.

Chapter 6

Experiment 2: Novel Cognitive Tasks

The third and final goal of this thesis is to use an experimentally validated BCI platform toward exploration of new cognitive tasks for BCI control and for neuroscientific verification of said cognitive tasks. The following experiment mirrors as closely as possible the motor imagery experiment described in Chapter 5 exploiting the idea that solid task analysis can be performed if nothing changes more than the tasks used for BCI control. More specifically, the same protocol used in the motor imagery experiment is recycled with only slight variation. These variations are delineated throughout the chapter but it is useful to think of this experiment as a repeat of the experiment outlined in Chapter 5with the slight modification of substituting two cognitive tasks for the left-hand and right-hand motor imagery tasks. Another noticable difference between these two experiments is how the novelty of these two tasks changes the analytical goals from the previous verification approach to a more exploratory approach. In other words, the results of this experiment are unable to be compared against a baseline of past research thereby making it difficult to confirm anything with certainty. This experiment provides first research into the use of these two cognitive tasks for BCI use and for using a BCI framework to validate their previous neuroscientific findings.

The remainder of this chapter is assembled in the precisely the same way as Chapter 5. To avoid repeating information, references will be made where methods and techniques used are precisely the same as in the previous experiment. We begin by again outlining the experimental methods tracing the experiment from its origin of capturing data from subjects to the storage of data for offline analysis. Next, we discuss again the preprocessing techniques used on the EEG data as well as the features extracted from the raw EEG signals. The same classification paradigms were applied yielding a series of results that mirror the lines of analysis taken in the motor imagery experiment. Care is taken in the final section to avoid too much discussion for incongruent results. These further discussions are left for the complementary Section 7.3.

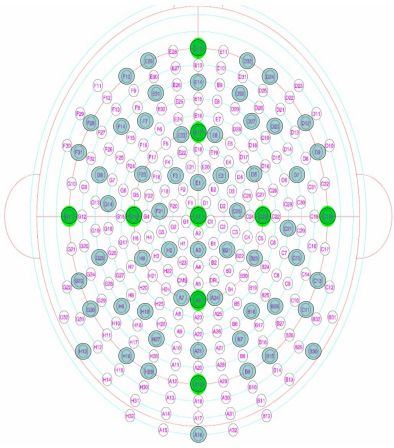

Figure 6.1: The electrode placement for 64 channels of the 256-channel Biosemi cap used for six of seven subjects in the cognitive imagery study. The subset of electrodes used are highlighted in blue.

6.1 Methods

Subjects

An EEG experiment with seven healthy, adult subjects (five males, two females) was performed. All of the subjects participated in the previous motor imagery-based BCI experiment. Five of the subjects participated in this experiment in a separate EEG session, weeks after involvment in the first experiment. Two of the subjects participated in the two experiments back-to-back. The current experiment consisted of one EEG session lasting between 30-40 minutes, with interjected breaks every 5-7 minutes. All subjects volunteered to participate and provided written, informed consent. Data for all seven subjects is included in the following analyses and results.

Data Collection

Data was captured using 64 active Ag/AgCl electrodes and a Biosemi ActiveTwo AD amplifier and converter. It was decided to use 64 channels for the cognitive experiment to obtain a higher spatial resolution for two reasons. First, brain regions important in classification of the two novel tasks is not as well-known and second, the spatial distribution activity may prove to be more spread out for cognitive tasks than in motor imagery tasks. Electrodes were placed atop a layer of electroconductive gel according to the closest-fit of the 64-channel, 10/20 Biosemi cap inlayed on the 256-channel, ABC layout Biosemi cap. For six subjects these 64 of 256 channels were used for recordings and for one subject the closest-fit of 64 of the 192-channel, 10/20 Biosemi cap were used (see Figure 6.1). To interpolate electrooculargrams (EOG) for later artifact detection, four flat electrodes were placed on the face: two above and below the left eye, one to the left of left eye, and one to the right of the right eye. Impedence in the electrodes was verified to be at an acceptably low level before beginning the experiment. EEG data for 64 scalp and 4 EOG electrodes were digitized at 256Hz and stored for later analysis.

Experimental Design

This experiment was conducted in under precisely the same conditions as the motor imagery experiment. As all subjects had participated in the first experiment, they were all familiar with the general paradigm. They were seated in the same room and led through a number of practice trials while being explained what each stimulus meant. Subjects were given the analogy to the prior experiment, namely that everything remains the same with the exception to the two tasks that they are asked to perform. These two tasks were accompanied by new visual stimuli and were described as follows:

- *Inner-speech command (ISC)*
  The subject is to internally command the virtual avatar to "go away" or "get out of here". All subjects are minimally bilingual, and were thus instructed to use whatever language they felt most comfortable with and whatever particular command they felt like using. The key requirements were that they were to imagine detesting or being sick of the the avatar's presence and to command it to leave.
- *Own Body Transformation (OBT)*
  The own-body transformation consists of projecting one's body outward from their current perspective into the physical position of another. The subjects were told to physically project themselves into the shoes of the avatar. This is a nonintuitive task so several methods were discussed. Possibilities included to imagine directly "hopping" into the avatars shoes and to imagine seeing oneself in its place or to envision physically leaving ones body, rotating 180 degrees, and finishing the projection and rotation in the place of the avatar. The subject was to select the method that made most sense to them. Methods were informally controlled by asking how the subject chose to perform the action between experimental blocks.

These two tasks were particularly chosen to be analogous to the left and right-hand motor imagery tasks from the previous experiment. Namely, previous studies show that inner-speech activates areas localized primarily in left hemisphere and SMA and that OBT task initiation results in right hemispheric, TPJ activation [4, 32]. Task selection was performed because of this left/right hemispheric distinction and the distinctively localized functional brain areas involved in the two tasks. Adding the "detest" and "compassion" elements was nothing more than an additional flavor that could hopefully help further separate these two cognitive tasks.

A block consisted of 200-300 trials (half ISC, half OBT). Each trial begins with a 2s fixation cross, as before. Following this comes a "warning" stimulus which lasts for 1s and consists of a 3-dimensional standing avatar with an overlayed green circle on its chest. This green circle was kept for consistency with the previous experiment. Next, a randomly selected task cue is presented in the form of a rotating arrow or a talk box. The rotating arrow indicated that the subject should perform the OBT task whereas the comic-like talk box indicated the ISC task. A graphical representation of a trial with its accompanying visual stimuli can be found in Figure 6.1.

6.2 Data Preprocessing

Most of the common preprocessing performed in the motor imagery experiment was replicated. It was decided again to keep all trials for analysis. All subjects were initially recorded at 256Hz so

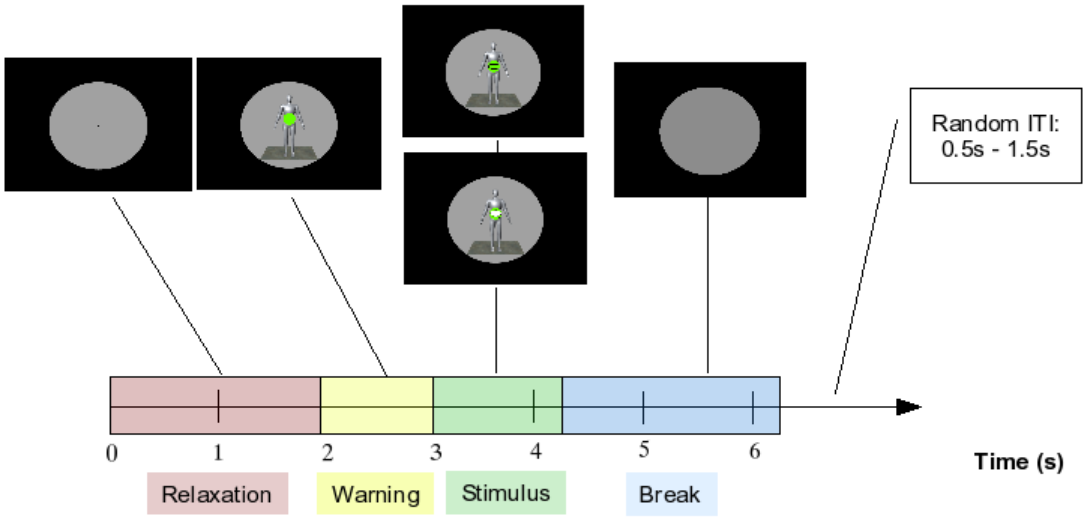

Figure 6.2: A single trial of the cognitive imagery study with accompanying visual stimuli. Subjects are asked to relax, take note of the stimulus warning, and then perform the mental task of either imagining to give the command "go away" to the avatar or to imagine physically placing one's body in the avatar's shoes. Afterward, the subject is given a few seconds of rest and another trial begins.

no downsampling was necessary. The continuous, raw EEG data was split and packaged in the same way as was done for the motor imagery experiment, resulting in 2 seconds of EEG data per trial representing the baseline and task-related data for the trial. The raw data was then baseline normalized.

