



HELSINKI UNIVERSITY OF TECHNOLOGY  
Department of Electrical and Communications Engineering

**Pasi Jylänki**

# **Classification of single-trial EEG for online brain-computer interface**

In partial fulfillment of the requirements for the degree of Master of Science

Supervisor: Academy Professor Kimmo Kaski

Instructor: Dr. Tech. Jukka Heikkonen

Espoo 5th June 2006

<b>Author:</b>	Pasi Jylänki	
<b>Title:</b>	Classification of single-trial EEG for online brain-computer interface	
<b>Date:</b>	5th June 2006	<b>Number of Pages:</b> 46
<b>Department:</b>	Department of Electrical and Communications Engineering	
<b>Professorship:</b>	S-114, Computational Engineering	
<b>Supervisor:</b>	Academy Professor Kimmo Kaski	
<b>Instructor:</b>	Dr.Tech. Jukka Heikkonen	
<p>A brain-computer interface (BCI) classifies brain activity during different tasks into different categories. BCIs could enable motor-disabled persons to control a computer application. The few existing online BCI studies with good results have required several weeks or even months of training. Furthermore, only very few online experiments have been done with motor-disabled patients. In this thesis a new online BCI approach based on single-trial classification of electroencephalogram (EEG) is presented. The aim was to construct a system that could be used without extensive training. Online experiments with ten healthy subjects and five tetraplegics were carried out to test whether satisfactory performance could be achieved during a single experiment. The subjects had no previous experience of BCIs.</p> <p>Healthy subjects performed real left or right finger movements and tetraplegics attempted hand movements to move a circle on the screen. Brain activity was measured with EEG. Recognition of the different tasks was based on large number of features related to different brain activation patterns. During the experiments feature transformations and the classifier were updated after each prediction using the correct class information. This enabled realistic feedback to the subject from the beginning of the experiment. Thus, both the subjects and the system were able to learn throughout the experiments.</p> <p>Six out of the ten healthy subjects achieved at least 75% classification accuracy after only ca. 20 min training. This has been suggested as the minimum accuracy required to control assisting applications. One healthy subject achieved information transfer rate of 21 bits per minute, which is comparable to the best results obtained after extensive training with existing online BCIs. Unfortunately, motor-disabled subjects performed significantly worse. Only two of them exceeded 60% accuracy. This indicates that methods that are effective with healthy subjects do not necessarily work with motor-disabled. Future research should concentrate more on the potential user group.</p>		
<b>Keywords:</b> electroencephalography, brain-computer interface, classification of brain signals, movement related potentials, tetraplegic patients.		

<b>Tekijä:</b>	Pasi Jylänki	
<b>Otsikko:</b>	Yksittäisten tehtävien tuottamien EEG-signaalien luokitteluun perustuva aivokäyttöliittymä	
<b>Päivämäärä:</b>	5. kesäkuuta 2006	<b>Sivumäärä:</b> 46
<b>Osasto:</b>	Sähkö- ja tietoliikennetekniikan osasto	
<b>Professori:</b>	S-114, Laskennallinen tekniikka	
<b>Valvoja:</b>	Akatemiaprofessori Kimmo Kaski	
<b>Ohjaaja:</b>	TkT Jukka Heikkonen	
<p>Aivokäyttöliittymä luokittelee aivotoimintaa eri tehtävien aikana eri kategorioihin. Sen avulla liikuntakyvyttömät henkilöt voisivat ohjata tietokonesovelluksia. Harvoissa hyviin tuloksiin johtaneissa tutkimuksissa koehenkilöt ovat joutuneet harjoittelemaan useita viikkoja tai peräti kuukausia. Lisäksi kokeita tehdään valitettavan usein ainoastaan terveillä liikuntakykyisillä henkilöillä. Tässä diplomityössä esitellään uusi tosiaikaisesti toimiva aivokäyttöliittymä, joka perustuu yksittäisiin ajatuskäskyihin liittyvien elektroenkefalografia (EEG) -signaalien luokitteluun. Tavoitteena oli suunnitella systeemi, jonka käyttö ei vaatisi pitkää harjoittelua. Järjestelmällä tehtiin kaksi koetta, joilla selvitettiin voidaanko yhden kokeen aikana saavuttaa tyydyttävä suorituskky. Koehenkilöinä oli kymmenen tervettä ja viisi nelirajahalvaantunutta, joilla ei ollut aiempaa kokemusta aivokäyttöliittymistä.</p> <p>Terveet koehenkilöt tekivät vasemman ja oikean sormen liikkeitä siirtääkseen palloa tietokoneruudulla. Vastaavasti potilaat yrittävät liikuttaa käsiään. Aivojen toimintaa liikkeiden aikana mitattiin EEG:llä. Eri tehtävien tunnistus perustui suureen joukkoon eri aivoaktivaatioihin liittyviä piirteitä. Kokeissa sekä muunnokset piirreavaruudessa että luokittelija päivitettiin jokaisen ennustuksen jälkeen käyttäen oikeata luokkatieta. Tämä mahdollisti todenmukaisen palautteen antamisen koehenkilöille jo kokeen alussa. Täten sekä koehenkilöt että systeemi pystyivät oppimaan alusta asti.</p> <p>Kuusi kymmenestä terveestä koehenkilöstä saavutti vähintään 75% luokittelutuloksen noin kahdenkymmenen minuutin harjoittelun jälkeen. On ehdotettu, että tämä olisi vähimmäistarkkuus avustavan laitteen ohjauksessa. Yksi terve koehenkilö saavutti 21 bits/min laskennallisen informaationsiirtokapasiteetin. Tämä on verrattavissa pitkäkestoisella harjoittelulla saatujen tuloksien parhaimmiston. Halvaantuneet koehenkilöt suoriutuivat huomattavasti huonommin. Vain kaksi heistä ylitti 60% tuloksen. Tämä osoittaa, että terveillä koehenkilöillä hyvin toimivat menetelmät eivät välttämättä toimi liikuntakyvyttömällä. Tulevaisuudessa tutkimuksen pitäisi keskittyä enemmän pääasialliseen käyttöjärjelmään.</p>		
<b>Keywords:</b> elektroenkefalografia, aivokäyttöliittymät, aivosignaalien luokittelu, motoriset herätevasteet, nelirajahalvaantuneet potilaat.		

# Foreword

This work was done in the Laboratory of Computational Engineering (LCE) at the Helsinki University of Technology (HUT) as a part of the EU-funded project, Non Invasive Brain Interaction with Robots - Mental Augmentation through Determination of Intended Action (MAIA). LCE was selected as the Academy of Finland's centre of excellence for years 2006-2010. The supervisor of this work was Academy Professor Kimmo Kaski and the instructor Academy Fellow, Doctor of Technology Jukka Heikkonen.

Here I would like to thank Prof. Kimmo Kaski for the opportunity to work in this fascinating project. Also I would like to thank all the partners of this project with whom I have had the privilege to work with. I would like to thank Dr.Tech Jukka Heikkonen for instructing my work and Dr.Tech. Harri Valpola for giving invaluable ideas. I am also deeply indebted to M.Sc. Janne Lehtonen for his hard work to make the online BCI operational and the online measurements possible. Especially I would like to thank M.Sc. Laura Kauhanen for helping to write this thesis and also for her expertise regarding BCI research more generally. I also wish to thank M.Sc Tommi Nykopp for his comments and advices regarding statistical modeling issues as well as being pleasant company in the night of Helsinki. And last but not least, I would like to express my gratitude to my parents and my friends who helped me whenever I needed it.

In Espoo, 5th June 2006

Pasi Jylänki

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The brain . . . . .	3
1.2	Electroencephalography . . . . .	3
1.3	Sensorimotor cortex . . . . .	5
1.4	Review of BCI systems . . . . .	7
<b>2</b>	<b>Materials and Methods</b>	<b>10</b>
2.1	Experiments . . . . .	11
2.1.1	Subjects . . . . .	11
2.1.2	Experimental setup . . . . .	11
2.1.3	Recording . . . . .	14
2.2	Features . . . . .	15
2.2.1	Computation of features . . . . .	15
2.2.2	Feature selection . . . . .	21
2.3	Online classifier . . . . .	23
2.3.1	Whitening and dimensionality reduction . . . . .	24
2.3.2	Linear discriminants for classification . . . . .	26
2.4	Evaluation of the BCI performance . . . . .	30
<b>3</b>	<b>Results</b>	<b>32</b>
3.1	Classification results and BCI performance . . . . .	32
3.2	Results of the feature selection . . . . .	35
3.3	Correlation matched ERPs for patients . . . . .	37
<b>4</b>	<b>Discussion</b>	<b>39</b>

# List of Figures

1.1	Different parts of the brain . . . . .	3
1.2	International 10-20 system for EEG electrodes . . . . .	4
1.3	The rhythmic activities of the brain . . . . .	5
1.4	The human motor cortex . . . . .	6
1.5	Operational present-day independent BCI approaches . . . . .	8
2.1	Online feedback given to the subjects . . . . .	12
2.2	Structure of the online BCI experiment . . . . .	13
2.3	BCI experiment with a tetraplegic patient . . . . .	14
2.4	Computation of features . . . . .	20
2.5	Average event related potentials for both classes . . . . .	22
2.6	Kolmogorov-Smirnov test statistic . . . . .	22
2.7	Kolmogorov-Smirnov test statistic for multiple features . . . . .	23
2.8	Operation of the online classifier . . . . .	24
3.1	Averaged ERPs for S1-S5 . . . . .	36
3.2	Averaged ERPs for S6-S10 . . . . .	36
3.3	Averaged ERPs for P1-P5 . . . . .	37
3.4	Correlation matched ERPs . . . . .	38

# List of Tables

2.1	Patient information for tetraplegic subjects . . . . .	11
3.1	Classification accuracies [%] for subjects S1 - S10 . . . . .	33
3.2	Bitrates for subjects S1 - S10 . . . . .	33
3.3	Classification accuracies for patients P1 - P5 . . . . .	34
3.4	Bitrates for patients P1 - P5 . . . . .	34
3.5	Application performance S1 - S10 . . . . .	35
3.6	Application performance P1 - P5 . . . . .	35

# Abbreviations

BCI	Brain computer interface
DFT	Discrete Fourier transform
EEG	Electroencephalography
EOG	Electro-oculogram
ERP	Event-related potentials
FFT	Fast Fourier transform
fMRI	Functional magnetic resonance imaging
MEG	Magnetoencephalography
P300	Positive peak in the ERP 300 ms after an event occurred
PCA	Principal component analysis
SCP	Slow cortical potentials
VEP	Visual evoked potential

# Chapter 1

## Introduction

In the last ten years, much effort has been put into research exploring whether signals measured from the brain can be used to control some computer application. This kind of communication systems are called brain-computer interfaces (BCIs) [44]. So far the existing BCI systems are able to interpret reliably only few, less than six, commands, and only rather simple applications can be controlled [44]. Thus, the main user group for BCIs are motor-disabled persons who are unable to move their hands. This group includes e.g. patients with spinal cord injuries as well as patients with motor neuron diseases like Amyotrophic Lateral Sclerosis (ALS). In the most recent studies, motor disabled have been able to control a hand orthosis [30], move a cursor on the screen [45] and choose letters from virtual keyboard [35].

Brain activity can be measured both invasively, e.g. electrode implants in the brain tissue, and noninvasively from the surface of the scalp. Invasive methods provide more accurate spatial information about the brain activity, even activity from single brain cell can be detected [44], but placing intracranial sensors require expensive surgeries that can be dangerous. So far most invasive BCI research has been conducted with monkeys [25]. Noninvasively brain activity can be measured with e.g. electroencephalography (EEG), magnetoencephalography (MEG), near infrared spectroscopy (NIRS), and functional Magnetic Resonance Imaging (fMRI). Most BCIs use electrical activity of the brain measured from the scalp with EEG because the measurement devices are much smaller and less expensive than with the other methods [44].

Many BCI studies have been made offline, i.e. data is first collected in a separate recording session and all the signal processing and classification is done afterwards. In an online experiment subjects are given real-time feedback about the performance of the BCI. With the help of online feedback the subjects learn to control the BCI more efficiently with their brain activity [45]. Also more generally, feedback is essential when learning any new task, consider e.g. learning to drive a bike without a bike. In addition to the need for more online studies, most BCI studies are made with healthy subjects. More emphasis should be put to the potential user group, motor disabled patients. Furthermore, subjects often need extensive training periods of several weeks or even months to achieve acceptable control of the BCI [19, 30, 45].

In this thesis a new online BCI approach based on single-trial classification of EEG is presented. The aim was to construct an online system that could be used effectively without extensive training. Online experiments were carried out with both healthy subjects and paralyzed patients to test whether satisfactory performance could be achieved during a single experiment. The classifier was trained online after each task performed by the subject. Thus, online feedback could be given to subjects right from the beginning of the experiment to shorten training periods. The performance of the subjects are compared with those of other BCI groups.

The rest of this chapter provides a brief introduction to some concepts required to understand the methods used in the experiments. The chapter ends with a short review of operational present-day BCI approaches. Detailed description of the online BCI and the experiments are presented in chapter 2. Results of the experiments as well as some offline analysis of the collected data are presented in chapter 3. Chapter 4 summarizes the discussions and final conclusions about the results.

## 1.1 The brain

The human brain consist of three parts, cerebrum, cerebellum, and brain stem [33]. The cerebellum controls the excution of learned movements and the brain stem connects the cerebrum to the spinal cord. The cerebrum is divided into a left and right hemisphere. Both hemispheres consist of four lobes, frontal, parietal, occipital and temporal, see figure 1.1. Each hemisphere is related to the opposite side of the body. For example, when left hand moves related activity can be seen in the right hemisphere. The outer surface of the cerebrum is called a cortex which is a thin layer of grey matter, i.e. cell bodies.

The different senses, visual, auditory, somatosensory, olfactory, and sense of taste, are processed in different parts of the cerebrum, e.g. the visual cortex is located in the posterior part of the occipital lobe [33]. Our main focus is to measure activity both from the primary somatosensory and motor cortex, which are more commonly known as the sensorimotor cortex. The primary motor cortex is situated in the posterior part of the frontal lobe, anterior of the somatosensory cortex.

