Examination and Comparison of Methods to Increase Communication Speed of Paralysed Patients by Brain-Computer Interfaces

Dissertation
der Fakultät für Informations- und Kognitionswissenschaften
der Eberhard-Karls-Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Dipl.-Inform. Michael Bensch
aus Nürtingen

Tübingen
2010

Tag der mündlichen Qualifikation:	3. November 2010
Dekan:	Prof. Dr.-Ing. Oliver Kohlbacher
1. Berichterstatter:	Prof. Dr. Wolfgang Rosenstiel
2. Berichterstatter:	Prof. Dr. Martin Bogdan (Universität Leipzig)

Acknowledgements

First and foremost I would like to thank Prof. Dr. Wolfgang Rosenstiel and Prof. Dr. Martin Bogdan for placing their trust in me and guiding me throughout my work. Prof. Dr. Rosenstiel has given me the opportunity to participate in his research group, which I greatly appreciate. I am truly indebted to Prof. Dr. Bogdan for his patience and time spent in discussions with me. He has been extremely responsive to any queries I had, no matter what day of the week or time of day.

I would also like to express my appreciation towards my close colleagues in the NeuroTeam group, Michael Tangermann, Lothar Ludwig, Elena Sapojnikova, Thomas Hermle and Dominik Brugger for their valuable advice and company during my time at the Computer Engineering Department. Many colleagues have become friends, and I thank Prakash Mohan Peranandam, Pradeep Kumar Nalla, Djones Lettnin and Julio Oliveira Filho for their continued fellowship throughout these years.

I wish to convey my sincere gratitude to Prof. Dr. Niels Birbaumer and Prof. Dr. Andrea Kübler for their expert opinion, guidance and support concerning the work with patients. I am equally indebted to Dr. Hubert Preißl for giving me the opportunity to use the MEG facilities for my studies. Without the assistance of Dr. med. Michael Schulze, Sonja Kleih, and Martin Spüler, some particular studies would not have been possible — their effort is greatly appreciated.

Many of the illustrations in this work were created with a software framework developed in a fervent display of brilliance by Jeremy Hill. Amongst further colleagues of the Tübingen BCI research group, Barbara Wilhelm, Femke Nijboer, Suzanne Martens, Jürgen Mellinger, Sebastian Halder and Ander Ramos Murguialday have contributed to the quality of this work and have supported me in persevering.

The cooperation of all the ALS patients participating in various studies was a prerequisite for this dissertation. Thank you for your interest and patience.

I affectionately thank my wife Jessica Kay Bensch for her loving moral support and never ending faith in me. Even in the most difficult moments, her belief and comforting words gave me new strength.

Lastly, I wish to thank my parents for their love and for instilling in me the courage and faith to go forward on this path.

Contents

List of Figures

List of Tables

1. Introduction

Since the first published report of electroencephalogram (EEG) measurement by Hans Berger in 1929 [12], neuroscience, medicine and other research fields have adopted the idea of obtaining information from the brain by measuring its electrical activity. Applications include groundlaying research work in the field of neuroscience, medical diagnosis, and more recently, entertainment.

Over the past twenty years, brain-computer interfacing has emerged as an interdisciplinary field with the aim of understanding a person's intention by the interpretation of his or her[1] EEG signal. Although there is fundamental research in this field concerned with understanding cerebral mechanisms on a neuronal level, most researchers are driven by the prospect of deciphering the intentions of completely paralysed persons who are unable to otherwise communicate. Ideally, the translation of intention into letters or actions by the computer takes place online for two reasons. Firstly, the user obviously wishes to express his intent immediately without waiting for hours until the computer has deciphered it. Secondly, the online response allows the possibility of contingent biofeedback, which has been shown to promote user learning in the context of brain-computer interfaces (BCIs). A BCI may be designed in such a way that user intent is recognised immediately and without prior training of the user, or it may be based on the principle of user training by operant conditioning with biofeedback. The operant conditioning approach gives sensory feedback or positive reinforcement when appropriate and was shown to lead to self-regulation of the EEG as early as 1969 [61]. Roughly ten years later, self-regulation of the sensorimotor or μ-rhythm [126] and of slow cortical potentials (SCPs) [36] was investigated for the first time. These potentials were later used in the first BCI experiments with cursor control on a computer screen based on μ-rhythm control [139] or SCPs [64].

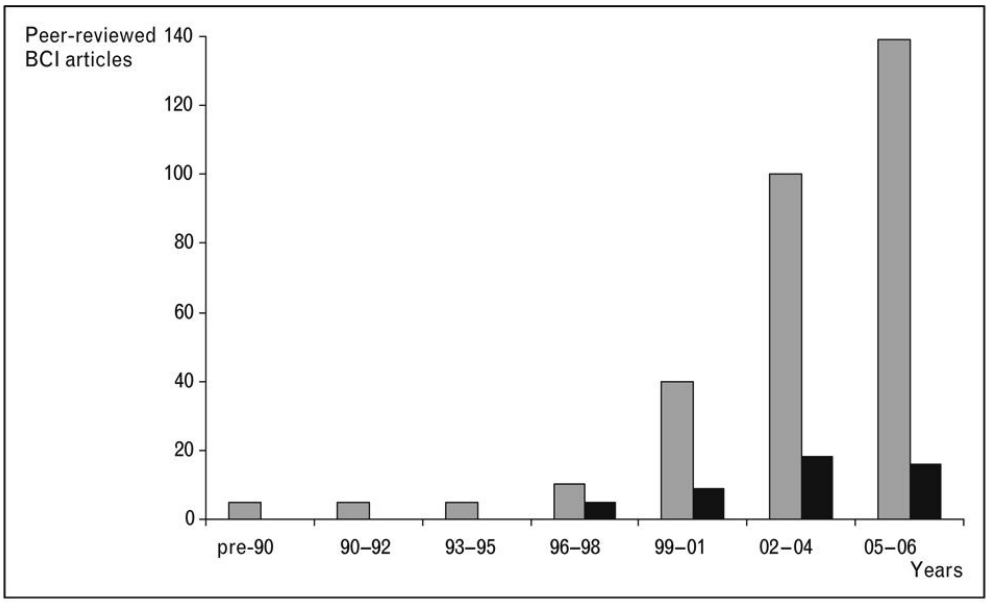
The foremost use of BCIs, which will be the centre of discussion in this thesis, is to enable completely paralysed people to communicate with others. To this end, the user must imagine specific movements or perform other mental tasks (*imagery*), or merely focus on particular letters displayed on a computer screen. An algorithm attempts to recognise the user's intent and translates it into a cursor movement or displays a letter on the screen. By the aforementioned principle of operant conditioning, the user can improve his adeptness at generating the correct brain signal. In addition, the algorithm can improve its function to discriminate between the tasks over time. Human and machine learning takes place simultaneously and when a mistake is made, it is hard to decide who is at fault.

Many types of BCIs have been developed over the past three decades. They can be classified under various criteria, such as with or without user training, based on evoked or spontaneous brain signals, presented via a visual, auditory or tactile medium, synchronous or asynchronous operation, etc. The most relevant BCI methodologies will be introduced in Chapter 2 after the goal of this thesis has been motivated and its main contributions have been given in the following.

1.1. Motivation

Historian Tony Judt describes the effects of amyotrophic lateral sclerosis (ALS) as "progressive imprisonment without parole". Apart from the successive paralysation, this debilitating disease successively robs the afflicted of their ability to communicate. The aim of using a BCI is to restore patients' ability to at least inform caregivers of their most urgent needs [69]. The locked-in syndrome (LIS) itself will be covered in more detail in Section 2.3.2.

[1] For the remainder of this dissertation, masculinum gender will be used to increase readability. No implicit preference is intended.



The search was limited to publications in English and spanning from 1 January 2005 to 31 December 2006. The results from pre-90 to 02–04 are from Jonathan R. Wolpaw, Wadsworth Center, NY State Department of Health, Albany, USA, with kind permission; the search terms were 'brain–computer interface' and 'brain–machine interface'. ▒ All BCI articles; ■ BCI articles with patients.

Figure 1.1.: Number of peer-reviewed BCI articles alongside BCI articles with patients (personal communication by A. Kübler).

Although BCIs have been around for some time, the "quantum leap" in technology that would allow any paralysed patient to use such a communication system effortlessly and efficiently has yet to occur. There are numerous factors that have hindered this development: Standard recording techniques are often cumbersome and hard to apply to patients, most studies are performed with healthy subjects (and the results of those studies often not easily transferred to patients), and theoretical results from the lab are hard to realise in practice when confronted with the multitude of problems when working with patients.

While it is understandable that algorithms must first be tested on healthy subjects, the experimental design and algorithms should consider patients as eventual users of the BCI right from the outset. The current gap between the number of studies performed with patients as compared to healthy subjects is evident (see Figure 1.1). While the total number of publications in the BCI field has increased dramatically, papers including patients have not yet followed the trend.

The main factor preventing daily use of BCIs by patients is speed. Communication speed is directly linked to accuracy, as is explained in Section 2.5. This thesis therefore presents ideas on how to increase communication speed whilst keeping in mind the applicability to patients. Studies with healthy subjects are presented to validate these ideas. Studies with paralysed patients are presented for those cases where the algorithms have been successful with healthy subjects, but not yet applied to patients.

Essentially, the thesis can be stated as: "Brain connectivity analysis, nonmotor imagery and the detection of error-related potentials can increase the communication speed of patients using a BCI."

Many afflicted with successive paralysis lose the ability to produce a control signal strong enough to communicate. Increased spatial resolution and signal amplitude of electrocorticogram (ECoG) make this recording technique a candidate for reinstating communication. The unique opportunity to work with two

LIS patients was realised to collect evidence pertaining to the nonmotor imagery part of the abovementioned thesis.

In addition, a cognition detection battery is for the first time used with chronically implanted ECoG grids after the patients reached the complete locked-in state (CLIS), and various methods of interpreting the results are studied. The goal is to investigate changes in evoked responses and frequency spectrum of the ECoG as the patients progress into CLIS and to discuss which changes may be attributed to daily fluctuations in vigilance as opposed to changes caused by disease progression. As no CLIS patient has been able to use a BCI to date [68], the overall aim is to find out why this has not been possible.

1.2. Main Contributions for Patients

Before giving an outline of this dissertation, the main contributions of my work toward the goal of improved communication with LIS patients are listed below.

- An error correction system (ECS) based on error-related potentials (ErrPs) applied to a P3 brain-computer interface (BCI) which was tested with patients as well as healthy individuals [Bensch et al., in preparation]. Most patients were able to increase their communication speed with this system.
- Application of methods for cognition detection in complete locked-in state (CLIS) patients when standard BCI methodologies fail [Bensch et al., in preparation].
- First measurement of the magnetoencephalogram (MEG) of a CLIS patient. A report of the findings, potential benefits and risks of such an undertaking is given here.
- Recommendations from a multiclass study with healthy subjects for designing faster BCIs with nonmotor tasks [11].
- Publication of methods and open source software for internet browsers based on SCP, μ-rhythm and P3 control for paralysed patients, adapted for slow and error-prone input [10, 82, 85, 86].

The ECS is the most valuable contribution, as it was successfully tested with ALS patients. The cognition detection gives insights into the cognitive state of a CLIS patient, but was only applied to two patients using ECoG recordings, which are still uncommon. Similarly, the MEG measurement was performed with a single CLIS patient only. The knowledge gained is that the results in this single case did not justify the enormous organisational and financial effort. The multiclass study contributed to the understanding of nonmotor imagery tasks and their combination with one or more motor tasks to improve BCI speed. The multiclass BCI was not tested with patients, yet bears promise. The availability of internet browsers for BCI users has lead to a number of studies with patients testing these browsers. Many users have given positive feedback.

1.3. Structure of the Dissertation

The motivation to increase communication speed for paralysed patients has been given above. The following chapter will explain the fundamental principles of BCIs. After an explanation of the electrophysiological processes giving rise to the brain activity which is measured, an overview of common recording methods is given. The target groups for BCIs are introduced. Section 2.4.1 delves into the signal processing and feature extraction methods necessary for algorithms to “recognise” particular brain states. The focus is on the methods used and refined in this thesis. The comparison of methods is made possible by the concepts explained in Section 2.5. Finally, the concept of ErrPs is introduced (Section 2.6) as an effect which will be exploited to increase communication speed in patients.

Chapter 3 surveys the current state of the art of BCI systems as pertains to the fields of connectivity features, multiclass approaches, cognition detection and ErrPs. Particular attention will be given to distinguish between results of healthy subjects and patients. It will become evident that the above thesis statement is truly bold, as these topics have rarely been discussed in the literature on LIS patients up to now.

Indeed, the initial investigation of so-called connectivity features (experimental methods presented in Section 2.4.2) had to be supplemented with two studies concerning multiclass classification (Section 4.2) and cognition detection (Section 4.3). Experimental paradigms and methods of said studies are consequently presented under the banner of connectivity in Chapter 4.

The following chapter describes experimental paradigm, setup and methods for the examination of ErrPs in patients and healthy subjects (Chapter 5).

Results for the three major studies performed are given in Chapter 6.

The summary (Chapter 7) discusses the actual realistic benefit for patients in light of the obtained results. Care will be taken to highlight implications which may lead to a more detailed understanding of the numbers presented. Also, hypotheses will be ventured in those cases where the results appear unintuitive. In conclusion, an outlook on possible improvements of the methods presented in this work for patients in need of a faster communication method is given.

2. Fundamentals of Brain-Computer Interfaces

A brain-computer interface (BCI) allows users to communicate or to control external devices by measurement of brain signals without requiring the user to move any muscle. This is particularly helpful for those who have lost all motor control and suffer from complete paralysis. The definition given by Wolpaw et al. [141] in their well-known review of the topic reads as follows:

> "A BCI is a communication system in which messages or commands that an individual sends to the external world do not pass through the brain's normal output pathways of peripheral nerves and muscles."

A more recent review by van Gerven et al. [128] offers an updated and concise view of BCIs. The current chapter introduces concepts vital to the understanding of the dissertation. To this end, it is advisable to begin with a high-level description of the parts that constitute a BCI. Figure 2.1 depicts these parts together with arrows representing the information flow between them. A number of topics surrounding each of these constituents will be discussed more deeply in this work. The reader may refer to Figure 2.2 for a better understanding of how these topics are interrelated and linked to improving communication speed in BCIs.

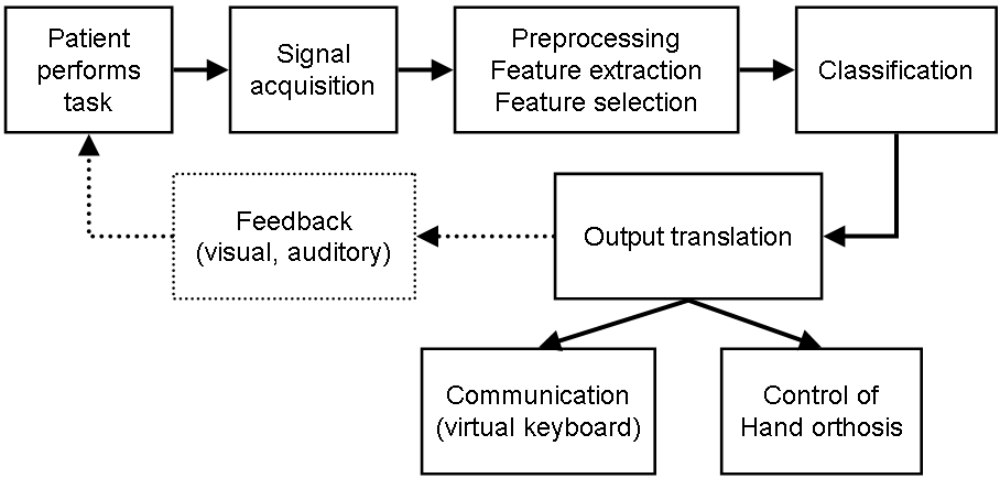


Figure 2.1.: Constituents of a BCI. All parts are required for the system's operation, except for the feedback. Also, the *communication* and *control of hand orthosis* blocks can be replaced with other output options.

2.1. BCI Paradigms

Whenever BCI researchers describe a recording *paradigm*, they refer to the set of parameters governing the stimulus presentation, brain signal recording, and feedback presentation for a BCI setup. These parameters largely depend on the type of brain signal being analysed, to be explained in Sections 2.1.2 and 2.1.3. The recording paradigm implicitly defines an upper bound on the achievable communication speed (see Section 2.5).

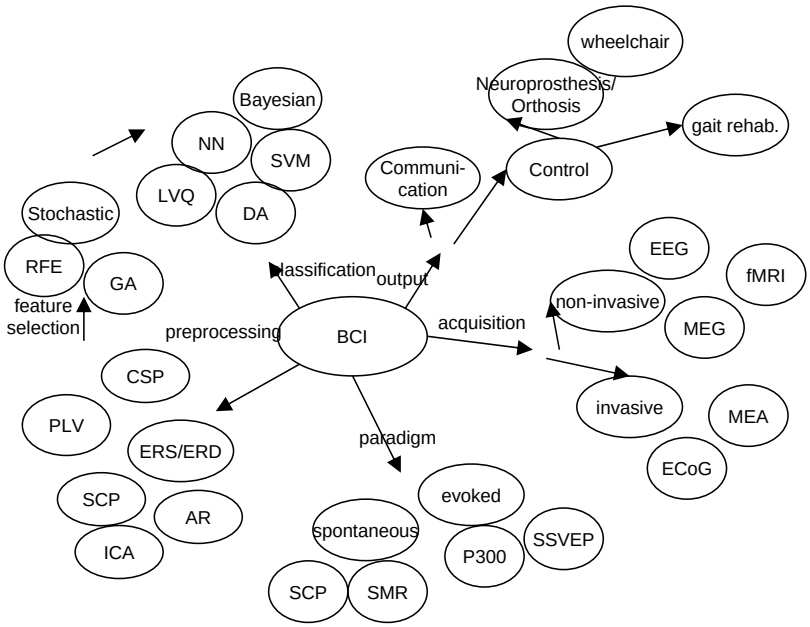


Figure 2.2.: BCI taxonomy showing the interrelation between concepts vital to the understanding of the thesis and further topics. The concept "paradigm" should be seen on a meta-level excluded from the BCI cycle.

An important distinction to make when designing paradigms and evaluating BCIs is between *online* and *offline* BCIs. In most cases, certainly when a BCI is to be used for communication, only the online paradigm is useful. This means that classification takes place during use of the communication interface. Scientific papers often report offline results. Reasons for this include easier data collection, availability of more data, and more time to train a classifier according to the specific data collected. However, online results should be given preference as they portray more closely the actual use of a BCI. Online results are reported wherever possible in this dissertation.

Another dichotomous categorisation in BCI paradigms is made by the timing of the classification interval: It can be either *synchronous*, whereby stimuli notify the user in regular intervals of when they have to engage in mental imagery tasks, or *asynchronous*, allowing the user to set the pace of communication himself. An example is the selection of a letter once a predefined time or amplitude threshold is reached.

For an understanding of BCI paradigms and recording techniques, Section 2.1.1 explains the origin of cortical electrical fields. Sections 2.1.2 and 2.1.3 give two examples of common BCI paradigms — these will be expanded upon in the further course of this dissertation.

2.1.1. Cortical Origin of Electrical Fields

To understand the origin and conscious modulation of electrical brain activity, it is worthwhile to study the functioning of the human cortex. The reader is assured that the ensuing explanations will be kept as simple as possible, considering only those aspects that are relevant to BCI research.

The human cortex is the outer layer of brain tissue, about 0.5 cm thick, consisting of so-called white matter and grey matter (see Figure 2.3). The grey matter is an agglomeration of roughly 10^9 to 10^{10} specialized cells called neurons. Information is transferred within neurons in the form of electrical impulses (action potentials)

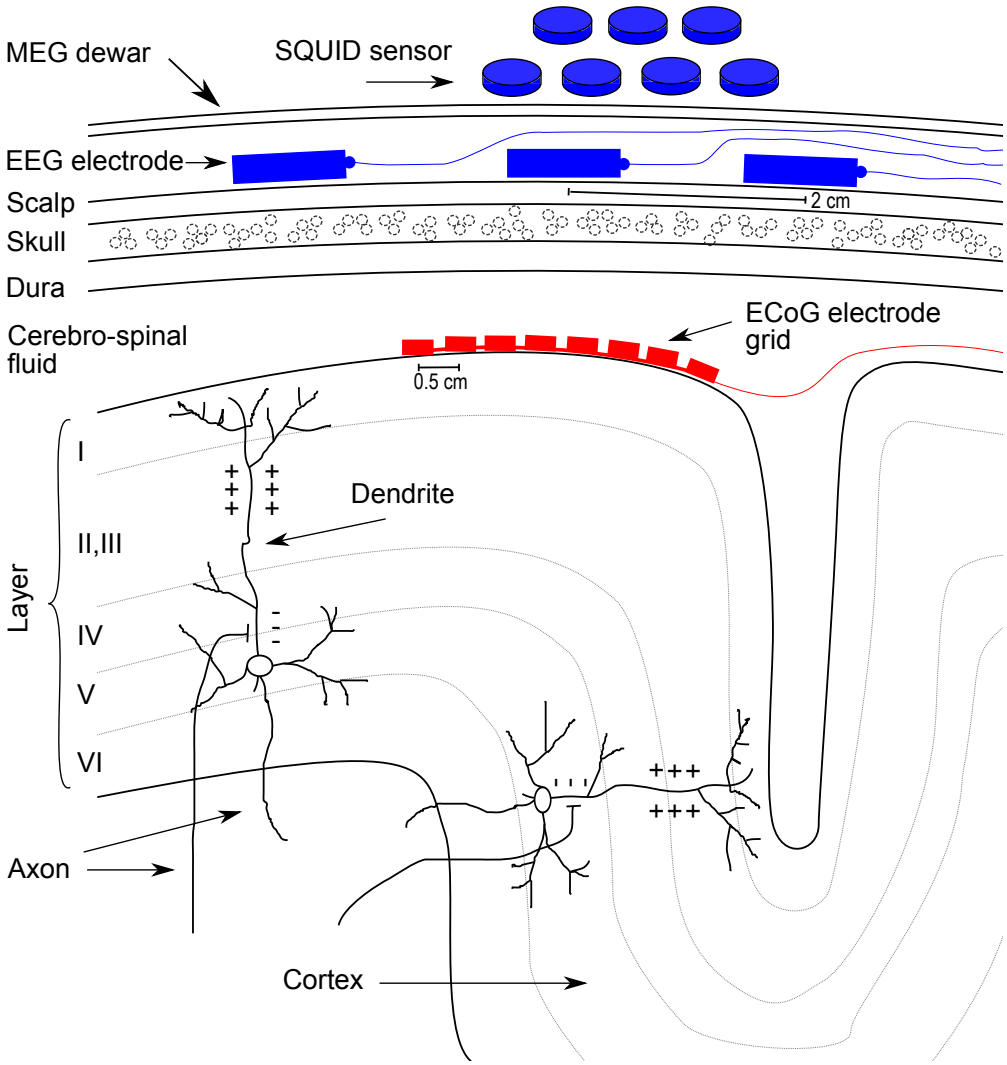


Figure 2.3.: Magnified view of human cortex with electroencephalogram (EEG), magnetoencephalogram (MEG) and electrocorticogram (ECoG) recording methods displayed (adapted from [97]).

lasting 0.5–2 ms on the order of 60–100 mV. The impulse is generally passed to neighbouring neurons by a chemical process. The majority of excitatory synapses in the cortex are connections between pyramidal cells. In fact, the dominant feature of the cortex is the pyramidal cell-to-cell connection, represented by 75% of all synapses [19]. Roughly half of the connections between cortical neurons occur on the order of a few millimetres, with the other half of the pyramidal neurons having axons which leave the cortex and re-enter in distant locations. This makes long-range functional connectivity possible.

The cortex can be partitioned into the frontal, central, parietal, occipital and temporal lobes (see Figure 4.3). The advent of cortical mapping by electrostimulation was brought about by Penfield [102] in the 1950's. Since then, certain areas on the cortex, such as the motor cortex and the sensory cortex, have been explored in much detail. Many BCIs use control signals from the motor cortex, yet other specialised centres exist in the cortex. BCIs often function by detecting specific oscillations of electrical potential in such functionally localised areas. This task is impeded by the fact that no two brains are alike. In ALS patients, the task is more difficult as the brain is affected by the disease, e.g., by cortical atrophy.

Some exemplary pyramidal neurons producing electric and magnetic activity, together with three relevant recording methods, can be viewed in Figure 2.3. All methods shown record a sum potential of neuronal activity, i.e., tens of thousands of neurons contribute to the signal recorded by one sensor. The activity measured on the scalp surface can be the result of deflections as it passes through cerebro-spinal fluid, bone and tissue layers. To obtain evidence for a particular hypothesis, it is therefore often interesting to localise the source of electrical activity (cortical or even subcortical). To this end, the localisation algorithm named low resolution electromagnetic tomography (LORETA) [100] is used in later parts of this work.