A channel reduction from 64 channels to the nearest-fit of the 32-channels used in the previous experiment was performed after initial analyses showed the dimensionality of the problem to be a limitation in classification (see the related discussion in Section 7.3). The 32 channels were a proper subset of the 64 channels, so comparisons between studies and across subjects can be safely made. After throwing away half of the channels, all remaining 32 channels were used for analysis in all subjects. Following this, specific data preprocessing was performed based on the eventual feature representation of the signal. The two main preprocessing techniques are outlined below.

1. **Manual Channel Reduction and Interpolation**
   We begin by performing a similar operation as was done during the motor imagery experiment preprocessing stage. First, a CAR spatial filter was applied to all channels (see Section 3.1.2). Based on our current understanding of the neurophysiology of these two cognitive tasks, two electrodes were first selected topographically selected. One electrode located near the left-hemispheric interior frontal gyrus was selected (F3) for its relation to internal speech. A second electrode was selected located near the right-hemispheric temporal-parietal junction (C15) known for its known functional role in own-body transformations. Next, a small, reduction Laplacian spatial filter was applied consisting of manual selection of these two channels and a small Laplacian using the finite difference method. Particularly, these two channels are interpolated by subtracting from them the mean signal at their five-nearest neighboring electrodes. At the end of this preprocessing stage we have further reduced 32 raw EEG signals to two Laplacian-interpolated signals.

2. **Feature Selection: Channel and Frequency Approach**
   The second preprocessing style is precisely the same as the second preprocessing technique in the motor imagery study. The distinction is again made between the "channel" perspective and the "frequency" perspective. If the destination feature representation is the "all-channel" approach, all channels are kept and subjected to a CAR spatial filter.

6.3 Feature Extraction and Data Classification

We are now faced with the nontrivial task of representing the two seconds of preprocessed data to uniquely describe the two tasks. It was elected to again go with SMR-based feature representations, namely all of the following representations are grounded in ERD. This choice reflects one of many possible ways to represent the signals and may not be optimal (see Section 7.3 for further discussion). The two feature representations are graphically portrayed in Figure 6.3.

1. **Hand-Crafted Feature Representation**
   As a first attempt at uniquely representing the signals coming from these two cognitive tasks we take inspiration from both the Graz-like features used in the motor imagery experiment and from current neurophysiological and EEG frequency domain-based research.

	Sub 1	*Sub 2*	*Sub 4*	*Sub 7*	*Sub 8*	*Sub 9*
Low Band	13-17	9-13	5-9	17-21	29-33	13-17
Middle Band	17-21	21-25	25-29	49-53	53-57	49-53
High Band	95-99	49-53	91-95	79-83	79-83	83-87

Table 6.1: The three hand-picked frequency bands (in Hz) for each subject used in computation of the manual selection features for use in classification of the cognitive OBT and ISC tasks. Note that the categories of low, middle and high band could loosely be labeled with μ, β, and γ. In some subjects the selected frequency bands do not span all of these classes and instead use several bands from the same class of bands.

More specifically, we use the two interpolated channel preprocessing scheme and represent these two channels by four ERD values, as in the motor imagery experiment. The four ERD values for each of these two channels are again computed by calculating the ERD between the first 250ms of task-related data against the first 250ms of baseline activity, the second quarter second against corresponding baseline segment, and so on. These ERD values are also computed over MEM-estimated power at three frequency bands, in contrast to the two bands used in the motor imagery experiment. These three bands were hand-picked for each subject based on preliminary classification results across all bands. One noteworthy difference in these bands is the fact that they often lie in high-frequency γ regions and are no longer limited to μ andβ bands. Evidence shows that higher-frequency γ bands play a role in cognitive processing, so it was desired to reflect this in the signal representation. Table 1 highlights the hand-picked frequency bands used for this feature representation.

2. **All-Channel Approach: All Channels, Individual Frequencies**
The all-channel representation is derived as in the motor imagery experiment. The average power estimation at a single frequency across the four quarter-second periods in the baseline is made. Then, the ERD for each channel is computed by comparing this baseline average against the average at the task-related data segment. The net effect is that each trial is summarized by one ERD value per channel.

As before, the feature matrix is unfolded into a single vector for each trial and an associated trial label (ISC or OBT) is stored for comparison. For example, the hand-crafted feature representation for each trial is 2 channels x 4 ERD-values x 3 frequency bands and is flattened to a single 24x1 vector.

6.4 Results

Results are provided for both classification and feature selection using 10-fold cross validation across two types of feature representations as in the motor imagery experiment. Effort was made to follow the same linear line of analysis as was followed in the previous chapter. That is to say, all figures and tables in this section have counterparts in the motor imagery experiment results.

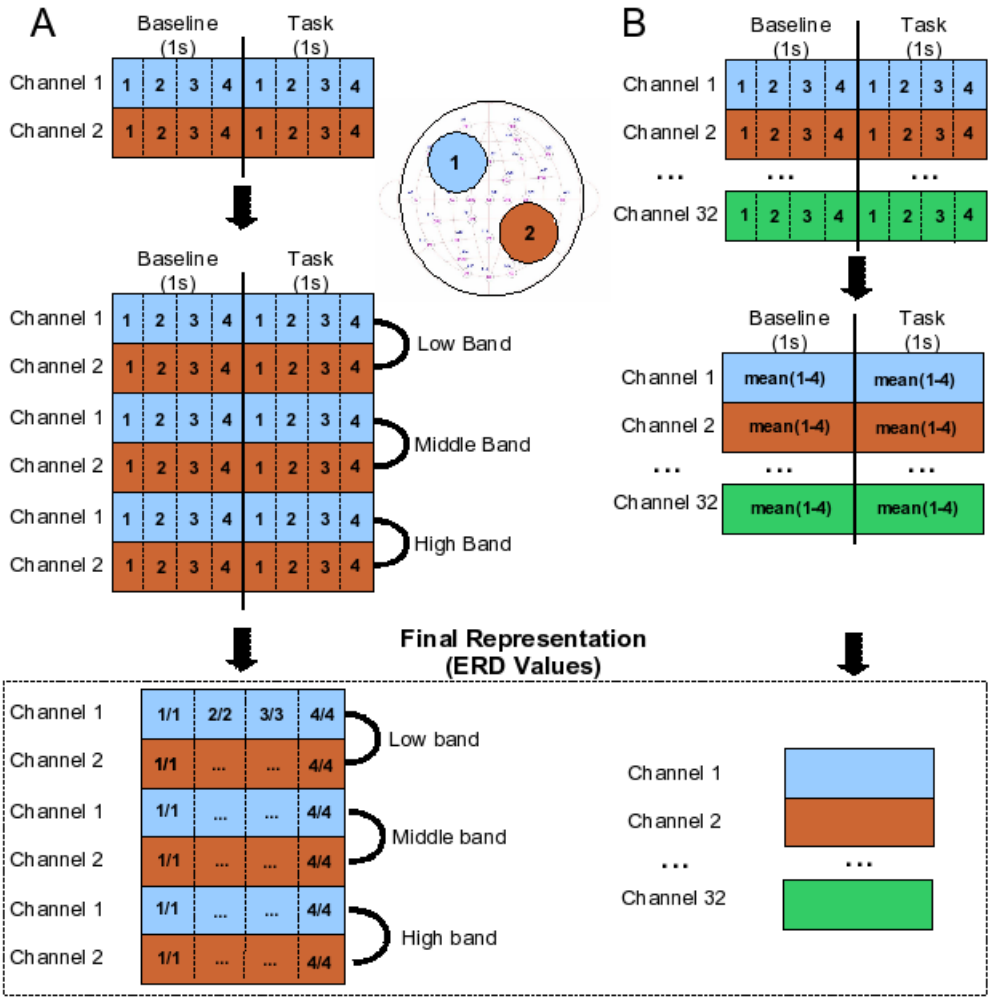

Figure 6.3: *Left (A).* Hand-crafted features extracted from preprocessed EEG signals . First, each trial is broken into eight windows of 250ms for the two channels located over left frontal and temporal-parietal areas. Spectral estimations are then made across thre hand-pickedfrequency bands. Finally, ERD is computed by matching the estimates from first 250ms period of task-related data to the first 250ms period of the baseline period, the second to the second, etc. This leaves one ERD per 250ms for two channels and three frequency bands (a 24-dimensional vector). *Right (B).* Features extracted for eventual automated channel feature selection. Each trial is broken into eight windows of 250ms for all 32 channels. Spectral power estimations are made for each 250ms and the average is taken for the baseline period and the task-related period. Finally, a single ERD value is computed using the averaged spectral powers leaving one ERD per channel (a 32-dimensional vector).