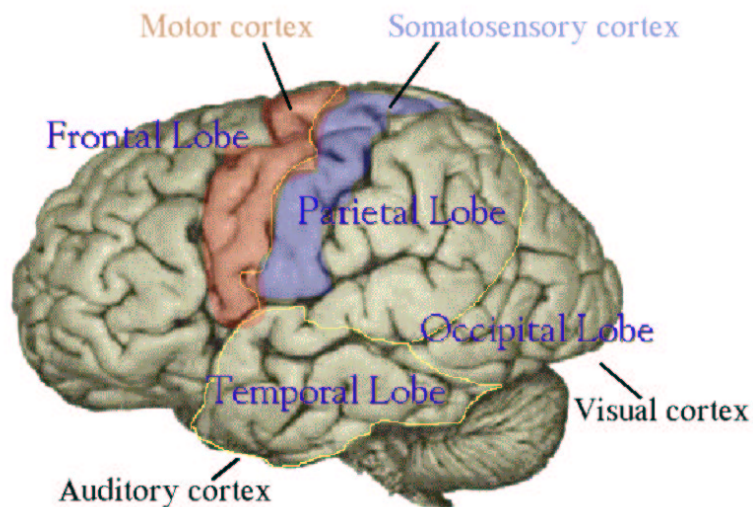


Figure 1.1: Different parts of the brain. Modified from [41].

## 1.2 Electroencephalography

The brain consists of hundreds of billions of neurons, i.e. brain cells, each connected to ca. 10000 other neurons [33]. The neurons communicate with each other

by sending nerve impulses. Simultaneous activity of thousands of neurons generate currents that can be measured from the surface of the scalp with electroencephalography (EEG). The most important neurons for our purpose are the pyramidal neurons of the cortex, because they lie perpendicular to the cortical surface and their activation can therefore be detected with EEG [26].

In EEG measurements the potential difference, i.e. the voltage, between scalp electrodes is recorded [26]. Measured EEG signals are usually in microvolt range. Usually voltages between several electrode pairs are measured simultaneously. The positions of the electrodes on a measurement cap are standardized according to the international 10-20 system, see figure 1.2. Names of the electrodes correspond to different lobes of the brain. Uneven numbers are on the left hemisphere and even numbers are on the right. For example electrode O1 is located over the left occipital lobe. Usually all electrodes are referenced to one common electrode. This electrode can be located on the nose, behind the ears, or on top of the head. An alternative way is to measure several bipolar electrode pairs without a common reference.

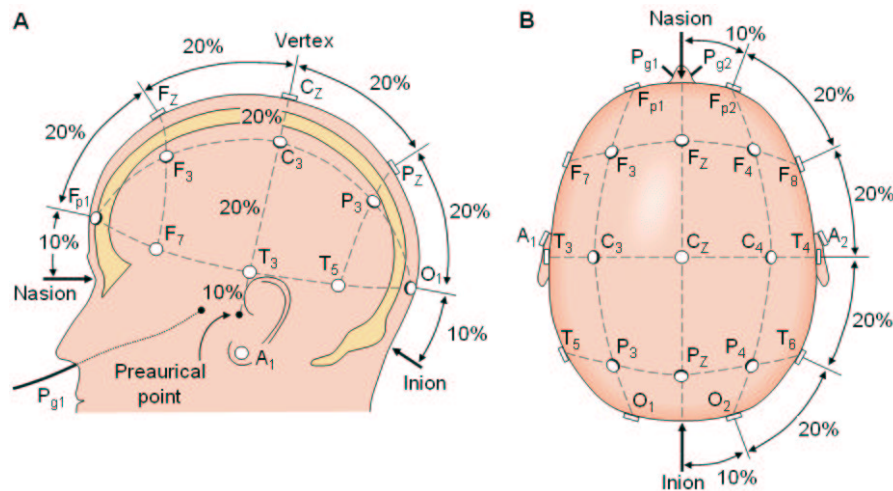


Figure 1.2: Placement of EEG electrodes according to the international 10-20 system [21].

Four different phenomena, brain waves, of different frequency content have been defined in EEG [15], see figure 1.3. These were named in the order they were discovered starting from alpha rhythm. Alpha rhythm can be seen over the occipital lobe when subject is eyes closed, relaxed but awake. Beta rhythm can be detected over the frontal and central regions when subjects are alert. In adults, theta and delta rhythms are mostly seen during sleep.

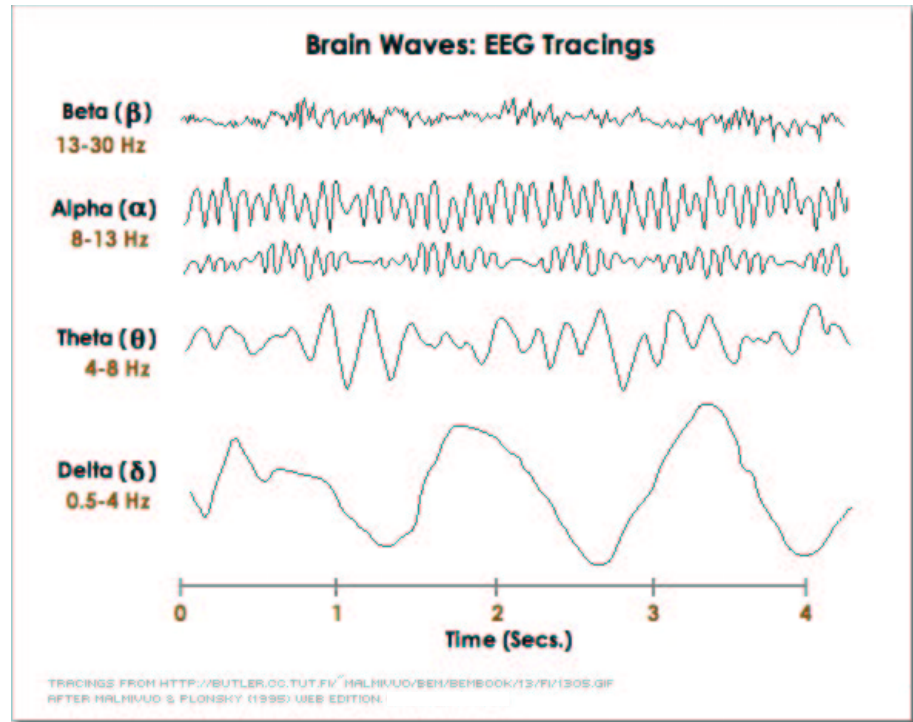


Figure 1.3: The rhythmic activities of the brain [14].

In addition to the brain waves, voltage fluctuations related to some phenomena at a particular time instant, usually called event related potentials (ERPs), can be detected with EEG [32]. In basic research the ERPs are studied by averaging segments of EEG over several occurrences of the same phenomena. For example movement related potentials can be recorded over motor cortex when subject performs hand movements.

### 1.3 Sensorimotor cortex

Many BCI research groups prefer to use signals generated in the sensorimotor cortex to control computer applications [44]. This is a natural approach to develop assisting technologies for motor-disabled persons who no longer use their motor cortex to perform movements. Different parts of the sensorimotor cortex control different voluntary muscular movements of the body [33], see figure 1.4. Activity related to left and right hand movements are more preferable for BCI control because corresponding areas are further away from each other in the motor cortex than e.g. areas controlling legs. Furthermore, muscles that control fine movements, such as hand and face movements, have larger representation in the motor cortex.

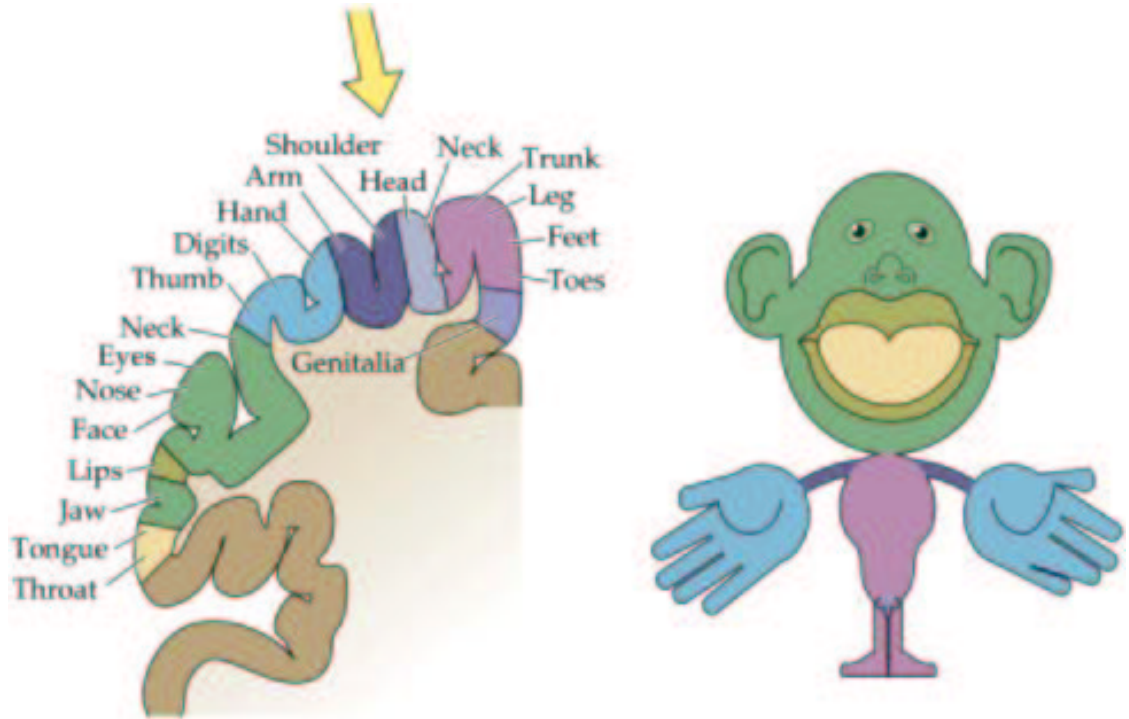


Figure 1.4: The human motor cortex [42].

The sensorimotor cortex generates rhythmic activity called the (rolandic)  $\mu$ -rhythm when subjects are not performing any movements [9, 13]. This activity consists of two frequency bands around 10 and 20 Hz which are both contralaterally suppressed 1-2 s before movement onset. This suppression becomes bilateral just before the movement onset [28, 24, 37]. A contralaterally dominant fast recovery and rebound of the rhythms can be detected after the movement ends [40].

Behaviour of the sensorimotor cortex during movements has been studied extensively with healthy subjects whereas there are only a few studies with paralyzed patients. Most of these are done with fMRI and only a couple with EEG. fMRI studies show that sensorimotor cortices of paralyzed patients are activated during attempted hand and foot movements [34, 36]. Very similar activation patterns, except weaker, were found when healthy control subjects performed real movements. However, with patients the activation patterns differed between imagined and attempted movements, suggesting that real movements should be used in studies with healthy subjects. Also a recent MEG study showed that sensorimotor cortices of tetraplegics respond to attempted hand movements [18]. In addition, contralateral motor potentials were detected with EEG when both paraplegics and tetraplegics attempt finger movements [10, 11].

## 1.4 Review of BCI systems

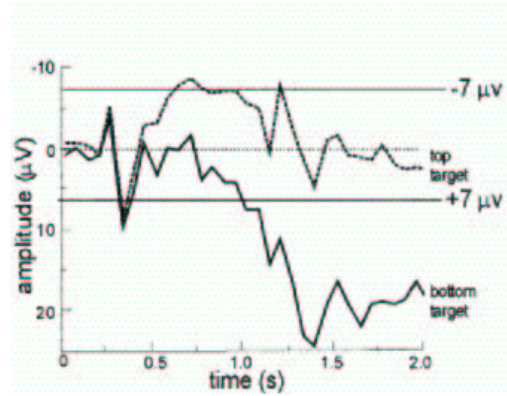
A brain computer interface is a system that translates mental tasks or different brain states into computer commands. At the first international BCI conference, BCI was defined as a communication system in which messages or commands that the user sends to the external world do not pass through the brain's normal output pathways of peripheral nerves and muscles [43]. A BCI could help e.g. motor disabled persons to operate a wheelchair, an arm prosthesis or a computer application for communication. Translating mental states into commands interpretable by a computer consists of four main phases. First, brain activity has to be measured and digitized with some method. EEG is most commonly used due to the portability and relatively affordable prices of the devices.

In the second phase the measured signals are processed to obtain features that characterize the mental states as well as possible. For example, spectrum estimates could be used to extract features that describe the rhythmic behaviour of the brain [45]. In the third phase the features are categorized using some classification method. Usually the objective is to assign a probability  $p(\mathcal{C}_k|D, \mathbf{x})$  for each mental state  $k$ , i.e. class  $\mathcal{C}_k$ , given a feature vector  $\mathbf{x}$  and some training data  $D$  collected previously from the user. There exists many different classification methods ranging from simple linear models to more complex support vector machines and neural networks [3]. The main issue is to find distinctive features that separate the classes optimally because even an arbitrarily complex classifier can not categorize the mental states with bad ones. With good features a simple model is sufficient to achieve high classification accuracy.

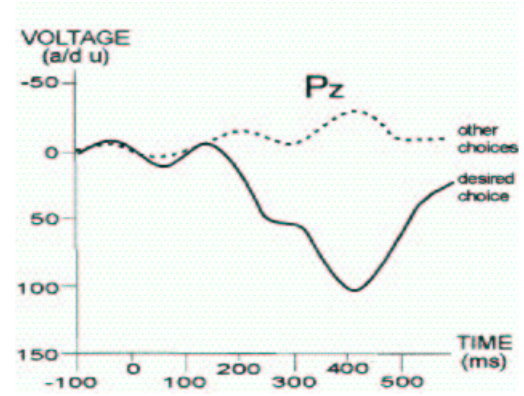
According to Wolpaw et al. [44], present day BCIs can be divided into five groups based on the electrophysiological signals they use. Systems in the first group use visual evoked potentials (VEPs) that can be detected over the visual cortex when the user directs his gaze to a desired object on the screen. Each selectable object on the screen flashes at different rate and the activity patterns of the visual cortex are compared to the corresponding VEP templates previously defined for the current user. Using this approach subjects have been able to operate a word processing program at 10-12 words/min [38].

VEP based BCIs are considered dependent because they require muscular control of gaze direction and thus does not fulfill the definition of a BCI presented in the beginning of this section. The systems belonging to the other four BCI groups,

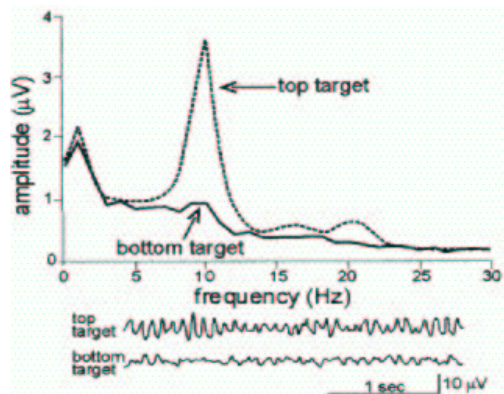
based on slow cortical potentials, P300 evoked potentials,  $\mu$ - and  $\beta$ -rhythms, and cortical neuronal action potentials respectively, are assumed to be independent of any muscular or perihairal nerve activity although comprehensive confirmation has not been presented.