2.1.2. Spontaneous Brain Activity

An apt definition of *spontaneous* in the sense of brain activity is "having no apparent external cause or influence; self-acting" (Webster's New World College Dictionary, online). Such a signal is not evoked by external stimuli, but rather the result of an internal process. During BCI operation, there is usually a rest period, followed by an instruction, e.g., mental imagery of limb movement, followed by a classification interval during which the user performs the mental imagery. A pause follows before the next trial. Some trial configurations for the measurement of spontaneous brain activity are shown on page 32. This was done in the multiclass studies associated with the μ-rhythm and nonmotor imagery. In contrast, the cognition detection and ErrP studies rely on evoked response potentials.

2.1.3. Evoked Response Potentials

An evoked response potential (ERP) can be elicited by visual, auditory or tactile stimulation. The user does not need to be "trained", as the paradigm is completely passive, i.e., it relies solely on the selective attention of the user to the stimuli. Visual ERPs are most widely used for BCIs. They can be categorised into normal and steady-state visual evoked potentials (SSVEPs). SSVEPs are elicited by viewing elements on a display which flash at a predefined frequency. The same frequency is then measured over the cortex. The procedure is not further discussed in this dissertation, as SSVEPs currently play a minor role in the plight to make BCIs more accessible for patients.

The most widely-used BCI paradigm concerning ERPs is the so-called *P3 speller*, whereby the user focuses his attention visually to a letter on the screen to be written, or to a letter or number presented auditorially. Farwell and Donchin [38] were the first to test this paradigm of selective attention for brain-computer communication. The user was presented with a 6-by-6 matrix of letters on a display. Columns and rows of letters were flashed in random order. The user selected a letter by focussing his attention on the appropriate letter on the screen. The user was instructed to count the number of times the letter was flashed to increase his attention. A trial consisted of one flash for every column and row, lasting 100 ms apiece. A depiction of such a speller matrix is given on page 47 as presented by the BCI2000 software which is able to record, preprocess, and classify the signal. Each flash of the target letter elicited a so-called P3 response, a positive deflection in the EEG roughly 300 ms after stimulus presentation. The latency varies from person to person, and, as we shall see later, more so for patients. We will thus prefer the term *P3* above the more common *P300* to imply that we are talking about a component of varying latency. Normally, multiple trials are needed to average the EEG signal before a P3 component can be recognised. The number of repetitions can be set dynamically according to the subject before the measurement begins.

More recently, many adjustments to the original design have been made and higher information transfer rates (ITRs) achieved (e.g., [105, 127]). Also, BCIs based on auditory ERPs are being developed [43]. This is useful for those patients who have lost the ability to focus their eyes or to see at all.

Tests seeking to acquire information on the intactness of cortical processing are used in late-stage ALS patients when "standard" BCI systems fail [52, 65]. They too employ auditory ERPs, although not exclusively,

and their primary goal is not the establishment of direct communication with the patient. If successful, they can however be modified to allow slow communication. The electrophysiological cognition tests used in this dissertation are taken from the test battery originally described in [88]. Two tests use simple tones to elicit an ERP, others present words or sentences which require semantic processing. The first criterion to watch for is the N1/P2 component, a short-lived negative peak appearing 100 ms after an important tone, which turns positive around 200 ms post-stimulus. The potential signifies functioning of the auditory system.

The *Mismatch Negativity (MMN)* test measures an involuntary and attention-independent response (N2a) to an occasional deviation in an auditory stimulus sequence, expressed at a latency of 100–250 ms. The *Standard Oddball* test evaluates the patient's ability to perceive a high-pitched tone as a target (deviant) embedded in a sequence of low-pitched tones (standards). A P3 component to deviants is expected with intact attentional processing. This P3 component consists of a frontally maximal P3a component and a parietally maximal P3b component [79]. As only the P3b component is task-relevant, the term P3 will imply P3b in this work. Where a distinction is necessary, the terms P3a and P3b will be used. The ERP components N2a and P3b are both are associated with the cognitive processes of perception and selective attention, defined in [101].

The *Priming* test presents semantically congruent and incongruent word pairs; the N400 component is usually suppressed when the second word is primed by the semantic context of the first word. In other words, a N400 deflection is expected for word pairs with different semantic contexts (e.g., "arm – fish"). A variant of the Priming test uses one word with two syllables, whereby the second syllable either forms a correct word or leads to a nonsensical pseudoword. The advantage of this test is that it can be strengthened by the use of the patient's own name [75]. The *Semantic Oddball* test presents 60 words from one of five semantic categories (plants, animals, tools, professions, body parts) whilst the patient is asked to count the number of words in a particular category. Semantic comprehension is indicated by elicitation of a delayed P3 component (P600–700 in healthy subjects) for the body parts.

2.2. Recording Techniques

It is important to understand the signal recording process and to ensure that the recorded signal is free of artefacts (the term *Garbage in, Garbage out* applies well here). Artefacts can be of biological origin, caused by the recording electrodes, or by electromagnetic disturbance from the electrical mains or medical equipment at the bedside of the patient. Once the signal is picked up by an electrode or a sensor, it is amplified and digitised. Digitisation occurs two-fold: the voltage amplitude, as well as the time (sampling rate) become discrete values. Amplifier resolution, range, and sampling rate have to be tuned to the requirements of the experiment so that no important information is lost. The arrangement of electrodes on the scalp is referred to as a *montage*. The placement of electrodes/sensors in relation to the cortex is shown in Figure 2.3.

Popular software used in this dissertation for stimulus presentation and recording is BCI2000[1] [111] and the BrainVision Recorder (Brain Products, Gilching). For rapid prototyping and online testing of algorithms, both programs can forward blocks of recorded data to an external program such as Matlab®.

As the brain signal's origin and recording paradigms have been explained in the previous sections, we now delve into the specific properties and merits of using EEG, ECoG and MEG in BCI recordings.

2.2.1. Electroencephalogram

The electroencephalogram (EEG) is the simplest and most cost-effective way of recording cortical neural activity. Berger was the first person to do this in 1924 [12]. Usually, a cap with 8, 16, 32 or 64 electrodes

[1] `www.bci2000.org`

is positioned on the user's head. Electrical contact is established by applying abrasive paste followed by conductive gel between the scalp and the electrode. Advantages of recording EEG over the other methods are: low cost, portability, high time resolution, no health risk because recording is passive and non-invasive. Disadvantages are low spatial resolution, dependence of the signal on the placement of the reference electrode, and altered frequency resolution with signal refraction due to cerebro-spinal fluid and bone/tissue layers acting as a lowpass filter. Spatial resolution can be improved by using a high-density electrode cap with 128 electrodes at the cost of prolonged preparation time. Patients normally rely on EEG due to its low cost and portability.

As with the ECoG in the next section, the voltage measured by EEG is a potential difference between two electrodes. One electrode is determined to be the *reference*, which should be placed in an electrically inactive area (often the mastoid, ear lobe or nose). The reference signal is then subtracted from every other electrode. Often, a common average reference (CAR) is obtained afterwards by subtraction of every channel from the mean of all channels, to reduce the noise common to all electrodes. This is particularly useful for μ-rhythm-BCIs [81]. Both referencing methods are used throughout this work.

2.2.2. Electrocorticogram

In an effort to increase spatial resolution, signal bandwidth and signal-to-noise ratio (SNR), the implantation of electrocorticogram (ECoG) grids is becoming increasingly popular. The electrode grid can be placed on the cortex directly (subdural) or epidurally in a procedure which is often termed "minimally invasive". Most often, BCI experiments are performed with epilepsy patients whose grids have been implanted to find the herd of epileptic activity before surgical resection [45, 58]. In these cases, the electrodes are explanted after two weeks, leaving a short timespan for BCI experiments. As the ECoG implant is often not placed over the motor cortex, the possibility exists to test nonmotor paradigms such as auditory imagery [137]. Reports of chronic ECoG implants used for BCI experiments are scarce [56].

2.2.3. Magnetoencephalogram

As with the EEG, the source of the magnetoencephalogram (MEG) signal is electric neuronal activity. However, as electric and magnetic fields are perpendicular to each other, and radial magnetic fields are not picked up by the MEG sensors, the MEG allows improved detection of neural activity stemming from the sulci of the cortex. It is thought that in EEG, neural activity from the gyri partially masks activity from the sulci. The magnetic field's tangential orientation to the skull allows an unhindered passage. Thus, measuring magnetic field solves the problem of deflection and attenuation experienced in EEG due to the skull (also referred to as *smearing*, because this makes the sources harder to localise). Another factor giving the MEG higher spatial resolution is the vast amount of nearly 300 superconducting quantum interference device (SQUID) sensors which are used in newer models. Short preparation time for measurements is often named as an advantage of MEG over EEG, since no electrodes have to be placed, checked for impedance, and filled with gel. This argument is weakened as time is lost due to head positioning and affixation of head positioning coils. The MEG signal is reference-free, which can be seen as a great advantage over both EEG and ECoG. The choice of an incorrect reference electrode (e.g., close to a source of artefacts) can render a recording useless.

However, as the reader might have assumed, no single method is perfect: The MEG system is not portable, i.e., many patients reaching the stage of immobility are not able to use it. As there is no connection between scalp and sensors, the slightest head movement can cause a signal previously picked up by sensor a to be subsequently picked up by another sensor b. Lastly, the fact that only tangential fields are measured can also

Table 2.1.: Comparison of three recording techniques for BCIs. While the first column mentions physical distance between sensors, the second column gives the author's estimate of how well the measured activity can be localised in the cortex. The third and fourth columns are the author's estimate of suitability for subjects and patients.

Method	Channels	Spatial Resolution	Source Loc.	For Subjects	For Patients	Invasive
EEG	1–128	1–5 cm	difficult	yes	yes	no
ECoG	4–128	0.5–1 cm	good	limited	limited	yes
MEG	151–275	0.2 cm	good	yes	limited	no

be construed as a drawback, even though by the shape of the cortex most of the sensorimotor fields produced should be tangential to the scalp.

Reports of BCI studies using MEG are becoming more common [11, 83].

The merits and drawbacks of the three recording techniques with regard to their use in ALS patients are listed in Table 2.1.

2.3. User Groups

Development of BCIs mainly targets the restoration of communication between paralised persons and medical supervisors or family and friends. Although BCIs are also being developed as alternative input devices for healthy people playing computer games, that user group will not be of interest to us. Apart from locked-in patients, we will focus on stroke and epilepsy patients and healthy individuals solely in their role as test persons for algorithms ultimately designed to serve patients.

Obvious differences between healthy subjects and patients are age, amount of medication taken, and the ability to move. Less obvious factors influencing performance are amount of sleep, the ability to understand the instructions given by the supervisor on how to operate the BCI, concentration, and the ability to stay vigilant throughout the course of a communication session. These factors contribute to the difficulty of designing a BCI for patients.

The progressive deterioration of classification rates is thus not a problem that can be solved by the introduction of tuned classifiers alone. Rather, ways to "broaden" the worsening communication channel between the patient and the classifier have to be found — one of the goals of this dissertation. New features, different preprocessing, direct cortical recordings, second-level information to help correct the classifier if it makes a mistake — these are the methods that will be presented further on.

The term BCI *user* will be used whenever a distinction between healthy subject and patient seems counterproductive. Whenever the term *patient* is not further qualified, ALS patients are meant.

2.3.1. Healthy Subjects

Ironically, healthy subjects are the largest user group of BCIs today (see Figure 1.1). In part, this is understandable, since any new algorithms that are tested should first be tested on healthy subjects — however, the application to patients should not be lost from sight as we are still far from an everyday system which can be used in paralysed person's homes on a daily basis.

When perusing studies with healthy subjects, the reader should also be aware that most often, subjects are recruited from the local university's pool of students, i.e., subject age is considerably lower than that of the average ALS patient.

Age-related performance has been investigated in a small number of studies. One study demonstrated larger P3 amplitudes and shorter latencies in younger subjects compared to elderly subjects in auditory ERPs [130]. Differences in scalp topography between young and older adults were found in a meta-analysis of seven studies including the auditory and visual modality [42]. A BCI study using steady-state visual evoked potentials has shown age to correlate negatively with performance [4] and sensorimotor signal strength has been shown to correlate positively with age [108]. Thus, age can play a confounding role in BCI study results, so healthy controls should ideally stem from the same age group as the patients. This was accomplished in the error correction study (see Section 5.3).

2.3.2. Locked-in Patients

This section intends to supply the reader with background information on a group of people who currently stand to benefit the most from improvements in BCI research: people stricken with amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's Disease.

ALS belongs to the group of Motor Neuron Disorders. Though first described by Charcot and Joffroy [26] more than 140 years ago, its early diagnosis remains difficult and its progression unstoppable. Many discrepancies between published ALS studies remain. For example, it is not even known whether regular exercise would benefit or harm patients [31]. Conflicting views on cognitive function abound (see Section 3.3). ALS is a progressive, usually fatal, paralysis of adults caused by systemic motor neuron degeneration. This often goes along with emotional lability. Sooner or later there is chronic weakness with atrophy of affected muscles [93]. A–myo–trophic–lateral–sclerosis may be translated as absence of–muscle–nourishment–side (of spine)–hardening. The spinal muscles (outer limbs) are first affected by the hardening due to absence of movement, while the internal organs remain unaffected. In later stages, all voluntary muscle control is affected, including the lung. An increasing number of studies concerns medical treatment of the disease, yet only a single drug has been shown to mildly slow progression [98]. Patients have to decide between a fate of death by suffocation or tracheotomy with artificial ventilation, whereby artificially ventilated patients can survive longer than five years [55].

Bauer et al. [9] was the first to differentiate between LIS as total immobility except for vertical eye movements and blinking and total LIS as total immobility with undisturbed cortical function. In this work, we follow the LIS/CLIS definition of Birbaumer [15] as he writes

> [...] "the disease progresses until the patient loses control of the last muscular response, which is usually the eye muscle or the external sphincter. The resulting condition is called completely locked-in state (CLIS) complete locked-in state. If rudimentary control of at least one muscle is present, we speak of a locked-in state (LIS) locked-in syndrome."

Disease progression is usually documented with the ALS functional rating scale (ALSFRS), on a scale of 0–40 [24], or the revised ALSFRS (ALSFRS-R) [23] on a scale of 0–48. The ALSFRS-R score will be given for all patients in this work. A weakness of the current rating scales is their failure to differentiate between LIS and CLIS: both stages receive a score of 0.

It is a well-known phenomenon that the ability to communicate with a BCI is usually upheld as the disease progresses, yet lost completely in CLIS [68, 121]. Reasons for this that come into consideration are: neuronal degeneration/reorganisation due to the illness, generally reduced vigilance, varying degrees of alertness from day to day, medical complications which hamper the ability to concentrate on the BCI task. Even the extinction of goal-directed thinking has been suggested [68]. If proven correct, this theory could lead to a loss of interest in BCI research unless some method of restoring goal-directed thinking is found. Yet, some of these points can be addressed by research.

Other neurological diseases such as Duchenne muscular dystrophy, stroke, high spinal cord injury, or Guillain-Barré syndrome may lead to motor paralysis in the form of LIS/CLIS as well.

Tübingen has a long history of working with ALS patients. In the ten years 1996–2006, 33 patients have participated in the BCI research programme, of which 28 were ALS patients and 17 were artificially ventilated. Thirty-one participated in SCP training, 13 did sensorimotor rhythm (SMR) training, and 12 did P3 training [68]. For the reader to gain an insight into the difficulty of creating a working system for patients, their success with the system regarding communication is outlined shortly: chance level was not reached by 11 patients, 5 had significant control, 10 obtained enough control to communicate via a speller (defined as 70% accuracy), and 6 achieved independence by being able to control further environmental switches and being able to communicate well through the system [68]. The time of writing of this dissertation coincided with an attempt by the Tübingen BCI group to regain communication with two late-stage ALS patients by long-term (chronic) ECoG recordings. These patients are described in Section 4.1.

2.3.2.1. Cognitive Function

It is vital to consider the decline of cognitive function in ALS, for obvious reasons: if a person is unable to comprehend or react to instructions any more, there is no way of achieving communication with a BCI. It is important to gain more insight into what type of cognitive processing the brain is still capable of in the CLIS, as patients are generally purported to be cognitively intact enough to follow the instructions of a BCI paradigm, yet have failed to achieve communication in this stage. To blindly continue BCI testing in a patient with no more muscular communicative abilities without doing a cognition detection test would be a case of putting the cart before the horses. The assessment of cognitive functions by electrophysiological methods will be referred to as a *cognition test* [88].

2.3.3. Stroke and Epilepsy Patients

The aim of this section is to briefly introduce two further groups of people who are contributing to BCI research.

The first group is stroke patients with hemiplegia. These people are participating in movement imagery studies using EEG and MEG for rehabilitation purposes. Feedback of recognised movement imagination is given by an orthosis strapped around the hand or on a display in the form of an opening/closing hand. Although this dissertation is devoted to enabling communication, the prospect of controlling a limb with a BCI should be equally interesting in future. The knowledge about sensorimotor cortex function gained in these experiments, lately also being undertaken with ECoG recordings [133], is valuable to communication experiments.

Epilepsy patients contribute to BCI research by participating in ECoG studies with the prospect of ECoG being used more readily with ALS patients in future. Due to open ethical questions, health risks and issues with informed assent by persons with limited communicative abilities, chronic ECoG grid implantations are currently only being undertaken in a small number of patients on a case-by-case basis.

2.4. Signal Processing

The focus now shifts from recording of brain activity to technical and information theoretic concepts of BCIs relevant in the further course of this work. We begin with the four important aspects of preprocessing, feature extraction, feature selection and classification (refer to Figure 2.1).

Preprocessing concerns the spectral or spatial filtering and artefact removal (or correction) of the raw recorded signal after it is amplified and sampled digitally. Typical steps are removal of the so-called DC

component by linear detrending or high-pass filtering at around 0.5 Hz. Depending on the further use of the signal, a notch filter is applied to remove electrical mains interference or a low-pass filter is used to attenuate high frequencies not interesting to the analysis. Commonly, the signal will be downsampled after application of the low-pass filter to simplify online processing, but saved to disk with the original sampling rate for later offline analysis. Artefact removal should be done not only for non-physiological artefacts (line noise, ventilation machines, etc.), but especially for myographic activity (neck muscle activity, clenched jaw, tongue movements, swallowing) and eye movements and blinks in particular, known as the electrooculogram (EOG). Reference to the channels containing the horizontal and vertical eye movement signal will be made with the acronyms hEOG and vEOG, respectively. The electromyogram (EMG) and EOG signals drown out the brain signal due to their far higher amplitude. A current review of EMG and EOG artefacts underlines the importance of checking for these in BCI lest they interfere with the neurological control signal [39]. The online independent component analysis (ICA) algorithm developed at our department [50] will be used in Chapter 5 to filter possible artefacts.

2.4.1. Feature Extraction

Feature extraction is the art of extracting information from the recorded signal which can be used to differentiate between brain states in the best sense possible. The extracted features should be as succinct and informative as possible for the classifier to be accurate and quick. Univariate features are extracted from each recording site separately, whereas bivariate and multivariate features express a particular relation between two or more recording sites (see Section 2.4.2). Reviews report on the diversity of signal processing algorithms for BCI [8, 51].

Feature selection is often neccessary to prevent confusing the classifier with too much irrelevant information. Processing speed is also improved this way. To this end, a small number of highly relevant features is selected. It is not easy to determine the exact cardinality of the best feature set when generalisation is required. In this work, the *alpha estimate* is used. When reducing the original feature set one by one, there generally comes a point characterised by a sharp increase in classification error. After cross-validation (CV), the lowest number of features where the mean classification accuracy p is still within two standard errors (SE) of the lowest overall error estimate is used. That p is called the alpha estimate.

In machine learning, an instantiation of a training data point is mostly spoken of as a training sample. We will use the term *training vector* in this dissertation to prevent any confusion with the term sample as it is understood in the domain of signal processing. A *sample* in our case will denote a discrete amplitude value in µV/mV/fT at a discrete time point within a time series.

2.4.1.1. Time Domain Features

Time domain features are straight-forward to use, as every sample in time (representing a voltage difference) is just that, a feature. These features are commonly used in the P3 speller introduced in Section 2.1.3 and, as we will discover in Section 6.3, they are equally useful in the detection of ErrPs. The signal is usually downsampled to reduce the number of time domain features to classify (to avoid the "curse of dimensionality"). In addition, the voltage amplitudes are scaled to a range suitable for the classifier, often [-1,1]. When ERPs are elicited, a *baseline* period immediately preceding the stimulus is defined (usually of 200 or 100 ms duration). The whole trial is then shifted to obtain a zero mean signal during the baseline period, i.e., for each channel separately it is guaranteed that the sum of all samples equals zero. This increases visibility of the ERP after averaging over multiple trials.

2.4.1.2. Spectral Features

Spectral bandpower features are attained by transforming the raw EEG signal into the frequency domain by a Fast Fourier Transform (FFT) or wavelet transform. The Welch method is used extensively to this end in this work. The use of spectral bandpower features facilitates the analysis of particular frequency bands well known to convey a certain meaning. In this respect, the μ-rhythm band will play an important role in later chapters. The frequency ranges are named as follows: 0.1–4 Hz = delta band, 4–8 Hz = theta band, 8–13 Hz = alpha band, 14–30 Hz = beta band, >30 Hz = gamma band. The μ-rhythm is found at 10 Hz and 20 Hz and should not be confused with the alpha rhythm which has its spectral peak somewhere within the alpha band. The μ-rhythm is functionally related to the motor cortex (and somatosensory cortex), whereas the alpha rhythm has a more posterior topography [90].

Spectral bandpower is a popular feature due to its simplicity and efficiency regarding classification of imagined movement. Other types of spectral features, such as singular spectral entropy and spectral profile [143], have been sidelined in the past.

2.4.1.3. Autoregressive Coefficients

Autoregressive coefficients are helpful to express a time series in a compact way. Three coefficients can suffice for a whole time series without affecting the accuracy of a BCI. The coefficients can be thought of as a hybrid representation; they often clearly represent a particular frequency, yet are calculated from the time series. The Burg method of calculating autoregressive (AR) features was used in this work.

2.4.2. Connectivity Features

Functional connectivity within the cortex has received little attention in the context of BCI research to date (see Section 3.1). What follows here is an explanation of connectivity and a motivation for its use in BCI. Bivariate connectivity can be expressed as a number, for instance in the range [0,1], representing the degree of coupling or causality between two cortical locations. A high value speaks for information exchange taking place between these locations. A BCI can differentiate between two cognitive tasks if the regions or degree of information exchange varies from task to task.

Functional connectivity has been associated with cognitive tasks before [109, 129], making it an a priori viable candidate for BCI features. There are some points to consider, however. Firstly, connectivity features require enough data, e.g., one second, to deliver an accurate value. Secondly, volume conduction can artificially inflate connectivity values. Thirdly, the data should not be re-referenced to a common average. Lastly, because each feature relates to multiple channels, the total number of features to consider is the binomial coefficient

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!} \tag{2.1}$$

with n being the number of observed channels and k being the order of the relation. This simplifies to $\frac{n(n-1)}{2}$ for bivariate features where $k = 2$. Nonetheless, the quadratic complexity makes feature selection inevitable. Feature selection is the intelligent reduction of mostly redundant features, to aid the classifier in making a swift and accurate decision. The actual feature selection method used in this dissertation is explained in Chapter 4.

We focus on bivariate connectivity and present the concepts of coherence and phase synchrony to be investigated in this dissertation in the following two sections. The methods are discussed in more detail in [103].

2.4.2.1. Coherence

Spectral coherence characterising the linear dependence of two channels in a given frequency band f is calculated as follows:

$$\gamma^2_{ab}(f) = \frac{|< C_{ab}(f) >|^2}{< C_{aa}(f) >< C_{bb}(f) >} \quad , \tag{2.2}$$

where $C_{ab}(f)$ is the cross-power spectrum between two signals $a(n)$ and $b(n)$ in a given frequency band f and $C_{aa}(f)$ is the autospectrum of signal $a(n)$. The $< >$ operator implies that the spectra should be mean values across multiple time series. In single trial classification this is not possible, so the available time has to be segmented into multiple windows for the calculation.