	Sub 1	*Sub 2*	*Sub 4*	*Sub 7*	*Sub 8*	*Sub 9*	*Subject Average*
LFD	**57**	**57**	**59**	**57**	52	61	57
SLPM	52	51	51	56	51	**66**	55

Table 6.2: Classification performance (in %) for two classifiers (LFD and SLPM) using the hand-crafted feature vectors described in Section 6.3. The best per-subject performance is in bold. At first glance, it can be seen that LFD outperforms SLPM for every subject with exception to subject 9, who stands out as the most promising subject for discrimination. Note also that performance is consistently lower than the motor imagery experiment counterpart.

6.4.1 Classification Results

The subsequent classification results represent classification performance for 10-fold cross validation for six of the seven participants in this study. Results are split depending on the underlying feature representation.

1. **Hand-Crafted Feature Representation**
   Using a hand-crafted feature representation of two interpolated channels located over Broca's area and the temporal-parietal junction across three subject-dependent frequency bands, we compute the classification error for LFD and SLPM using a 10-fold cross validation scheme. Table 1 presents these results. For the first round of data analyses, it can be seen that classification performance is on the whole lower than in the motor imagery experiment. On the other hand, there are a couple of subjects (4 and 9) who stand out as having classification performance at levels considerably above chance. This motivates further investigation and leads us to now explore what happens if we classify with features extracted from all 32 channels.

2. **All-Channel Approach: Individual Channel, All Frequencies**
   Though classification results of the the hand-crafted representation indicate that the OBT and ISC tasks are somehow discriminable, using all channels and looking across a larger band of frequencies is a more exploratory tactic. Since the important channels and frequencies for discriminating these two tasks are not known *apriori*, it is informative to use the all-channel representation across 1Hz-wide windows. By computing the classification performance at each 1Hz for all channels gives us a large 2-dimensional results matrix with one classification performance per frequency (see Appendix A.1). We begin by looking at classification performance using the all-channel representation for LFD and SLPM. The results are in Table 6.3 and clearly demonstrate that again, the all-channel representation better distinguishes between the two classes. Furthermore, SLPM handles this high, 32-dimensional representation better than LFD as all subjects' top performances are found by SLPM. Subject 9 receives up to 67% classification accuracy using this simple feature representation.

   Since SLPM better handles this feature-representation we zoom in on its performance for each subject across each 1Hz-wide frequency band between 5-100Hz. The whiskerplot in Figure 6.4 demonstrates the mean classification error as well as maximum and minimum classification errors at each frequency. Again, note that this figure is in terms of classification error, not performance. The results show that for most subjects, classification rates

	Sub 1	***Sub 2***	***Sub 4***	***Sub 7***	***Sub 8***	***Sub 9***	***Subject Average***
LFD (*low*)	54	56	53	51	55	58	55
LFD (*med*)	54	62	51	51	54	58	53
LFD (*high*)	53	50	53	50	53	56	53
SLPM (*low*)	55	**63**	57	**55**	**60**	66	59
SLPM (*med*)	**55**	**63**	**59**	**55**	**60**	**67**	60
SLPM (*high*)	52	49	53	53	**60**	**67**	56

Table 6.3: Summary of classification performance (in %) for two classifiers (LFD and SLPM) in low and high frequency bands using the all-channel feature vectors in the motor imagery experiment. The low and high frequency bands were selected according to Table 5.1. The best per-subject performance is in bold. SLPM outperforms LFD for all subjects using the "all-channel" feature vectors.

are around chance (50%). The lowest average errors (40-45% error) tend to fall between 20-40Hz. There are a few subjects, however, that obtain good classification performance in the higher frequencies as demonstrated by low classification error for frequencies higher than 50Hz.

To circumvent the anonymity of good candidates in the whiskerplot, we instead look at these results in a tabular, averaged format for 4Hz-wide frequency bands. Table 6.4 provides summarized results highlighting the frequency bands yielding best performance for each subject. A full table can be found in appendix in Table A.3. These results further demonstrate that the important bands for classification performance do indeed fall into three frequency categories, a low band (˜5-25Hz), a medium band (49-57Hz) and a high band (79-91Hz). It is clear that high classification accuracy can also be obtained in this high band, perfectly aligning with evidence that cognitive tasks cause coherency in γbands [11]. This is in direct contrast with the motor imagery study where the best classification rates were always found in frequency bands less than 30Hz. This full exploration shows us that classification rates of up to 67% are achievable using a simple, SMR-based feature representation and a sophisticated classifier. Certain subjects (e.g. 2, 8 and 9) clearly perform better than the others, whose classification results tend to hover around 50% (pure chance).

6.4.2 Automated Feature Selection with SLPM

Selected Channels

The automated channel selection results are given via the average subject scalp plot for two low band, one medium band, and one high band frequency (Figures 6.5, 6.6and 6.7, respectively). The selected bands were chosen at 9-13Hz, 21-25Hz, 49-53Hz, and 79-81Hz based on classification performance from the previous section's findings. The results are mixed and do not clearly show selected channels coming from the hypothesized regions. This fact, as well as possible explanations of their meaning is discussed at length in the subsequent discussion chapter in Section 7.3.

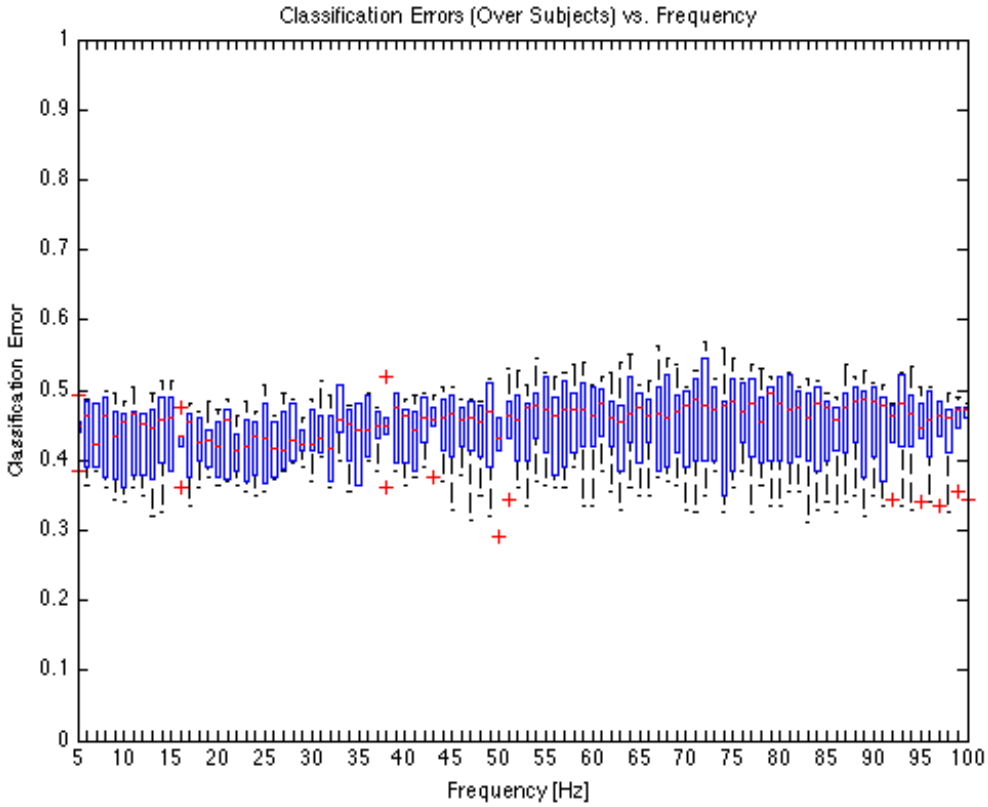

Figure 6.4: Whisker plot of classification *error* for different 1Hz-wide frequency bands for all subjects in the cognitive experiment using the "all-channels" feature representation. The box at each frequency shows error variance as well as the average classification error (red line). Dotted lines extend to classification errors within two standard deviations of the mean. Red crosses indicate statistical outliers. Upon inspection, the best average classification error lies between 20-40Hz. Most subjects obtain classification rates around chance (50%) hereafter. For several subjects, however, high frequencies (>65Hz) yield low errors. The lowest classification error, for instance, was at 49Hz.