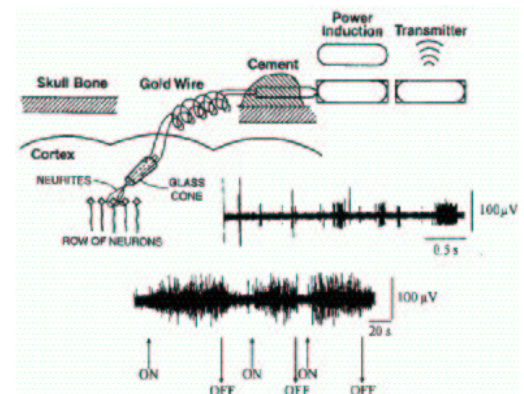
(a) Slow cortical potentials



(b) P300 Evoked potential



(c) Sensorimotor rhythms



(d) Cortical neuronal activity

Figure 1.5: Operational present-day independent BCI approaches [44].

Slow shifts of the brain activity occurring over 0.5 – 10 s, called slow cortical potentials (SCPs), have been used as control signals for BCIs. Subjects can learn to generate more positive and negative SCP values after several weeks or even months of training [2]. Each selection requires a 2 s baseline period to measure the initial activation level after which the amplitude shift is measured during a 2 s time period, see figure 1.5(a). Using SCPs subjects have been able to write 2 – 36 words per hour by a series of two-choice selections with accuracies of 65 – 90%.

When subjects are presented with infrequent or particularly significant auditory,

visual, or somatosensory stimuli, a positive peak, P300 evoked potential, can be detected 300 ms after the stimulus in the EEG, see figure 1.5(b). If subject concentrates on a particular symbol in a grid, P300 peak can be detected if either the corresponding row or column flashes rarely compared to other flashes. By flashing all rows and columns randomly and monitoring the EEG activity the symbol of subject's interest can be determined. Online experiments and offline simulations suggest that communication rate upto 7.8 characters per minute with 80% accuracy per choice can be achieved using P300 [7]. The advantage of the P300 approach is that theoretically it requires no training.

The most promising online BCI systems are based on user's ability to control his sensorimotor cortical  $\mu$ - and  $\beta$ -rhythms, corresponding to 8 – 12 Hz and 18 – 26 Hz frequency bands respectively, e.g. see figure 1.5(c). Computing power spectral estimates from one or several electrodes over the sensorimotor cortex and using linear combinations of the  $\mu$ - and  $\beta$ -band activities from these locations user's brain activity can be translated into a control signal for BCI. Using this approach four subjects achieved significant control of a two-dimensional real-time cursor movement [45]. After a minimum of five weeks training subjects were able to move the cursor to a right target with 82% mean accuracy. In another study a tetraplegic patient achieved almost error-free control of a hand orthosis with imagined foot movements after three months of training [30]. Each classification required 8 s time period. The desynchronization of the  $\mu$ -rhythm during imagined movements was compared to a 4 s baseline activity prior to the movement onset.

In addition to the noninvasive methods described earlier some BCI groups measure cortical neuronal action potentials invasively, i.e. from electrodes placed on top or inside brain tissue, see figure 1.5(d). Most of these studies have been made with monkeys because suitable intracortical electrodes for human use were not available until recently [44]. In a study by Nicolelis et al. a monkey was able to control 3D movement of a robot arm using signals recorded from intracortical electrodes [25].

# Chapter 2

## Materials and Methods

Our aim was to construct an online BCI that could be used effectively without extensive training. We decided to rely on single-trial classification of real finger movements with healthy subjects and attempted movements with paralyzed patients. Brain activity related to real finger movements can be detected in healthy subjects with EEG over the sensorimotor cortex [28, 29]. Furthermore, similar activation patterns can be found when motor-disabled patients attempt movements [17]. Activity related to high level mental tasks such as cube rotation used by some BCI groups, see e.g. [22], can be harder to detect because it is not spatially as localized as motor tasks. In addition this kind of tasks are more difficult to perceive and perform regularly. Successful use of them usually requires a lot of training.

We preferred a synchronous single-trial approach because continuous control needs extensive training periods to be effective [45]. Approximate time information of the movement makes classification easier because we know where to look for the activity in the EEG signals.

The main requirement for our BCI was the ability to give feedback to the subject right from the beginning of the experiment. This requires fast learning and good generalization capabilities from the classifier. Thus, we decided to use a linear classifier with principal component analysis (PCA) for dimensionality reduction, see e.g. [3].

## 2.1 Experiments

Two online BCI experiment were carried out using the same paradigm. Ten healthy subjects participated in the first experiment and five motor-disabled tetraplegic subjects in the second one. This section describes the subjects, the experimental setup and the EEG recordings in detail.

### 2.1.1 Subjects

Ten healthy right-handed subjects S1 - S10 (five males, aged 20-24) participated in the first experiment and five male tetraplegic subjects in the second one. Table 2.1 displays the age, time since injury and cite of injury of the subjects as well as the lowest site where some (less than normal) movement could be detected. Cite C4 corresponds to the elbow flexors, cite C6 to the wrist extensors and cite C7 to the elbow extensors. When asked, subject P4 reported being left handed and the rest reported being right-handed. The tetraplegia in P1-P4 was caused by trauma induced spinal-cord injury (SCI) and in P5 by Syndroma Guillain Barre (SGB). All subjects were first-time BCI users.

Table 2.1: Patient information for tetraplegic subjects

	Age	Time since injury	Cite of injury (movement detected)	Cause of tetraplegia
P1	26	4 months	C4(C5)	SCI
P2	47	4 months	C5(C4)	SCI
P3	50	3 months	C5(C7)	SCI
P4	63	35 years	C5(C6)	SCI
P5	59	1.5 years	C4(C7)	SGB

### 2.1.2 Experimental setup

The experiments were performed with a BCI system developed at the Laboratory of Computational Engineering, Helsinki University of Technology. The same experimental setup was used with both the healthy subjects as well as the patients.

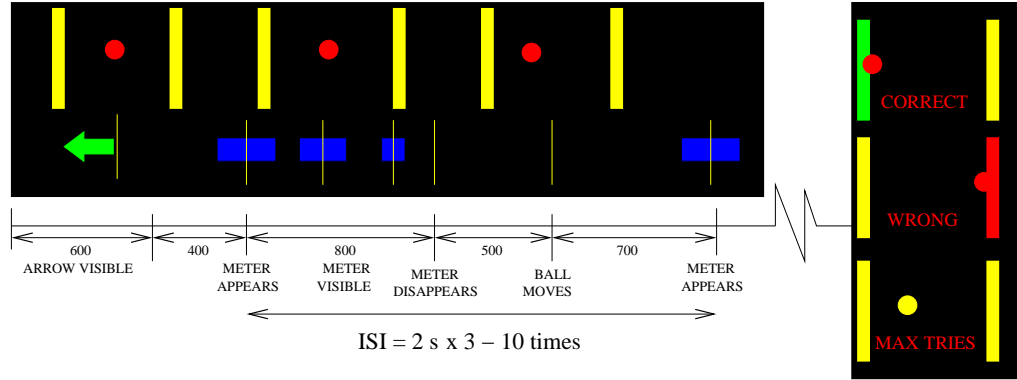


Figure 2.1: Online feedback given to the subjects

Figure 2.1 shows a screenshot of the program used to give tasks and feedback to the subject. The task was to move a circle to the target located on the left or right side of a computer screen by means of electrical brain activity measured with EEG. Healthy subjects performed fast real left or right index finger movements. The tetraplegic subjects were instructed to attempt fast left and right hand movements. Subject P1-P2 tried to close their fists. Subject P3 tried to lift his index fingers and subject P4-P5 attempted to pinch their index finger and thumb together.

One classified movement is called a trial. The target could be reached in three trials. If the trial was classified correctly the circle moved to the target direction, otherwise it moved to the opposite direction. Thus, it was also possible to reach the wrong (opposite) target if too many incorrect classifications were made. The position of the circle was updated according to  $k[P(C_2|\mathbf{x}, \mathbf{X}) - 0.5]$ , where  $P(C_2|\mathbf{x}, \mathbf{X})$  is the output of the classifier in a two class problem, i.e., the posterior probability of class 2 given a feature vector  $\mathbf{x}$  and training data  $\mathbf{X}$ . Parameter  $k$  is a distance measure in pixels adjusted according to the size of the screen. In other words, the larger the probability given by the model the bigger the step the circle moved. The target could be reached in two trials assuming 100% output probability for correct class but in practice a minimum of three trials was required.

As indicated in figure 2.1, each game started with an arrow indicating which target the subject should try to reach. Each game consisted of 2-10 trials and ended when the subject either reached one of the targets, or a maximum of 10 trials was exceeded. Each game took on average about 15 s. During the game, the subjects saw a visual trigger every two seconds. The subjects were instructed to perform the task shown by the arrow at the beginning of the game after the visual trigger disappeared. The trigger was a rectangle that got smaller and smaller and

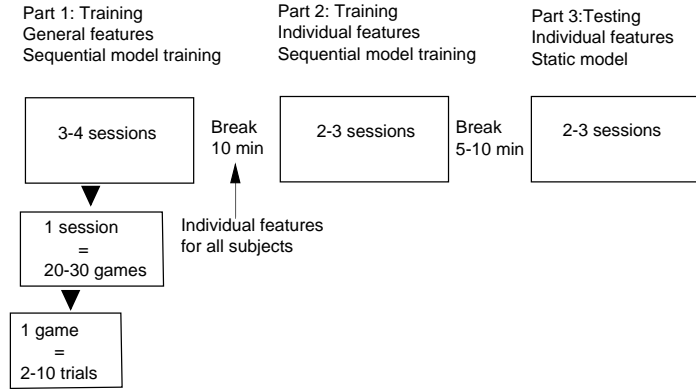


Figure 2.2: Structure of the online BCI experiment

disappeared 0.8 s after it was displayed. It has been shown that the preparation of finger movements, real or attempted, can be seen in the EEG signals as a slowly changing potentials [4]. If we wish to use these potentials for classification accurate time information of the movements is required. The visual trigger enables the subjects to prepare for the task and allows them to time the movement to the disappearance of the meter with as little variance as possible across trials.

Figure 2.2 displays the overall structure of the experiment. Data was collected in about three and a half minute sessions. There were short breaks, approximately 30 s long, between the sessions to avoid the subjects getting tired. Each session consisted of about 20 games depending on the subject's performance. The whole experiment consisted of three parts. Each part consisted of one to four sessions depending on how the subjects felt and how well they performed. Longer breaks were kept between the parts because time was needed to find the individual features of the subjects.

No separate training and testing parts were needed because the classifier was trained online using the correct information about the task of the moment in the first two parts. In the first ten trials of the experiment the circle moved in the correct direction because there was not enough samples to train the model. In the rest of the experiment the circle moved in the direction indicated by the classifier. The classifier was retrained after each trial (movement) with the previous samples. Training of the classifier is described in more detail in the next section. Because supervised training of the classifier is not possible in real applications (the intent of the subject is not known) the model was not trained with new samples in the third part. The performance of the classifier is evaluated based on the classification accuracy in the third part.



Figure 2.3: BCI experiment with a tetraplegic patient

### 2.1.3 Recording

For the first experiment recordings were made inside an electrically shielded room designed for EEG measurements at the Helsinki University of Technology. Recordings with the tetraplegic patients were made at the Käpylä Rehabilitation Centre in Helsinki. We set up the equipment in the patients' rooms. During the measurements, two to three additional people were in the room. The patient sat in the wheelchair in front of the computer screen. All efforts were made to reduce electrical interferences. Equipment such as lights, TVs and electrical beds were turned off.

The data acquisition and BCI softwares were ran on a 3 GHz Pentium 4 PC with 1 GB of RAM. EEG was measured with 32-channel electrode cap and BrainProducts Standard amplifier. Electrode impedance was below 10 kOhm for those electrodes that were used in the analysis. The impedance was checked in the beginning of the experiment and during the longer breaks. EEG was measured from 14 locations Fp1, Fp2, F3, F4, C3, C4, Cz, Fc1, Fc2, Cp1, Cp2, Fc5, Fc6, and Fz, according to the international 10-20 system. All electrodes were referenced to an electrode on top of the head situated in the middle of electrodes Cz and Pz. The recording passband was 0.1-200 Hz and the sampling frequency 500 Hz.

In addition to the EEG channels both vertical and horizontal electro-oculograms (EOG) from the eyes was measured with three additional electrodes. One electrode was placed beneath the left eye and one electrode on the right side of the right eye and another on the left side of the left eye.

## 2.2 Features

We classified EEG activity related to single trial finger movements both real and attempted. In the online experiments, the first four sessions were carried out using the same set of features for all subjects because they were first-time BCI users (see figure 2.2). These features were chosen based on preliminary online experiments with 10 healthy subjects and one patient. Based on the data gathered during the first part, individual features for each subject were chosen automatically. This combination of features was kept unchanged for the rest of the experiment.

Section 4.1.1 describes in general how the features are computed from the raw EEG signals. Only the parameter values and settings that remain unchanged for all subjects are given. Details of the feature selection, i.e. determination of subject specific frequency bands etc., are addressed in section 4.1.2.

### 2.2.1 Computation of features

In the online experiment total of 14 channels were recorded with 500 Hz sampling rate  $f_s$ . To reduce computational efforts only 6 of them C3, C4, Cp1, Cp2, Fc1, and Fc2 located over the sensorimotor cortex were used.

Figure 2.4 illustrates the computation of features for one channel but all six channels were processed similarly. All features used in the experiments were extracted from a one second segment of EEG signals starting 0.6 s before and ending 0.4 s after the visual cue. In the pictures cue is marked with green vertical line. Depending on the feature selection several narrow frequency bands (width  $\leq 2$  Hz) between 1 Hz and 30 Hz were extracted from these six channels. For example, the figure shows two frequency bands filtered from channel C4. The actual features were computed by averaging amplitude values from different frequency bands over short periods of time. Feature selection enabled the use of arbitrarily long time windows but only

ones of length 50 ms and 100 ms were used.