2.4.2.2. Phase Synchrony

In contrast to coherence, phase calculations disregard the signal amplitude. The phase-locking value (PLV) characterising the stability of phase differences between two channels a and b in a time window with N samples is calculated as follows [72]:

$$PLV(f) = \frac{1}{N} \mid \sum_{n=1}^{N} e^{j(\varphi_a(n) - \varphi_b(n))} \mid \quad , \tag{2.3}$$

where $\varphi_a(n) - \varphi_b(n)$ is the instantaneous phase difference of two signals at time n. Here, the instantaneous phase is calculated by Hilbert transform [106]. The f indicates that PLV is a function of frequency; the signal should be bandpass-filtered to obtain a physiologically meaningful result. The PLV adheres to the inequality $0 \leq PLV \leq 1$ with a PLV value of 1 signifying high synchronisation between channel a and b. To obtain the Hilbert Transform, one proceeds as follows: A signal $x(n)$ can be written as a combination of its amplitude and phase, $\zeta(n) = A(n) \cdot e^{j\varphi(n)} = x(n) + j\tilde{x}(n)$. This is known as the analytical signal, introduced by Gabor in 1946. The imaginary part of this complex signal is defined as the Hilbert Transform of the real part, i.e., $\tilde{x}(n) = x(n) * (1/\pi n) = \frac{1}{\pi} P.V. \int \frac{x(\tau)}{t-\tau} d\tau$, where $*$ is the convolution operator. One then obtains the phase by $\varphi(n) = \arctan \frac{\tilde{x}(n)}{x(n)}$.

2.4.3. Feedback

Feedback is a fifth (optional) step in the BCI processing pipeline.

Feedback related to the EEG signal recorded can be visual, aural, or haptic. Visual feedback is used in most cases where the user's eyesight is in order. ALS patients (see Section 2.3.2) experience a deterioration of vision in later stages of the disease, making auditory feedback the method of choice. Haptic feedback has also been used in previous experiments. Whichever feedback modality is chosen, it is important to distinguish between *continuous feedback* and *feedback of results*. *Continuous feedback* is given during a trial, i.e., when EEG relevant to the classification is recorded. It can effect a short-term change of strategy to correct a wrong classification in the making; and it effects long-term changes in strategy on behalf of the user (learning). The primary goal of *feedback of results*, which happens at the end of a trial, is to inform the user of the classification result. This can be vital information for the user to consider his next step. Here, the learning of the user is effected in a much more indirect manner, as the contingency between mentation and feedback is lost.

Let us have a closer look at feedback in the visual modality, as this is the most common form of feedback in BCIs. Whilst it is obvious that feedback of results, as explained above, can only be beneficial to the user (the chosen letter is displayed on the screen, the chosen internet link is highlighted in a web browser),

continuous feedback is a contentious issue. The advantages mentioned above can only be considered valid if a method to prevent or correct for eye movements has been put in place. Several studies have received criticism as the eye movements made to track the feedback can overlay the EEG signal, indirectly influencing the classification outcome. In addition, one has to ensure that the feedback is valid and timeous. The studies described later do without continuous feedback.

2.5. Evaluation Criteria and Classification

Classification actually precedes feedback in the BCI pipeline (Figure 2.1). As it is such an important topic and best combined with the discussion of evaluation criteria, a whole section is devoted to this theme.

2.5.1. Bit Rate

The reader will probably agree that speed of operation and accuracy are the vital evaluation criteria of any BCI suitable for communication or environmental control. Both criteria are usually mentioned in publications; they can be elegantly combined into one variable, the so-called *bit rate* or information transfer rate (ITR) (the number of error-free bits transferred per unit of time). A similar, more applied measure (limited to communication) is the number of correct letters per minute.

It is difficult to find a measure which allows a fair comparison between the plethora of configurations and paradigms that exist in the BCI world today (Section 2.1 mentions but a few). Yet it can be said that the bit rate measure allows a fair comparison between BCIs of the same type. However, the use of such a universal measure brings the drawback of inexact bit rates in very particular cases.

An example should serve the purpose: If a user wishes to write an error-free message with a P3 speller, the commonly used Wolpaw bit rate [140]

$$b_W = \log_2(M) + p * \log_2(p) + (1-p) * \log_2 \frac{1-p}{M-1} \tag{2.4}$$

merely represents an upper bound on the actual bit rate, i.e., it is overoptimistic (M is the number of classes, p is accuracy). Some further restrictions apply to b_W (e.g., equiprobability of classes in desired and actual selection). Thus, in this particular example of a spelling interface, a bit rate definition representing the number of correct letters spelled represents a more exact measure. Example:

$$b_S = \frac{b_L}{T_L} = \frac{\log_2(M-1)(2p-1)}{c} \tag{2.5}$$

where $b_L = \log_2(M-1)$ is the number of bits conveyed per letter, and $T_L = c/(2p-1)$ is the time taken to spell a letter (see [30] for the derivation). Here, c is the trial length in seconds.

Put simply, the total number of correct letters written on the display is the difference of correct trials and incorrect trials, since each incorrect trial places the user a step further away from the goal (whether it be an erroneously spelled letter, which has to be deleted, or an inadvertent deletion of a correct letter) and a correct trial places the user a step closer to his goal. Rephrased in terms of the average accuracy of a P3 speller, p_{P3},

$$T_L = \frac{1}{2p_{P3} - 1} \tag{2.6}$$

is the average number of trials required per letter. Thus,

$$b_M = log_2(M-1) \cdot (2p-1) \tag{2.7}$$

		outcome	
		error	correct
predicted	error	TP	FP
	correct	FN	TN

		feedback letter displayed	
		incorrect	correct
ErrP prediction	positive	ErrP classifier recognises an error. A backspace is invoked. The letter is indeed incorrect (best case)	ErrP classifier recognises an error. A backspace is invoked, but the letter displayed is actually correct (worst case)
	negative	ErrP classifier does not recognise an error potential, even though the letter displayed is incorrect.	ErrP classifier does not recognise an error potential. The letter is indeed correct.

Figure 2.4.: Confusion matrix for a binary problem (left) and applied to error correction (right). TP = true positive, FP = false positive, FN = false negative, TN = true negative. The matrix is M-by-M for an M-class problem.

is the normal bit rate.

Whatever the measure, the most vital piece of information which has to be recorded during an experiment is the *confusion matrix*. All other evaluation criteria can be inferred from this. A confusion matrix for a two-class problem is shown in Figure 2.4.

Some variables related to the confusion matrix are as follows:

$$p = \frac{TP + TN}{TP + TN + FP + FN} \qquad \text{(accuracy)} \tag{2.8}$$

$$r_C = \frac{TN}{TN + FP} \qquad \text{(specificity, true negative rate)} \tag{2.9}$$

$$r_E = \frac{TP}{TP + FN} \qquad \text{(sensitivity, true positive rate)} \tag{2.10}$$

To consider multiclass classification, a more general notation for accuracy is introduced [117]: $p = \sum_{i=1}^{M} n_{ii}/N$, where N is the number of test data vectors classified. This allows us to define the proportion of data vectors for which agreement is expected by chance as $p_e = (\sum_{i=1}^{M} n_{:i} n_{i:})/N^2$. We are now in a position to state Cohen's kappa coefficient [28] as a measure of performance comparison between BCIs with varying number of classes:

$$\kappa = \frac{p - p_e}{1 - p_e} \quad . \tag{2.11}$$

Finally, a word of warning: Published bit rates may be misleading due to the exclusion of intertrial intervals and similar pauses required by the BCI paradigm. Also, systematic misclassification may even increase the bit rate. There are authors who report extremely high values of b_W for P3 spellers accompanied with accuracy levels below 50%. In practice, this would result in too many backspace operations, and the user would not be able to spell anything (this is discussed in Sellers and Donchin [120]). As the evaluation criterion b_S equals 0 when accuracy is below 50%, it is a far more realistic measure in the context of a P3 spelling interface and is used to in the evaluation of the error correction study in Chapter 5.

2.5.2. Receiver Operating Characteristics

An evaluation criterion which is used extensively in this dissertation due to its flexibility is the receiver operating characteristics (ROC), or more specifically, the area under curve (AUC) [7] in the range [0,1]. Although ROC curves are chiefly used for binary problems, multiclass extensions exist. The AUC value of

0.5 stands for no discrimination between two classes. Values of 0 and 1 stand for perfect discrimination and are equally preferable.

The maximum AUC score shown in the figures (e.g., Figure 6.3) should never be considered on its own. The score is highly dependent on the number of trials and features, as well as the data itself. For example, if there are a few high-amplitude artefacts in the data, this could already lead to a high AUC score. The effect is exacerbated if a low number of trials or high number of features is used. Thus, ROC plots shown in the results are furnished with additional information to allow a more realistic estimate of class discriminability:

- A random plot using random data and randomised labels
- A plot using the actual data and randomised labels
- A plot showing a baseline period, e.g., the inter-trial interval (ITI) and the classification interval (CI)
- An AUC 5 % confidence interval, which means that values outside of this interval are statistically significant

Figures with randomised labels serve as an indicator for the reliability of the AUC plot for a particular dataset. For example, if a dataset has too few data points in a particular class, both the correct and the randomised plot could display high AUC values due to statistical variation. This indicates low informative value of the AUC score.

Because 100 % classification accuracy is hardly achieved with BCI, the colour scale in the figures saturate to white at the AUC values 0.1 and 0.9, thereby increasing the colour range and readability of the figures. Both ends of the scale are "good" values i.e., turquoise and yellow are equally desirable.

2.5.3. Statistical Test For Two Distributions

The simplest method of testing two distributions for discriminability is a statistical test of difference of means or medians of the distributions. This is, of course, not of any use in an online BCI (where *single-trial* methods are required). However, for those patients whose offline data is studied in the course of this work, the statistical test is regarded a safe method when testing for the existence of ERPs [88].

The t-test is used to test for difference of means, and the Wilcoxon Rank test is used for difference of medians. When using the t-test, it is important to ensure that the data points follow a normal distribution.

It can also be used to group a large number of recordings to test whether a patient has been able to respond to the BCI commands or not. Statistical significance can be achieved in this way, even if communication on a session-to-session basis fails due to the classification rate being too low to operate a communication interface.

Furthermore, to remove the possibility of a methodological error in calculating the statistics, the test can be repeated multiple times with randomly assigned labels in a Monte Carlo fashion. For a particular dataset, the number of statistically significant results and their standard deviation (SD) give an indication of how meaningful the original test was.

The t-test will also be used to test for significant differences between results or other information of various populations of patients or healthy subjects. In such standard use, the t-test will be performed at a significance level of $\alpha = 0.05$ unless otherwise stated.

2.5.4. Binary Classification

Binary classification belongs to the bread-and-butter tasks of BCI research. The multiclass methods introduced in the following section and used within this dissertation (see Section 4.2.4) will rely on the algorithms explained here.

Binary classification concerns the assignment of a class label $y \in \{-1, 1\}$ to an uncategorized data vector x. A classifier decides based on a set of labelled *training data* vectors, as well as any a priori information given.

Examples of binary classifiers are linear discriminant analysis (LDA), step-wise linear discriminant analysis (SWLDA), support-vector machine (SVM) or artificial neural networks. Most of these have *hyperparameters* which must be adjusted to the data set at hand. The adjustment is usually done in an n-fold cross-validation (CV) loop, whereby the data is divided into a training set of $n-1$ blocks and a validation set of 1 block. Then n training runs are performed, so that each block of training vectors is used as a validation set once, and the n results are averaged to obtain the CV result. The process is repeated for as many combinations of hyperparameters as is feasible to test; however, the search should not be too fine-granular to avoid overfitting.

A recent review of classification algorithms [78] states that the SVM is popular due to its good results in BCI. Generally, the radial basis function (RBF) kernel is used. However, the LDA is also used successfully when speed is important (e.g., in online classification).

The SVM used here is a soft margin classifier and has a RBF kernel. The best hyperplane separating the training patterns is found by minimising the expression

$$\frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi_i \tag{2.12}$$

subject to the two constraints $\xi_i \geq 0$, where $i = 1, ..., n$ and $y_i(\langle \mathbf{x_i}, \mathbf{w}\rangle + b) \geq 1 - \xi_i$, for $i = 1, \ldots, n$. The first term represents the inverse of the width of the margin. The second term "punishes" training patterns inside the margin, so minimising this term prevents too many margin errors. The regularisation constant C weights this term.

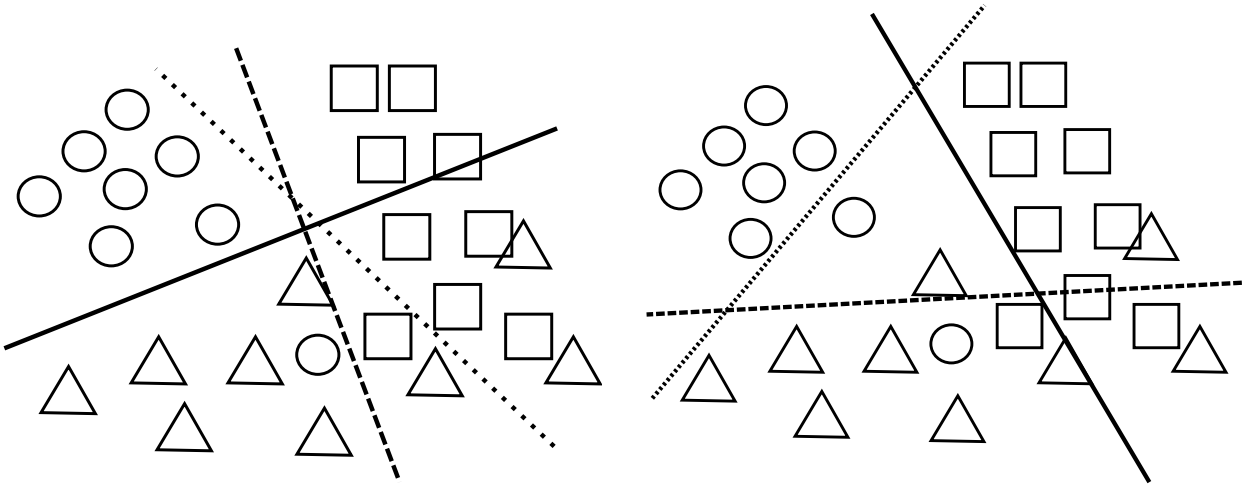
2.5.5. Multiclass Classification

In a multiclass problem with M classes, the set of class labels y contains M elements. Multiclass BCIs can be compared with the b_W measure if they have the same number of classes, and with κ if the number of classes differs. Accuracy cannot be used to compare results obtained with different numbers of classes, as the chance level $1/M$ varies dramatically depending on M.

Any intrinsic binary classifier can be generalised to the multiclass case by one of two popular methods: A *one-vs-rest* or pairwise *one-vs-one* configuration can be used to generate multiple binary decisions which aggregate into one answer by a voting procedure. It is clear that the one-versus-one (OVO) configuration requires $M \cdot (M-1)/2$ classifications, whereas the one-versus-the-rest (OVR) method only needs M decisions. Even though the OVR method is hampered by unequal number of training vectors, it is the method of choice for most BCI researchers [78]. The OVO system has the additional benefit that it is usually easier to separate two classes than to separate one class from all the others. We will return to this multiclass method in Section 4.2.4. The concept of the two methods is shown in Figure 2.5.

2.6. Error-Related Potentials

The first error-related potentials (ErrPs) were already reported twenty years ago. A classification of various types of ErrPs, depending on the task and feedback, has since been undertaken. More recently in 2000, ErrPs were measured by Schalk in a BCI environment [110]. Before going into detail about ErrPs and their use in BCI (Section 3.4), we shall introduce the general concept here.



(a) Classification according to the OVO method.

(b) Classification according to the OVR method.

Figure 2.5.: Example of multiclass classification for three classes. Possible decision surfaces are shown for both variants.

It is well established that the brain reacts to detection of errors with a characteristic potential referred to as error-related negativity [37]. It has been repeatedly shown that the error-related negativity is elicited by erroneous selections with a BCI [40, 110]. However, BCI research has shown this potential to contain a positive element as well. To ascribe for this, it is referred to as error-related potential (ErrP) in this work. Ferrez and del R Millán [40] describe the same effect as an "interaction error potential".

The question of how to maximise the overall communication speed with such an error correction system (ECS) is almost a philosophical one. Theoretically, an ECS should offer an error correction strategy which maximises the bit rate. It is easy to train the ErrP classifier accordingly. Secondary effects of this strategy (e.g., false positives) will, however, almost certainly lead to non-optimal results. From a human point of view, it is understandable that the removal of correct letters would lead to frustration, impatience or similar emotions of the user, degrading his performance in subsequent trials.

Thus, the indeterminate psychological factor has to be taken into account, even if there is no a priori optimal solution. For example, the number of expected false positives can be bounded by appropriate training of the classifier to prevent user frustration (leading a priori to a non-optimal bit rate increase). Due to less interference with the system, though, this could result in the best bit rate overall.

A further problem with ErrP classification is highly unbalanced data, which often degrades classification performance. There exist some methods to mitigate the effect, details of which will be supplied in Section 5.5.

Some say that the ErrP could in fact be a P3 response. If that were the case, it would be more prominent in those subjects who commit less errors. Within the context of BCI error correction, the ErrP has to be classified after a single trial, whereas the P3 is usually averaged over 5–15 trials.

An overview of current state-of-the-art methods of extracting ErrPs from EEG for use in a BCI is given in Section 3.4.

3. State of the Art

The BCI fundamentals required for an understanding of the state of the art were covered in the previous chapter. We now turn to an overview of the state of research as it stands pertaining to BCI studies performed with and for patients. Thereafter, current research covering bivariate features, multiclass BCIs, cognition detection and ErrPs is covered. To enable a fair comparison, bit rates will be reported as b_W (Wolpaw bit rate, see (2.4)) or b_S (see (2.5)). Whenever there is insufficient information to calculate the bit rate, accuracy will be reported instead. Care will be taken to distinguish between results reported for healthy subjects and patients, and whether an online or offline method was used (see Chapter 2.1 for an explanation).

As the stated focus of the thesis is on patients, some general results by researchers working closely with patients are presented first.

In 1999, one of the first successes in working with patients was made public, a system which allowed two ALS patients to choose letters on a display by modulating their SCPs [14]. The accuracy in the two-class experiment was around 75 %, enough to spell 2 letters/min. A second single-case study with a cerebral palsy patient achieved roughly 1 letter/min by β-band event-related desynchronisation (ERD) [89]. As these were single-case studies and the modulation of SCPs and β-band ERD required lengthy training, researchers set out to improve on these first successes.

Kübler et al. ([70], $N_p = 4$)[1] went on to show that ALS patients were able to operate a BCI by modulating their µ-rhythm in 2005. A training phase of 3–7 months was needed to achieve an average of 78 % accuracy. The patients learned to modulate their µ-rhythm by hand or foot movement imagery.

The same year, Hill et al. studied five paralysed subjects (two afflicted with ALS) who were unable to communicate, i.e., in CLIS [56]. They failed to surpass chance level using a BCI based on the µ-rhythm. These patients had progressed much further in the disease compared to Kübler et al.'s patients, and the results are based on a single recording session.

A number of studies with the ERP paradigm followed, as the ERP paradigm promises high bit rates. In 2006, Piccione et al. tested the P3 response in a study including one ALS patient ([105], $N_h = 7$, $N_p = 5$). The ALS patient managed the task of fixating one of four flashing arrows on the screen with 80 % accuracy. The other patients and healthy subjects were all able to use the P3 as a control signal, yet the bit rate achieved was relatively low due to the study design. Sellers and Donchin in 2006 achieved 0.7 bit/min with three ALS patients, also using a four-choice (yes/no/pass/end) P3 BCI [120].

Hoffmann et al. tested their P3 BCI with disabled subjects in 2008 and were able to substantially increase the bit rate to 19 bit/min ([59], $N_p = 5$). However, only one patient had ALS. The others were wheelchair bound with cerebral palsy, multiple sclerosis, traumatic brain injury and post-anoxic encephalopathy. The higher bit rate might stem from the fact that a six-choice visual paradigm was used and a shorter trial duration was used (400 ms).

Still in 2008, Nijboer et al. presented the first P3 speller results for people with ALS, claiming that BCI performance, as well as P3 amplitude and latency, remained stable over several months. Three of their patients achieved an average of 11.3 bit/min during free spelling ([91], $N_p = 6$). The correlation between degree of impairment and BCI performance was not statistically significant.

[1] To give an indication of significance of a study, N_p and N_h shall denote the number of ALS patients and healthy participants, respectively.

A first longitudinal study on BCI performance in ALS patients was published in 2009 by Silvoni et al. [121]. Although 21 patients participated, only five could be tested one year later. The test did not reveal any difference in BCI performance compared to the first measurement. Also, no significant correlation between BCI performance and level of disability was found. Mean accuracy was 78 % (N_p = 21) in a four-class paradigm.

Patients often cannot focus or see properly in LIS or CLIS stage, preventing the use of a regular visual P3 paradigm. Kübler et al. introduced an auditory P3 BCI in 2009 ([67], N_p = 4). Two patients achieved above chance accuracy of 25 %, which is insufficient for communication.

The results indicate stable BCI performance for most part of the disease, yet a sudden loss of BCI control once the CLIS stage is reached. This finding is reiterated by a recent meta-analysis by Kübler and Birbaumer ([68], N_p = 29). Their patients were in various stages of physical impairment. No correlation between impairment and BCI performance was found when excluding CLIS patients, but the correlation was negative upon inclusion of the CLIS patients.

3.1. Connectivity

The general research area of cortical connectivity is vast. We shall limit ourselves to the discussion of a few publications relevant to the recognition of mental imagery for BCIs.

This brings us to the work of Gysels et al., who found that synchronisation measures, specifically the PLV, combined with the classical power spectral density (PSD) features can lead to improved results compared to using one of the measures exclusively ([49], N_h = 3). A later paper employs feature selection ([48], N_h = 5), but does not investigate a combination of classical and bivariate features. Moreover, the results are difficult to interpret as the classifier used an "undecided" class. Song, in collaboration with Gysels, found discriminative information in motor tasks by extracting phase synchrony rate features ([124], N_h = 5). The synchrony rate feature is, however, a simplification of the PLV feature and consequently did not perform as well as PLV. An ensuing publication ([125], N_h = 5) compared covariance as a bivariate feature, applied to spatial and temporal filtering, with the spatial filtering methods CSSP [77] and CSSSP [35]. Song's method performed in the same range as the other two methods, with a mean accuracy of 90 % (offline CV).

The achievement of online control using PLV features in a three-class study by Brunner ([20], N_h = 3) marked a major step forward. With PLV features alone he achieved κ = 0.19–0.43; combined with PSD features κ = 0.52–0.66. The classification accuracy of 46–62 % is not satisfying, but the online control was a first. Wei et al. achieved higher classification accuracy in their study of amplitude and phase coupling ([134], N_h = 5). But the accuracy was still slightly lower than their result for AR coefficients and was achieved in 10-fold CV. Both of these studies did not test nonmotor imagery, which might have been interesting.

The aforementioned results are summarised in Table 3.1. A detailed overview of coupling methods and their impact on classification for BCI is given in [13]. In the MEG study where subjects had to alternate between an overt movement task and a resting task, the combination of PSD and coupling features was beneficial only in the higher frequency bands such as β and above. Connectivity has also been investigated in ECoG and MEG. For instance, phase locking can be used to find coupling between low and high frequency bands in ECoG, possibly representing communication in the brain during cognitive processing [22]. In MEG data, coupling between frontal and posterior areas during motor imagery of hand movements was also found [33].

Table 3.1.: Overview of important BCI studies employing cortical connectivity features. The first study was a three-class study.

Authors	Year	N_h	mode	Accuracy	Comment
Gysels and Celka [49]	2004	3	offline	62.5 %	PLV + PSD better than PSD alone
Song et al. [124]	2005	5	offline	90 %	PS better than CSP for some subjects
Brunner et al. [20]	2006	3	online	48–57 %	online control possible
Wei et al. [134]	2007	5	offline	87–93 %	

3.2. Multiclass BCI

A first report of classifying three different EEG patterns came out in 1990 ([62], $N_h = 5$). In that work, Keirn and Aunon tested five nonmotor mental tasks (rest, mental arithmetic, 3D object rotation, mental letter composing and visual counting) and achieved a mean accuracy of 84.6 % with OVR classification. The result seems formidable for the relatively low number of trials that were used — however, a naïve classifier achieves 80 % in such a case where the majority class is four times larger than the minority class. As a comparison, Anderson and Sijerčić reported two-class accuracy between 38 % and 71 % on the same dataset six years later ([5], $N_h = 4$). Still in 1996, Kalcher et al. were able to separate the overt movements (left index finger, right index finger, foot) with an accuracy of 60 % and imagined movements with a mean accuracy of 51 % ([60], $N_h = 4$).

More related to today's understanding of brain-computer interfacing, Obermaier et al. investigated discriminability of up to four classes ([94], $N_h = 3$). The authors found that $M = 3$ classes offer higher bit rates than two or four classes. However, as the number of participants in the study was low and the two-class result (0.58 bit/trial) was very close to the three-class result, the reader should be cautious not to overrate this result.

A study using two motor (opening and closing the left or right hand) and two nonmotor imagery tasks (spatial navigation and auditory imagery of a familiar tune) concludes that nonmotor task pairs can be discriminated easier than others ([29], $N_h = 10$). The navigation/auditory imagery pair reached 74 % accuracy whereas the left motor/right motor pair reached 71 %. Generally, participants found the nonmotor tasks easier to perform. A thorough investigation of accuracy and bit rate for two to six classes by Dornhege et al. followed ([34], $N_h = 10$). He too arrived at the conclusion that the use of three classes was optimal.