Frequency Band	*Sub 1*	*Sub 2*	*Sub 4*	*Sub 7*	*Sub 8*	*Sub 9*	*Subject Average*
5-9Hz	52	**63**	57	53	53	60	56
9-13Hz	53	**63**	54	52	55	65	57
13-17Hz	**55**	61	54	51	54	66	57
17-21Hz	**55**	62	55	**55**	54	64	57
21-25Hz	53	**63**	58	53	56	63	58
25-29Hz	52	61	**59**	53	58	61	57
29-33Hz	51	56	58	53	**60**	61	56
49-53Hz	52	49	53	**55**	59	**67**	56
53-57Hz	52	48	50	54	**60**	64	55
79-83Hz	49	49	51	53	**60**	66	55
83-87Hz	51	50	52	53	**60**	**67**	55
87-91Hz	51	49	51	52	**60**	66	55

Table 6.4: Per-subject SLPM classification results (in %) for the cognitive experimentfor 4Hz-wide frequency bands using the "All-Channels" feature represntation. Only those frequencies leading to some subject's best classification rate are included. Subject IDs refer to the same subject as used in the motor imagery study. Unlike in the motor imagery study, the best results come from a wide range of frequencies including high γ frequencies.

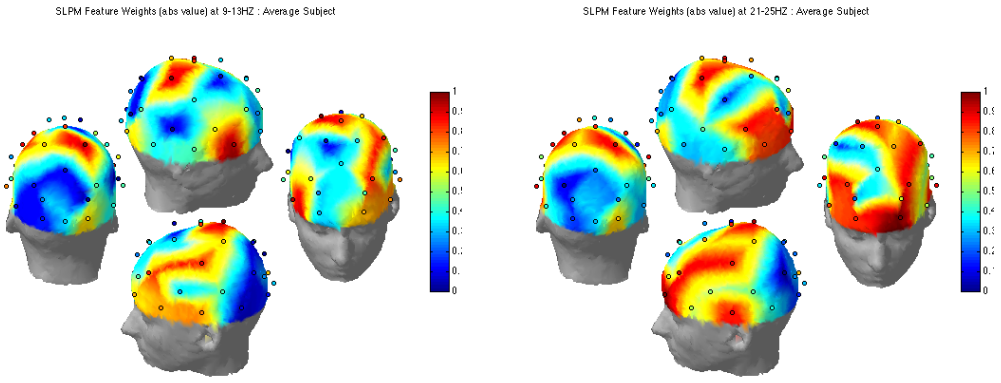

Figure 6.5: Topographic scalp plots for the average subject in two "low" bands (9-13Hz and 21-25Hz) with superimposed feature weights for each channel as generated by the SLPM joint classification and feature selection algorithm. The selected channels lie primarily over motor cortex but are well distributed along the entire scalp.

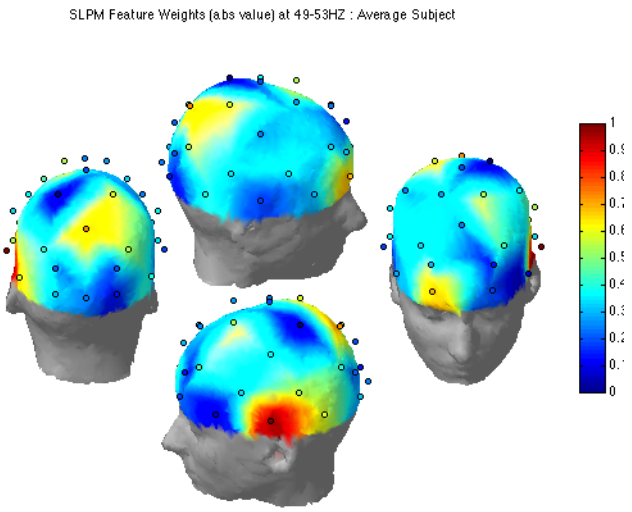

Figure 6.6: Topographic scalp plots for the average subject in a "middle" band (here 49-53Hz) with superimposed feature weights for each channel as generated by the SLPM joint classification and feature selection algorithm. The selected channels are cleanly isolated to left frontal-temporal regions. This strong weight preference is located near Broca's area but being at the periphery could also reflect electromyographic (EMG) muscle noise from the neck.

Selected Frequencies

Frequency selection was performed exactly as it was in the motor imagery experiment, namely subject weights were summed for each frequency and normalized to produce Figure 6.4.2. This shows two strong peaks of prefered frequencies, those around 45-50Hz and those between 80-95Hz. Interestingly, several of the best classification performances were found within these regions. This also is in line with the ongoing hypothesis that these cognitive tasks can be distinguished on low, middle, and high band features.

Recapitulation

This chapter has focused on presenting a cognitive BCI-based EEG experiment from scratch to its results. Two novel cognitive tasks inspired by previous cognitive neuroscientific research were used as the control signals in the BCI. More specifically, subjects were either asked to mentally command an avatar to "go away" (the inner-speech command task) or to envision projecting themselves into the avatar's position (the own-body transformation task). Using the same BCI framework whose correctness was verified by experimental results during the motor imagery study, data were analyzed in precisely the same ways as the previous motor imagery study. This mirrored analysis provides insight as to how sensorimotor-based features classify these tasks and how close the automatically-selected frequency bands and channels align with our hypotheses.

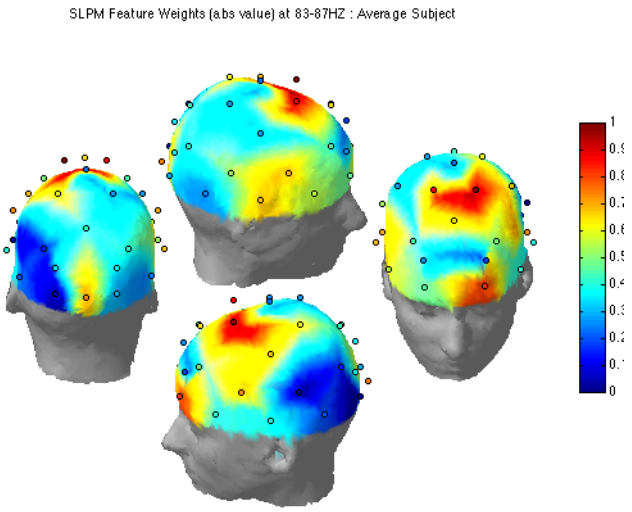

Figure 6.7: Topographic scalp plots for the average subject in a "high" band (here 83-87Hz) with superimposed feature weights for each channel as generated by the SLPM joint classification and feature selection algorithm. The selected channels are mostly around frontal and SMA regions with a preference to the left hemisphere. Another favored electrode is directly above the right eye and may indicate that this channel is driven by noise or artifacts.

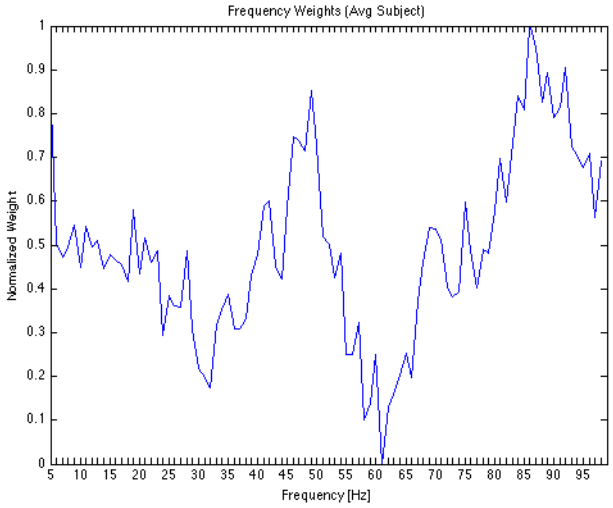

Figure 6.8: Min-max normalized weights for each frequency in the cognitive experiment using joint automated feature selection and classification with SLPM. Note that the selected frequencies (strongest weights) lie between 8-11Hz and 15-22Hz and align with the expectation of sensorimotor rhythm importance.

The results indicate that these two tasks can indeed be well-distinguished with rates of up to 67%. As in the motor imagery study, only several subjects seemed to perform well whereas the others all had classification rates at and around chance level. Using the SLPM joint classification and feature selection algorithm "important" channels and frequencies in the classification of these signals were distinguished. Though the channel selection does not directly align with the hypothesis that the most important channels are around Broca's area and the temporal-parietal junction, this can explained by previous research indicating that cognitive processes are highly distributed. In short, this experiment provides a first crack at analyzing these types of cognitive data in a BCI framework. Important frequencies were found to come in three wide bands: a μ band around 9-13Hz, a βband around 21-30Hz, a γband around 45-55Hz, and a high γ band around 85-95Hz. There are plenty of improvements to be made and alternative ways of analyzing the data, the most noteworthy of which is further investigation into how to represent these type of EEG signals for discriminative purposes.

Chapter 7

Discussion

We now enter a more in-depth discussion of the results presented from the two experiments. We begin this discussion with general observations that apply to both of the experiments and their results. This covers topics such as how we can compare data across the experiments and the types of results should generally expect from any BCI-like experiment. Next, we discuss some of the results in greater detail for both the motor imagery and cognitive experiments and refer to several analytical techniques that were left out of the above results sections. Finally, a number of potential future research directions are presented as motivation for continued work. Many of the following discussions have their origins in the results sections of the two experiments. It should be thus noted that this chapter's flow is a bit different than preceding chapters in the sense that many of the following discussions are only related by a general categorical topic and do not cleanly flow from one to another.