Before any frequency analysis the linear trends and means are removed from all channels (top right in figure 2.4). Detrending removes a linear aperiodic low frequency component from a signal retaining approximately the stationary component. All the six channels can be detrended by removing the best linear fit from each column of  $6 \times 500$  signal matrix

$$\mathbf{G} = \begin{bmatrix} g_1[0] & \dots & g_6[0] \\ \vdots & \ddots & \vdots \\ g_1[N-1] & \dots & g_6[N-1] \end{bmatrix}, \quad (2.1)$$

where  $g_i[n]$  denotes the discrete signal from channel  $i$  sampled in  $N = 500$  discrete time instants. The best linear fit is found by minimizing the squared error  $\mathbf{e}^2$  in

$$\mathbf{G} = \mathbf{A}\boldsymbol{\beta} + \mathbf{e}, \quad (2.2)$$

where matrix of regressors

$$\mathbf{A} = \frac{1}{N} \begin{bmatrix} 1 & N \\ 2 & N \\ \vdots & \vdots \\ N & N \end{bmatrix}. \quad (2.3)$$

The least squares solution  $\hat{\boldsymbol{\beta}}$  for regression coefficients is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{G} \quad (2.4)$$

and matrix  $\tilde{\mathbf{G}}$  of approximately stationary signals is obtained by subtracting the best linear fit

$$\tilde{\mathbf{G}} = \mathbf{G} - \mathbf{A}\hat{\boldsymbol{\beta}}. \quad (2.5)$$

From now on the detrended signals are denoted simply  $g[n]$  because all the channels are processed in the same way.

To obtain features from different frequency bands the discrete Fourier transform (DFT) is computed

$$G[k] = \sum_{n=0}^{N-1} g[n] e^{-i \frac{2\pi}{N} kn}, \quad k = 0, \dots, N-1, \quad (2.6)$$

where  $g[n]$  is the original signal,  $G[k]$  its Fourier transform and  $N$  number of samples [27]. Relation between the discrete frequency  $\Omega = 2\pi k/N$  [rad/sample] in 2.6 and continuous frequency  $f_c$  [Hz] is

$$f_c = \frac{1}{2\pi} f_s \Omega = \frac{f_s k}{N}, \quad k = 0, \dots, N - 1. \quad (2.7)$$

The frequency resolution of the Fourier transform is  $f_r = f_s/N = 1/T$ , where  $T$  is the length of the time window. The product of time resolution and frequency resolution remains constant implying a trade-off between the two properties. Thus, with longer time windows we get features with more accurate frequency information but we are averaging more over time. One second time window results in 1 Hz frequency resolution and was found sufficient based on preliminary offline simulations. Absolute lower limit for sampling rate  $f_s$  is determined by the highest frequency component  $f_{max}$  used in feature extraction and the Nyquist sampling theorem  $f_{max} < f_s/2$ .

Different frequency bands are extracted by first adjusting undesired fourier components to zero according to

$$G[k] = 0 \quad \text{for} \quad k = \begin{cases} 0, \dots, N \frac{f_l}{f_s} - 1 \\ N \frac{f_u}{f_s} + 1, \dots, N - N \frac{f_u}{f_s} - 1 \\ N - N \frac{f_l}{f_s} + 1, \dots, N - 1 \end{cases}, \quad (2.8)$$

where  $f_l$  and  $f_u$  are the lower and upper bounds of the frequency band respectively, and computing the inverse DFT

$$g[n] = \frac{1}{N} \sum_{k=0}^{N-1} G[k] e^{-i \frac{2\pi}{N} kn}, \quad n = 0, \dots, N - 1. \quad (2.9)$$

In equation 2.8 terms  $N f_l/f_s$  and  $N f_u/f_s$  have to be rounded appropriately to integer values. Only small number of components are retained, i.e. narrow frequency bands are chosen, because features tend to get worse when number of nonzero components increase. For example, setting  $G[k] = 0$  for  $k = 0, 4, 5, \dots, 496$  with  $N = 500$ , a low frequency band 1 – 3 Hz is obtained.

Figure 2.4 shows an example of two frequency bands extracted from channel C4. Signal from the lower 0.5 – 3 Hz band is obtained by setting the undesired Fourier components to zero according to 2.8 and computing the inverse transform of 2.9. For the higher 19 – 21 Hz band the instantaneous amplitude of the so called analytic

signal is computed because we are only interested in the envelope of the fast varying signal (bottom left graph).

The analytic signal can be computed easily using the discrete Fourier transform of the original signal with undesired frequency components set to zero. This is illustrated in the following paragraph for a continuous-time signal from which the discrete-time analogy can be inferred.

For a real continuous-time signal  $g(t)$  the corresponding analytic signal  $z(t)$  can be computed as

$$z(t) = g(t) + i\hat{g}(t) \stackrel{\mathcal{F}}{\Leftrightarrow} G(\omega) + \text{sgn}(\omega)G(\omega), \quad (2.10)$$

where

$$\text{sgn}(\omega) = \begin{cases} 1 & \text{for } \omega > 0 \\ 0 & \text{for } \omega = 0 \\ -1 & \text{for } \omega < 0 \end{cases} \quad (2.11)$$

and  $\hat{G}(t)$  and  $G(\omega)$  are the Hilbert and Fourier transforms of  $g(t)$  [27]. According to 2.10 the spectrum of  $z(t)$  is zero for negative frequencies and twice the spectrum of  $g(t)$  for positive frequencies. The analytic signal contains all information about  $g(t)$  because the positive frequency spectrum is sufficient to represent a real signal. Linearity of Fourier transform implies that

$$\hat{g}(t) \stackrel{\mathcal{F}}{\Leftrightarrow} -i \text{sgn}(\omega)G(\omega). \quad (2.12)$$

It can be seen that the Hilbert transform causes  $-\pi/2$  phase-shift for positive frequencies and  $\pi/2$  phase-shift for negative frequencies. Because the inverse Fourier transform of  $-i \text{sgn}(\omega)$  equals  $1/\pi t$  and multiplication in Fourier domain is equivalent to convolution in time domain it follows that

$$\hat{g}(t) = g(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(\tau)}{t - \tau} d\tau, \quad (2.13)$$

where  $*$  denotes convolution. The integral is Cauchy's principal value integral which extends the class of functions (limited by pole at  $t = \tau$ ) for which 2.13 exists. Equation 2.13 is congruent with definition of Hilbert transform. The absolute value  $A(t)$  of the analytic signal is called instantaneous amplitude, i.e.

$$z(t) = A(t)e^{i\varphi(t)}. \quad (2.14)$$

The instantaneous amplitude  $A(t)$  is used to extract features from signals composed

of higher than 3 Hz frequency components. Less than 3 Hz signals produce better features when the slowly varying time domain signal is subsampled whereas for higher frequencies the fast varying signal would average to zero. Furthermore, it is shown that the 0.5 – 3 Hz band contains information about the preparation of finger movements, which can be seen as a slow negative amplitude drift preceding the movement onset, also known as the *Bereitschaftspotential* [4].

For discrete  $g[n]$  the negative frequency components of the spectrum are set to zero and the positive ones are multiplied by two analogously with 2.10. The discrete analogue of the hilbert transformer of 2.12 for even  $N$  is given by

$$H[k] = \begin{cases} -i & \text{for } k = 1, \dots, N/2 - 1 \\ 0 & \text{for } k = 0, N/2 \\ i & \text{for } k = N/2 + 1, \dots, N - 1. \end{cases} \quad (2.15)$$

The Hilbert transform is computed via multiplication in the Fourier domain and the discrete analytic signal  $z[n]$  is obtained using the inverse DFT of 2.9, given by

$$G(k) + iH[k]G(k) \xrightarrow{DFT^{-1}} g[n] + i\hat{g}[n] = z[n]. \quad (2.16)$$

The lower left graph in figure 2.4 shows how the instantaneous amplitude  $|z[n]| = \sqrt{g^2[n] + \hat{g}^2[n]}$  follows the envelope of the fast varying signal. Using Fast Fourier Transform (FFT) features from both lower (less than 3 Hz) and higher frequency bands can be computed very quickly during online experiments.

Finally, the features are computed by averaging amplitude values from different frequency bands over short time windows. Length and position of the time window can differ between bands. Multiple features (time windows) can be computed from a single band. In figure 2.4 only one feature per frequency band is extracted. The method used to select the best set of features, i.e. the most distinguishing frequency bands and time windows, for each subject is addressed in the next section.

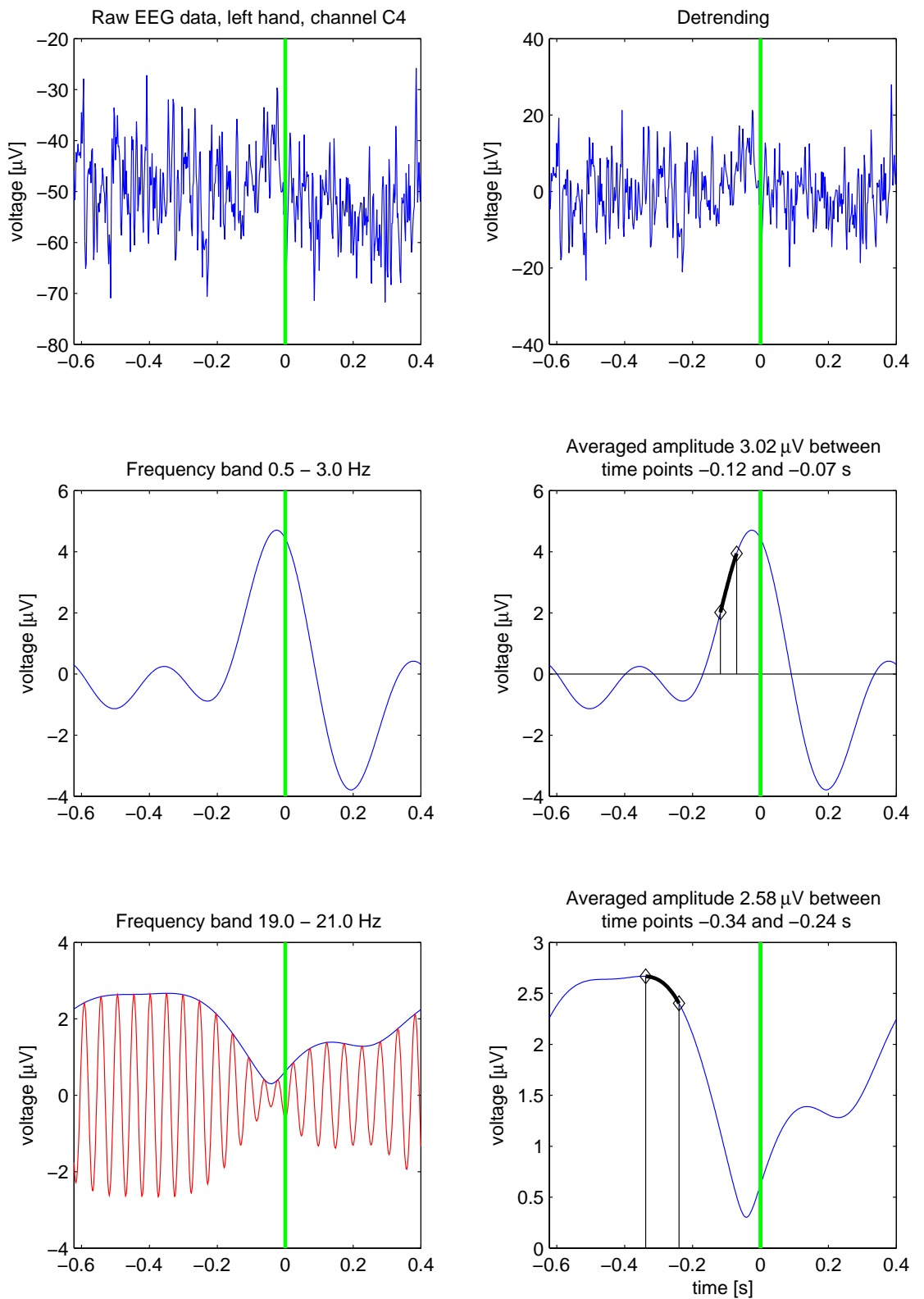


Figure 2.4: Computation of two features, 0.5 – 3 Hz averaged amplitude and 19 – 21 Hz averaged instantaneous amplitude, from one EEG channel C4

### 2.2.2 Feature selection

The frequencies that respond to finger movements, and motor intents in general, alter between subjects, see e.g. [29]. This was also seen in our preliminary online and offline experiments. Thus, individual feature selection is required for efficient use of the BCI. The main purpose of the first part of the experiment was to collect enough data so that individual features could be chosen.

In the first part of the experiment, the same selection of features was used for all subjects, see figure 2.2. We chose these features based on preliminary experiments with 10 healthy subjects and one patient. These were not necessarily the best features for each subject but they seemed to work somewhat for everyone. All features were computed from a low frequency band, 0.5 – 3 Hz for the six EEG channels. Features were computed from seven adjacent time windows of length 100 ms starting from 400 ms before and ending 300 ms after cue. Thus, total of 42 features were used (6 channels  $\times$  7 time windows).

During the first break, a large number of features was computed from the data collected in the first part of the experiment. Although the features are not independent they were considered independently. The Kolmogorov-Smirnov (KS) test statistic was used as a difference measure between the distributions of the classes for every feature independently [5]. The KS-test statistic  $D$  is defined as the maximum absolute difference between the cumulative distributions of two random variables, and in our case

$$D = \sup_x |F_1(x) - F_2(x)| = \sup_x |P(X_1 \leq x) - P(X_2 \leq x)|, \quad (2.17)$$

where  $X_i$  is a random variable corresponding to the value of feature  $x$  given it belongs to class  $i$ .  $F_i$  is the cumulative distribution of random variable  $X_i$ . Empirical (i.e. sample) cumulative distributions were used as estimates in feature selection. Figure 2.6 shows an example of the sample cumulative distributions for both classes for one particular feature. The maximum difference is marked with a vertical line.

Averaged event related potentials (ERPs) ( $\pm$  standard deviation) for both classes are shown in figure 2.5. The ERPs are computed from the same channel and frequency band as the feature in figure 2.6. The ERPs differ the most around 50 ms before the cue. In figure 2.6 this can be seen as a difference between the cumulative distributions of the feature computed from [-48,2] ms time interval.

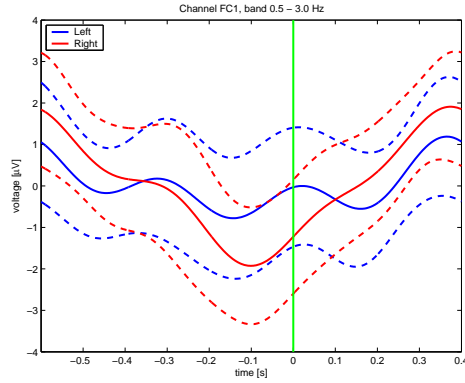


Figure 2.5: Average event related potentials for both classes

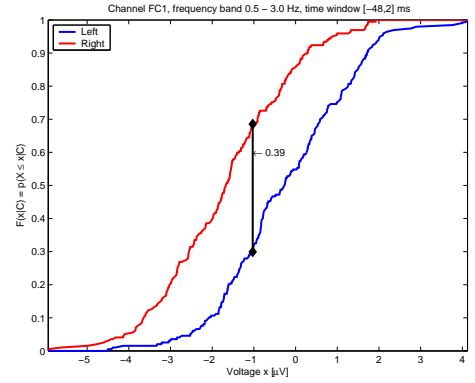


Figure 2.6: Kolmogorov-Smirnov test statistic

Figure 2.7 shows the KS test statistic computed in multiple adjacent 50 ms time windows for the 6 channels. Channel FC1 seems to separate classes best ca. 50 ms before the cue. Larger value of the KS test statistic seems to indicate that the feature separates the classes better.