In addition to Obermaier's publication in 2001, the Graz lab has been quite active in testing multiclass paradigms. A three-class asynchronous BCI using three motor tasks for spelling was shown to produce an average of 1.99 letters/minute [112], but is difficult to compare with synchronous BCIs. A more complete overview of the Graz three-class BCI is given by the same author three years later [113], however, no bit rates are given. The Pfurtscheller group did however publish some 4-class results in 2005 that showed the SVM to be superior to other classifiers when classifying adaptive AR parameters ([116], $N_h = 5$).

More recently, a 4-class BCI was introduced which used two binary classifiers in parallel to classify left hand and right hand movement separately [44]. Together with the resting state, this results in four classes. The results were offline only and not compelling.

An overview of important publications for multiclass BCIs is given in Table 3.2. Overall, the results are heterogeneous — even with a standard bit rate some publications do not offer enough information to calculate that bit rate. The majority of studies were undertaken by measuring EEG; modalities such as ECoG and MEG are as yet underrepresented. This work will present a multiclass study with brain activity measured by MEG.

An important aspect from a computer science point of view is the investigation of classification methods used in multicategory classification. If a committee of binary classifiers is used, the strategy of mapping multiple outputs onto one classification result also becomes relevant. Lotte et al. [78] state that the OVR

Table 3.2.: Overview of important BCI studies employing multiclass classification. M denotes the number of classes investigated and b_w/trial is the Wolpaw measure of ITR. If not enough information was given to calculate b_w, accuracy is shown.

Author	Year	N_h	mode	M	b_w/trial	Comment
Obermaier et al. [94]	2001	3	offline	5	0.59	3-class better than 2 or 4
Curran et al. [29]	2003	10	offline	4	74 %	nonmotor tasks easy to discriminate
Dornhege et al. [34]	2004	6	offline	6	0.77	3-class, see paper
Scherer et al. [112]	2004	3	online	3	-	asynchronous, 2 letters/min
Schlögl et al. [116]	2005	4	offline	4	$\kappa = .51$	SVM better than rest

method is generally preferred over the OVO method of combining classifiers. This issue will also be investigated in the further course of this work.

3.3. Cognition Detection

This section will highlight the state of research as regards cognition detection in ALS patients. The role that possible deterioration of cognitive ability plays in the late stages of ALS is unclear (see below). Specifically, it is open whether the decline in BCI performance that patients present after transitioning to the complete locked-in state has to do with their cognition being affected by the disease. Another suspected cause for the decline in performance is the patient's lack of movement and inability to influence the outside world [68].

In the LIS cognitive functioning can be assessed by established active or passive cognitive stimulation paradigms and recording brain signals [57, 65, 88]. Little is known about cognitive functioning in completely locked-in patients due to ALS and other progressive neurological diseases. Several authors reported neuropsychological [80], morphological [1, 2] and electrophysiological [52, 107, 131] signs of cognitive impairment in ALS. Others, however, did not find any signs of cognitive deterioration even in advanced stages of ALS [73]. Among possible reasons for the divergent results, there may be numerous artefacts related to recording conditions in locked-in ALS patients: artificial ventilation and tracheotomy and other life support equipment exclude testing in a MR-environment and electrical and mechanical movement artefacts from the devices influence the recording environment. Intervention by nurses (e.g., regular suction of saliva through the tracheotomy, change of position, temperature adjustment due to excessive sweating) further impedes EEG measurements. A major difficulty is the changed circadian cycle and sleep-waking patterns in patients living in such a condition: patients sleep at a different rate and experience episodes of low arousal [76] precluding regular recording conditions, which is particularly important for cognitive tests.

However, evidence is lately accumulating in favour of the cognitive impairment theory. The ALSFRS-R showed a positive correlation between grey matter volume reduction in a magnetic resonance imaging (MRI) study by Grosskreutz et al. [46, N_p =17, N_h =17]; Boyajian et al. [18, N_p =7, N_h =8] believe to have found cortical dysfunction indicated by a correlation of density of MEG slow wave dipole sources in cingulate gyrus with ALSFRS; Ogawa et al. [95, N_p =19, N_h =19] discovered cognitive deficits in ALS patients with an ERP test battery. A large-scale study by Witgert et al. [138, N_p =225] links cognitive impairment to behavioral dysfunction in ALS. To round off the argument, an evidence-based review by Miller et al. [84] concludes that a significant proportion of ALS patients demonstrate some cognitive impairment.

In essence, many studies speak of cognitive deterioration, while others refute it. Yet, even if patients are affected by cognitive deterioration, there is no proof that previously acquired BCI skills lose effect, as studies have reported constant BCI performance over time for patients up to the CLIS stage. Up to now no test has been reported of applying the cognition test to completely locked-in ALS patients in ECoG.

3.4. Error-Related Potentials

The concept of error-related potentials (ErrPs) was introduced in Section 2.6. This chapter will consider the advances that the use of ErrP, specifically feedback ErrP, have brought to the BCI community up to this point in time.

It was Schalk et al. [110] who for the first time in 2000 suggested that ErrP could be useful to speed up communication in BCIs. He measured a positive potential at Cz at a latency of 180 ms after the trial (up/down cursor movement task) ended. His results are not conclusive, as the effect could be based on error anticipation (movement of cursor was visible throughout the trial) and an offline simulation of bit rate increase by application of a threshold criterion to recognise ErrP was used. His offline analysis projects an average bit rate increase of 0.03 ± 0.036.

Blankertz et al. [16, 17] followed suit in 2002 and 2003 . The study, however, was based on a forced-choice task rather than a BCI task. N_E was found at a latency of 10–60 ms at a fronto-central location and P_E was discovered at a latency of 220–260 ms. $85\,\%$ of the errors could be detected with a false positive (FP) rate of $2\,\%$ in 7 of 8 subjects.

Likewise, Parra et al. [99] use a forced-choice visual discrimination task to detect ErrP. As with Blankertz, this entailed an overt movement (pressing a button). A fronto-central negativity was observed within 100 ms after the user's response. Correct trials led to a more prolonged posterior positivity.

A further key-press study was conducted by Buttfield et al. [21], whereby one of two buttons had to be pressed and the interface simulated a wrong button press 20 % of the time. This setup ensured that the error was made by the interface rather than the user himself. Using a bit rate model presented in [41], a 72 % increase in bit rate was achieved by assuming a theoretical accuracy of 80 % for the BCI. However, the work was never integrated into a BCI system. A more elaborate publication from the same lab [40] investigates, once again, ErrP elicited by manual button presses. Intriguingly, this study consisted of two sessions which were three months apart. Using the first session as training data, the gaussian classifier achieved $84\,\% \pm 5.3$ specificity[2] with a sensitivity of $79\,\% \pm 7.3$.

Finally, at the Graz 2008 BCI workshop, Visconti et al. [132] presented an ErrP recognition system built into a P3-based BCI. It was programmed for an error rate of 20 %. 4 of 5 subjects benefitted from that system as long as the accuracy of the P3 speller was $< 75\,\%$. The specificity and sensitivity values for the 4 best subjects were $86\,\% \pm 3.6$ and $75\,\% \pm 4.9$, respectively. Areas to improve on this study would be to enlarge the sample size of healthy subjects to 10, and to test the system with patients. Also, the results were obtained with a 3-fold CV. This implies that the system was not working online. The first online ECS using ErrPs was presented in 2009 by Dal Seno et al. [30]. ErrPs were detected above chance level in 2 of 3 subjects. For those two subjects, sensitivity and specificity were $62\,\% \pm 5.7$ and $68\,\% \pm 2.1$, respectively.

The current state-of-the-art might be summed up as in Table 3.3. All studies agree that the Cz/Fz electrodes are best for picking up the ErrP. Note that none of those studies were undertaken with patients.

In summary, we can assert that offline cross-validation results predict bit rate improvements due to ErrP classification in healthy subjects, yet two recent online studies have not yet delivered on that promise, as the ErrP classification accuracy did not exceed the P3 classification accuracy. Another weak point of the latest studies is the low N_h . Also, the most important step — application to the patients — has yet to be taken.

The dissertation at hand successfully addresses these shortcomings.

[2] The objective of most error correction systems based on error-related potentials is to maximise specificity, thereby minimising false positives. See page 21 for an explanation.

Table 3.3.: Overview of studies employing error correction systems (ECSs) based on error-related potentials (ErrPs). The column N_h denotes the number of healthy subjects that participated. Type BP is given for Button Press studies. Mode of the studies is given as online (on) or offline (off).

Author	N_h	Type	Latency	Mode	b_S	Δb_S
Schalk et al. [110]	4	µ-rhythm	180	off	0.56 ± 0.13	0.03 ± 0.036
Blankertz et al. [16, 17]	8	BP	220–260	off		yes
Parra et al. [99]	7	BP	$<$100	off		yes
Buttfield et al. [21]	3	BP	270 / 350–450	off	0.62 ± 0.06	0.02 ± 0.06
Ferrez and del R Millán [40]	5	BP	250/320/450	off	0.63 ± 0.06	0.03 ± 0.06
Visconti et al. [132]	5	P3	N300 / P400	off	3.43 ± 0.33	0.08 ± 0.1
Dal Seno et al. [30]	3	P3	N300 / P400	on	1.73	none
Kreilinger et al. [66]	3	µ-rhythm		on	0.48 ± 0.05	0.08 ± 0.01

4. Connectivity Methods

We begin the description of methods used and expanded upon in this dissertation with an explanation of the sections to come. The reader might be asking why the concept of *connectivity* is taken to encompass two very different studies, namely the multiclass study (see Section 4.2) and the cognition detection study (see Section 4.3). Let us first recapitulate our understanding of connectivity as defined in Section 2.4.2: In the widest sense, the exchange of information between different areas within the brain can be measured as connectivity betweem two or more sensors located outside of the brain.

The thesis states that connectivity (i.e., bivariate or multivariate features) can contain interesting information to increase bit rate of BCIs. This is the reason for investigating connectivity in the upcoming two sections, even though the primary objectives of the below mentioned studies were different.

In other words, each study was able to serve two purposes, and connectivity is seen as the common element in both studies — reason enough to group them in this chapter. The description of the experiments in the ensuing sections will therefore encompass methods applied and expanded upon pertaining to connectivity, the multiclass concept and cognition detection.

Five major datasets will be discussed during the course of this work, whereby each participant will be referenced with a letter subscripted by the dataset number, i.e., Ha_1. Healthy participants shall receive the prefix *H*, whereas patients with ALS will be prefixed *P*.

4.1. Patients

The recordings performed with two late-stage ALS patients who received chronic ECoG implantations are summarised in Table 4.1. These two patients deserve particular attention due to the unique situation that they both transitioned into CLIS during the observation period. What follows is a description of the ECoG recordings and signal quality. One ECoG grid was implanted subdurally in 2004 (patient Pa_3, aged 44, diagnosis 1999, female) and the second patient received an implant epidurally in late 2007 (patient Pb_3, aged 39, diagnosis 1997, male). Both implants were positioned over the left motor cortex. Both patients had an ALSFRS-R score of 0 at the time of implantation. A critical appraisal of the merits of chronic ECoG measurements follows in Section 7.1.

It was found that the number of electrodes that were delivering useful signals declined over time, a phenomenon not experienced on the same scale with the standard procedure of two-week implants for epilepsy patients. This required flexible algorithms which adjusted to the currently available set of channels. An impression of the channel loss can be obtained by a look at Figure 4.1.

4.2. Multiclass Studies

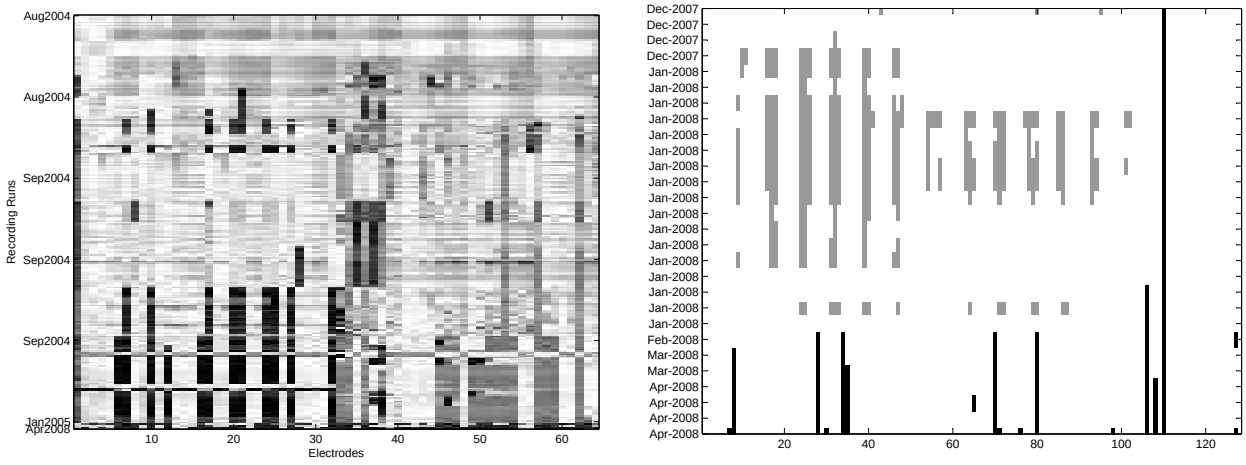
The general goal of increasing bit rate by increasing the number of tasks (classes) was motivated in Section 2.5.5. This chapter describes an offline analysis done on prerecorded[1] data to investigate the benefits of using connectivity features as compared to standard features (the prerecorded study's participants are denoted with a subscript of 1). Thereafter, the author's study is described in Chapters 4.2.2 to 4.2.4 (*multiclass* experiment, subjects subscripted with 2). An important point of the experimental setup in the multiclass experiment

[1] Datasets recorded by colleagues from the Tübingen BCI research group prior to the time of writing will be referred to as *prerecorded*.

Table 4.1.: Overview of the 9 relevant measurements with patient Pa_3 (top) and the 23 most relevant measurements pertaining to this dissertation out of a total of 210 measurements undertaken by the Tübingen group with patient Pb_3 (bottom).

Patient Pa_3								
Date	Method	Paradigm	Session	Runs				
Before 2004			Reached complete locked-in state (CLIS)					
04.06.2004	EEG	cogdetect	1	4				
18.06.2004	EEG	cogdetect	2	3				
23.06.2004	EEG	cogdetect	3	5				
04.08.2004				Implantation				
06.09.2004	ECoG	cogdetect	23	3				
07.09.2004	ECoG	cogdetect	24	1				
09.11.2004	ECoG	cogdetect	27	1				
15.02.2005	EEG	cogdetect	51	2				
06.12.2007	ECoG	cogdetect	57	7				
19.04.2008	ECoG	cogdetect	58	4				

Patient Pb_3								
Date	Method	Paradigm	Session	Runs	Heart Rate	Temp	Blood Pressure	Glucose
23.10.2007	EEG	best3	1	5				
24.10.2007	EEG	best3	2	8				
25.10.2007	EEG	best3	3	6				
01.11.2007	EEG	best3	5	7				
19.11.2007	EEG	best3	15	7				
12.12.2007	EEG	best3		1	114	36.6	125/70	
13.12.2007	EEG	best3		4	115	37.1	140/70	156
10.12.2007	MEG	best3	1	6	93-122		108/79–181/96	
21.12.2007				Implantation				
16.01.2008	ECoG	cogdetect	1	10	91	37	106/64	190
07.02.2008	ECoG	4class-screen	1	6	79	36.3	119/77	159
09.02.2008	ECoG	4class-screen	2	10	92	36.7	108/72	106 (5am)
11.02.2008	ECoG	4class-screen	3	10	85	36.8	126/83	
14.02.2008	ECoG	4class-screen	4	3				
16.02.2008	ECoG	cogdetect	2	9	87	36.2	106/63	120
02.03.2008	ECoG	4class-screen	6	12	102			
05.03.2008	ECoG	4class-screen	7	9	96			
17.03.2008			Reached complete locked-in state (CLIS)					
18.03.2008	ECoG	cogdetect	3	6	86			208
03.06.2008	ECoG	sing-vs-rest	1	10	104			
29.06.2008	ECoG	sing-vs-rest	2	10	105		89/56	
29.06.2008	ECoG	cogdetect	4	3	114		126/72	
29.06.2008	ECoG	instrument	1	1				



(a) Patient Pa_3. Development of standard deviation on channels. The ordinate represents one recording run per line. Dark lines represent channels containing too much electrical mains interference to obtain a useful signal.

(b) Patient Pb_3. Incidence of problematic channels. The ordinate represents one recording session per line. Grey lines represent high mains interference, black lines represent no signal being measured.

Figure 4.1.: Changes in ECoG electrode characteristics in the two chronically implanted ALS patients. Date is shown on the ordinate, electrodes are shown on the abscissa.

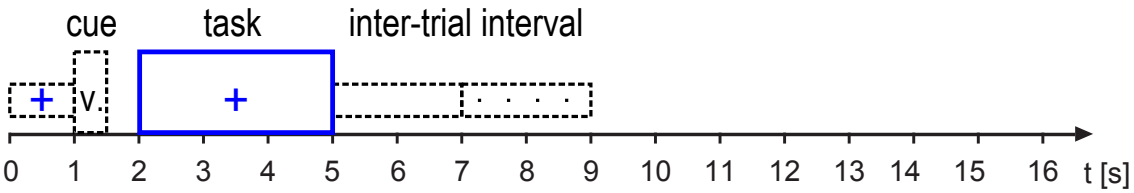
is that the subject's theoretical bit rate was estimated for all two, three and four-class task combinations available. This gives a choice of the best multiclass combination for each individual subject. The multiclass recordings were done with the MEG. The advantages of using the MEG were described in Section 2.2.3, the most prominent in the particular context of our multiclass study being the high spatial resolution of the MEG. The aim was firstly, to determine cortical areas that are modulated by nonmotor tasks such as navigation, mental subtraction, or visualising a scene and secondly, to achieve a good seperability of the signals due to the high spatial resolution. In contrast, the motor tasks accompanied by μ-rhythm desynchronisation are well studied and can be located on the motor cortex without great difficulty [70, 104, 142].

A caveat when using the MEG is that it is not portable and hard to use with late-stage ALS patients. Notwithstanding this disadvantage, the idea was to locate interesting tasks with healthy participants and then target the appropriate areas on the cortex when using EEG electrodes with patients. This can circumvent the need to place electrodes over the whole scalp and thereby greatly reduce preparation time.

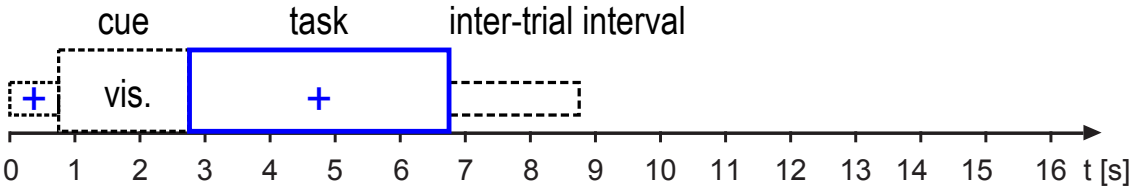
No multiclass recordings were undertaken with patient Pa_3. She had already been in CLIS for some years, so that the focus with her was primarily on cognition detection (Section 4.3). The multiclass recordings with patient Pb_3 were done with the MEG (Section 6.1.3.1) and ECoG (see Sections 4.2.5.2 and 4.2.5.3). To enable the reader a comparison of the timing of the recordings, which is important for bit rate calculations, the trial structure for each recording is shown in Figure 4.2.

4.2.1. Prerecorded Data

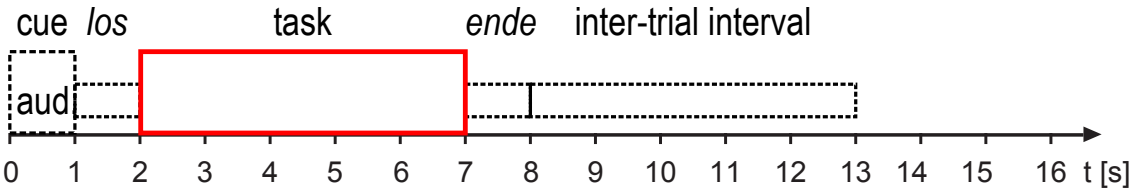
Prerecorded data from a previous BCI experiment using MEG data was analysed to aid in deciding on experimental parameters for the current multiclass experiment. The data had been labelled with two tasks — movement imagery of the little finger and imagined tongue movement. Further details of the experimental setup are given in [74]. Connectivity features were extracted (see Section 2.4.2) and classified according to the nested cross-validation (CV) procedure described by Lal et al. [74]. The nested CV is used throughout this work to estimate accuracy for various types of features, and is therefore explained in more detail. When



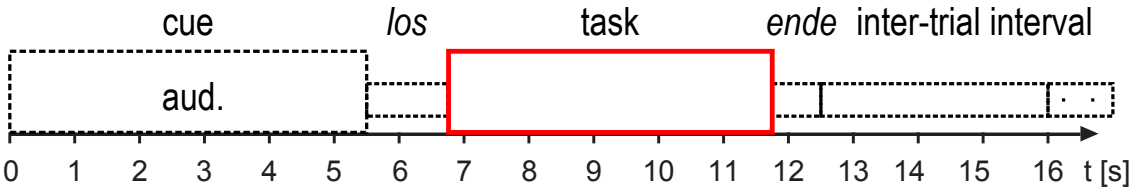
(a) Prerecorded study with imagined finger/tongue movement, two classes, subjects Ha_1– Hj_1, visual cues.



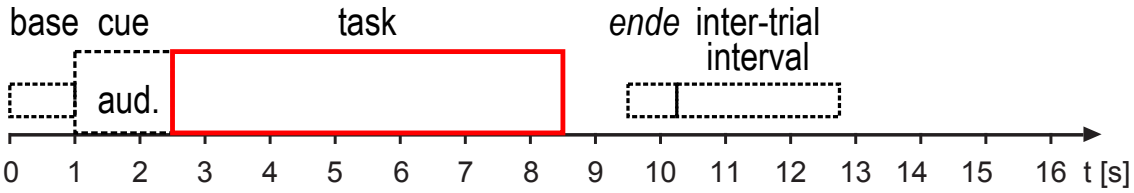
(b) Multiclass study with healthy subjects Ha_2– Hj_2, visual cues.



(c) Three-class MEG session, patient Pb_3, *best3* study, auditory cues.

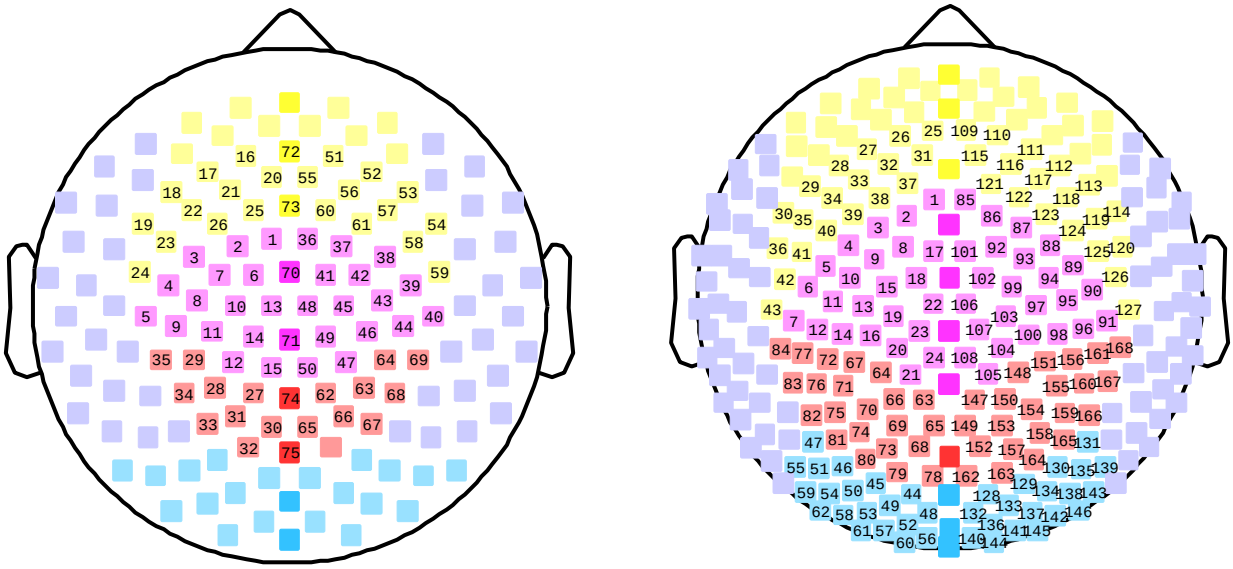


(d) Four-class ECoG screening sessions, patient Pb_3, auditory cues.