7.1 General Observations

Comparing Classification Results

Cross Validation Results

Though cross validation provides a way of generalizing the classification error by averaging classification performance over several splits of the data, the outgoing results can vary each time depending on the data splits. The variance in the results can lead to inconsistencies while comparing across classifier type or feature representation. Throughout this thesis great care has been taken to provide an accurate estimation of classification performance. One other possibility is to use leave-one-out (LOO) cross validation as described in Section 4.1. LOO cross validation provides more "fair" intra-subject results while varying classifiers and feature representations as each sample is used in both the training set and the test set. For sake of computational time, it was opted to use 10-fold cross validation in place of LOO cross validation. LOO cross validation is, however, available in The Singletrial Matlab Toolbox.

Comparing Results Across The Two Experiments

One of the interesting aspects of mirroring the two experiments and using the same subjects is that we can make cross-experimental observations. Perhaps one of the most interesting concepts to study is how individual subjects performed at the two sets of different tasks. By comparing Tables 5.3 and 6.3 we can clearly see that those subjects who demonstrated the best classification performance with motor imagery tasks were not necessarily the ones responsible for the highest classification performance between the cognitive tasks. These differences beg the question of what constitutes an ideal candidate for a given set of tasks. That is, how can we select a set of tasks that are well-suited to each subject? Unfortunately the picture is more complex than this. Since the two experiments were done on separate days, subject fatigue, mental state, and concentration can also play a role. In any case, the evidence points to no direct correlation between those who are talented at performing motor imagery and those who are able to perform the ISC and OBT tasks.

Classification Results: Is This What We Expect?

A quick glance at the classification rates of the majority of the subjects for the two experiments can be discouraging. It seems that only a few subjects have the ability to be well classified for the two experiments whereas others show nothing more than chance-like levels. This question of what to expect in terms of classification results for a general set of naïve BCI users has been explored in various contexts . Large studies of show only about 10% of naïve BCI users can achieve classification rates greater than 80% on their first try. The majority of the users, however, achieve classification rates of only 55-65%. It is also claimed that a certain population of users is "BCI-illiterate", that is, they are forever incapable of using BCIs for control. These facts encourage the results from the present work which show that 2-3 out of 8 subjects could control the BCI with success rates approximately between 75-90%. The second experiment shows that for 2-3 users out of a small group of 6 subjects could control the BCI with 60-70% accuracy.

Because of these well-known differences amongst naïve subjects and their abilities to control a BCI, often times a series of pre-screening sessions are performed across many subjects, amongst whom only the top three or four subjects are invited for continued participation [28]. This tactic generally produces better classification results as the screening process takes out poorer subjects and results only reflect the best 25-35% of the results. Appendix A.2 provides classification results for the best 3 subjects to further demonstrate that the motor imagery results compare with previous findings.

Computational Time and Offline Analyses

All of the analyses in this work took place offline. When considering a move to an online setup the data structures and algorithms used for classification play an important role in the interactivity of classification results. Subjectively speaking, the Linear Fisher Discriminant classifies data significantly faster than the Sparse Linear Programming Machine which must solve a linear optimization problem to obtain its decision hyperplane. For simple, low-dimensional feature vectors such as the Graz-like features used in the motor imagery experiment and the hand-crafted features used in the cognitive experiment, the difference in performance between LFD and SLPM is marginal. On the

other hand, the respective time it takes for each algorithm to come up with its decision notably differs.

It should be noted that increasing the complexity of the feature representation, such as the move to the "all-channel" representation, highlights SLPM's sophistication in distinguishing higher-dimensional data. All of the top results found in this paper were found by using SLPM and a higher-order feature representation. Though SLPM finds a better dividing hyperplane than LFD in higher-dimensional data, it still has its limitations. As one increases the "difficulty" of the problem, that is, when the two classes' feature representations are not unique, SLPM increases in the computational time it takes to find a solution. At the limit, the optimization problem cannot be solved and one simply cannot use SLPM for such complex machine learning problems. The reality of this consequence played a role in reducing the 64 captured electrodes to 32 electrodes in the cognitive experiment (see Section 7.3further below).

7.2 Motor Imagery Experiment

The results provided for the motor imagery experiment proved to be non-controversial. That is, classification performance was in the expected range for naïve BCI users and the automatically selected features corresponded well to previous findings. Instead of repeating the concordant findings, this section describes a couple of discussion points started in the section and covers a couple side-issues left out of the initial analysis.

Preprocessing Observations

It should be mentioned that other preprocessing techniques were tried before selecting the current set of temporal preprocessing strategies. The preprocessing is heavily dependent upon how you eventually wish to classify the data. First, zero-mean normalizing the data had very little effect on the classification performance. This can be easily explained by looking at the types of classifiers that were used in the analysis. A zero-mean normalization simply causes the data to linearly shift and both LFD and SLPM can adapt their decision hyperplanes to adapt to linear transformations of the underlying data. On the other hand, unit variance normalization had a dramatically negative effect on classification performance. The unit variance is a nonlinear transformation of the underlying data and is not adapted to by the LFD and SLPM models. Unit variance normalization is generally used if the data are going to be submitted to a neural-network classification algorithm such as Learning Vector Quantization (LVQ) or Multi-layer Perceptrons (MLPs).

Classification Results

One of the key findings is in comparing the classification performance using the same classifiers across different feature representations. The results show how absolutely critical signal representation is for distinguishing between the two tasks. A second factor is the strength and sophistication of the classification technique. These are both highlighted by the initial results which show classification performance for Graz-like BCI features. Both LFD and SLPM perform at about the same

level given this feature representation. The benefits of the complex distinctions that the SLPM can make are not noticed in this context. However, simply moving to the "all-channel" feature representation showed a marked increase in performance (up to 15-20%) for SLPM classification. LFD performance also tended to increase by changing feature representations, but of critical note is that five of the eight subjects had their best-found classifications as found by SLPM in the second feature representation. The difference in performance with respect to the two feature representations shows us how fragile of a state of affairs things are. Manual selection of the expected channels and frequencies (ie: Graz-like BCI features) had little benefit according to the results of this experiment.

Another particular result is worth comment: the classification results for subject four using the Graz-like BCI feature representation. The model built for subject 4 was consistently "wrong" in its discriminative output. Since the there are only two classes to distinguish between, this consistency was exploited and outgoing classifier labels were inversed to obtain results. The reasons behind this behavior remain unknown. Futher BCI experiments must be done with this subject to determine if it was a session-based quirk or if the subject should be excluded.

Automated Feature Selection

Selected Channels
The scalp plots provided in Section definitively show what we expected, namely that the most important channels for classification were the bilateral, motor cortex regions above the hand area of the motor homunculus. These are nicely isolated in the frequencies we expect them to be (e.g. sensorimotor rhythms) and are not present in higher, γband frequencies.

Selected Frequencies
Since a single classification run was performed for each 1Hz bin, the magnitudes in the weight vector for a particular frequency differ amongst frequencies, even for the same subject. To solve for this, some type of normalization has to be performed to make each weight's relative importance on the same scale. For the current study it was decided to divide each weight by the maximum weight for each classification run. That is, for each particular frequency, for each subject, all weights were normalized to be maximally as large as the largest value. Then, the aggregate frequency weight plot was computed by summing these normalized weights for all subjects and then again normalizing the result with a min-max normalization. It is currently unknown what the best way to compare the magnitudes of the weight vectors for the SLPM across classification trials.

Joint Channel/Frequency Selection
Given that we have feature weights for all channels and for all frequencies, we can construct a 2-dimensional portrait of how important each frequency is for each channel. These channel vs. frequency plots were left out of the above analysis because though they exhibit several expected results, namely that the weights are stronger for the channels we expect them to be, there are several incongruencies that could be caused due to the aforementioned weight-normalization issues. A couple of these channel-frequency plots are provided in Appendix A.1.

7.3 Cognitive Experiment

The results from this section are difficult to definitively discuss. Though we have several hypotheses about the important underlying phenomena of the two cognitive tasks investigated in this experiment, there are several deficits in our knowledge. First, the way in which to properly extract features from such tasks for BCI control is unknown. For a first exploration we repeated the representations that worked in the above motor-imagery paradigm. This may be far from the correct way of representing such signals for distinguishing between the two tasks. Next, we have no clear picture of what the important frequencies nor associated neurophysiological regions are important for distinguishing between these two tasks. Both tasks are, however, inspired by a number of studies that have begun investigating the important functional brain regions and the frequency-domain characteristics of the corresponding EEG signals. This second hole in our knowledge signifies that though we have hints of where to look for automated feature selection of channels and frequencies, there is no golden standard to compare results against. Ideally this experiment is an aid in the search for uncovering more information about these two tasks from another perspective, namely through the lens of a BCI framework.