All features were arranged according to the KS test statistic. Then features were chosen one by one corresponding to the largest KS-value so that there were no overlapping frequency bands or time windows from the same channel. This was done to decrease the redundancies among the features. It is clear that amplitude values of the same signal from nearby time instants are correlated. The same applies to frequency bands containing partially same frequency components. Choosing features independently according to the value of the KS test statistic would result in multiple redundant features corresponding to the same maximum difference between the classes. For example in figure 2.7 two features close to each other would be chosen from channel FC1. As a result many other good features may not be chosen because the dimensionality of the feature space must be kept reasonable. Basically redundancies among the features may result in better classification performance but according to offline simulations choosing features related to apparently different characteristics of the signals results in better accuracies.

When a particular frequency band and time window was chosen from one channel the corresponding feature (the same band and time window) was also chosen from the other five channels. Eight different combinations of one frequency band and time window were selected resulting in  $6 \times 8 = 48$  features altogether. This was done to acquire more information about the same event from different areas of the scalp.

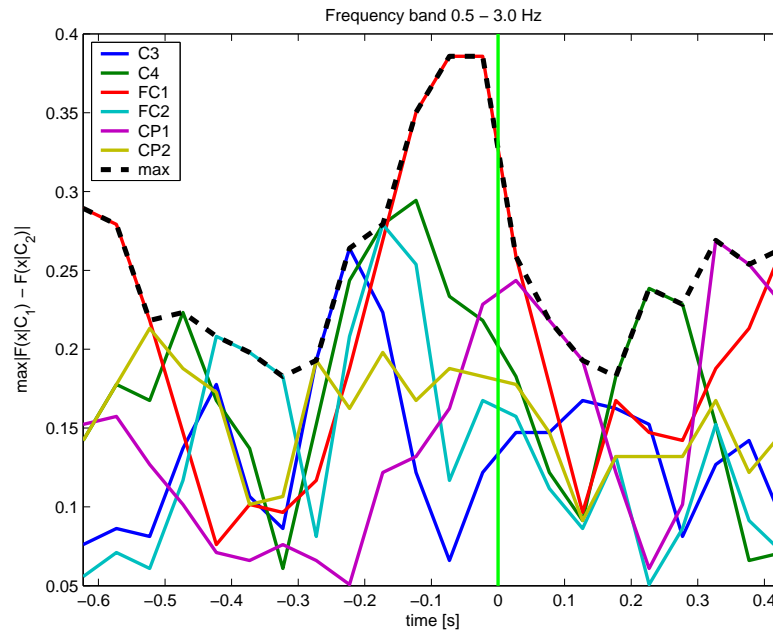


Figure 2.7: The values of KS test statistic in multiple adjacent 50 ms time windows from 6 channels

## 2.3 Online classifier

Figure 2.3 illustrates the operation of the online classifier during training periods. After feature extraction and normalization each new sample is classified using the existing model. Based on the result feedback is given to the subject. The classifier is updated using correct class labels. This is a kind of supervised sequential retraining approach to online learning, not a real adaptive classifier based on a dynamic model. Online training of the model is started once a minimum number (was set to 10 in the experiments) of samples (features) is acquired from every class. Thus, feedback can be given to the subject already from the beginning from the experiment. Furthermore the generalization capability of the classifier can be monitored from early on.

In addition to all other necessary model parameters, same number of previous feature vectors from every class are stored in the memory during the online experiment. After classification the oldest sample from the corresponding class is replaced with the new one. This makes regularization of the model easy, i.e., access to a whole history of previous samples helps avoiding overfitting when the model is optimized.

During testing periods the classifier was kept static. The classifier could have been retrained also during the testing periods even though information about the correct

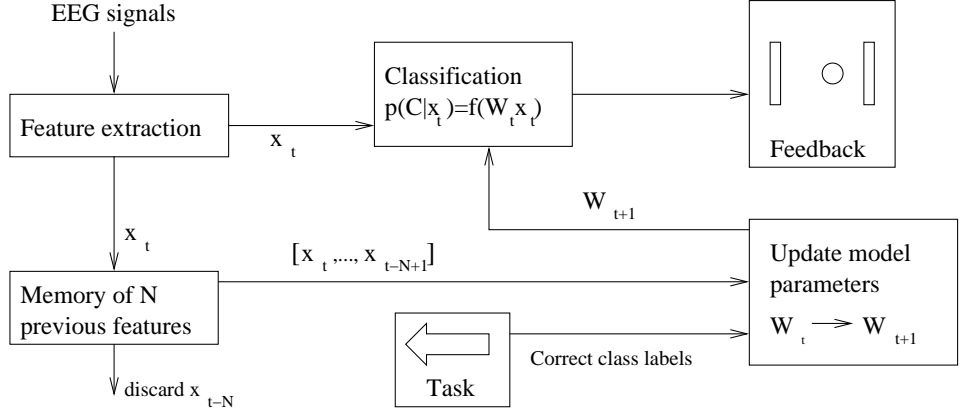


Figure 2.8: Operation of the online classifier

class would not be available. Each new sample could be associated with a class based on the output probability of the classifier. A new sample would be added to the class only if the posterior probability was high enough.

We denote the  $d$ -dimensional feature vector with  $\mathbf{x}$ . Total of  $N = N_1 + \dots + N_c$  feature vectors from previous trials are stored in the columns of matrices  $\mathbf{X}_1, \dots, \mathbf{X}_c$  where  $c$  is the number of classes. Maximum number of samples from each class is  $N_{max}$ , that is  $N_k \leq N_{max}$  for  $k = 1, \dots, c$ . Size of matrix  $\mathbf{X}_k$  is therefore  $d \times N_k$  where  $N_k$  is the number of samples collected from class  $k$ . Note that in the early parts of the experiment we have less than  $N_{max}$  features stored in some of the matrices  $\mathbf{X}_k$ . Parameter  $N_{max}$  controls the speed at which the classifier is able to 'adapt' to possible changes in the subject's brain activity. Because EEG signals include lots of undesirable information related to background brain activity and measurement noise, parameter  $N_{max}$  has to be set quite big. This implies restrictions on the maximum learning speed of the classifier. In the experiments,  $N_{max}$  was set to 200. Thus, with the rate of one classification in two seconds it takes at least 14 minutes to collect an entirely new set of training samples and to adapt the classifier to a totally new mental task.

### 2.3.1 Whitening and dimensionality reduction

Relatively high number of features (48) were used in the online experiments to compensate for the lacks in the feature selection. Selecting features independently does not lead to an optimal combination of features. Some individually very poor features may separate the classes extremely well when used together. Many features

computed from several different channels and frequency bands were used in hope to possibly include some good feature combinations. However, it is impossible to make accurate inferences (e.g. compute correlations) from such a high dimensional feature space without thousands of samples. Linear whitening transformation was used to reduce the dimension and to produce uncorrelated features [16]. This was done by retaining only the largest principal components of the feature space and normalizing the resulting features independently.

First, data from all classes is put together in a single  $d \times N$  matrix  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_c]$ . Then every feature (component of  $\mathbf{x}$ ) is normalized by subtracting sample mean and dividing the result with corresponding standard deviation

$$x_i^n = \frac{x_i^n - \bar{x}_i}{\hat{\sigma}_i}, \quad n = 1, \dots, N \quad (2.18)$$

where

$$\bar{x}_i = \frac{1}{N} \sum_{n=1}^N x_i^n \quad (2.19)$$

$$\hat{\sigma}_i = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (x_i^n - \bar{x}_i)^2} \quad (2.20)$$

for  $i = 1 \dots d$ , that is for each row of  $\mathbf{X}$  independently.  $\bar{x}_i$  and  $\hat{\sigma}_i$  are stored so that new samples can also be normalized during on-line classification of new samples.

Covariance matrix  $\mathbf{R}$  for the whole data (all classes together) is computed and diagonalized using eigenvalue decomposition

$$\mathbf{R} = \frac{1}{N-1} \mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad (2.21)$$

where  $\mathbf{\Lambda}$  is a diagonal matrix of eigenvalues and  $\mathbf{U}$  matrix of corresponding orthonormal eigenvectors. If number of samples  $N$  is much less than the dimension  $d$  it is more efficient to compute the singular value decomposition of matrix  $\mathbf{X}$  and taking the square of the resulting singular values.

Usually only the sample mean is subtracted from each component of the data before computing the correlation matrix [16], i.e.

$$\mathbf{R} = \frac{1}{N-1} \sum_{n=1}^N (\mathbf{x}^n - \mathbf{m})(\mathbf{x}^n - \mathbf{m})^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T. \quad (2.22)$$

where

$$\mathbf{m} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}^n. \quad (2.23)$$

This is the same as omitting equation 2.20. Each feature was normalized to unit variance because lower frequencies have much greater power in EEG. Computing correlation matrix according to 2.22 and retaining only the largest eigenvalues would lead to features describing only slow changes in the EEG.

The whitening matrix is computed using only  $d'$  largest eigenvalues and corresponding eigenvectors (columns of  $\mathbf{U}$ ) of the correlation matrix  $\mathbf{R}$

$$\mathbf{W} = \mathbf{\Lambda}_{(1:d',1:d')}^{-1/2} \mathbf{U}_{(1:d,1:d')}^T. \quad (2.24)$$

The resulting matrix is of size  $d' \times d$ . Number of retained principal components  $d'$  determines the level of regularization and it was set to  $a \cdot \min\{N_k; k = 1 \dots c\}$ , i.e. lower number of samples leads to greater reduction of dimension. It could also be selected so that  $\sum_{i=1}^{d'} \lambda_i = p \sum_{i=1}^d \lambda_i$ , with e.g.  $p = 0.9$ . This means that 90% of the total variance of the original feature space is explained with the new features. Based on offline cross-validations no significant differences in classification results were found between these two approaches. The first was chosen and parameter  $a$  was set to 0.1 leading to 20 whitened features.

### 2.3.2 Linear discriminants for classification

Using the whitening transformation of equation 2.24, we can compute uncorrelated features with manageable dimensionality. Utilizing the uncorrelatedness, we aim to find linear transformations of the whitened features that would be distinctive between the classes. First, features in matrices  $\mathbf{X}_k$  for  $k = 1, \dots, c$  are assumed to be normalized according to 2.18. Then the whitening transformation of equation 2.24 is applied

$$\mathbf{Y}_k = [\mathbf{y}_k^1, \dots, \mathbf{y}_k^{N_k}] = \mathbf{W} \mathbf{X}_k, \quad k = 1, \dots, c \quad (2.25)$$

where  $\mathbf{Y}_k$  is  $d' \times N_k$  matrix. If samples from all classes are considered together components of  $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_c]$  are clearly uncorrelated i.e.,

$$\frac{1}{N-1} \mathbf{Y} \mathbf{Y}^T = \mathbf{W} \mathbf{R} \mathbf{W}^T = \mathbf{\Lambda}_{(1:d',1:d')}^{-1/2} \mathbf{U}_{(1:d,1:d')}^T \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \mathbf{U}_{(1:d,1:d')} \mathbf{\Lambda}_{(1:d',1:d')}^{-1/2} = \mathbf{I}, \quad (2.26)$$

where  $\mathbf{I}$  is  $d' \times d'$  identity matrix. This means that whitened features are spherically distributed around the origin in  $d'$ -dimensional space with unit variance. However, features belonging to class  $\mathcal{C}_k$  (rows of  $\mathbf{Y}_k$ ) are hopefully not uncorrelated. They may also have nonzero mean. If i.e., the class  $\mathcal{C}_k$  has nonzero mean,  $\mathbf{m}_k$  must also the other classes have nonzero mean because  $\mathbf{m}_k + \sum_{i \notin \mathcal{C}_k} \mathbf{y}_i = \mathbf{0}$ . This means that classes can be distinguished based on their means. That is why we compute Fisher's linear discriminant from the whitened features [3]

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_2 - \mathbf{m}_1) \quad (2.27)$$

where

$$\begin{aligned} \mathbf{S}_W &= \sum_{k=1}^c (N_k - 1) \mathbf{R}_k = \sum_{k=1}^c \sum_{n \in \mathcal{C}_k} (\mathbf{y}^n - \mathbf{m}_k)(\mathbf{y}^n - \mathbf{m}_k)^T \\ \mathbf{m}_k &= \frac{1}{N_k} \sum_{n \in \mathcal{C}_k} \mathbf{y}^n. \end{aligned}$$

Vector  $\mathbf{w}$  defines the direction that maximizes the between-class covariance  $\mathbf{S}_B$  of the data divided by the total within-class covariance  $\mathbf{S}_W$  i.e.  $\mathbf{w}$  maximizes the Fisher's criterion  $J(\mathbf{w})$  [3]

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}, \quad (2.28)$$

where

$$\mathbf{S}_B = \sum_{k=1}^c N_k (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^T. \quad (2.29)$$

It also separates the classes best in the least squares sense. Equations 2.27 and 2.28 are valid only for two classes but they could be extended to multiple classes.

In addition to the Fisher's discriminant also other linear combinations of the whitened features  $\mathbf{Y}_k$  are used to discriminate between the classes. Equation 2.26 (uncorrelatedness) implies that if one class has not got a diagonal covariance matrix at least one other class must have nondiagonal covariance matrix. Consequently, if the samples in the class  $\mathcal{C}_k$  have large variance in direction of  $\mathbf{u}$  i.e.  $\mathbf{u}^T \mathbf{R}_k \mathbf{u}$  is large, then other classes must have smaller variances in direction of  $\mathbf{u}$ . Thus, we choose the directions that maximize the within-class variance for each class as discriminative features. As in calculation of the whitening matrix, principal component analysis is used to determine these directions. To ensure that nonzero class means will not inflict errors during classification the principal components are modified so

that they are orthogonal to the Fisher's discriminant  $\mathbf{w}$ . During online classification the mean of all the data is subtracted from the new samples according to 2.18 whereas during model training the within-class covariances are computed using the exact class means  $\mathbf{m}_k$ .