(e) Two-class ECoG sessions, patient Pb_3, *sing-vs-rest* study, auditory cues.

Figure 4.2.: Overview of timing for the multiclass recordings described in this section with healthy subjects and patient Pb_3. The classification intervals for healthy subjects and patients are represented by the blue and red rectangles, respectively.



(a) Prerecorded data, 75 channels analysed.

(b) Multiclass study, 168 channels analysed.

Figure 4.3.: Positions of the MEG sensors. The analysed channels are numbered. Colour-coding represents functional areas of the brain: frontal lobe (yellow), central (pink), parietal (red), occipital (turquoise), temporal (grey).

optimising a hyperparameter of the SVM in conjunction with feature selection, there is a danger of overfitting the selected features to each fold of data. Therefore, a separate feature selection is performed in a nested CV within each outer fold, which should give a more realistic estimate of online performance. This procedure allows a direct comparison of the AR coefficients used in [74] and the connectivity features used here (Section 6.1.1).

The prerecorded MEG signals stem from ten healthy subjects (Ha_1– Hj_1) and were recorded at 625 Hz sampling frequency with 150 channels. Cue duration was 500 ms. The classification interval began after a further 500 ms and lasted for 3 s, followed by an inter-trial interval of random length (see Figure 4.2a).

The extracted connectivity features were phase-locking value (PLV) and coherence (Coh). The outermost sensors on the perimiter of the MEG helmet can be prone to artefacts caused by electromyographic activity in the shoulder or neck muscles. These channels are also unlikely to pick up much discriminative information about the tasks, especially for the motor tasks. When it became evident that a reduction in the number of channels would significantly speed up the calculations, it was decided to exclusively use the central 75 MEG channels (resulting in 2775 channel combinations). These channels are shown in Figure 4.3a. Three frequency bands were investigated: 8–16 Hz, 16–24 Hz (known to be linked to motor tasks by the μ-rhythm) and the 8–40 Hz band, which encompasses all frequencies known to be relevant to movement imagery. These bands were extracted with a linear phase finite impulse response (FIR) bandpass filter. Additionally, three time windows were extracted from each trial's classification interval, which results in a total of $3 * 75 * (75 - 1)/2 = 8325$ features per frequency band. It was evident from an inspection of the AUC that the 8-40 Hz frequency band works best for the majority of subjects. The results of Section 6.1.1 are consequently based on that frequency band. The most discriminative time window of the classification interval varied across subjects, so that all time windows entered the analysis.

The results with the prerecorded data were encouraging as concerns the combination of bivariate and univariate features (details are divulged in Section 6.1.1). However, a new study had to be designed to find

Table 4.2.: Categorisation of the mental imagery tasks.

Category	Class	Task	Instruction
Motor	1	Foot	Rotate both feet
Motor	2	Left hand	Open and close hand in repetitive motions
Motor	3	Right hand	Open and close hand in repetitive motions
Motor	4	Tongue	Lick ice-cream with tongue in repetitive motions
Nonmotor	5	Subtraction	Start at 99 and repeatedly subtract 7 from answer. Do not visualise numbers, do the calculation each time. Start at a new number once the sequence has been memorised too well.
Nonmotor	6	Navigation	Walk around a well-known location (e.g., the house you grew up in). Recognise objects in the rooms.
Nonmotor	7	Visual scene	Imagine a green scene, e.g., a lawn or green landscape.

out whether a multiclass BCI (potentially employing bivariate features) would be more efficient than a binary BCI. The new study was to investigate connectivity in motor and nonmotor tasks. Additionally, the feasibility of using connectivity features online needed to be evaluated, as the high number of features require more processing time. The upcoming sections describe the multiclass study.

4.2.2. Experimental Setup

Data was recorded with BCI2000 in a 275-channel whole-head MEG (VSM MedTech Ltd.) at a sampling rate of 586 Hz from 10 healthy individuals aged between 24 and 34 years, on two different days for each subject. Each of the two sessions included three runs of data acquisition, with seven mental imagery tasks being presented in block-randomised order. Each of the three runs per session lasted 17 minutes — enough to make most subjects feel tired after every run. A break of up to five minutes was granted between runs. Cues were given in textual form (column three in Table 4.2) on a screen positioned in front of the subject seated in the MEG chamber. Subjects were instructed to look at a fixation cross in the centre of the screen during the cue and imagination phases to prevent eye movement artefacts. The trial structure was as follows: display of fixation cross ("get ready", 0.7 s) - text cue overlayed on fixation cross ("prepare for task", 2 s) – fixation cross ("mental imagery", 4 s) – blank screen (ITI, 2 s), (see Figure 4.2b). Section 2.1 includes some thoughts on the advantages and disadvantages of presenting online feedback. To prevent head movements caused by the feedback and due to the exploratory nature of this study, the decision was made to omit visual feedback. The focus was not to encourage learning by feedback, but to compare the subjects' performance with respect to the number of tasks that were used. Also, it was unclear how to give feedback to such tasks as subtraction, navigation, or visual scene imagination.

Seven mental imagery tasks were chosen as a trade-off between testing a maximal number of tasks per participant and being able to collect enough data for each task to train the classifier with sufficient robustness. Seven tasks allowed the collection of 102 trials per task and kept the actual recording time per session below two hours. A description of the imagination tasks is given in Table 4.2. As suggested in [29], functional motor tasks (some related to a subject's specific skills) were endorsed to make the task more interesting. For example, subject Hj_2 enjoys drumming, and chose drumming motions for tasks 2 and 3.

Some subject-specific details are given in Table 4.3. The only female was subject Hf_2. The high number of first-time (so-called "naïve") BCI users is a strong point of the study. Naïve users are interesting, because the competing factor of user experience plays no role in the evaluation of results. Handedness was surveyed

Table 4.3.: Healthy subjects participating in the MEG multiclass study. The second column indicates the number of times the subject had participated in previous BCI experiments. The *Hardest task* column lists the task that the subject reported as the most difficult to perform. The *Mood* column represents the subject's mood on a scale from 0 (bad mood) to 1 (good mood).

Subject	Experience	EMG recorded	Hardest task	Handedness	Mood S01 / S02
Ha_2	0	Session 1	5	right	0.75 / 0.68
Hb_2	0	Session 2	-	right	0.85 / 0.43
Hc_2	0	none	-	right	0.85 / 0.75
Hd_2	4	Session 1	7	right	0.80 / 0.80
He_2	0	Session 2	6	right	0.70 / 0.60
Hf_2	0	Session 1	7	right	0.68 / 0.78
Hg_2	0	Session 2	7	right	0.73 / 0.68
Hh_2	0	none	-	right	0.75 / 0.80
Hi_2	>5	Session 1	-	right	0.65 / 0.68
Hj_2	0	Session 2	5,6	left	0.83 / 0.78

with the Edinburgh Handedness Inventory [96]. Handedness can explain hemisperic differences in location of the tasks. Mood data was collected with a subscale of the Scales for the Assessment of Quality of Life [6] which comprises a total of 10 items to be answered on a five point scale to be correlated with performance. The answers were normalised to values in the range [0, 1].

The subject's head position was measured before each run. To recreate the previous head position, and thus guarantee a consistent measurement, subjects were asked to reposition their head according to a head position display (CTF Systems). EMG was measured in eight subjects to control for overt movement artefacts during the motor tasks. In the sessions shown in Table 4.3, one pair of EMG electrodes was placed on each forearm on either the lateral or medial antibrachial muscle. Due to the high preparation time, EMG was recorded in one of the two recording sessions only.

Overt movements would disqualify the results as inapplicable to patients who have to modulate their brain signal by pure mental imagery. This underlines the importance of artefact checks. The methods that were used to check for artefacts are discussed in the following section.

4.2.3. Artefact Rejection

Two types of artefacts were considered in this study. Firstly, there are eye movement artefacts and those caused by tension of neck or shoulder muscles. To minimise their effect on the recordings, the outer MEG channels close to eyes and neck muscles were precluded and a subset of 168 inner channels was analysed (shown in Figure 4.3b). Subjects were also instructed to place the back of their head against the MEG helmet, thereby increasing the distance between the frontal sensors and the eyes. Secondly, unintentional overt movements during motor task imagery possibly contribute to the fact that healthy subjects mostly achieve higher bit rates than patients when using BCIs. Trials containing overt movements were registered by the EMG electrodes and a comparison of classification performance with and without the artefactual trials was made.

The EMG time series was high-pass filtered at 0.5 Hz (FFT, removal of unwanted frequency bands, followed by inverse FFT) to remove low-amplitude drifts caused by sweating, strain on electrode cables, etc. To find trials containing EMG artefacts, we used a threshold-based algorithm on the EMG time series. Because the noise level of the EMG signal varied with time (due to strain on electrode cables and other

Table 4.4.: Number of artefacts, listed separately for runs 1–3 of the EMG session (columns 9–11) and for each task. F, LH, RH, T are the motor tasks (F)oot, (L)eft hand, (R)ight hand, (T)ongue. S, N, V are the nonmotor tasks (S)ubtraction, (N)avigation, (V)isual imagery. Note that the tasks LH and RH, being the hand motor imagery tasks, are those that were expected to reveal most artefacts, as the EMG was measured on the arms.

	Motor Task			Nonmotor Task				Total			
Subject	F	LH	RH	T	S	N	V	R01	R02	R03	All
Ha_2	8	4	8	7	8	8	9	20	11	21	52
Hb_2	3	4	2	4	2	4	5	7	14	3	24
Hd_2	0	16	2	0	2	1	0	7	9	5	21
He_2	0	2	2	3	3	3	2	1	6	8	15
Hf_2	1	0	0	1	0	3	3	1	1	6	8
Hg_2	1	0	1	3	5	6	0	4	7	5	16
Hi_2	7	4	5	2	7	2	5	7	16	9	32
Hj_2	7	2	4	1	5	5	8	13	12	7	32
Sum	27	32	24	21	32	32	32	60	76	64	200

effects), we used a sliding window to determine the noise level for each trial separately and standardised the trial accordingly. The window size was 250 samples (0.43 s) and the step size was 50 samples. For each window, the average of the 10 highest peaks was computed. The window with the lowest value was assumed to be free of artefacts and therefore used as noise level for the scaling of the whole trial. This is a reasonable method if one assumes that artefacts were produced unintentionally, and were thus unlikely to last over the whole trial. If the maximum peak pk in the standardised signal of a particular trial was further than d standard deviations from the mean, i.e., $pk > |\mu - \sigma d|$, the trial was labelled as an artefact. After a visual inspection of some artefacts, d was set to 3 for Hd_2, Hf_2, Hg_2, Hj_2 and to 6 for the remaining four subjects.

The number of contaminated trials found by this method is listed in Table 4.4. Refer to Table 4.3 to see in which sessions the EMG was measured. No significant difference of means in the distributions of the artefacts across the tasks (analysis of variance (ANOVA), $p = 0.96$) or runs (ANOVA, $p = 0.76$) was found, indicating that the artefacts will in general not influence the classification results . There is one exception: subject Hd_2 activated his brachial muscle extremely often during *left hand* imagery.

4.2.4. Feature Extraction and Classification

As mentioned above, the number of channels was first reduced to 168. Even though the prerecorded data contains less channels, the locations of signal modulation in conjunction with the tasks remains comparable between the studies. Then, multiple cross-validation runs were done with the following features which were extracted from the data: autoregressive (AR) coefficients exclusively, power spectral density (PSD) features exclusively, phase-locking value (PLV) features exclusively, AR combined with PSD, AR combined with PLV, PSD combined with PLV.

A model order of two was chosen for the AR coefficients because firstly, model order three was tested and delivered poorer results; and secondly, model order two was found to be the best option for the prerecorded data [74], which is similar to the current data.

A nested CV with a feature selection in the inner loop was performed. The outer 5 folds consisted of a random split of the data into 80 % training set and 20 % testing set. The inner 10 folds entailed a recursive

feature elimination (RFE), similar to the method described in [74]. The Matlab® toolbox *Spider* [135] offers an implementation of the one-versus-the-rest SVM with ridge regularisation, which was used to obtain the three-class classification rates.

The feature selection algorithm treated all types of features equally where multiple feature types were used. The total number before feature selection was 336 AR coefficients and 8325 PLV features. The mean number of features selected was 7.7 ± 5.4 with binary classification and 44 ± 27 with ternary classification.

The best offline results were obtained by using solely the AR features and will be presented in Section 6.1.2.

4.2.5. Patient Recording Sessions

After the offline analysis described in the preceding chapters was completed, the opportunity arose to work with a patient in the final stage of ALS. The patient will be referred to as Pb_3— (see Section 4.1) for details of the patient's aetiology. Patient Pb_3 had never had particularly high BCI scores and standard motor-task paradigms were not bearing fruit. The decision was made to perform a set of multiclass screening recordings of nonmotor tasks combined with motor tasks. The goal was to reproduce the encouraging results (see Section 6.1.2) of the aforementioned multiclass motor/nonmotor recordings in a CLIS patient.

The opportunity to measure Pb_3's brain activity by MEG was unique. Although ALS patients in early stages of the disease have been studied in the MEG before, no report exists of a CLIS patient having been measured in the MEG. With this in mind, the description of the experiment in Section 4.2.5.1 mentions details which could be of interest to other researchers planning a similar measurement. The primary goal of the measurement was to locate brain areas that display a task-dependent modulation of electric/magnetic activity. The BCI team would use this and other information to decide on the location of the ECoG electrodes to be implanted for improved communication with the patient.

The multiclass screening is reported in Section 4.2.5.2 and was undertaken in a series of sessions lasting from 7 February 2008 until 20 March 2008. The singing sessions described in Section 4.2.5.3 were undertaken on 3 and 29 June 2008.

4.2.5.1. Magnetoencephalogram Session

In addition to the primary goal just mentioned, a secondary goal was to find out whether the use of a high-resolution system such as the MEG would improve Pb_3's classification rate compared to his EEG recordings. An improvement of classification accuracy with a high-resolution recording would strengthen the case for an implantation of ECoG electrodes. Conducting a MEG recording with an artificially ventilated, non-communicative patient was a daunting enterprise. The main issues to solve were:

- Obtaining informed consent
- Transportation of the patient from the hospital ward to the MEG centre, into the MEG chamber and back
- Artificial ventilation and monitoring of the patient's vital signs during the measurement
- Design of a paradigm which would test a large amount of communication tasks, yet generate enough data per task in the short time span available

The guidelines concerning ethical procedures with LIS patients postulated in [53] were adhered to. It is unclear to what extent the general informed consent of Pb_3's legal representative included "special" recordings such as the MEG recording. However, patient Pb_3 was informed about the aim and the planned course of the recording on multiple days when communication was possible by eye movements. After the explanations, the

Table 4.5.: Patient Pb_3. Overview of data recorded during the MEG recording session on 10 December 2007.

Time	Run	Paradigm	Trials	Heart Rate	Blood Pressure
16:32	1	best3	30	131/69	119
16:45	2	audiospeller	30	120/75	121
17:05	3	best3	30	148/64	117
17:16	4	best3	30	-	-
17:23	5	audiospeller	30	145/75	122
17:45	6	best3	30	-	-
17:55	7	best3	30	-	-
18:05	8	free speller	30	129/71	122

patient was asked whether he would like to proceed with the measurement and gave informed assent by eye movements. The answers were verified by repetition of the questions.

The patient was transported to the MEG centre by Red Cross transport (mandatory for insurance reasons), disconnected from the portable ventilator, carried into the MEG chamber with a carrying cloth, and connected to a ventilator positioned outside of the chamber by 5 m air pipes. The head anaesthesiologist of the university clinic supervised and monitored the patient from the time of his arrival. A second person stayed in the chamber throughout the measurement to repeat any instructions given by loudspeaker directly to the patient and to monitor the patient's well-being during breaks between runs. Unfortunately, the patient was unable to control his eye movements on the day of the measurement.

The patient was carried out of the chamber after 1 hour and 53 minutes. By this time, Pb_3 was sweating profusely on the head. Experience from the multiclass MEG study with healthy subjects showed that concentration drops rapidly after 60–90 minutes in the chamber. Among the reasons for this are: the dimly lit chamber, the silent atmosphere due to environmental noise being blocked by the chamber, and heating up of the head due to the insulation in the MEG helmet. This means that a patient would probably start losing concentration even before 60 minutes have passed, causing the quality of the recording to drop. This should be taken into consideration when working with the data.

As is good practice, the recording paradigms were tested with a healthy subject some days prior to the patient recording and found to be operational. The data from Pb_3's recording session is shown in Table 4.5. The trial structure for the *best3* runs is given in Figure 4.2c on page 32.

The *audiospeller* and *free speller* runs will not be discussed in any more detail, as they belong to the class of ERP recordings. The analysis of ERP recordings for direct communication by means of an *auditory speller* interface are not part of this dissertation. Considering the *best3* runs, we can gather from Table 4.5 that a mere 50 trials per class were recorded in the available timespan. Even with healthy subjects, the goal normally is to obtain 100 training vectors per class for classifier training.

A preview of class discriminability was achieved by processing the three-class data offline as follows: A classification interval (CI) and an inter-trial interval (ITI) following the CI were extracted from the BCI2000 data. The linear trend was removed from each interval. The data was bandpass-filtered from 1–40 Hz (by FFT, multiplication by a Hanning window centred over the 1–40 Hz range, followed by an inverse FFT). Thereafter, the data was downsampled to 80 Hz from the original sampling rate of 586 Hz. The filtering and downsampling ease further processing, as MEG data is prone to memory problems due to the high number of sensors.

To gain an insight into the discriminability between the classes, a transformation into the frequency domain was performed with the Welch method and the PSD was collected over 20 frequency bins of logarithmically increasing width from 0–40 Hz. The logarithmic width increase compensates for the exponential falloff

typically encountered in frequency spectra. Visualisation of these values is achieved with a plot displaying channels on the ordinate and frequency bins on the abscissa. Topographical PSD plots have the disadvantage that only one frequency bin can be shown.

More effective discrimination can generally be achieved by extracting AR coefficients. Therefore, AR coefficients were additionally calculated using the Burg method (model order 2). AUC scores were obtained by grouping the data in a one-versus-the-rest fashion. In contrast to PSD, AR coefficients lend themselves to topographical display and each coefficient was projected onto a topographical view of the head. All these results can be found in Section 6.1.3.1 on page 59.

4.2.5.2. Four-class Screening Sessions

After implantation of the ECoG grid electrodes, including two electrode strips, the Tübingen BCI group attempted to re-establish communication with Pb_3 by using ECoG paradigms. Figure 4.4b displays the electrodes used in the analysis. This section reports on the recording and analysis of a set of seven multiclass screening sessions undertaken within two months (refer to Table 4.1). The tasks were labelled as follows: (1) Left hand imagery (motor task); (2) Right hand imagery (motor task); (3) Aim-and-shoot (nonmotor task); (4) Singing (nonmotor task).

The *aim-and-shoot* task was chosen as this was a task the patient was accustomed to because of his profession. It was expected that aiming should lead to high theta activity [54] and so be easily discriminable from the other tasks.

Stimulus presentation and recording was accomplished with PyBCI [118], a Python wrapper to BCI2000 which offers greater flexibility in the experimental design than BCI2000 itself. The stimulus presentation script was a modified version of an existing script and resulted in the trial structure shown in Figure 4.2d.

Data preprocessing was performed as follows: All trials of a run were read into memory and seconds 7–10 of the CI with seconds 0–3 of its preceding ITI were extracted to allow for the comparison between tasks and baseline signal. A surface Laplacian spatial filter was applied to focus signal modulation on each electrode. The data was converted to single precision to further conserve memory. The Welch method was applied to each signal segment to calculate PSD. This was repeated for each run of each session. Four ROC plots were generated (page 60), one comparing each task to the baseline.

The rationale behind the choice of the abovementioned tasks was to obtain easily differentiable brain states between the motor and nonmotor tasks. Again, the first step was to visualise the AUC score for PSD. PSD was chosen here to find frequency bands which were being modulated by the patient, which is not possible when viewing AR coefficients directly. The next step was to build a two-class BCI with the two most discriminable tasks. Once the patient would use that BCI successfully, this would be extended to three classes.

4.2.5.3. Singing Sessions

In the analysis of the screening data mentioned in the previous section, the *sing* task seemed very promising for patient Pb_3. The author thus designed a *sing-vs-rest* paradigm with PyBCI (Figure 4.2e displays the trial structure) and managed to record two sessions of ten runs each before Pb_3 had to leave Tübingen. The patient was instructed to actually sing or shout at the top of his voice (once again, a task he was probably used to because of his profession). The intention was that this would lead to stronger signals than mere imagined singing. For the *rest* task, Pb_3 was instructed to relax.

The signal was detrended and referenced to CAR. Thereafter, it was bandpassed from 1–40 Hz and downsampled to 80 Hz as explained in Section 4.2.5.1. AR coefficients (order 5) were calculated and submitted to the CV algorithm introduced in Section 2.5.4.

4.3. Cognition Detection Study With Two Patients

The two patients Pa_3 and Pb_3 with chronic ECoG implants for communication participated in this study. Further information on the patients can be found in Section 4.1 and at the beginning of this chapter.

The main objective of the psychophysiological cognition tests was to investigate changes in cognitive functions at regular intervals during the transition from LIS to CLIS, by measuring evoked cortical responses to auditory stimuli and changes in the ECoG signal spectrum. A secondary objective was to evaluate the patients' cognition and correlate this with BCI performance to see whether there is a connection. Care had to be taken to ensure that the patients were not asleep during the test. This was checked by evaluating the spectrum of the signal. A dominant frequency below 5 Hz was declared as signifying drowsiness or sleep (see [123]).

The hope was that many of the problems described with EEG recordings in LIS/CLIS patients could be circumvented by recording the ECoG. The study is unique due to the availability of evoked response potential (ERP) recordings from early stage LIS and the transition to the CLIS.

4.3.1. Prerecorded Data

Spectral plots of data recorded before the cognition detection tests started served as a baseline measure to determine the sleep phases of the patient. Also, the averaged P3 responses to (visual and) auditory oddballs, which were recorded in similar experiments before the cognition detection tests started, were examined. This allowed a comparison of latency, duration and amplitude of N1/P2, N2a and P3 components between the LIS and CLIS phase. Specifically, increased latency (or decreased amplitude) of N2a and P3 in the late stages of ALS would lead us to believe in decreased cognitive ability. The indication would be even stronger if the N1/P2 complex were to be affected in the same way.

Patient Pa_3 showed some control of a BCI before the LIS stage. Also, she was able to signal her informed assent to receive the ECoG implant with a salivary pH-test [136]. Patient Pb_3 also showed some control of a P3-based BCI before becoming locked-in [67] and gave his informed assent to the ECoG implant with eye movements [87]. BCI performance dropped to random after several weeks of LIS without reliable eye or external sphincter control, suggesting a possible decline of cognitive functioning. Some P3 oddball ERP tests were conducted with both patients before CLIS. In CLIS, two cognition detection tests were carried out with Pa_3 and four with Pb_3 at monthly intervals, except for the last test which was done after a break of three months (due to logistic and medical complications). The motivation for focussing on these tests rather than the BCI experiments in this section is that no information could be gleaned from the BCI sessions as to the reason for the drop in performance in CLIS.

The prerecorded data was analysed as follows: Session 1–3 of Pa_3's EEG oddball data were read from disc, 200 ms before up to 800 ms after the stimulus tone were extracted. The pre-stimulus interval of 200 ms was defined as the baseline and the 800 ms thereafter corrected to obtain a zero mean baseline. A visual examination revealed that some trials still contained nonstationarities (signal drift), so each trial was additionally detrended. Another irregularity was found during the visual examination. The correlation between channels Cz and hEOG was unusually high, indicating a problematic recording. The oddball data of session 2 was therefore excluded from further analysis. The standards (516 trials) and deviants (84 trials) were averaged separately and plotted (page 67); no reduction in the number of standard trials was undertaken. A second visual of discriminability between the two conditions was obtained by plotting the AUC values in a channels-by-samples fashion. As this data merely contained six EEG channels, the topographic plot was not considered meaningful enough to show. The tests requiring semantic processing were evaluated with an additional step. After removal of linear trends on a per-trial basis, the EEG channels were corrected for eye movements [115].

Table 4.6.: Overview of ECoG cognition detection sessions. Heart rate and blood pressure are shown for beginning and end of each session. Letters in brackets represent the instruction to (c)ount deviants or (i)gnore all tones. Numbers in brackets are number of deviant stimuli collected in a particular session.