Task Selection

All subjects were asked how difficult the different tasks were. Since all subjects had participated in both experiments, an informal ranking was made where all subjects believed the most difficult task to be OBT followed by left-hand and right-hand motor imagery, and the simplest being the internal-speech command. This informal ranking tells us a lot about task selection for BCI control. It is clear that people do not consciously perform the OBT task on a daily basis, but people do however internally speak for extended periods every day. The familiarity of the task can help users directly immerse in the environment and not become stressed that they are not properly performing the desired task. The dramatic difference in task difficulty between the OBT and ISC tasks could be a point of contention for this study, but these tasks were selected for their neuroscientific relevence.

Though motor imagery provides strong, differentiable, and well-studied EEG signals with particular characteristics, the informal ranking also shows us that motor imagery is not entirely natural either. The utility of moving toward cognitive tasks in BCI control is that subjects can use mental tasks that they are familiar with on a daily basis and that require no training.

Feature Representation

As mentioned above, both the underlying feature representations (*i.e.* the "hand-crafted" and "all-channel" feature representations) are based on the premise that event-related desynchronization (ERD) plays a roll in distinguishing between cognitive tasks. Though this premise has been shown to be true with imagined motor movements, it is not verified for cognitive task use. On the other hand, it is not too unrealistic to imagine that there is a measurable ERD while performing the ISC and OBT tasks. There are certain to be motor components associated with internal speech as well as body placement.

The difficulty of SLPM to distinguish between these two cognitive tasks can provide important insight. The most flagrant sign of the "difficulty" of distinguishing between the two tasks was the inability for SLPM to find a dividing hyperplane for 64 channels. Even though data was captured at 64 channels, it had to be reduced to 32 for analysis. This defeats the initial purpose of exploring at a higher spatial resolution. Furthermore, while using the hand-crafted features it took significantly longer for the SLPM algorithm to solve its optimization problem than it did for the motor imagery task. These incidents point to one of two causes. First, it could be the manner in which the features are extracted from the signals. Using a different feature representation may cause the two tasks to be better separable. More dramatically, it could be the case that the tasks simply contain too much overlap and they are too difficult to segregate regardless of the feature representation.

Classification Results

The average classification performance for the cognitive experiment is lower than in the motor imagery experiment. This can be for a number of reasons including feature representation, task selection, or even the few number of subjects that took place in this experiment. Knowing that only about 12% of naïve BCI users perform motor imagery tasks levels greater than 80%, we would need at least ten subjects before we can statistically expect to see such results. Conversely, there are already three subjects with a classification performance between 60-70%. Though there were no stellar results, these preliminary findings indicate that these two tasks can in fact be well discriminated.

Automated Feature Selection

Selected Channels
The automatically selected channels do not correspond with the loose hypothesis that the most important channels would be over Broca's area and the TPJ. This does not mean, however, that nothing can be gleaned from looking at the results. The low-band features demonstrate that the motor cortex, and areas around Broca's area do play a certain importance in classification distinction. In the 9-13Hz bin, for instance, there is bilateral selection around the frontal interior gyrus (Broca's area). The 21-25Hz bin shows heavy frontal electrode selection at the periphery of the EEG cap. This could be reflective of the cognitive tasks or of an artifact driving the signal. In the middle-band 49-53Hz channel selection is nicely isolated near Broca's area, but again it lies at the periphery and could indicate EMG noise from the neck muscles. Finally, the high frequency band at 83-87Hz shows SMA channels to be the most important in distinguishing. This could be the case as these cognitive tasks may include associated motor activity.

Selected Frequencies
The most important information to be gleaned from the automated feature selection of frequencies is that there are consistently three clusters of frequencies that appear to be important in classification of the ISC and OBT tasks. The first piece of evidence comes from classification performance whose best results lie in three distinct clusters. The next piece of evidence is in the frequency weight plot who shows three distinct peaks. One in SMR bands around 9-27Hz, one in a 45-55Hz range, and one in the high gamma range of 75-90Hz. Further investigation is needed to explore

the potential of these high frequencies in task discrimination of cognitive tasks used in BCI control and to understand what role these frequencies play in the underlying cognitive processes.

7.4 Future Research

In this section we describe a scattershot set of ideas for future research directions and ideas. A number of these ideas have already been put into action, that is, continued research is exploring these concepts.

Data cleaning

In both experiments it was opted to leave all trials in for analysis to gauge how well sophisticated methods could pick through imperfect data. Furthermore, the EOG was captured but not used to detect eye artifacts. It is well known that these artifacts can have detrimental effects on the classifiers, as the models reflect noise instead of the true underlying characteristics of the tasks. In spite of this, the results are successful. It would be informative to manually inspect all trials and "clean" the data to see the effect of artifact-free data on classification performance.

Hyperparameter Search Using Nested Cross Validation

Nested cross validation was referred to a number of times throughout the present work. The SINGLETRIAL MATLAB TOOLBOX has been implemented to support hyperparameter search with a nested cross-validation scheme, but there was no time to analyze the data using this computationally lengthy operation. Ideally nested cross validation will improve classification rates by automatically choosing the model order parameter for the MEM estimation and the regularization parameter for the SLPM. It will be interesting to compare and contrast results for classification and feature selection obtained with the optimal hyperparameters versus those manually set throughout the present work.

A similar idea is to manually tweak the regularization/sparsity hyperparameter for the SLPM to observe how feature selection is affected. Since this controls the sparsity of the weight vectors, it should ideally hone in on the most important channels and frequencies, nullifying those who play a marginal role. The effect of this parameter on the selected features is to be determined.

Cognitive Experiment Improvements

There are a number of possibilities to further improve the classification performance between the ISC and OBT tasks. First, more data should be collected. Having more subjects will allow us to notice the trends. Next, different feature representations (not based on ERD) should be explored. This may permit the data from 64 channels to become classifiable by SLPM. In any case, moving toward finding a representation that allows higher-dimensional feature vectors to be well-classified is of particular import. Finally, investigation into the origin and importance of these higher-order frequency bands in the ISC and OBT tasks is needed.

Towards and Online BCI

One future goal of this project is to additionally move in the direction of an online BCI. This move will change the types of algorithms that are used in the analysis, favoring those who can react in a real-time environment. There are a number of technical challenges involved in building an online BCI and there exist few (if any) non-commercial, online BCIs available to the research community. Again, the burden lies on interested research groups to build and implement an online BCI themselves. Preliminary steps have been taken to move the SINGLETRIAL MATLAB TOOLBOX in this direction.

Recapitulation

In this concluding chapter we have discussed the finer details of the experimental results while making higher-level observations of the outcome of the current work. In general, it has been demonstrated that all of the three goals set forth in the introductory sections were achieved by first building and implementing a full-functioning BCI framework. This framework was put to the test by running a traditional motor imagery-based BCI experiment and using the Single-trial Matlab Toolbox to analyze the results. The results conclusively confirmed previous findings and classification rates of up to 87% were found for simple, sensorimotor-based feature representations. Furthermore, sophisticated feature selection algorithms were used to naïvely select important channels and frequencies while classifying between left-hand and right-hand motor imagery. The selected channels correspond to the predicted neurophysiological areas and the selected frequencies reconfirm previous studies showing the importance of sensorimotor rhythms.

After the success of the motor imagery experiment, a new experiment was proposed that mirrored the experimental design but substituted two cognitive tasks, that of providing an inner-speech command and that of transforming the position one's body, in place of the previous motor imagery tasks. Preliminary results show that these two cognitive tasks can in fact be well-distinguished using a similar feature representation as was used in the motor imagery study with rates of up to 67%. Furthermore, exploratory analysis provides evidence that these cognitive tasks are more spatially distributed across the scalp and that the important frequencies for task distinction lie in three clusters: a μ band around 9-13Hz, a βband around 21-30Hz, a γband around 45-55Hz, and a high γ band around 85-95Hz. These findings provide a solid foundation for the exploration of novel cognitive tasks for control in BCIs and for concurrent use of the BCI as a neuroscientific research tool.

Appendix A

Further Analyses

A.1 Further Directions of Analysis

All-Frequency Approach: Individual Channel, All Frequencies

This is an alternative feature representation in the same spirit as the "All-Channel" representation. The incoming preprocessed data is for one channel. Since we eventually wish to select features at important frequencies, we replicate the signal for a single channel for numerous frequency bands and separately compute a MEM spectral estimation for each desired frequency band. These frequency bands can be arbitrarily defined and are generally 1Hz-wide bands. For instance, for a single channel we separately compute the band power at 8Hz, 9Hz, 10Hz, etc. Now the ERD value for each frequency band at this channel is computed as above. The resulting feature representation is one ERD value per frequency band for a single channel. This third feature representation has been left out of the results sections. Preliminary analyses from this "angle" showed the feature selection algorithms to improperly function. It is hypothesized that using a single channel and multiple frequencies is not in fact the way in which to derive automated features using SLPM. Further investigation is needed.