To find the principal components estimates for the within-class covariance matrices have to be computed for each class

$$\mathbf{R}_k = \frac{1}{N_k - 1} \sum_{n \in \mathcal{C}_k} (\mathbf{y}^n - \mathbf{m}_k)(\mathbf{y}^n - \mathbf{m}_k)^T. \quad (2.30)$$

The covariances are "deflated", so that the variance in the direction of  $\mathbf{w}$  is zero, i.e.,  $\mathbf{w}^T \mathbf{R}_k \mathbf{w} = 0$ . Rank of matrix  $\mathbf{R}_k$  is decreased by one resulting into only  $d' - 1$  nonzero eigenvalues. The eigenvalue decompositions are computed for these deflated matrices

$$\mathbf{R}_k - (\mathbf{w}^T \mathbf{R}_k \mathbf{w}) \mathbf{w} \mathbf{w}^T = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^T \quad (2.31)$$

for  $k = 1 \dots c$ , where  $\mathbf{U}_k = [\mathbf{u}_k^1, \dots, \mathbf{u}_k^{d'}]$ . In equation 2.31  $\mathbf{w}$  must be normalized according to

$$\mathbf{w} = \frac{\mathbf{w}}{\sqrt{\mathbf{w}^T \mathbf{w}}}. \quad (2.32)$$

Because this deflation is not numerically very accurate and because the eigenvalue decompositions are calculated for both classes separately the resulting eigenvectors will not be exactly orthogonal. Entirely orthogonal set of vectors can be generated e.g. by first computing the full rank eigenvalue decompositions

$$\mathbf{R}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^T \quad (2.33)$$

for each class  $\mathcal{C}_k$  and orthogonalizing them to  $\mathbf{w}$  using Gram-Schmidt procedure [1]. Vectors  $\mathbf{u}_k^i$  are assumed to be in descending order (according to corresponding eigenvalues) as  $i = 1, \dots, n$ . The largest principal components  $u_k^1$  for each class can be orthogonalized using

$$\tilde{\mathbf{u}}_k^1 = \mathbf{u}_k^1 - (\mathbf{u}_k^1{}^T \mathbf{w}) \mathbf{w} - \sum_{l=1}^{k-1} (\mathbf{u}_k^1{}^T \tilde{\mathbf{u}}_l^1) \tilde{\mathbf{u}}_l^1 \quad (2.34)$$

and

$$\tilde{\mathbf{u}}_k^i = \frac{\tilde{\mathbf{u}}_k^i}{\|\tilde{\mathbf{u}}_k^i\|^2} \quad (2.35)$$

for  $k = 1, \dots, c$ . Set of  $n$  orthogonal eigenvectors per class can be generated by

continuing the procedure repeatedly for  $i = 2, \dots, n$  using

$$\tilde{\mathbf{u}}_k^i = \mathbf{u}_k^i - (\mathbf{u}_k^{i,T} \mathbf{w}) \mathbf{w} - \sum_{l=1}^{k-1} \sum_{j=1}^i (\mathbf{u}_k^{i,T} \tilde{\mathbf{u}}_l^j) \tilde{\mathbf{u}}_l^j - \sum_{j=1}^{i-1} (\mathbf{u}_k^{i,T} \tilde{\mathbf{u}}_k^j) \tilde{\mathbf{u}}_k^j \quad (2.36)$$

and 2.35 for all  $k$ . No more than  $d' - 1$  linearly independent vectors  $\tilde{\mathbf{u}}_k^i$  in addition to  $\mathbf{w}$  can be found. Based on offline simulations the inaccuracies in 2.31 do not seem to affect the classification results significantly compared to equations 2.34 - 2.36. Therefore 2.31 was used due to its simplicity.

Only  $n$  eigenvectors  $\mathbf{u}_k^i$ ,  $i = 1 \dots n$ , from each class ( $k = 1, \dots, c$ ) corresponding to the  $n$  largest eigenvalues are chosen. In the online experiments  $n$  was set to one. We combine these eigenvectors with Fisher's discriminant and form the final  $(1 + cn) \times d$  transformation matrix  $\mathbf{F}$

$$\mathbf{F} = [\mathbf{w}, \mathbf{u}_1^1, \dots, \mathbf{u}_1^n, \dots, \mathbf{u}_c^1, \dots, \mathbf{u}_c^n]^T \mathbf{W}. \quad (2.37)$$

During the online experiment a new feature vector  $\mathbf{x}$  is first normalized according to equation 2.18 and then transformation matrix is applied,  $\mathbf{y} = \mathbf{F}\mathbf{x}$ . Because vectors  $\mathbf{u}_k^i$  describe the direction in which class  $\mathcal{C}_k$  varies the most we are only interested in the magnitude of the corresponding components of  $\mathbf{y} = [y_1, \dots, y_{cn}]^T$ . Thus the absolute value of components  $y_i$  for  $i = 2, \dots, cn$  is taken. The resulting features are used as an input to a linear classifier with nonlinear output. A logistic output function is used for a two class problem

$$\hat{p} = (1 - e^{-[\mathbf{y}^T \mathbf{1}] \mathbf{v}})^{-1}, \quad (2.38)$$

where  $\hat{p} \in [0, 1]$  and  $\mathbf{v}$  is a vector of model parameters. A softmax function is suitable for a multiple class problem. For gaussian features  $\hat{p}$  can be interpreted as the posterior probability of class 2 given input  $\mathbf{x}$ , and training data  $\mathbf{X}$ , i.e.,  $\hat{p} = p(\mathcal{C}_2 | \mathbf{x}, \mathbf{X})$ , where it is assumed that probability for class 1 is  $1 - \hat{p}$ . Model parameters  $\mathbf{v}_t$  are optimized with the iteratively reweighted least squares algorithm [12]. After each new sample, only a couple of iterations are required to update the previous parameters  $\mathbf{v}_t$  to new ones  $\mathbf{v}_{t+1}$ .

## 2.4 Evaluation of the BCI performance

Comparing different BCI systems is difficult because different mental tasks require different time to perform leading to different frequency of predictions  $f_p$ . The classification accuracies are more difficult to compare when there are different number of classes. Thus it has been proposed that BCI's theoretical performance could be measured by determining its available bit rate as a communication channel [44].

We consider the BCI as an input-output system with input class, i.e. mental task,  $x$  and output class  $y$ . The information transferred by the BCI in one trial can be expressed as the mutual information  $I(x, y)$  between the inputs and outputs

$$I(x, y) = H(y) - H(y|x), \quad (2.39)$$

where  $H(y)$  is the entropy of the output class and  $H(y|x)$  the entropy of the output given input  $x$ , see e.g. [39]. Entropy is a measure of the uncertainty of a random variable and hence equation 2.39 describes the reduction of uncertainty about  $y$  when input  $x$  is known. The larger the reduction of uncertainty about the outputs given the inputs, the smaller the uncertainty about the task and the better the performance.

The mutual information is given by

$$I(x, y) = \sum_{x \in X} P(x) \sum_{y \in Y} P(y|x) \log_2 \left( \frac{P(y|x)}{P(y)} \right) dx, \quad (2.40)$$

where  $P(y) = \sum_x P(y|x)P(x)$  and summations are taken over all mental states. The mutual information is given in bits per trial. If the probabilities for correct classifications, the probabilities for error classifications, and the input probabilities are assumed to be the same for all classes  $\mathcal{C}_i$ ,  $i = 1, \dots, K$ , i.e.  $P(y = \mathcal{C}_i|x = \mathcal{C}_i) = P$ ,  $P(y = \mathcal{C}_i|x = \mathcal{C}_j) = (1 - P)/(K - 1)$  when  $i \neq j$ , and  $P(x) = 1/K$ , the following approximation for the mutual information is obtained [44]

$$I(x, y) \approx \log_2 K + P \log_2 P + (1 - P) \log_2 \frac{(1 - P)}{K - 1}. \quad (2.41)$$

The problem with this approximation is that the classification accuracy can differ significantly between classes. The mutual information of equation 2.40 can be approximated more accurately using confusion matrix  $\mathbf{C} = [c_{ij}]$ , where  $c_{ij}$  is the number of trials from class  $i$  categorized to class  $j$ . If  $P(y = \mathcal{C}_i|x = \mathcal{C}_j) \approx$

$c_{ij} / \sum_j c_{ij} = p_{ij}$  and  $P(x = C_i) = 1/K$ , i.e. all classes all equally probable at the input, we get

$$I(x, y) \approx \sum_{i=1}^K \frac{1}{K} \sum_{j=1}^K p_{ij} \log_2 \left( \frac{p_{ij}}{\frac{1}{K} \sum_{i=1}^K p_{ij}} \right). \quad (2.42)$$

Channel capacity of a communications channel is defined as the maximum mutual information over all input probabilities [6]. The channel capacity, i.e. the theoretical maximum bitrate, of a BCI could be determined by maximizing 2.40 with respect to  $P(x)$ . However, real input probabilities depend much on the application. Thus, it is reasonable to assume that all classes are equally probable at the input. If the frequency of predictions made with the BCI is  $f_p$ , given in predictions per minute, then bitrate can be expressed as  $f_p I(x, y)$ , [bits/minute].

# Chapter 3

## Results

In this chapter classification accuracies as well as the corresponding bitrates from the two online experiments are presented. Results of the feature selection are illustrated graphically and compared with the classification rates to explain the performance of each subject. Furthermore, offline analysis of the data collected during the experiments with the paralyzed patients was made to find out if distinct task related activations could be found.

### 3.1 Classification results and BCI performance

Table 3.1 shows the classification accuracies (CA) in the three testing sessions and the average accuracy over all trials for the healthy subjects S1-S10. Accuracies for different tasks, left hand (L) and right hand (R), are also displayed. Each session contained ca. 100 trials resulting in about 300 testing trials altogether. The average accuracies varied between 56 – 95 %. S1 - S6 achieved over 70 % performance which has been suggested as the minimum requirement for use of a binary language support program described in [2]. Average accuracy over all subjects is 74 % Subject S1 achieved very good accuracy (95 %) whereas subject S10's performance was near the chance level of 50 %.

Table 3.2 shows the corresponding bitrates calculated according to equation 2.42. The bitrate is presented both in bits per minute and bits per trial. Because the rate of classifications in the experiments is one every two seconds, the maximum

Table 3.1: Classification accuracies [%] for subjects S1 - S10

	Session 1			Session 2			Session 3			Average		
	L	R	CA	L	R	CA	L	R	CA	L	R	CA
S1	92	90	<b>91</b>	96	100	<b>98</b>	98	94	<b>96</b>	95	95	<b>95</b>
S2	94	74	<b>84</b>	92	80	<b>86</b>	87	78	<b>83</b>	91	77	<b>84</b>
S3	80	81	<b>81</b>	86	76	<b>81</b>	86	81	<b>83</b>	84	79	<b>82</b>
S4	76	76	<b>76</b>	72	78	<b>75</b>	76	85	<b>81</b>	75	80	<b>77</b>
S5	80	67	<b>73</b>	67	75	<b>71</b>	92	71	<b>81</b>	80	71	<b>75</b>
S6	92	69	<b>80</b>	84	72	<b>78</b>	87	47	<b>66</b>	88	62	<b>75</b>
S7	65	86	<b>75</b>	57	75	<b>65</b>	63	73	<b>68</b>	61	78	<b>69</b>
S8	71	51	<b>61</b>	61	75	<b>67</b>	62	65	<b>64</b>	65	64	<b>64</b>
S9	71	62	<b>66</b>	73	44	<b>59</b>	69	46	<b>57</b>	71	51	<b>61</b>
S10	55	57	<b>56</b>	49	50	<b>50</b>	69	58	<b>63</b>	57	55	<b>56</b>

available bitrate is 30 bits/minute. The average rate over all subjects is 7 bits/min. The bitrate for S1 (21.2 bits/min) is twice as high as the bitrate for S2 whereas the classification accuracy is only about 10 percentage units higher. For a two task application, a 10 unit increase in classification accuracy corresponds approximately to a twice as high bit rate. 70 % accuracy, which has been considered adequate for a two task application, corresponds to a rate of 0.12 bits per trial.

Table 3.2: Bitrates for subjects S1 - S10

	Session 1		Session 2		Session 3		Average	
	bits/ trial	bits/ min	bits/ trial	bits/ min	bits/ trial	bits/ min	bits/ trial	bits/ min
S1	0.56	16.7	0.88	26.3	0.76	22.8	0.71	21.2
S2	0.39	11.8	0.42	12.7	0.34	10.1	0.38	11.4
S3	0.29	8.7	0.30	9.0	0.36	10.7	0.32	9.5
S4	0.20	6.1	0.19	5.7	0.29	8.8	0.23	6.8
S5	0.17	5.0	0.13	3.9	0.33	9.9	0.20	5.9
S6	0.31	9.4	0.25	7.4	0.10	3.1	0.20	6.1
S7	0.21	6.2	0.08	2.3	0.10	2.9	0.12	3.6
S8	0.04	1.1	0.09	2.8	0.06	1.7	0.06	1.8
S9	0.08	2.5	0.02	0.7	0.02	0.5	0.04	1.1
S10	0.01	0.3	0.00	0.0	0.06	1.7	0.01	0.3

Table 3.3 shows the average classification accuracies for subjects P1 - P5. Because patients P1, P3, and P4 grew tired during the experiment only two testing sessions were carried out. The average accuracy over all patients is 57 % which is significantly worse than with the healthy subjects. For P3 - P5 the classification

accuracies are in the chance level and none of the patients achieved 70 % accuracy. Table 3.4 shows the corresponding bitrates. P1 performed equally well in two sessions whereas P2 improved towards the end of the experiment.

Table 3.3: Classification accuracies for patients P1 - P5

	Session 1			Session 2			Session 3			Average		
	L	R	CA	L	R	CA	L	R	CA	L	R	CA
P1	67	71	<b>69</b>	66	71	<b>69</b>	-	-	-	66	71	<b>69</b>
P2	55	57	<b>56</b>	61	62	<b>62</b>	58	71	<b>65</b>	58	63	<b>61</b>
P3	47	60	<b>53</b>	52	50	<b>51</b>	-	-	-	50	55	<b>52</b>
P4	57	45	<b>51</b>	61	44	<b>52</b>	-	-	-	53	50	<b>52</b>
P5	54	41	<b>48</b>	37	62	<b>49</b>	70	33	<b>51</b>	53	45	<b>50</b>

Table 3.4: Bitrates for patients P1 - P5

	Session 1		Session 2		Session 3		Average	
	bits/ trial	bits/ min	bits/ trial	bits/ min	bits/ trial	bits/ min	bits/ trial	bits/ min
P1	0.11	3.2	0.10	3.1	-	-	0.10	3.1
P2	0.01	0.3	0.04	1.2	0.07	2.0	0.03	1.0
P3	0.00	0.1	0.00	0.0	-	-	0.00	0.1
P4	0.00	0.0	0.00	0.1	-	-	0.00	0.0
P5	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.1

Table 3.5 shows the number of right (hit) and wrong (miss) targets reached in the three testing sessions for subjects S1 - S10. The number of games, where the maximum number of trials was exceeded (max), as well as the percentages of targets hit and missed of all the games are also shown. Table 3.6 shows the corresponding results for the paralyzed patients. The percentages are not fully comparable because only two testing sessions were carried out with patients P1, P3, and P4.