Patient Pa_3			
Session	S51	S57	S58
Date	15.2.2005	6.12.2007	19.4.2008
Time	14:00–16:00	22:10–23:20	11:00–11:50
Total Nr. of Runs	2	7	4
MMN Runs	0	4(c)	4
Standard Oddball Runs	1	1(i),6(c)	1–3
Priming Runs	2	3,5	0
Semantic Oddball Runs	0	2	0
Amplifier	BrainAmpMR	BrainAmpMR	BrainAmpMR
Nr. of Channels	64	64	64
Sampling Rate [Hz]	1000	200	200
Reference Electrode		G01	G01
Ground Electrode		forehead	forehead
Heart Rate [bpm]		83–73	94/80
Last Communication	6 months ago	3 years ago	3.5 years ago

Patient Pb_3							
Session	S01		S02		S03		S04	
Date	16.1.2008		16.2.2008		18.3.2008		29.6.2008	
Time	16:35–19:20		10:30–13:45		15:50–17:00		17:30–18:00	
Total Nr. of Runs	10		9		6		3	
MMN Runs	1+6	(140)	9(i)	(70)	6(i)	(70)	2(i)	(70)
Oddball Runs	8(c)	(45)	1(c),5(c)	(90)	1,4(i) 2,5(c)	(180)	1,3(c)	(90)
Priming Runs	4+9		3+7		0		0	
Semantic Oddball Runs	2+7	(120)	2+6	(120)	3	(60)	4	
Amplifier	QuickAmp		BrainAmpMR		BrainAmpMR		BrainAmpMR	
Nr. of Channels	121		64		64		65	
Sampling Rate [Hz]	500		500		500		500	
Reference Electrode	5		38		43		43	
Ground Electrode	48		54		54		53	
Blood pressure [mmHG]	106/64–100/73		106/63–110/71		n.a.		126/72–120/60	
Heart Rate [bpm]	91–85		87–84		86–86		111–105	
Glucose [mg/ml]	188		120 (at 7am)		208		n.a.	
Lying on...	back, fairly upright		back		back/left		back	
Last Communication	1 day ago		32 days ago		2 days ago		105 days ago	

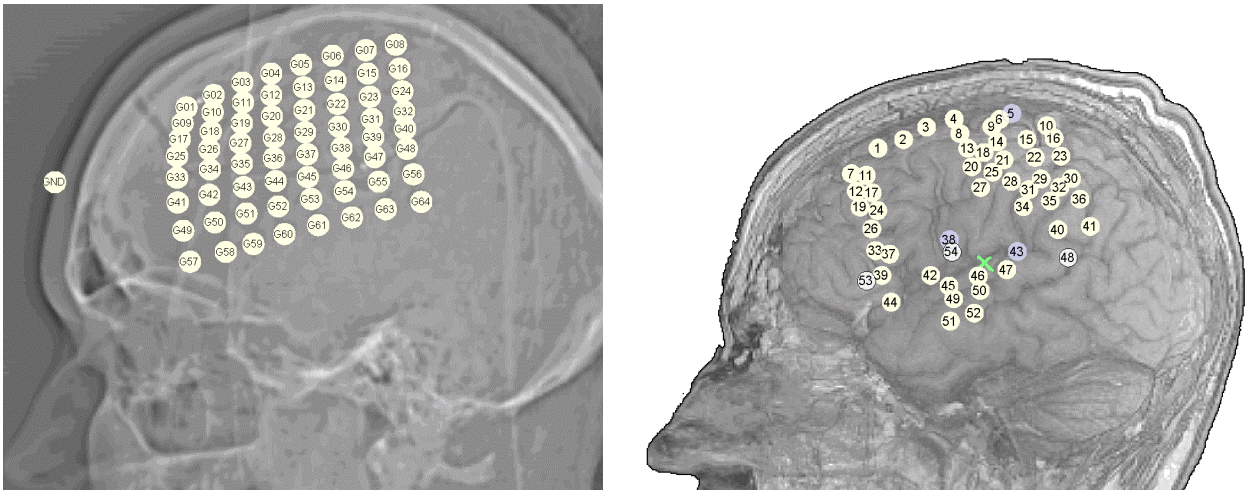
4.3.2. Experimental Setup

For communication purposes, an ECoG grid of 64 (Pa_3) and accordingly 128 (Pb_3) platinum electrodes had been implanted epidurally over the left fronto-central cortex (see Figure 4.4). An electrode strip was placed at the ventral edge of Pb_3's grid. Details of the recording sessions can be found in Table 4.6.

The ECoG was amplified before being written to disc together with trigger signals generated by the

Table 4.7.: Overview of the ECoG recording paradigms with the expected ERP component and the expected cortical location. The ERP component's latencies were determined visually. The inter-stimulus interval (ISI) is the time between two beeps or word endings. The number of standard and deviant tones presented per run, as well as their length, is given in the last column.

Description	Component	Expected Location	ISI	Standards / Deviants
Mismatch Negativity	N2a 0.10–0.50 s	sup. temp. gyrus	0.4 s	630 / 70 70 ms / 30 ms
Oddball Tones Pa_3	P3b 0.25–0.75 s	inf. parietal lobule	0.85 s	255 / 45 100 ms / 100 ms
Oddball Tones Pb_3	P3b 0.40–0.90 s	inf. parietal lobule	0.85 s	255 / 45 100 ms / 100 ms
Priming	N400 0.50–1.0 s	post. temp. cortex	1.2 s	120 word pairs
Semantic Oddball	P600 0.50–1.0 s	parietal	1.0 s	5 times 60 stimuli



(a) Patient Pa_3, X-ray image. An EEG electrode on the forehead was used as the ground. Electrode numbers are prefixed with *G* to prevent confusion with the second patient.

(b) Patient Pb_3, MR image. The electrodes in the centre of the grid were not used because of intermittent signal failure. The green cross marks the expected location of the primary auditory cortex in the posterior superior temporal gyrus.

Figure 4.4.: Positions of the ECoG electrodes used in the analysis. Shaded electrodes were used as reference, circled electrodes were connected to ground as indicated in Table 4.6.

software STIM, a separate DOS-based auditory stimulus presentation system by Neuroscan Inc. The auditory stimuli were presented via canalphones to block environmental noise and set to a sound pressure level of 70 dB above the mean auditory threshold of a healthy control group. The number of deviant stimuli collected per run is given in Table 4.7.

The ERP component latencies in Table 4.7 reflect the time windows used in the analysis, with the exception that the P3 time window for Pa_3 was set to 0.25–0.75 s, because the visual analysis indicated an earlier P3 response for that patient.

4.3.2.1. Data Analysis

The data was analysed offline. Channels which were not recording physiological signals (channels containing 50 Hz line noise and flatlines) were removed. Thus, 13 channels remained for Pa_3 and 46 channels remained for Pb_3. The signal was downsampled to 100 Hz and bandpass filtered with a FIR filter (Hamming window,

order 500, passband 0.5–9 Hz). The cutoff was chosen at 9 Hz to remove the 10 Hz alpha oscillations which were quite strong in some of the sessions (thereby increasing the signal-to-noise ratio of any ERP). The bandpass filter order of 500 was needed to filter high-amplitude drifts in the signal on the order of $1/35$ Hz. Signal windows of 1 s length following each stimulus were evaluated. In all figures, the number of trials of standard/deviant tones was balanced by removing a random subset of standard tones. If this is not done, the mean time series for the standard tones appears smoother because of the higher number of standard tones. However, all data was used in the statistical tests. The average ERP for each channel was calculated relative to a baseline of 100 ms before the auditory stimulus. Filtering and subsequent averaging of all trials attenuated signal amplitude by a factor of about 3.

Long-term ECoG recordings in humans may lead to an increasing number of noisy channels, possibly due to detachment of the rigid ECoG grid from the dura. Because of this, the recording reference was changed from session to session to obtain a good signal on as many electrodes as possible. The analysis concentrated on a subset of all available channels and recordings were rereferenced to electrode 44 (because of its distance from the auditory cortex) or a CAR. Rereferencing can cause the polarity of the ERPs at some channels to switch. The CAR causes about half of the channels to have either polarity and thus does not reflect the true ERP polarity. Figure 4.4 shows the subset of analysed channels, plotted on an X-ray for Pa_3 and a MRI of Pb_3's brain, including positions of the reference and ground electrodes. No MRI is available for Pa_3, as her stay at the clinic was short.

4.3.2.2. Significance of Evoked Responses

Viewing averaged ERPs can be misleading, as no indication of the variance of the distributions is given. Statistical testing is thus vital. However, especially in patients, statistical power (p value) can be weak because not enough data is available or the task cannot be performed as well as expected. So it was decided to additionally visualise effect size of the ERPs to obtain a bigger picture. Both methods used to track cognitive deterioration over the period of six months are explained here.

To determine the statistical test to be used, a Lilliefors goodness-of-fit test at significance level 0.05 confirmed that 93 % of all samples in all channels stem from a normal distribution. This was considered sufficient for the use of a parametric test. Thus, Student's t-test was chosen to find out which electrodes' cortical responses to standard and deviant tones differed significantly, i.e., in which cortical regions cognitive abilities on a relatively low level were intact. The null-hypothesis was that the population means of standard and deviant tones were equal. The t-test was calculated on time windows defined according to a priori knowledge of the expected ERP components (see Table 4.7). This ensured that the N1/P2 component would not lead to a "false positive". The N1/P2 component is too early after the stimulus to signify a conscious cognitive effort to recognise a particular tone or word. Non-paired, two-tailed t-tests were calculated. The two-tailed test was used because it is unclear when using an ECoG signal compared to EEG whether the polarity of a component is the same on all electrodes. When using a CAR, for example, it was evident that about half of the channels showed inverted components. The test was repeated for each time point of the predefined window (Table 4.7) and performed on unbalanced data (all *standard* trials included), taking into consideration the unequal variances of the two populations. Due to the high autocorrelation of the ECoG signal (found to be in the range 0.93–0.97 for patient Pb_3), a component was only flagged as significant if the criterion $p < 0.05$ was attained in at least $M = 12$ consecutive time points. M was determined by cubic extrapolation from the values that Guthrie and Buchwald [47] suggest for signal autocorrelation of up to 0.9. Randomised t-tests were calculated to check the validity of the t-test under the condition $M = 12$. The Monte Carlo method used here repeats the t-test 100 times with trial labels randomised according to their original probability distribution and aggregates the result. The mean and standard deviation are displayed in square brackets in the tables with statistical tests.

Because of session-to-session variability in average signal strength, the effect size E was used as a continuous measure (Cohen's d, standardised to pooled variance)

$$E_{d-s} = \frac{1}{2r+1} \sum_{t=t_p-r}^{t_p+r} \left(\frac{X_{d,t} - X_{s,t}}{\sqrt{((N_s-1)V_{s,t} + (N_d-1)V_{d,t}) \,/\, (N_s+N_d)}} \right) \quad , \tag{4.1}$$

where $X_{c,t}$ and $V_{c,t}$ are the mean and variance over all trials of condition c at time point t. Condition $c \in \{s, d\}$ can be either standards (s) or deviants (d) for MMN or standard oddball. A 50 ms window centred at the time point of maximum absolute difference of means (t_p) between standards and deviants was chosen to calculate E. The number of samples r represents 25 ms in this case. N_c is the number of tones presented for condition c. To visualise the effect sizes topographically (results on page 66), a single time point has to be chosen at which to calculate d. It turned out that t_p varied considerably between electrodes within the defined interval (Table 4.7). This is a direct consequence of very weak or nonexistent ERPs, combined with high-amplitude low-frequency oscillations that mask the signal of interest. As signal properties varied from session to session, manual correction of t_p was considered the best choice. This was accomplished by visual inspection of t_p, t-values, and time series plots of standards and deviants for all channels. The manual correction was only used to align peaks on some electrodes that were too far apart to produce a meaningful distribution of E.

4.3.2.3. Latency of Evoked Response Potentials

To determine the latency and duration of the N1/P2 and P3 components, all trials in a session were averaged separately for each channel. Using the average deviant wave, the first local peak of the signal beginning 100 ms after stimulus was chosen. This and the following local peak were considered to be the N1/P2 component. The latency of the N2a or P3 component was defined as the time point of the highest local peak in the range 500–1000 ms post-stimulus. Note that a different time point (maximum difference of means) was used in the effect size calculation.

The delayed P3 response in Pb_3 led to the question whether the delay of the primary N1 response can be taken as a predictor of the P3 delay. Another goal was to determine whether a progression of the N1/P2 and P3 delays took place over the six-month study period, i.e., during disease progression. The prerecorded EEG data was downsampled to 25 Hz and standardised for that analysis, results of which are shown on page 68.

4.3.2.4. Spectral Analysis

A spectral decomposition is interesting for two reasons. From the technical point of view, the spectral analysis can help to elucidate possible artefacts to decrease both false negative and false positive findings. From the psychophysiological point of view, it can shed light on the relationship between cognition (as measured by the ERP tests) and PSD.

Spectral power of the ECoG signal was estimated in a three-minute interval of every run according to the Welch method after applying a FIR highpass filter (order 500, 0.1 Hz cut-off) and a common average reference. The frequency bins were 0.5 Hz wide, contained 20 PSD values each, and were aggregated into physiologically meaningful bands by averaging (see Section 2.4.1.2). Spectral peaks at multiples of 50 Hz were removed. Relative power was calculated by dividing each frequency band's result by the total spectral power in the range 0.5-70 Hz.

5. Error Correction Methods

Error-related potentials (ErrPs) were introduced in Section 2.6. Their positive effect on communication speed is evidenced by a number of publications, which are listed in Section 3.4. However, only a small subset of the current literature [30, 132] describes the use of ErrP in a real-world BCI environment with healthy subjects (P3 speller).

The current chapter for the first time exploits ErrPs to boost bit rate during the communication process in patients. Nonetheless, the study includes a cohort of 18 healthy subjects (as described in Section 5.3) to enable a comparison with the patients' results using the same methods. It is always difficult to compare with other publications because authors use different preprocessing and classification methods.

An important general rule to remember is that in all two-tier classification schemes where the second classifier C_{ErrP} depends upon the output of the first classifier C_{P3}, the overall performance can only be improved if the accuracy $acc_{ErrP} > acc_{P3}$. This highlights the inherent difficulty of error-related potential (ErrP) recognition in a P3 BCI context, as the ErrP is classified single-trial, whereas the P3 component which is classified is actually the average P3 response to multiple flash sequences.

The bit rate of the ECS is calculated by substituting T_L in (2.5) by

$$T_L = \frac{c}{p\, r_C + (1-p)\, r_E + p - 1} \quad , \tag{5.1}$$

where r_C is specificity, r_E is sensitivity and p is accuracy of the ErrP recognition (acc_{ErrP}).

A simpler formulation follows. The bit rate of the P3 speller with error correction system (ECS) can be defined with the help of the confusion matrix for a two-class problem (refer to Figure 2.4). The aim of the ECS is to convey less (incorrect) information compared to the P3 speller without ECS. This is the case if ErrPs are recognised (because the letter is deleted), i.e. in $Pos = (FP + TP)/(TN + FP + TP + FN)$ percent of the trials. Thus, the original bit rate (2.7) is multiplied by $(1 - Pos)$. In addition, for every TN a correct letter is transferred, and for FN an incorrect one, so the redefined accuracy is $p_{ECS} = TN/(TN + FN)$. Inserting these factors into the bit rate formula (2.7), we obtain

$$b_{ECS} = (1 - Pos) \cdot (log_2(M-1) \cdot (2p_{ECS} - 1)) \tag{5.2}$$

$$= \frac{TN + FN}{TN + FP + TP + FN} \cdot log_2(M-1) \cdot \frac{2TN - TN - FN}{TN + FN} \tag{5.3}$$

$$= log_2(M-1) \cdot \frac{TN - FN}{TN + FP + TP + FN} \quad . \tag{5.4}$$

A transformation of (5.1) proves the equality of (5.1) and (5.4).

The study presented in the current chapter additionally makes use of prerecorded EEG data of ALS patients, described in the next section, to develop the online ErrP classification method (Section 5.6).

5.1. Prerecorded Data

The prerecorded patients' data was recorded between 2004 and 2007 with the standard P3 speller paradigm without modifications. This group of patients shall be referred to as Group P1. At the time, nobody had the idea that the data might at some point be used to evaluate ErrPs; therefore no explicit instruction had been given to the patients to watch the screen intently until the letter that was being spelled had appeared. It is

Table 5.1.: Accuracy and demographic information of prerecorded datasets of patients for ErrP analysis. These patients shall be referred to as Group P1. Patients Pd_4, Pe_4 and Pf_4 used multiple electrode montages: the number "2" indicates the same montage as used for all other subjects (see Figure 5.3), suffix "3" indicates an eight-channel montage.

Subject	Train Trials	acc_{P3}	Age	Sex	Year	ALSFRS-R
Pa_4	515	0.86	70	m	2004–5	17
Pb_4	310	0.51	56	m	2004	23
Pc_4	396	0.50	42	f	2005	8
Pd_4	377	0.77	38	f	2004–5	14
$Pd2_4$	517	0.85	38	f	2006–7	14
$Pd3_4$	898	0.76	38	f	2007–8	14
Pe_4	304	0.67	52	m	2004–5	31
$Pe2_4$	366	0.72	52	m	2006–7	31
$Pe3_4$	292	0.68	52	m	2007	31
Pf_4	496	0.66	50	f	2004	8
$Pf2_4$	274	0.85	50	f	2006–7	8

however highly likely that patients were doing just that, given that they were interested to know whether their letter was correct or not.

Some aborted sessions and sessions containing signal artefacts were excluded from the analysis. The remaining datasets are shown in Table 5.1. These were used to determine parameters such as the type of classifier, evaluation time interval, type of signal filter, etc., for the later classification of online data. Including classifier hyperparameters, the parameter search space was too large to perform the search in one CV loop. It was thus decided to perform line searches on one parameter at a time whilst keeping the others constant.

5.2. Experimental Setup

The study consisted of two sessions on different days. Session 1 included a P3 training phase and an ErrP training phase. Session 2 contained an ErrP training phase and the ErrP testing phase. To obtain at least 50 training vectors in each class, it was deemed necessary to collect data over two sessions. We took into account that a user's concentration would drop after about two hours, especially in the case of patients. Also, less data can be collected from a patient in the same amount of time, as a patient normally requires a higher number of sequences to achieve the same accuracy as a healthy subject. The standard P3 speller paradigm was used as described in the fundamentals (see Section 2.1.3). In addition, one second after the flash sequence ended, an empty square was displayed in the centre of the screen to draw the user's attention to the imminent appearance of the classification result (see Figure 5.1). After a further second, the letter recognised by the P3 classifier was displayed (time t=0). The interval t = 100–800 ms served as the input to the ErrP classifier. On recognition of an ErrP, the interface would remove the letter assumed to be incorrect and ask the user to repeat the current letter. The experimental setup is shown in Figure 5.2. Distance from the eyes to the top of the display was 1 m. Before each session, an EOG calibration measurement of one minute was recorded. Before the ErrP Test phase began, subjects and patients were informed that some of the letters would be erased after the trial, and the user given a second chance to write that letter. They were told that either an incorrect or a correct letter could be erased, and that they were simply to follow the instructions. It is safe to assume that at least 50 % of the subjects were not aware that the deletion of letters was based on their own brain signal.

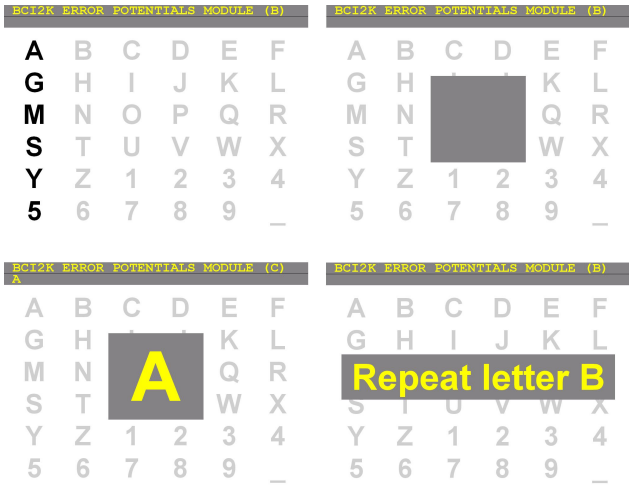


Figure 5.1.: Modified visual feedback for ErrP detection. Rows and columns flash as in the default BCI2000 configuration (top left). After P3 classification, a gray square draws the user's attention to the centre of the screen to prevent eye movements when the letter is shown (top right). The letter chosen by C_{P3} is displayed for one second, the classification interval is 100–800 ms after the letter appears (bottom left). If classified as a positive (error), the letter is removed from the text bar at the top of the screen and a message is shown to repeat that letter (bottom right). During free spelling, the letter to be repeated is not known and thus not shown.

All the ErrP studies mentioned in Section 3.4 use sham feedback to attain an error rate of 20 %. This study uses real feedback. This was considered more natural, as there was no risk that the user would find out about the sham feedback and consequently reduce his attention (or start experimenting with the system). Furthermore, we aimed to increase the validity of the online test with patients by foregoing sham feedback.

However, the number of sequences were adjusted individually to obtain an error rate of around 25 % ($acc_{P3} = 75\,\%$). Although healthy subjects often achieve close to 100 % accuracy, the aim was to test with an error rate typical of patients. Kübler et al. found a mean error rate of 34 % in ALS patients using the P3 speller [68, N_p=11].

The 16 EEG channels and 3 EOG channels were recorded with two g.USBamp amplifiers (an internal [0.5–30 Hz] order 8 Chebyschev bandpass filter was active) and digitised at 256 Hz. Ground and reference electrodes were placed at the left and right mastoid, respectively. In most cases, impedance of electrodes was below 5 $k\Omega$. Impedances were measured at the beginning and end of the recording and found to decline gradually towards the end of the session. The mean of the two measurements was compared between the two recording sessions. The position of the electrodes is shown in Figure 5.3.

After spelling "BRAIN POWER" to train C_{P3} (SWLDA), users were asked to repeatedly spell the sentence

```
1FRANZ JAGT|2IM KOMPLETT|3VERWAHRLOSTEN|4TAXI QUER|5DURCH BAYERN.
```

Horizontal bars indicate a break between runs. The aim was to spell the sentence four times per session. The four phases of the study can be summarised as follows:

- P3 Training. Ten minutes to record some data (without feedback) to train C_{P3}. There was no re-training in session 2.
- ErrP Training. Two hours of *copy spelling* to record the ErrP during feedback of the recognised letter in session 1, another 30 minutes of session 2. The C_{ErrP} was trained with all data collected in this

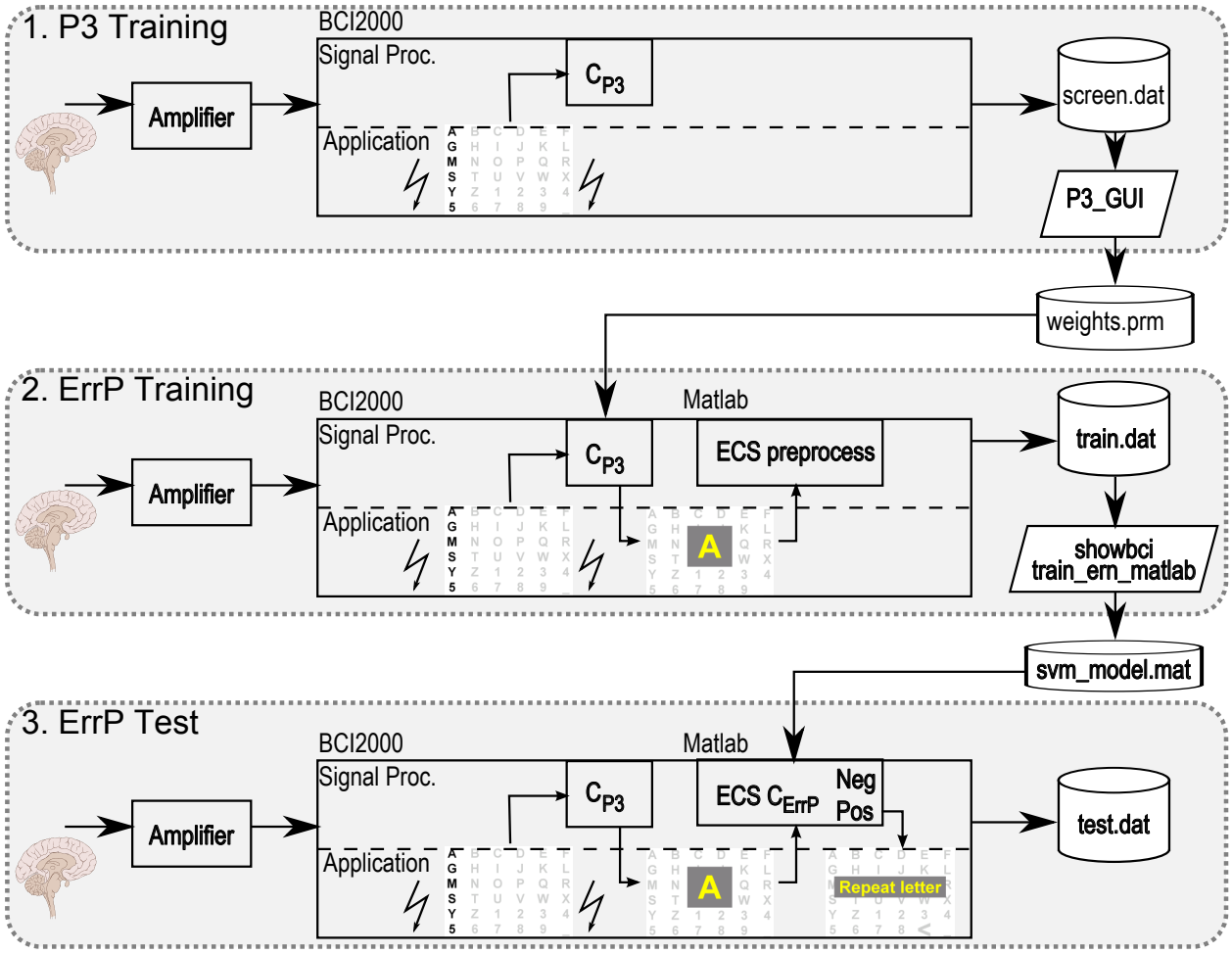


Figure 5.2.: Experimental setup showing the three phases of the study. (1) P3 Training: The participant's brain signal is amplified and passed through the signal processing pipeline of BCI2000 while columns and rows of the P3 Speller matrix flash, indicated by the flashes in the Application. The data is loaded into an existing Matlab GUI, P3_GUI, which allows extraction of relevant features for the SWLDA P3 classifier C_{P3}. (2) ErrP Training: The trained C_{P3} sends its classification result to the Application, which is modified to display the classified letter in large print. The signal up to 800 ms following this timepoint is later extracted the the tool `train_ern_matlab`, which additionally trains the SVM ErrP classifier C_{ErrP}. (3) ErrP Test: The ECS classifies the brain signal after appearance of the letter and if an ErrP is recognised, the participant is asked to repeat the letter.