Channel-Frequency Plots

One additional desired type of analysis was to build a two-dimensional portrait of the weight values corresponding to a particular channel/frequency pair. This analysis was performed by wrapping multiple classification runs around the two latter feature representations. More specifically, for the "all-channel" representation a series of analyses were performed where each individual analysis considered a frequency band of 1Hz for all 32 channels. The frequency of interest was then augmented and analysis for all 32 channels was repeated. The feature weights for a single analysis were normalized by dividing by the maximum value with respect to the current frequency and were stored for later comparison. For the "all-frequency" representation, the inverse was done. A series of analyses were performed with classification of an individual channel represented by 55

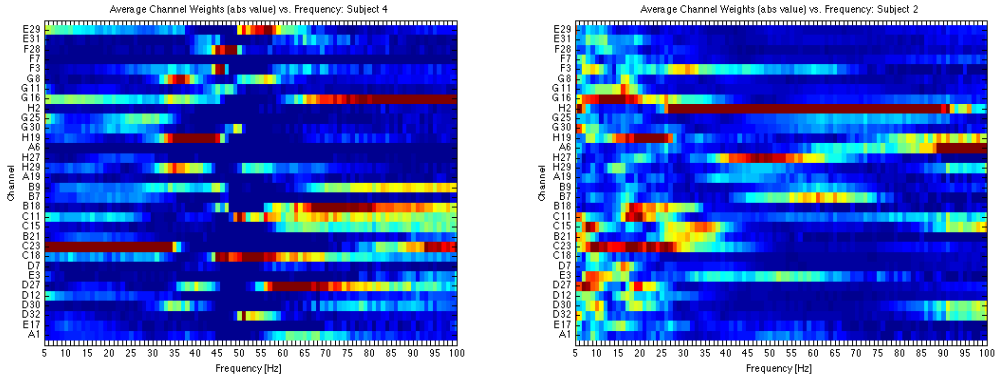

Figure A.1: Two subjects' frequency-channel plots for the motor imagery experiment. Results show that channels at and around C23 and G16 are amongst the most important and have high weight intensities for the expected sensorimotor rhythms.

	MEM Model Order	SLPM Regularization C
10-fold cross validation	3	10

Table A.1: Hyperparameter values for the k-fold cross validation schemes used in the motor imagery and cognitive experiments.

different frequency bands. Weights for each of the frequencies at that channel were normalized by the maximum weight for that run and the next channel was selected and likewise analyzed.

A pair of sample channel-frequency plots are provided in Figure A.1 for the motor imagery experiment. Though it can be seen that the expected channels (C23 and G16 are the equivalent of C3 and C4) have the highest weight intensities, the picture is a little muddied. Also, high weight intensities above 30Hz do not make much analytical sense as we do not expect these areas to be important in the distinguishing of the motor imagery tasks. Other channels have high weight intensities as well and at unexpected frequency bands. For the most part the 9-13Hz and 21-25Hz frequency bands also have high intensities attached to them. One way to potentially improve these plots is to play with the SLPM sparsity parameter in hopes to supress unimportant weights which find their way into being normalized to high values. This idea is proposed in Section 7.4.

A.2 Motor Imagery Experiment

Classification Scheme Parameters

Table A.1 displays all parameters used in the k-fold cross validation classification scheme. 10-fold cross validation was performed by manual selection of parameter values.

	Sub 2	*Sub 7*	*Sub 8*	*Best Subject Avg*
LFD (Graz)	69%	64%	61%	65%
SLPM (Graz)	68%	62%	59%	63%
LFD (Channel)	60%	64%	56%	60%
SLPM (Channel)	**87%**	**72%**	**76%**	**78%**

Table A.2: Classification results for the motor imagery experiment taking the best 3 subjects using two different feature representations. Results extracted from Tables 5.3,5.2, and2. Using the all-channel feature representation and subject-specific frequency bands leads to average classification error of 78% for the top three naïve subjects.

Classification and Automated Feature Selection Results for the Best 3 Subjects

As mentioned above in the discussion Section 7.2, displaying the top candidates for BCI use is a common tactic in the BCI literature. These results are summarized in Table A.2which shows the classification performance rates for the best 3 of 8 subjects. Note that other authors sometimes select even 1 out of 4 (˜25%) of the "best" results for further investigation.

A.3 Cognitive Experiment

Full SLPM Classification Results

A detailed, full report of the classification performance for all 4Hz-wide frequency bands is provided in Table A.3. This shows clearly that important frequencies with respect to classification performance fall within μ, β, and even γregions.

Classification and Automated Feature Selection Results for the Best 3 Subjects

Section 7.1 motivates the common practice of selecting the best BCI candidates for use in further experimentation and for results. These results are summarized in Table A.4which shows the classification performance rates for the best 2 of 6 subjects. Note that other authors sometimes select even 1 out of 4 (˜25%) of the "best" results for further investigation.

A.4 Underlying Neurophysiology: Expectations for Automated Feature Selection

Though the details have been spelled out in the related research sections, here we briefly summarize our expectations of the important channels and frequencies for the classification of the

Frequency Band	*Sub 1*	*Sub 2*	*Sub 4*	*Sub 7*	*Sub 8*	*Sub 9*	*Subject Average*
5-9Hz	52	**63**	57	53	53	60	56
9-13Hz	53	**63**	54	52	55	65	57
13-17Hz	**55**	61	54	51	54	66	57
17-21Hz	**55**	62	55	**55**	54	64	57
21-25Hz	53	**63**	58	53	56	63	58
25-29Hz	52	61	**59**	53	58	61	57
29-33Hz	51	56	58	53	**60**	61	56
33-37Hz	51	55	55	53	59	62	56
37-41Hz	51	55	53	53	59	62	56
41-45Hz	52	53	54	52	59	63	55
45-49Hz	52	53	53	52	59	66	56
49-53Hz	52	49	53	**55**	59	**67**	56
53-57Hz	52	48	50	54	**60**	64	55
57-61Hz	51	48	51	53	59	64	54
61-65Hz	51	47	53	53	59	64	55
79-83Hz	49	49	51	53	**60**	66	55
83-87Hz	51	50	52	53	**60**	**67**	55
87-91Hz	51	49	51	52	**60**	66	55
91-95Hz	50	54	53	50	59	66	55
95-99Hz	52	53	52	51	58	66	56

Table A.3: Per-subject SLPM classification results (in %) for the cognitive experiment between 5-100Hz for 4Hz-wide frequency bands using the "All-Channels" feature represntation. Subject IDs refer to the same subject as used in the motor imagery study. Values in bold indicate the frequency band at which the highest classification performance was obtained. Frequency bands in which highest classification performance was achieved for some subject are additionally highlighted.

	Sub 2	*Sub 9*	*Best Subject Avg*
LFD (Hand-crafted)	57	61	59
SLPM (Hand-crafted)	51	66	58
LFD (All-Channel)	62	58	60
SLPM (All-Channel)	**63**	**67**	**65**

Table A.4: Classification results (in %) for the cognitive experiment taking the best 2 subjects of 6 using two different feature representations. Results extracted from Tables 6.3,1, andA.3. Using the all-channel feature representation and subject-specific frequency bands leads to average classification error of 65% for the top two naïve subjects.

left-hand/right-hand motor imagery and the ISC/OBT tasks. The expectations for the cognitive tasks are more poorly-founded and are more like guides than rules.

Motor Imagery Experiment

Channels
The channels on or around the motor cortex and the supplementary motor area (SMA). Particularly, left-hand motor imagery should cause right-hemispheric activation around the "hand" area of the motor homunculus and vice versa for right-hand motor imagery.

Frequencies
Frequencies centered around the sensorimotor rhythms. Particularly those in the 8-12Hz, 16-24Hz and 36-40Hz regions.

Cognitive Experiment

Channels
For the inner-speech command (ISC) task we expect regions around Broca's area (left-hemisphere frontal insular gyrus) as well as a potential contralateral activation. Motor areas and SMA may also be activated as thought of moving ones lips and vocal chords may cause motor imagery. For the own-body transformation we expect channels on or around the right-hemispheric temporal-parietal junction to be most important.

Frequencies
The important frequencies are less studied in these tasks. Cognitive tasks in general have shown coherence in high γ bands, so these high-frequency bands may be of importance. If there is spillover motor imagery the sensorimotor rhythms could also play an important role.