All subjects except P4 and P5 hit the wrong target very few times although the classification accuracies differed significantly between them. Because the circle moved on the screen according to the output of the classifier, it seems that even for subjects with poor performance (S10, P3) the class probability given by the classifier contains some information about the movements. Besides giving additional feedback about the performance of the classifier to subject the proportional movements of the circle also seem to improve the application performance. Improved performance may also boost training periods by preventing subject from getting

Table 3.5: Application performance S1 - S10

	Hit	Miss	Max	Misses/ games [%]	Hits/ games [%]
S1	83	0	0	0	100
S2	57	0	3	0	95
S3	58	0	2	0	97
S4	46	0	6	0	89
S5	46	3	6	5.5	84
S6	45	1	6	1.9	87
S7	38	0	7	0	84
S8	17	0	19	0	47
S9	20	2	14	5.6	56
S10	7	1	24	3.1	22

Table 3.6: Application performance P1 - P5

	Sessions	Hit	Miss	Max	Misses/ games [%]	Hits/ games [%]
P1	2	21	1	8	3.3	70
P2	3	18	1	19	2.6	47
P3	2	5	1	15	4.8	24
P4	2	2	5	16	22	9
P5	3	14	11	18	26	33

frustrated. Rejection of trials based on the uncertainty given by the classifier might also be useful for certain applications.

## 3.2 Results of the feature selection

In the online experiments 48 features that independently separate classes best, were selected for classification based on the first four training sessions. Figure 3.1 shows the averaged event related potentials corresponding to the best feature found by the automatic feature selection from the first four training sessions (train) for subjects S1-S5. For comparison, averaged ERPs of the best feature for each subject are also computed from the three testing sessions (test). The best feature is a combination of an EEG channel, a frequency band and a time window as explained in section 2.2. The average of potential related to left (blue) and right (red) finger movements over trials are drawn with continuous lines. The average  $\pm$  standard deviations are shown with dashed lines. The disappearance of the visual cue is marked with a green vertical line and the time window from which the feature was averaged with

two black vertical lines. The corresponding results are shown for subjects S6-S10 and patients P1-P5 in figures 3.2 and 3.3 respectively. For patients the potentials are related to attempted hand movements.

The averaged ERPs are comparable with the classification accuracies presented in tables 3.1 and 3.3. For S1-S3 the best features differ clearly between classes and the averaged potentials are similar between the training and testing sessions. For subjects S8-S10 and all the patients, classes are much closer to each other and differ significantly between training and testing sessions. Thus, worse classification results can be partly explained by the weaker features. Of course, there may exist a combination of features that separate the classes better but finding them is time consuming and difficult because approximating probability densities in several dimensions requires a huge number of training samples.

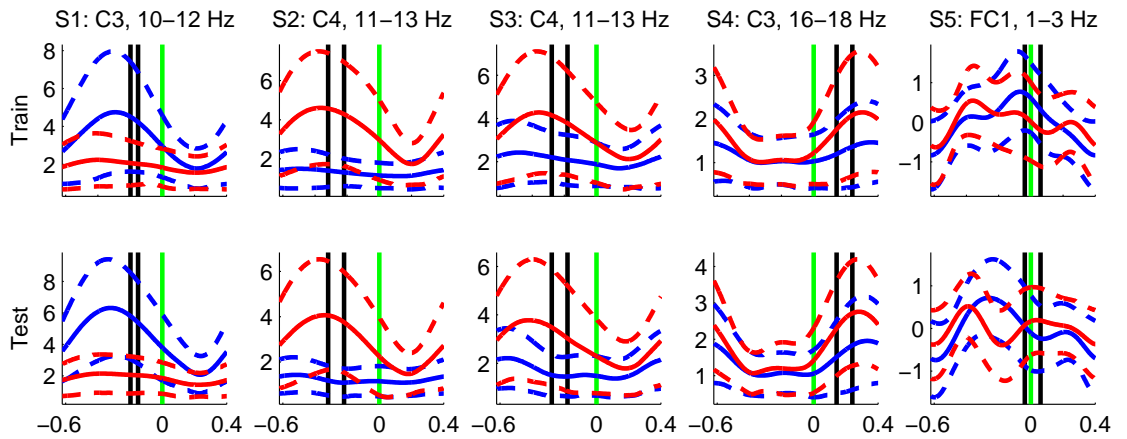


Figure 3.1: Averaged ERPs corresponding to the best feature for subjects S1-S5

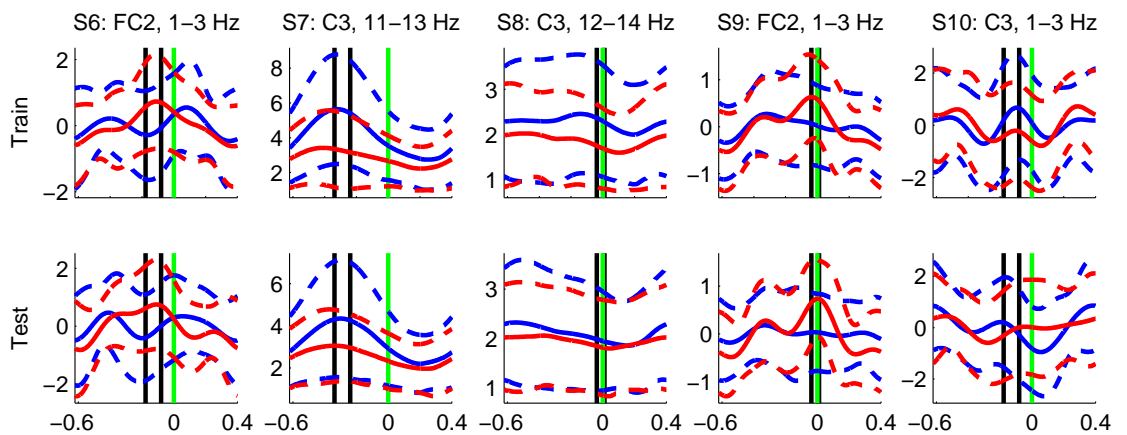


Figure 3.2: Averaged ERPs corresponding to the best feature for subjects S6-S10

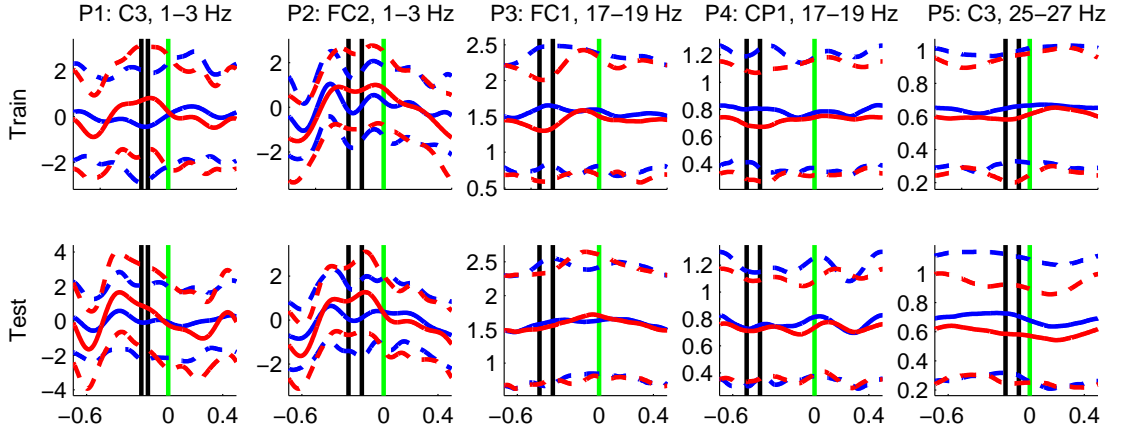


Figure 3.3: Averaged ERPs corresponding to the best feature for patients P1-P5

### 3.3 Correlation matched ERPs for patients

One definite weak point in our approach is the use of a visual trigger to synchronize the timing of the EEG data extraction with the single trial finger or hand movements. Using light port detectors in the online experiment with healthy subjects, we could see much jitter between the timing of the single trials. For some subjects, timing of the finger movements was more difficult than for others. This affects significantly the quality of the features because they were extracted from a short time window around the visual cue.

For patients we can not measure the timing of the attempted movements. Thus, the data from the online experiments was analysed offline to test whether the features would improve, i.e. more difference between the classes would emerge, by iteratively shifting each single trial in time so that the cross-correlation between each trial and the average over trials is maximized. First the data was filtered between 1 – 3 Hz, i.e. we are interested only in the slow cortical potentials. The iterative shifting of the ERPs was done separately for each class based on the data from the first four training sessions. Only channels C3 and C4 were used in the procedure because they are suitably located over the motor cortex.

At each iteration the mean of the ERPs over all trials is first computed for both channels. Then each trial is shifted so that the sum of the cross-correlations between each trial and the mean over both channels is maximized. Maximum shift for each trial was set to  $\pm 200$  ms relative to the cue. Figure 3.4 shows the results of the procedure after 20 iterations.

The graphs on the left were obtained by averaging 1 – 3 Hz potentials for both channels based on the visual cue (black line). Almost no difference between the classes can be seen. Classification accuracy based on these channels would be near chance level. The corresponding iteratively matched ERPs are shown on the right. Large differences between the classes can be seen in P1-P4 (larger than the within class standard deviation of the other class). No difference could be found between the classes for P5 even though some event related activity that is not class specific can be seen. The amplitude of the averages are bigger and the within class standard deviations are smaller than before shifting. Thus, it seems that the subjects had difficulties in timing the movements to the visual trigger.

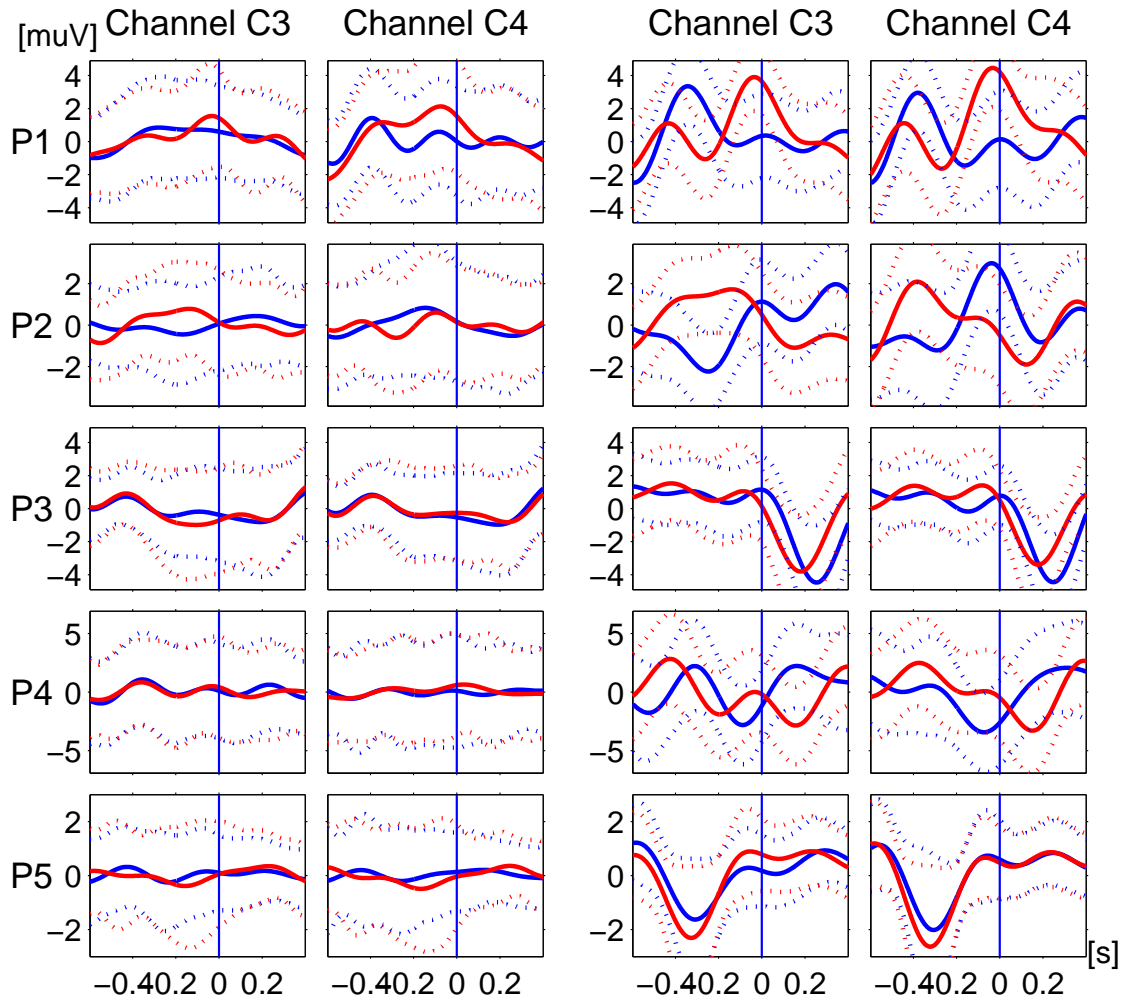


Figure 3.4: Left: Average ERPs from channels C3 and C4 for patients P1 - P5. The mean of the potentials related to the left (blue) and right (red) hand movements are shown with continuous lines. Mean  $\pm$  the standard deviation are drawn with dotted lines. Right: The ERPs are iteratively shifted so that the cross-correlation between each ERP and the mean of the ERPs is maximized.

# Chapter 4

## Discussion

In this thesis a new online BCI approach based on single-trial classification of EEG was presented. The aim was to construct a system that could be used without extensive training. Online experiments with ten healthy subjects and five tetraplegics were carried out to test whether satisfactory performance could be achieved during a single experiment. The subjects had no previous experience of BCIs.

In most of the present-day online BCIs, subjects require lots of training to reach high performance [19, 31, 44, 45]. It should be emphasized that most of the learning is done by the human user, not the system, because it is hard to find different task related patterns from the EEG. For the user to learn it is essential to give him online feedback of the performance. During the training subject tries to alter his brain activity according to the feedback. To maximize the amount of information conveyed to the user, we moved the circle proportionally to the estimated class probabilities. Thus, the user could see after each trial how well he succeeded and was able to improve for the next task. This kind of feedback encourages the subjects to try harder because the larger the classifier output the quicker the subject reached the goal.