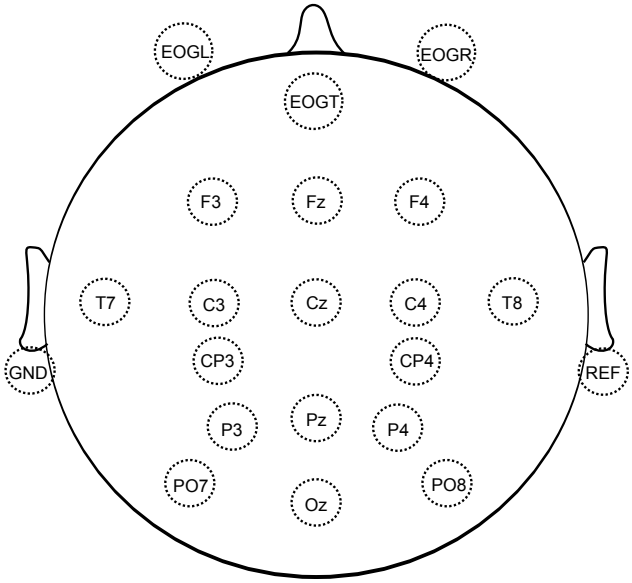


Figure 5.3.: Electrode montage. EOG electrodes were positioned as indicated according to the placement described in [115].

phase. (After copy spelling of the first sentence was completed, C_{P3} was re-trained with the 7 runs of available data to increase robustness).

- ErrP Test. One and a half hours of copy spelling with error correction active. If a trial was classified as positive, the current letter was deleted (application of an automatic *backspace*).
- ErrP Test with free spelling (optional). If the user was motivated and awake enough to continue, the final 2–3 runs were devoted to *free spelling*, i.e., the system did not prescribe what was to be spelled, and the ECS was still active.

One method of investigating why patient's results are mostly worse than those of healthy subjects is the direct group comparison. A second method which was employed is the assessment of mood and motivation and its a posteriori correlation with results, to find influencing factors on the results. To this end, two questionnaires for motivation and mood assessment were handed to participants. The "Questionnaire for Current Motivation" for BCI [92] comprises eighteen items divided into four subscales incompetence fear, mastery confidence, interest and challenge which have to be rated on a seven point scale. To assess mood, a subscale of the "Scales for the Assessment of Quality of Life" [6] which comprises a total of ten items to be answered on a five point scale was used.

5.3. Participating Subjects and Patients

This section describes the healthy subject data included in the evaluation. A comparison with the data and methods used for the patients is given in the following section. Because healthy subjects are mostly drawn from a population of young students (in this case aged 24.6 ± 2.3), an effort was made to include a second group of older subjects (aged 43.5 ± 4.3) in the study as well. This made it possible to study effects between patients and healthy subjects with matched age. Indeed, the age difference between Group H1 and the patients was significant (t-test, $p < 0.001$), while this was not the case for Group H2 (t-test, $p = 0.16$). Table 5.2 gives an overview of the recording sessions with healthy subjects. Inclusion criteria were right-handedness and the

Table 5.2.: Overview of recording sessions with healthy subjects (top sections) and patients (bottom section) for the ErrP study. Δ days is the number of days between the two measurements.

Subject	Age	Sex	EOG	Train Trials	Test Trials	Δ days
Ha_5	25	f	yes	335	332	5
Hb_5	27	m	no	350	220	1
Hc_5	24	f	S02	298	266	1
Hd_5	24	m	yes	311	234	1
He_5	25	f	no	332	357	1
Hf_5	25	m	yes	298	232	1
Hg_5	23	f	yes	358	384	2
Hh_5	28	m	yes	298	203	2
Hi_5	20	m	no	265	190	7
Hj_5	39	m	yes	298	241	1
Hk_5	51	m	yes	298	222	20
Hl_5	45	m	yes	249	211	1
Hm_5	39	m	no	313	314	4
Hn_5	40	f	yes	310	250	7
Ho_5	51	m	yes	248	117	8
Hp_5	47	f	yes	298	198	14
Hq_5	52	m	yes	276	201	1
Hr_5	42	f	yes	119	-	-

Subject	Age	Sex	EOG	Train Trials	Test Trials	Δ days	ALSFRS-R
Pa_5	54	m	yes	249	130	7	43
Pb_5	54	m	yes	298	129	17	21
Pc_5	36	m	no	298	205	4	8
Pd_5	63	m	yes	239	132	2	22
Pe_5	58	m	yes	239	119	9	39
Pf_5	42	f	no	179	97	7	8
Pg_5	72	m	yes	60	-	-	16

ability to concentrate and sit calmly for two hours. Contact-lens wearers were asked to wear glasses due to experiences of a co-investigator with excessive blinking.

Only subject Hm_5 had prior BCI experience. Of the 18 healthy subjects, 6 are affiliated with our lab as students or researchers.

Here some relevant details of the patients participating in the study are shown. All patients had used the P3 speller before. In contrast to the other patients, Pc_5 suffers from Duchenne muscular dystrophy. (As the ALSFRS-R is not fine-granular enough between 0 and 1, some additional details concerning remaining muscular functions are given in Table 5.2). The mean age of this group was 54.2 ± 13.3. Patient Pa_5 was able to drive to our institute; all other recordings took place in the patients' homes.

Patients generally require more sequences per trial than healthy subjects. This is reflected in the lower number of training trials (i.e., three sentences instead of four during session 1). It was clear from the outset that there would be less training data available from the patients, and one of the reasons why the study was broken into two sessions is that one session's data would not have sufficed to train the classifier robustly

enough (the aim was to collect > 50 training vectors from the *incorrect letter* class).

5.4. Treatment of Artefacts

Because P3-based BCIs generally work quite well and are used with many flash sequences, artefacts such as eye blinks are not critical. It can be argued that they become critical as the number of flash sequences per trial is reduced, which was done in this study. On the other hand, the study's main aim was not to optimise the P3 classifier's accuracy acc_{P3}, but the ErrP classifier accuracy acc_{ErrP}.

To keep the ErrP classifier robust towards blinks and similar artefacts, these were not removed from the training data. Rather, a visual check of the user's hit-minus-miss average was performed after the first 5 runs of ErrP training data were recorded. If the hit-minus-miss average was highest in a very frontal location, the user was instructed to avoid any type of blinking during the letter feedback and particular attention was paid to this in the further course of the session.

However, EOG was recorded in almost all sessions. To obtain a baseline EOG measure, some blinks and eye movements trials were recorded before the session began. The EOG electrodes were affixed as described by Schlögl et al. [115].

In the offline analysis, a number of steps were undertaken to collect evidence that the ECS is actually using the ErrP rather than eye movements for classification:

- Visual inspection in the time domain (latency, morphology)
- ROC plots of time samples as features
- Classification using single channels or EOG channels exclusively
- Regression-based EOG correction method
- ICA and removal of components exhibiting features typical of artefacts [50]

5.5. Feature Extraction and Classification

In the following, the methods used to obtain useful training data for ErrP recognition will be explained. Particular note will be taken of adjustments of the algorithms due to the patients. Figure 5.2 shows the three stages of the experiment, named P3 training, ErrP training, and ErrP testing. The ErrP training step includes a cross-validation (CV) to determine the hyperparameters of the SVM classifier.

The P3 classifier will not be mentioned in much detail, as it does not play a large role here. Important to remember is the fact that acc_{P3} should not change significantly from the ErrP train to the ErrP test phase, as the classifier C_{ErrP} is optimised to match the characteristics of C_{P3} (confusion matrix).

Prior to the decision about how to preprocess the data, some tests were conducted on the prerecorded patient data. By visual comparison of ROC plots it became evident that detrending the data and using a common average reference (CAR) are important prerequisites. In contrast, baseline correction of the ErrP interval did not improve results. Compared to AR coefficients and PLV values, the downsampled time series was found to be better to discriminate between correct and incorrect letters. By a 10-fold CV it was found that among many combinations of time intervals, the [100–800] ms interval offered the best classification accuracy. Various classifiers were compared (LDA, SWLDA, linear SVM, RBF-SVM), whereby the RBF-SVM delivered the best results. Regarding the number of EEG channels to use, a subset of 8 channels performed just as well as all 16 channels.

After the ERN training phase, a window of [100–800] ms following the letter feedback was extracted from the data. Analysis of the prerecorded data from ALS patients had determined the window size. The

EEG was bandpass filtered in the range [0.5–30] Hz (FFT, removal of unwanted bands, inverse Fast Fourier Transform (iFFT)) before being downsampled to 16 Hz. Thereafter, the signal was detrended and scaled by centering and mapping the absolute maximum value to ± 1. The time domain samples of all channels were used as input vectors to train the RBF-SVM classifier C_{ErrP}. Concerning the SVM hyperparameters, prior cross-validation on the prerecorded data showed that $C = 10$ is a reasonable choice. It was thus kept constant throughout the study to speed up the CV process during the second session (patient has to wait) and prevent overfitting this parameter to the data of session 1 (75 % of the ErrP training data is collected in session 1).

To combat the problem of highly unbalanced data (roughly 4 "correct" trials for every "incorrect" trial), and at the same time bound the number of FPs, a number of methods exist. The data itself may be balanced by *random undersampling* of the majority class or *random oversampling* of the minority class. The *synthetic minority oversampling technique* [27] is more intricate, but at the end of the day all three methods failed to bound the number of FPs sufficiently. They were thus abandoned in favour of modifying the SVM weights w_1 and w_{-1}. This is achieved by modifying the term $\frac{C}{n}\sum_{i=1}^{n}\xi_i$ in (2.12) to read

$$\frac{w_{+1}C}{n}\sum_{i|y_i=+1}\xi_i + \frac{w_{-1}C}{n}\sum_{i|y_i=-1}\xi_i \tag{5.5}$$

In fact, only the ratio of the weights is relevant, so $w_1 = 1$ was kept constant and $0 \leq w_{-1} \leq 1$ was determined by 10-fold CV (line search in increments of 0.1) on the ErrP training data. The CV output was a plot showing estimated sensitivity, specificity and accuracy as a function of w_{-1}. A user-specific value which maximised both accuracy and specificity was chosen. The a priori goal was to keep specificity > 0.9. Maximising accuracy alone is not a good option (see explanation on page 21).

As with the classification of connectivity features and the multiclass problem discussed earlier, the RBF kernel was found to be suitable for the ErrP classification in CV on the prerecorded data. The RBF kernel requires just one hyperparameter $\gamma = 1/(2\sigma)$, which was left unchanged from its default value in *libSVM* [25], which is the reciprocal of the number of features.

The trained SVM model was saved as a `.mat` file and loaded into the Matlab® classification module from within the BCI2000 graphical user interface (GUI).

5.6. Online Classification

Training of the classifier took place as described above. Before online use of the ECS, the trained classifier `svm_model.mat` is loaded (see Figure 5.2, ErrP Test phase). The ErrP classification was realised as a BCI2000 Matlab® filter[1]. Upon receiving a predefined number of data blocks (one second of data), the filter does preprocessing identical to the steps described in Section 5.5. Upon classification, the user is either granted a further two second break or instructed to repeat the current letter. Preprocessing and classification cause high CPU load for a few hundred milliseconds and cause the BCI2000 *roundtrip* parameter to surpass 1, but this is not critical as it happens during the 2 s break before the following trial. Online classification results are reported in Section 6.3.2.

In future, this module can be integrated into the BCI2000 filter pipeline as a C++ class derived from *GenericFilter* to enable the error correction method to run out-of-the-box without Matlab® dependency. This will further the error correction system's acceptance among fellow researchers, supervisors and patients. The Matlab® function currently calls *libSVM* routines from a dynamic link library (DLL) file and is thus easy to port to C++.

[1] `http://www.bci2000.org/wiki/index.php/Programming_Reference:MatlabFilter`

6. Results

Results for all three studies undertaken by the author will be presented in this chapter. A common theme in all of the results presented will be connectivity, i.e., care will be taken to focus on bivariate features as opposed to univariate features where reasonable. Equally important will be the influence these results could have in assisting patients to attain a higher information transfer rate (ITR). Therefore, care has been taken to describe the findings such that they can be implemented to benefit patients. Where appropriate, own implementations and methods leading to these results are described in Section 4.

6.1. Multiclass Studies

The multiclass studies serve multiple purposes: to compare univariate and bivariate (connectivity) features, to compare bit rates between binary, ternary and quaternary classification, and to compare the efficacy of motor versus nonmotor tasks. We will begin by summarising the ITR achieved for univariate and bivariate features. Thereafter, the focus is on the number of classification classes used to achieve the highest ITR. The final section is devoted to results achieved with patients in a number of smaller studies. Later on, the discussion includes an evaluation of the efficacy of motor versus nonmotor tasks.

6.1.1. Connectivity Features

The error estimation method used (nested cross-validation (CV)), employing a linear support-vector machine (SVM) and recursive feature elimination (RFE). The high number of features, resulting from the 151 MEG channels, forced a reduction of the number of outer CV folds to 20 (as compared to a previous report where 50 outer CV folds are used [74]). Additionally, a faster version of the RFE is used in this dissertation: The algorithm initially discards 50 % of the features per iteration instead of one feature per iteration. This might have a slight negative influence on the alpha error estimate (see Table 6.1). However, the trade-off can be justified by the speedup it generates. This estimate is for the number of selected features with a deviation of no more than two standard errors from the lowest error estimate during the CV. The number of selected features used for the alpha estimate is displayed in the column to its right. The errors were estimated for three time windows of the classification interval (CI). The accuracy of each subject's best time window is shown.

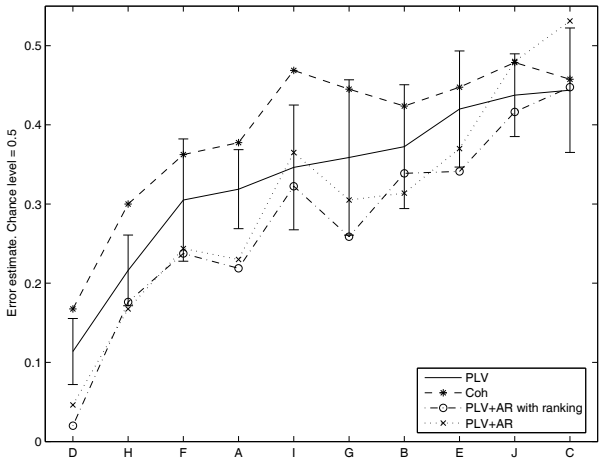
Results for the combination of PLV+AR features show an improvement over the accuracies for a single feature type (see Table 6.1). This result includes PLV features calculated over all three time windows ($2775 \cdot 3$ PLV + $150 \cdot 2$ AR features). Note that the AR features are not grouped channel-wise as is done in recursive channel elimination (RCE) [74], i.e., it is possible that any number of features belonging to a particular channel are selected. The RCE uses this "all-or-nothing" approach to simplify physiological interpretation of the chosen features. The rationale behind the decision to interpret each feature individually is that speed and accuracy in the online implementation are more important than physiological interpretation in the case of this dissertation.

A subject-specific comparison of PLV, coherence (Coh), and PLV+AR features is shown in Figure 6.1a. Subjects are shown sorted by the estimated PLV error, with error bars of one standard deviation shown for the PLV error only. The figure shows clearly the superiority of the combination PLV+AR to the use of PLV alone, whereas use of the Coh feature results in the highest error.

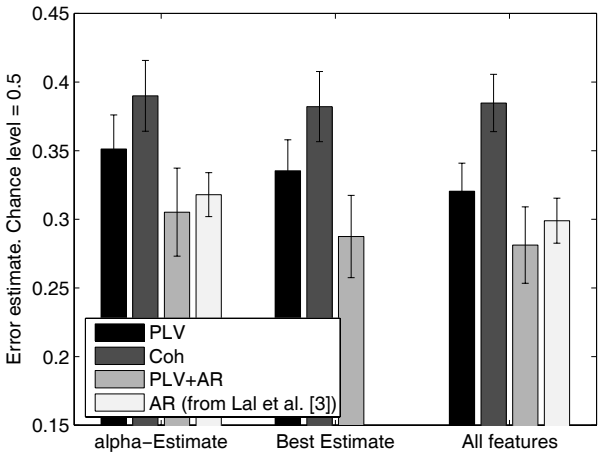
The bar graph shown in Figure 6.1b aggregates the result over all subjects while giving a visual comparison between the feature types and previous work. The PLV feature is superior to the Coh feature, which is

Table 6.1.: Classification error estimate using PLV features and combined PLV and AR features (cross-validation). PLV features are bandpass filtered in the range 8–40 Hz. The alpha estimate is taken from a selected set of features; its cardinality is shown in the adjacent column.

	PLV			PLV+AR		
Subject	All Features	Alpha	Nr. Feat.	All Features	Alpha	Nr. Feat.
Ha_1	0.32 ± 0.05	0.30 ± 0.08	7 ± 16	0.25 ± 0.06	0.23 ± 0.06	8 ± 11
Hb_1	0.37 ± 0.06	0.35 ± 0.10	2 ± 2.3	0.35 ± 0.07	0.31 ± 0.08	5 ± 9.6
Hc_1	0.42 ± 0.08	0.48 ± 0.06	11 ± 13	0.46 ± 0.06	0.53 ± 0.07	5 ± 6.3
Hd_1	0.12 ± 0.05	0.14 ± 0.05	4 ± 2.7	0.04 ± 0.03	0.05 ± 0.03	14 ± 7.4
He_1	0.38 ± 0.07	0.44 ± 0.09	4 ± 4.9	0.33 ± 0.05	0.37 ± 0.07	9 ± 8.0
Hf_1	0.28 ± 0.07	0.37 ± 0.06	8 ± 6.6	0.21 ± 0.05	0.24 ± 0.08	2 ± 1.8
Hg_1	0.35 ± 0.07	0.40 ± 0.06	6 ± 6.6	0.23 ± 0.09	0.31 ± 0.09	13 ± 13
Hh_1	0.22 ± 0.04	0.21 ± 0.06	1 ± 0.8	0.21 ± 0.06	0.17 ± 0.07	2 ± 0.7
Hi_1	0.35 ± 0.05	0.34 ± 0.08	5 ± 3.3	0.29 ± 0.06	0.37 ± 0.07	10 ± 9.8
Hj_1	0.40 ± 0.06	0.48 ± 0.07	7 ± 9.8	0.45 ± 0.06	0.48 ± 0.07	14 ± 21
Mean	0.32 ± 0.09	0.35 ± 0.11	6 ± 6.6	0.28 ± 0.13	0.31 ± 0.14	8 ± 9.0



(a) Inter-subject variation in estimated classification error using four combinations of bivariate features, sorted by PLV. The error bars show the variance in the cross-validation procedure, indicating that none of the methods significantly outperforms the other. Healthy participants are denoted on the abscissa.



(b) Comparison of four feature types/feature combinations. The two groups "alpha estimate" and "best estimate" use a reduced feature set. "Best estimate" is an average resulting from the use of the best feature set in each CV fold.

Figure 6.1.: Comparison of feature combinations for the prerecorded data. Estimated classification error as determined by cross-validation is shown on the ordinate.

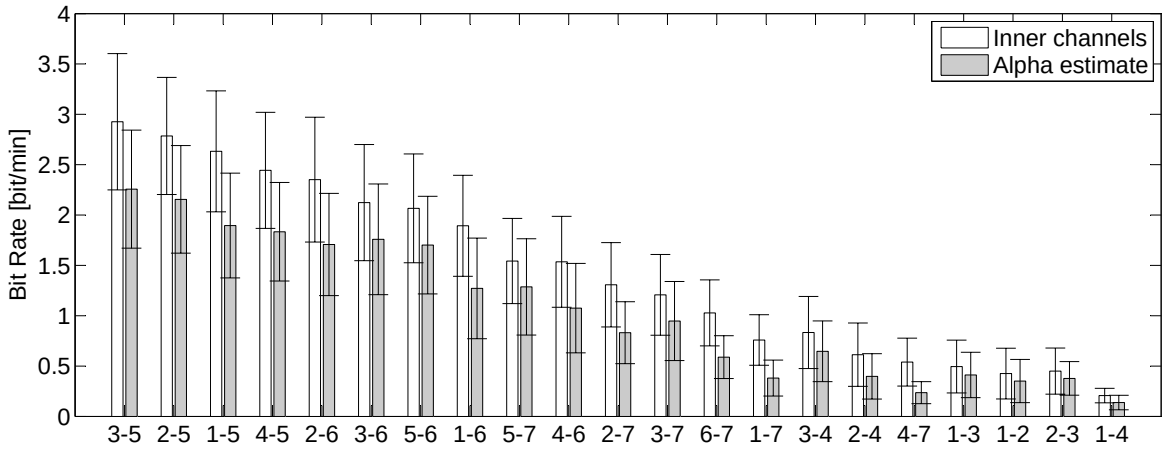


Figure 6.2.: Information transfer rate in b_W/min (mean over 10 subjects) for each two-class combination shown for all features and a feature subset, the alpha estimate. Error bars depict the standard error, i.e., the standard deviation across all subjects. For each subject, the mean result of the 10-fold CV goes into the overall mean. Bars are labelled with class combinations on the abscissa: 1 (foot), 2 (left hand), 3 (right hand), 4 (tongue), 5 (mental arithmetic), 6 (navigation), 7 (visual scene).

consistent with previous findings [49]. Although the combination of features PLV+AR produces the lowest error, the difference between using AR coefficients and PLV+AR is not significant (t-test, $p = 0.44$). The number of features was reduced to 8 on average using RFE. The feature selection increases the mean estimated error by less than 3 %.

Interestingly, the Coh feature shows a higher dependence on the chosen frequency band than the PLV: choosing the most discriminative frequency band for each subject would have improved the estimated mean accuracy by 2.5 % when classifying Coh features (result not shown). Overall, that does not make the Coh feature any more attractive.

6.1.2. Binary, Ternary and Quaternary Classification

This section compares the efficacy of multiclass BCI to binary BCI. Let us commence with binary combinations of the seven classification tasks, shown in Figure 6.2. In this figure, where task pairs are ordered from most effective to least effective, the commonly used hand-tongue pairs appear quite far on the right with a mediocre bit rate. A combination of mental arithmetic with any motor task seems to be the most effective, followed by task pairs including navigation. In a nutshell, the combination of a nonmotor task and a motor task bears the promise of much higher bit rates than the use of motor tasks alone.

Subject-specific results for the two-class and three-class error estimates can be seen in Table 6.2. The column description *Inner channels* refers to the fact that the outermost MEG sensors were not included in the analysis. The lower number of channels makes the system capable of online operation due to lower computing time. Feature selection is needed for the online case despite the a priori channel reduction to keep processing speed at a minimum. Results with feature selection are denoted *alpha estimate*. The use of feature selection produces a slightly lower accuracy (3–4 %), but because the effect is consistent over all subjects, it is significant (paired t-test, $p = 0.011$ ternary, $p < 0.001$ binary). Comparing the alpha estimates of ITR, all subjects except Ha_2 and Hi_2 perform better when using three tasks. Looking at the best task combinations, it is clear that the combination of mental arithmetic with a motor task works best for most subjects in the binary

Table 6.2.: Results for the best combination of three classes per subject, in terms of estimated classification error and bit rate b_W/min. The error estimate was obtained by 5-fold CV. The binary results are given as a comparison. The class combinations are: 1 (foot), 2 (left hand), 3 (right hand), 4 (tongue), 5 (mental arithmetic), 6 (navigation), 7 (visual scene). Three-class bit rates that represent an improvement over binary bit rates are printed in bold.