Appendix B

Implementation: The Singletrial Matlab Toolbox

The Singletrial Matlab Toolbox is the BCI framework implemented in Matlab for the current project. All data analyses and charts were produced using this toolbox. Great care was taken while engineering the toolkit to be flexible for future additions of preprocessing, feature extraction, task selection, classification, and plotting mechanisms. Detailed documentation, code, examples, and scripts can be found on an accompanying CD.

Bibliography

[1] C. Anderson and Z. Sijercic. Classification of eeg signals from four subjects during five mental tasks. *Solving Engineering Problems with Neural Networks: Proceedings of the Conference on Engineering Applications in Neural Networks*, pages 507–414, 1996.

[2] F. Babiloni, F. Cincotti, L. Lazzarini, J. Millan, J. Mourino, M. Varsta, J. Heikkonen, L. Bianchi, and M. Marciani. Linear classification of low-resolution eeg patterns produced by imagined hand movements. *Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Neural Systems and Rehabilitation]*, 8(2):186–188, 2000.

[3] J. Bayliss and D. Ballard. A virtual reality testbed for brain-computer interface research. *Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Neural Systems and Rehabilitation]*, 8(2):188–190, 2000.

[4] O. Blanke, C. Mohr, C. Michel, A. Pascual-Leone, P. Brugger, M. Seeck, T. Landis, and G. Thut. Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *Journal of Neuroscience*, 25(3):550, 2005.

[5] B. Blankertz, G. Dornhege, M. Krauledat, K. Müller, V. Kunzmann, F. Losch, and G. Curio. The berlin brain-computer interface: Eeg-based communication without subject training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):147–152, 2006.

[6] M. Blankertz, B. Krauledat, K. Müller, G. Dornhege, V. Kunzmann, F. Losch, and G. Curio. *Toward Brain-Computer Interfacing*, chapter The Berlin Brain-Computer Interface: Machine Learning-Based Detection of User Specific Brain States, pages 85–102. The MIT Press, 2007.

[7] J. Carmena, M. Lebedev, R. Crist, J. O'Doherty, D. Santucci, D. Dimitrov, P. Patil, C. Henriquez, and M. Nicolelis. Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biology*, 1(2):e2, 2003.

[8] E. Curran and M. Stokes. Learning to control brain activity: A review of the production and control of eeg components for driving brain–computer interface (bci) systems. *Brain and Cognition*, 51(3):326–336, 2003.

[9] E. Curran, P. Sykacek, M. Stokes, S. Roberts, W. Penny, I. Johnsrude, and A. Owen. Cognitive tasks for driving a brain-computer interfacing system: a pilot study. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Rehabilitation Engineering]*, 12(1):48–54, 2004.

[10] A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.

[11] S. Fitzgibbon, K. Pope, L. Mackenzie, C. Clark, and J. Willoughby. Cognitive tasks augment gamma eeg power. *Clinical Neurophysiology*, 115(8):1802–1809, 2004.

[12] C. Guger, H. Ramoser, and G. Pfurtscheller. Real-time eeg analysis with subject-specific spatial patterns for a brain-computer interface (bci). *Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Neural Systems and Rehabilitation]*, 8(4):447–456, 2000.

[13] E. Gysels and P. Celka. Phase synchronization for the recognition of mental tasks in a brain-computer interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Rehabilitation Engineering]*, 12(4):406–415, 2004.

[14] T. Hinterberger, N. Neumann, M. Pham, A. Kübler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor, and N. Birbaumer. A multimodal brain-based feedback and communication system. *Experimental Brain Research*, 154(4):521–526, 2004.

[15] T. Hinterberger, F. Nijboer, A. Kübler, M. T., A. Furdea, U. Mochty, M. Jordan, J. Mellinger, T. Lal, J. Hill, B. Schölkopf, M. Bensch, W. Rosenstiel, M. Tangermann, G. Widman, C. Elger, and N. Birbaumer. *Toward Brain-Computer Interfacing*, chapter Brain-Computer Interfaces for Communication in Paralysis: A Clinical Experimental Approach, pages 43–64. The MIT Press, 2007.

[16] A. Kübler and K. Müller. *Toward Brain-Computer Interfacing*, chapter An Introduction to Brain-Computer Interfacing, pages 1–25. The MIT Press, 2007.

[17] F. Lotte, M. Congedo, A. Lécuye, F. Lemarche, and B. Arnaldi. A review of classification algorithms for eeg based brain computer interfaces. *J. Neural Engineering*, 4:R1–R13, 2007.

[18] S. Makeig, A. Bell, T. Jung, T. Sejnowski, et al. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, 8:145–151, 1996.

[19] D. McFarland, L. McCane, S. David, and J. Wolpaw. Spatial filter selection for eeg based communication. *Electroencephalography and Clinical Neurophysiology*, 103:386–394, 1997.

[20] J. Millán, P. Ferrez, and A. Buttfield. *Toward Brain-Computer Interfacing*, chapter The IDIAP Brain-Computer Interface: An Asynchronous Multiclass Approach, pages 103–110. The MIT Press, 2007.

[21] K.-R. Müller, M. Krauledat, G. Dornhege, G. C. G, and B. Blankertz. Machine learning techniques for brain computer interfaces. *Biomedical Technology*, 49:11–22, 2004.

[22] E. Niedermeyer. The normal eeg of the waking adult. *Electroencephalography: basic principles, clinical applications, and related fields. 5th ed. Philadelphia: Lippincott Williams & Wilkins*, pages 167–92, 2005.

[23] W. Penny, S. Roberts, E. Curran, and M. Stokes. Eeg-based communication: a pattern recognition approach. *IEEE Transactions on Rehabilitation Engineering*, 8(2):214–215, 2000.

[24] G. Pfurtscheller, D. Flotzinger, and J. Kalcher. Brain-computer interface: a new communication device for handicapped persons. *Journal of Microcomputer Applications*, 16(3):293–299, 1993.

[25] G. Pfurtscheller and F. Lopes da Silva. Event-related eeg/meg synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110(11):1842–1857, 1999.

[26] G. Pfurtscheller, G. Müller, G. Bernhard, R. Scherer, R. Leeb, C. Brunner, C. Keinrath, G. Townsend, M. Naeem, F. Lee, D. Zimmermann, E. Höfler, A. Schlögl, C. Vidaurre, S. Wriessnegger, and C. Neuper. *Toward Brain-Computer Interfacing*, chapter Graz-Brain-Computer Interface: State of Research, pages 65–84. The MIT Press, 2007.

[27] G. Pfurtscheller and C. Neuper. Motor imagery activates primary sensorimotor area in humans. *Neuroscience Letters*, 239(2-3):65–68, 1997.

[28] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer. Eeg-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology*, 103(6):642–651, 1997.

[29] G. Pfurtscheller, C. Neuper, G. Müller, B. Obermaier, G. Krausz, A. Schlögl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, et al. Graz-bci: state of the art and clinical applications. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Rehabilitation Engineering]*, 11(2):1–4, 2003.

[30] M. Pregenzer and G. Pfurtscheller. Distinction sensitive learning vector quantization (dslvq) application as a classifier based feature selection method for a brain computer interface. *Artificial Neural Networks, 1995., Fourth International Conference on*, pages 433–436, 1995.

[31] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*, chapter Power Spectrum Estimation by the Maximum Entropy (All Poles) Method, pages 572–575. Cambridge University Press, 1992.

[32] E. Ryding, B. BraÅdvik, and D. Ingvar. Silent speech activates prefrontal cortical regions asymmetrically, as well as speech-related areas in the dominant hemisphere. *Brain and Language*, 52(3):435–451, 1996.

[33] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *Biomedical Engineering, IEEE Transactions on*, 51(6):1034–1043, 2004.

[34] E. Sellers, D. Krusienski, D. McFarland, and J. Wolpaw. *Toward Brain-Computer Interfacing*, chapter Noninvasive Brain-Computer Interface Research at the Wadsworth Center, pages 31–42. The MIT Press, 2007.

[35] R. Sitaram, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer. fmri brain-computer interface: A tool for neuroscientific research and treatment. *Computational Intelligence and Neuroscience*, 2007(2), 2007.

[36] A. Vallabhaneni, T. Wang, and B. He. *Neural Engineering*, chapter Brain Computer Interface, pages 85–122. Dordrecht: Kluwer, 2005.

[37] J. Wolpaw, N. Birbaumer, W. J. Heetderks, D. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. Quatrano, C. Robinson, and T. Vaughan. Brain computer interface technology: A review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8:164–173, 2000.

[38] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan. Brain computer interfaces for communication and control. *Clinical Neurophysiology*, 113:767–791, 2002.

[39] J. Wolpaw, D. McFarland, G. Neat, and C. Forneris. An eeg-based brain-computer interface for cursor control. *Electroencephalogr Clin Neurophysiol*, 78(3):252–9, 1991.