In most of the offline BCI related studies static classifiers are used [4, 8, 20]. This means that features and classifier parameters are fixed based on a training set and the generalization of the model is studied using an independent testing set. Because in online experiments the user learns with the help of feedback his brain activation patterns related to each mental task change during training. Activation patterns change not only due to subjects voluntary actions but also because of

his state of mind, affected by tiredness, concentration etc. Consequently also the most relevant features and the probability densities in the feature space for each class change. Thus, classifier parameters and possibly even feature selection should be made continuously. Only a few offline studies have explored the advantages of adaptive classifiers [17, 39].

The same consideration applies to online experiments. If the classifier is static, the features and the classifier parameters have to be fixed based on prior knowledge or some initial training data. During the training the subject adapts to these prior settings with the help of feedback. However, there may exist better features or the user might develop them. Also more distinguishable activation patterns given the features chosen a priori might emerge during interactive training. Thus, both the feature selection and the training of the classifier should be made online.

In our experiments, the feature selection was not made online because it is difficult and computationally demanding. However, we tried to incorporate as many features as possible and to compress them by removing correlations. Also individual feature selection was made after four training sessions to find subject specific features. Furthermore, we assumed that adjacent EEG electrodes are correlated. Therefore, if a particular frequency seemed distinctive for one channel we included the same band from the other channels also. This way insignificant features will not get selected so easily due to independent comparisons between features. Also we tried to avoid redundancies among the features by rejecting overlapping frequency bands and time windows. Independent comparison of features with these heuristics is far from optimal procedure but it is computationally inexpensive and produced acceptable results.

Comparisons between features were made independently because accurate estimation of the joint probability density of even few hundred features would have required at least thousands of samples and we had only about 400 of them. This follows from the curse of dimensionality [3]. Furthermore, there would not have been enough time for extensive cross-validations during the 10 min break. Longer breaks would have prolonged the overall duration of the experiment too much.

The classifier was retrained online after each trial during the training phase. Each trial was first classified using the previous model to give realistic feedback to the user. We did not use a real adaptive classifier, i.e. posterior distribution for the model parameters is not estimated and updated online as in [17, 39]. Furthermore

our classifier was reoptimized only during training periods using correct class labels. The main advantage of this was the ability to give dynamic feedback throughout the training period. Our approach is easy to regularize and implement. All the parameters can be easily configured and interpreted.

The interactive learning process for both the user and system should be started as early as possible. Naturally some prior assumptions concerning features and the classifier have to be made to be able to learn from the data. For example Wolpaw et al. chose one rhythmic activation pattern from two EEG electrodes based on preliminary sessions [45]. During extensive training periods, several months, subjects learned to control these rhythms with the help of online feedback and only the classifier parameters were adapted.

We wanted to study whether the subjects could gain control of a BCI with very little training. To shorten the training, we gave feedback to the subjects right from the beginning of the experiment. This enabled the subjects to start learning from early on. We had to make some assumptions about the initial set of features. The choice of low frequency features was not necessarily the optimal for all subjects, but worked sufficiently for most of them. If the initial features were not good there was no advantage of the feedback. This was especially true for subjects P4-P5. Furthermore, in the first training sessions predictions made by the classifier were quite poor due to the small number of training samples and high dimensionality of the feature space. Consequently the feedback is not very accurate in the early sessions.

Our approach has some similarities with BCI studies using a feature extraction method called Common Spatial Patterns (CSPs) [20, 23]. This method, just as our feature extraction, makes use of the correlation of the brain activity between neighbouring channels. The CSPs are calculated separately for different frequency bands. For each band, the covariances between channels are calculated by averaging over consecutive filtered time-points within a trial. Contrary to CSPs, we do not assume that the activity in the different time points or the activity in different frequency bands during one trial are independent. We calculate the covariance estimate of different channels, frequency bands and time-windows over trials. Despite the similarities we cannot compare our results with [20, 23] because they have used knowledge of the exact timing of the movement in the computation of CSPs. The features calculated using exact time information of movement onset results in much better features with less variance over trials. When developing methods for

motor-disabled persons we cannot assume that we know when the movements are performed.

The use of visual trigger and time dependent features are clear drawbacks in our approach. With the correct time information the ERPs could result into better features compared to spectral features used in BCI systems performing classification of continuous states e.g. [45]. In the absence of the timing information some kind of trigger synchronizing the tasks performed by the subject is required for accurate single trial classification. Even with healthy subjects there is much variance in the reaction times making the trigger approach unappealing. Effective BCI based on single-trial classification would require an accurate method for detection of the ERPs from the EEG. Same kind of approach as was described in section 3.3 could be used in online classification also. Average ERPs for each channel could be determined from a training session using the iterative approach. Online detection of the single trials could be done by computing the cross-correlation between each channel and the corresponding averaged ERP continuously. However, due to the poor signal-to-noise ratio of the measured signals this is very difficult. Furthermore, classification of continuous states of the signals, e.g. average spectral power, enables much higher frequency of predictions resulting into higher bitrates.

Six out of the ten healthy subjects achieved at least 75% accuracy after only ca. 20 min training. In addition three healthy subjects and two tetraplegics exceeded 60% accuracy. Because the time interval between trials was relatively short some of the bitrates were quite high. S1 achieved information transfer rate of 21.2 bits per minute, which is comparable to the best BCIs reaching 25 bits/min [44]. S1-S9 and P1-P2 hit the right target in at least half of the games and clearly more often than the wrong target. The application performance was high compared to the classification accuracies. This must result from the use of the class probability estimates in the application control.

The classification results of the healthy subjects are much better than those of the patients. It was also easier to find more separable features for the healthy subjects than for the patients. The patients showed surprisingly little movement related activity. We acknowledge that it is more difficult for the patients to attempt hand movements than for the healthy subjects to perform real movements. Considering the very short training periods compared to other studies we believe that with more training the subjects would improve their results. In addition, the healthy subjects could also more easily time their finger movements according to the visual trigger.

# Bibliography

- [1] Stephen Barnett. *Matrices Methods and Applications*. Oxford Applied Mathematics and Computing Science Series. Oxford University Press, 1990.
- [2] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor. The thought translation device (ttt) for completely paralyzed patients. *IEEE Trans Rehabil Eng*, 8(2):190–3, 2000. 1063-6528 (Print) Journal Article.
- [3] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [4] B. Blankertz, G. Curio, and K-R. Müller. Classifying single trial eeg: Towards brain computer interfacing. In Diettrich T. G., Becker S, and Ghahramani Z, editors, *Artificial Neural Networks - ICANN 2002*, pages 1156–1161, Madrid, 2002. Springer.
- [5] W. J. Conover. *Practical Nonparametric Statistics*. Wiley Series in Probability and Statistics: Applied Probability and Statistics Section. John Wiley & Sons, Inc., third edition, 1999.
- [6] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.
- [7] E. Donchin, K. M. Spencer, and R. Wijesinghe. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *IEEE Trans Rehabil Eng*, 8(2):174–9, 2000. 1063-6528 (Print) Journal Article.
- [8] G. Dornhege, B. Blankertz, G. Curio, and K. R. Muller. Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans Biomed Eng*, 51(6):993–1002, 2004. 0018-9294 (Print) Clinical Trial Journal Article Validation Studies.
- [9] H. Gaustaut. Étude électrocorticographique de la réactivité des rythmes rolandiques. *Rev Neurol*, 87:176–182, 1952.
- [10] J. B. Green, E. Sora, Y. Bialy, A. Ricamato, and R. W. Thatcher. Cortical sensorimotor reorganization after spinal cord injury: an electroencephalographic study. *Neurology*, 50(4):1115–21, 1998. 0028-3878 (Print) Journal Article.

- [11] J. B. Green, E. Sora, Y. Bialy, A. Ricamato, and R. W. Thatcher. Cortical motor reorganization after paraplegia: an eeg study. *Neurology*, 53(4):736–43, 1999. 0028-3878 (Print) Journal Article.
- [12] P. J. Green. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J. R. Statist. Soc. B*, 46(2):149–192, 1984.
- [13] R. Hari and R. Salmelin. Human cortical oscillations: a neuromagnetic view through the skull. *Trends Neurosci*, 20(1):44–9, 1997. 0166-2236 Journal Article Review Review, Tutorial.
- [14] V. W. Hevern. <http://www.has.vcu.edu/psy/psy101/forsyth/lobes.gif>.
- [15] M. Hämäläinen, R. Hari, R.J. Ilmoniemi, J. Knuutila, and O.V. Lounasmaa. Magnetoencephalography theory, instrumentation, and applications to non-invasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 1993.
- [16] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, Inc., 2001.
- [17] L. Kauhanen, T Nykopp, J. Lehtonen, P. Jylanki, J. Heikkonen, P. Rantanen, H. Alaranta, and M. Sams. Eeg and meg brain-computer interface research with tetraplegic patients. *IEEE Transactions in Rehabilitation Engineering*, in press.
- [18] L. Kauhanen, P. Rantanen, J. Lehtonen, I. Tarnanen, H. Alaranta, and M. Sams. Sensorimotor cortical activity of tetraplegics during attempted finger movements. *Biomedizinische Technik*, 49(1):59–60, 2004.
- [19] A. Kubler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer, and J. R. Wolpaw. Patients with als can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–7, 2005. 1526-632X (Electronic) Clinical Trial Journal Article.
- [20] Y. Li, X. Gao, H. Liu, and S. Gao. Classification of single-trial electroencephalogram during finger movement. *IEEE Trans Biomed Eng*, 51(6):1019–25, 2004. 0018-9294 (Print) Clinical Trial Journal Article Validation Studies.
- [21] J. Malmivuo and R. Plonsey. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995.
- [22] R. Millan Jdel and J. Mourino. Asynchronous bci and local neural classifiers: an overview of the adaptive brain interface project. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):159–61, 2003. 1534-4320 (Print) Journal Article.
- [23] J. Muller-Gerking, G. Pfurtscheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial eeg classification in a movement task. *Clin Neurophysiol*, 110(5):787–98, 1999. 1388-2457 Journal Article.

- [24] T. Nagamine, M. Kajola, R. Salmelin, H. Shibasaki, and R. Hari. Movement-related slow cortical magnetic fields and changes of spontaneous meg- and eeg-brain rhythms. *Electroencephalogr Clin Neurophysiol*, 99(3):274–86., 1996.
- [25] M. A. Nicolelis and J. K. Chapin. Controlling robots with the mind. *Sci Am*, 287(4):46–53, 2002. 0036-8733 (Print) Journal Article.
- [26] Ernst Niedermeyer and Fernando Lopes Da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, fourth edition, 1999.
- [27] A. V. Oppenheim and R. W. Schaffer. *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [28] G. Pfurtscheller. Central beta rhythm during sensorimotor activities in man. *Electroencephalogr Clin Neurophysiol*, 51(3):253–64., 1981.
- [29] G. Pfurtscheller. Spatiotemporal erd/ers patterns during voluntary movement and motor imagery. *Suppl Clin Neurophysiol*, 53:196–8, 2000. 1567-424x Journal Article Review.
- [30] G. Pfurtscheller, G. R. Muller, J. Pfurtscheller, H. J. Gerner, and R. Rupp. 'thought'-control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neurosci Lett*, 351(1):33–6, 2003. 0304-3940 Case Reports Journal Article.
- [31] G. Pfurtscheller, C. Neuper, G. R. Muller, B. Obermaier, G. Krausz, A. Schlogl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wortz, G. Supp, and C. Schrank. Graz-bci: state of the art and clinical applications. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):177–80, 2003. 1534-4320 (Print) Evaluation Studies Journal Article.
- [32] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, Jr. Johnson, R., G. A. Miller, W. Ritter, D. S. Ruchkin, M. D. Rugg, and M. J. Taylor. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, 37(2):127–52, 2000. 0048-5772 (Print) Guideline Journal Article.
- [33] D. Purves, G. Augustine, D. Fitzpatrick, L. Katz, A. LaMantia, J. McNamara, and M. Williams. *Neuroscience*. Sinauer, second edition, 2001.
- [34] P. Sabbah, S. S. de, C. Leveque, S. Gay, F. Pfefer, C. Nioche, J. L. Sarrazin, H. Barouti, M. Tadie, and Y. S. Cordoliani. Sensorimotor cortical activity in patients with complete spinal cord injury: a functional magnetic resonance imaging study. *J Neurotrauma*, 19(1):53–60., 2002.
- [35] E. W. Sellers and E. Donchin. A p300-based brain-computer interface: Initial tests by als patients. *Clin Neurophysiol*, 117(3):538–48, 2006. 1388-2457 (Print) Journal Article.

- [36] S. Shoham, E. Halgren, E. M. Maynard, and R. A. Normann. Motor-cortical activity in tetraplegics. *Nature*, 413(6858):793., 2001.
- [37] Jr. Stancak, A. and G. Pfurtscheller. Event-related desynchronisation of central beta-rhythms during brisk and slow self-paced finger movements of dominant and nondominant hand. *Brain Res Cogn Brain Res*, 4(3):171–83, 1996. 0926-6410 Journal Article.
- [38] E. Sutter and D. Tran. Communication through visually induced electrical brain responses. In *Computers for Handicapped Persons*, 1990.
- [39] P. Sykacek, S. J. Roberts, and M. Stokes. Adaptive bci based on variational bayesian kalman filtering: an empirical evaluation. *IEEE Trans Biomed Eng*, 51(5):719–27, 2004. 0018-9294 (Print) Evaluation Studies Journal Article Validation Studies.
- [40] C. Toro, G. Deuschl, R. Thatcher, S. Sato, C. Kufta, and M. Hallett. Event-related desynchronization and movement-related cortical potentials on the ecog and eeg. *Electroencephalogr Clin Neurophysiol*, 93(5):380–9., 1994.
- [41] Virginia Commonwealth University. <http://www.has.vcu.edu/psy/psy101/forsyth/lobes.gif>.
- [42] D. Uzwiak. [http://www.rci.rutgers.edu/~uzwiak/AnatPhys/NPSpringLect3\\_files/image010.jpg](http://www.rci.rutgers.edu/~uzwiak/AnatPhys/NPSpringLect3_files/image010.jpg).
- [43] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan. Brain-computer interface technology: a review of the first international meeting. *IEEE Trans Rehabil Eng*, 8(2):164–73, 2000. 1063-6528 Journal Article.
- [44] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clin Neurophysiol*, 113(6):767–91, 2002. 1388-2457 Journal Article Review.
- [45] J. R. Wolpaw and D. J. McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc Natl Acad Sci U S A*, 101(51):17849–54, 2004. 0027-8424 Journal Article.