	Ternary Classification						Binary Classification					
	Inner channels			Alpha estimate			Inner channels			Alpha estimate		
Subject	Error	Comb.	b_W	Error	Comb.	b_W	Error	Comb.	b_W	Error	Comb.	b_W
Ha_2	0.34	1-5-6	2.25	0.40	3-5-6	1.50	0.19	3-6	2.04	0.23	3-6	1.53
Hb_2	0.16	2-5-6	5.30	0.23	1-2-6	**4.00**	0.08	2-6	4.02	0.11	3-5	**3.37**
Hc_2	0.29	1-5-7	2.96	0.29	1-5-7	**2.98**	0.15	2-5	2.65	0.21	2-5	**1.77**
Hd_2	0.32	2-3-6	2.47	0.39	3-5-6	**1.61**	0.22	3-5	1.65	0.24	3-5	**1.41**
He_2	0.25	2-5-7	4.59	0.23	2-5-6	**3.94**	0.10	1-5	3.68	0.14	2-5	**2.90**
Hf_2	0.23	4-5-6	4.59	0.23	1-5-6	**3.94**	0.10	3-5	3.58	0.15	3-5	**2.65**
Hg_2	0.16	1-5-6	5.37	0.21	4-5-6	**4.39**	0.06	2-5	4.50	0.10	2-5	**3.68**
Hh_2	0.16	1-5-6	5.30	0.20	2-5-6	**4.70**	0.05	3-5	4.77	0.09	5-7	**3.90**
Hi_2	0.48	2-3-5	0.72	0.50	3-5-6	0.59	0.26	3-5	1.20	0.31	2-5	0.75
Hj_2	0.09	3-5-6	7.32	0.12	3-5-6	**6.47**	0.03	3-5	5.53	0.06	3-6	**4.64**
Mean	0.24		4.09	0.28		**3.41**	0.12		3.36	0.16		**2.66**
SD	0.12		1.95	0.12		1.74	0.08		1.43	0.08		1.25

case, and arithmetic/navigation/motor task works best in the three-class case. The best combinations are fairly robust against the feature selection; at most one task changes after feature selection in the three-class case.

Removal of trials containing artefacts increased the offline classification error by 0.03 to 0.35 (subject Hd_2) and by 0.02 to 0.11 (subject Hj_2) using all inner channels and by 0.02 to 0.41 (subject Hd_2, no change for subject Hj_2) with feature selection.

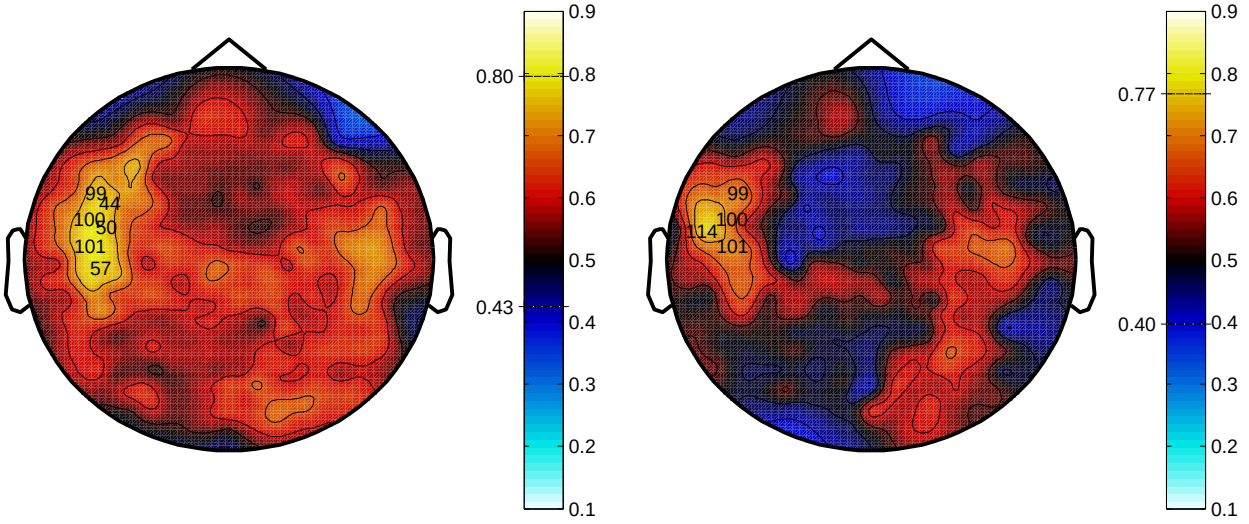
Due to problems with the software implementation for online operation and time constraints, there was no successful online test.

6.1.3. Patient Sessions

This section reports on multiclass results from various BCI paradigms measured in MEG and ECoG modalities with patient Pb_3.

6.1.3.1. Three-class Session (Magnetoencephalogram)

This three-class paradigm *best3* based on spontaneous MEG is described in Section 4.2.5.1. In a trial-by-trial check, some artefacts were found on outer channels, such as blinking (e.g., channel 25) and pulse at roughly two Hz (e.g., channel 128). The artefactual trials were not removed due to the low number of trials per task and their uniform distribution across the tasks. Task discriminability between imagined foot movement (*foot*), navigation between various locations in Pb_3's home town (*nav*), and aiming at a target (*aim*) is investigated. A visual check of task-related discriminability and its location over the cortex was performed for each of the five recording runs. The plots displayed high AUC scores in the eye area, indicating that Pb_3 might have been restless and unconcentrated in the beginning. Therefore, run 1 was omitted from the further plots. Average discriminability for runs 3, 4, 6 and 7 is shown in Figure 6.3. The discriminability measure is area



(a) *Foot* (40 trials) against *nav+aim* (80 trials). Second AR coefficient is plotted.

(b) *Nav* (40 trials) against *foot+aim* (80 trials). First AR coefficient is plotted.

Figure 6.3.: Patient Pb_3. Topographical ROC plots showing AUC values of AR coefficients (model order two) by collating tasks in a OVR manner. Bright values close to 0 or 1 signify high discriminability, whereas values close to 0.5 signify no discriminability. Minimum and maximum AUC scores are marked on the colourbar.

under curve (AUC) represented on a colour scale from black to bright blue or yellow, with AR coefficients as features.

Although some discriminability is evident in the left temporal/motor cortex area (the area surrounding channel 101) in both subfigures 6.3a and 6.3b, it is not clear from which task this results. To find out which task the patient is able to perform best, a comparison of each task CI with the preceding inter-trial interval (ITI) period (acting as a baseline without task-related signal, power spectral density (PSD) features) is shown in Figure 6.4. Now it is clear that the *nav* task generates discriminable activity in the 6–8 Hz range over the left motor cortex electrodes. Electrodes 100–102 express the strongest μ-rhythm modulation in the healthy subjects (study of Section 4.2.2, result not shown) which fits well with this observation.

To estimate how well a OVR classifier will be able to recognise one particular task amongst the other two without using baseline information, all three task combinations' AUC scores using AR coefficients are shown in Figure 6.5. The two right-hand columns show the result for randomised labels as a visual comparison. The first two rows (*foot* and *nav*) differ sufficiently from the randomised columns. However, the fact that *aim* is not recognised well indicates that this three-class BCI would not reach the desired classification rate online.

The information gained by the MEG measurement implies that the left frontal hemisphere would be a good location to implant the ECoG grid, at least for the *foot* and *nav* tasks.

6.1.3.2. Four-class Screening Sessions (Electrocorticogram)

After implantation of the ECoG grid, the six four-class screening sessions *4class-screen* (see Table 4.1) were evaluated offline to determine the most discriminative tasks for patient Pb_3. This was a neccessary measure, as the *foot* and *nav* tasks had not achieved an accuracy useful for single-trial classification , except for one session seven days post-implantation where 70 % was achieved (personal communication). The method of data analysis was mentioned in Section 4.2.5.2. Discriminability between preceding ITI and each task for

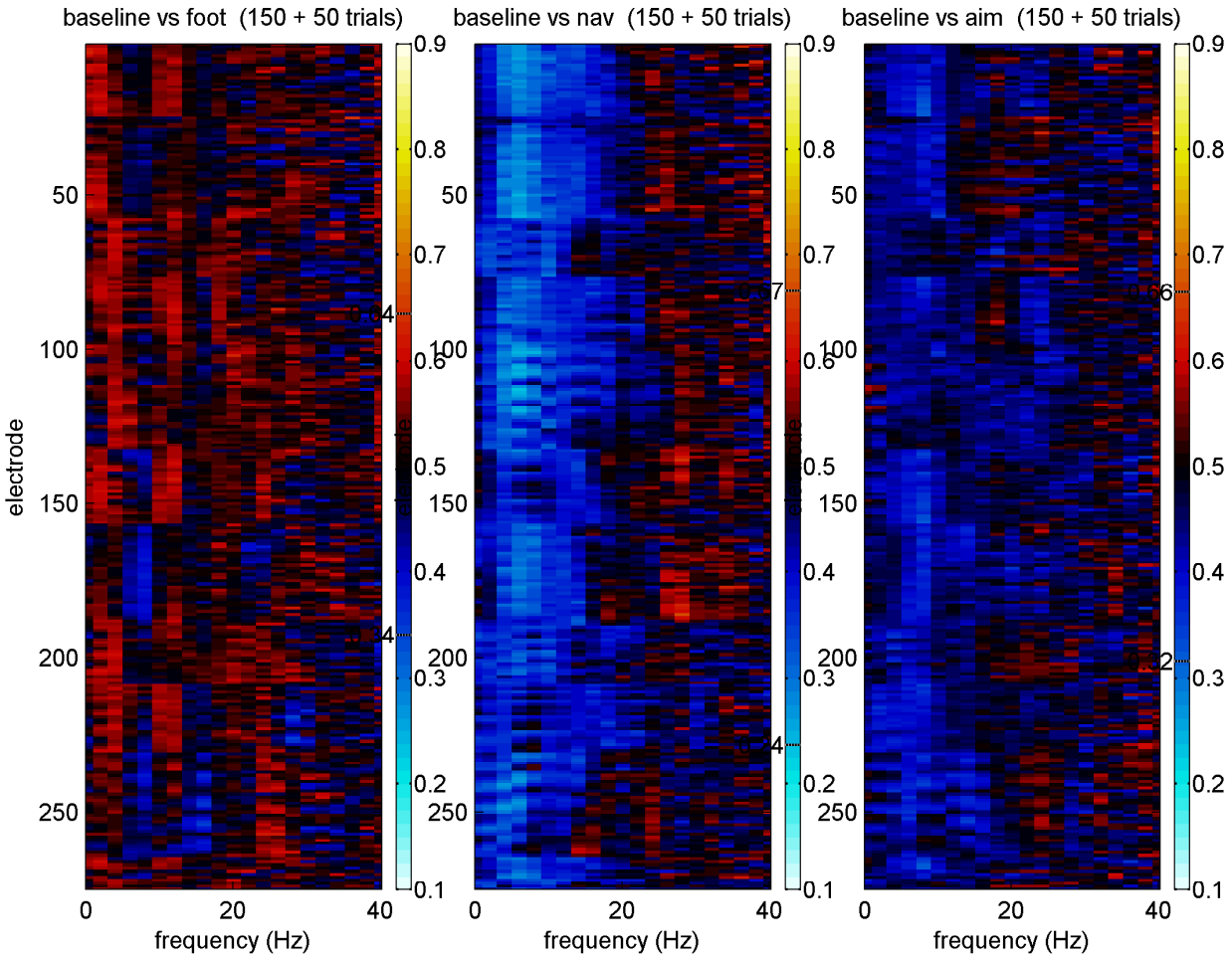


Figure 6.4.: Patient Pb_3. AUC values for the three task pairs baseline vs. *foot*, baseline vs. *nav* and baseline vs. *aim*. Each block represents a combination of channel and frequency band. Minimum and maximum AUC values are indicated on the colourbar.

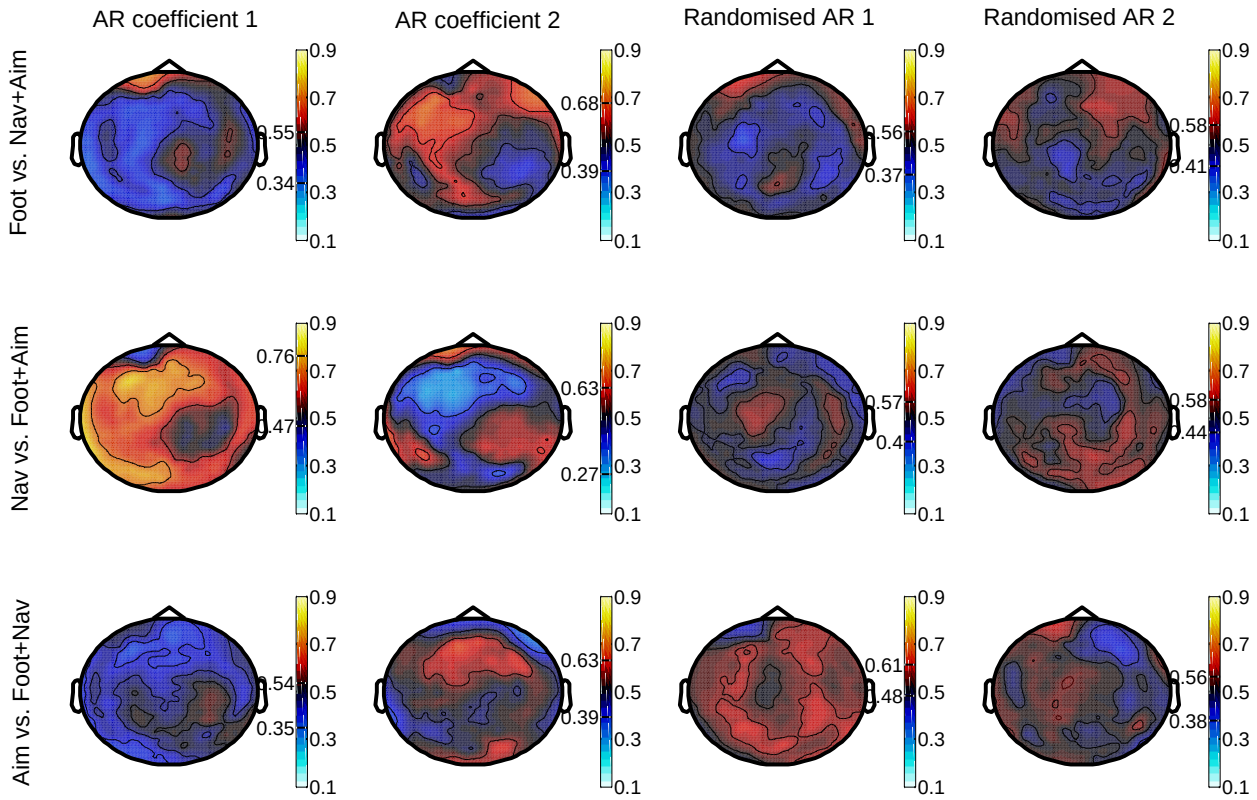
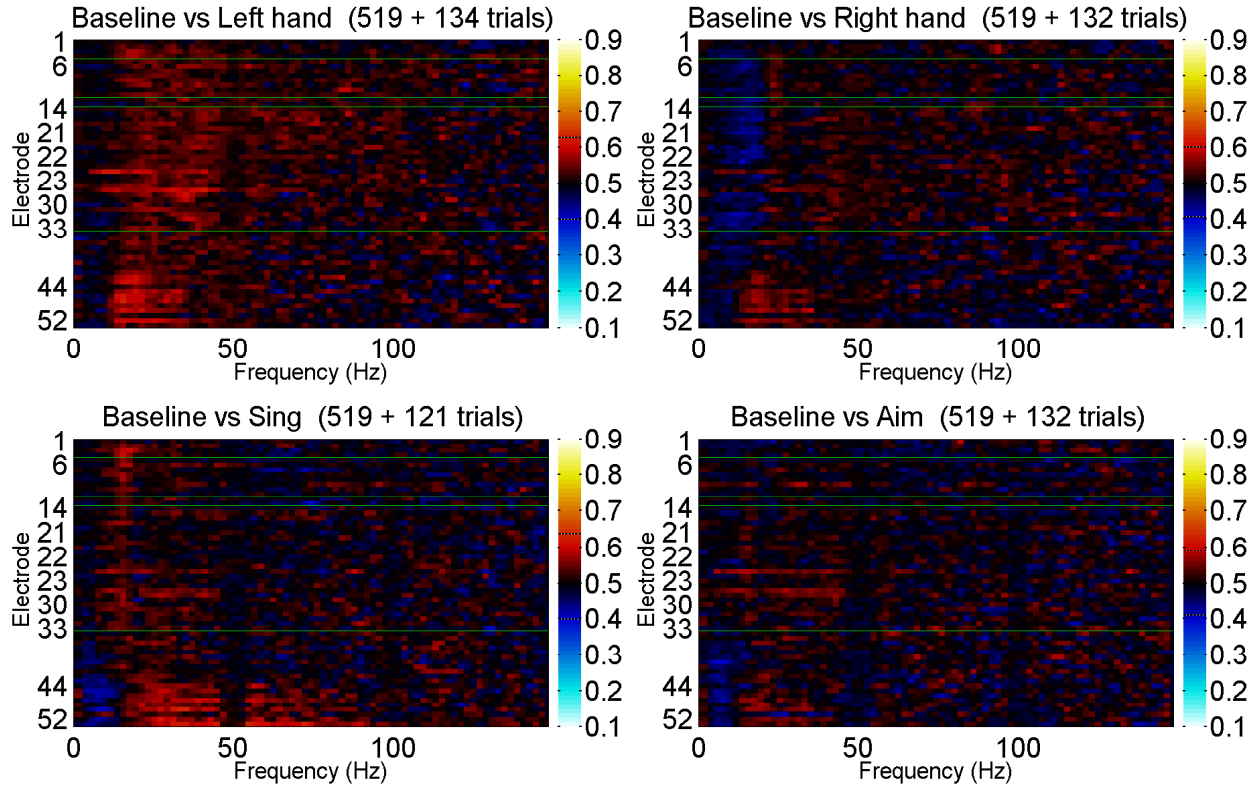
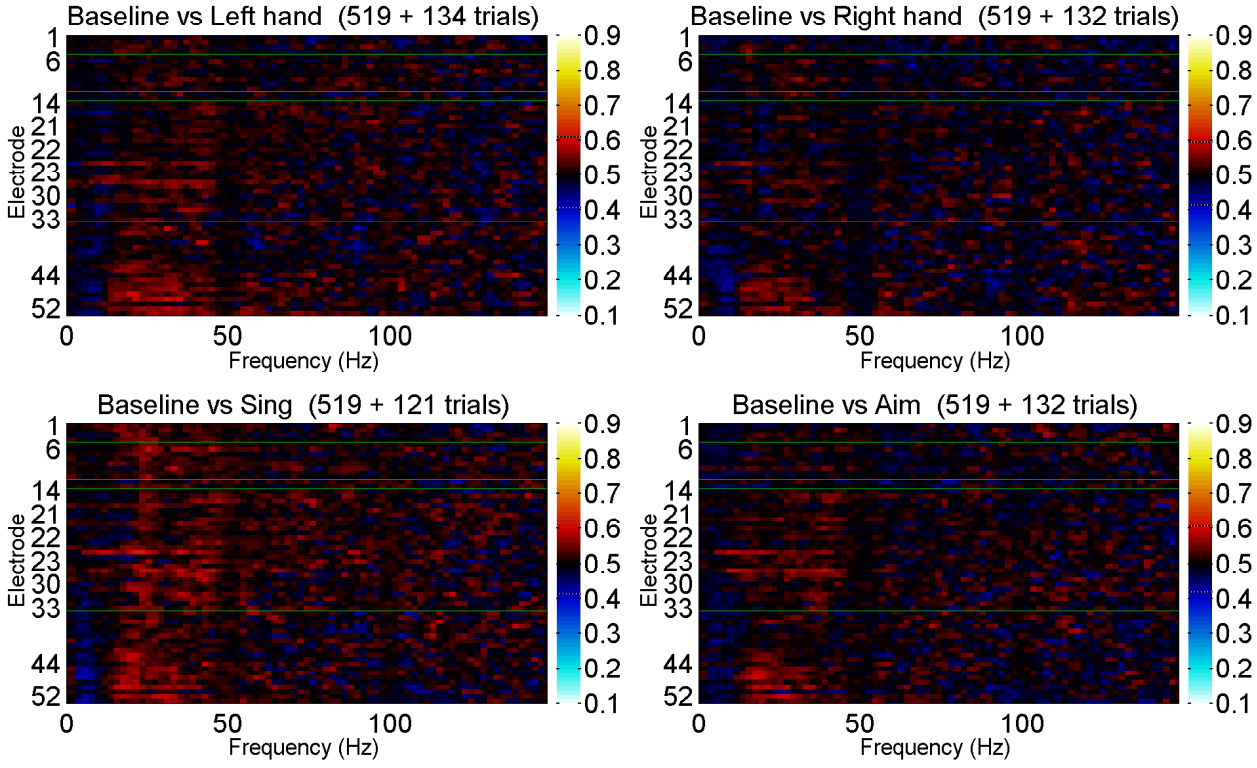


Figure 6.5.: Patient Pb_3. Topographical ROC plots showing AUC values of AR coefficients (model order two) by collating tasks in a OVR manner for all three combinations. First and second columns depict discriminability of first and second AR coefficient, respectively. Third and fourth columns represent first and second AR coefficient with randomised labels. These show the result for randomised labels, a method to estimate the maximum AUC score that can be achieved by a randomised assignment of the class labels, due to the number of trials and features.

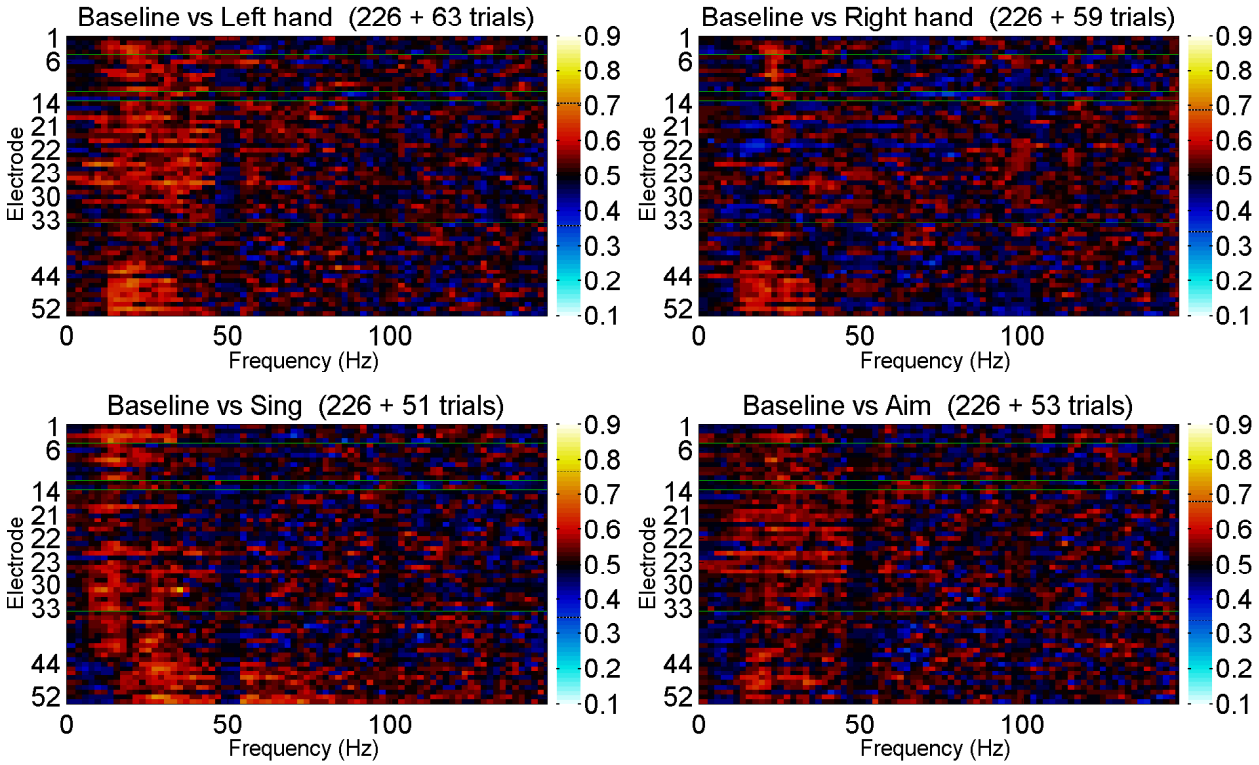


(a) Actual class labels.

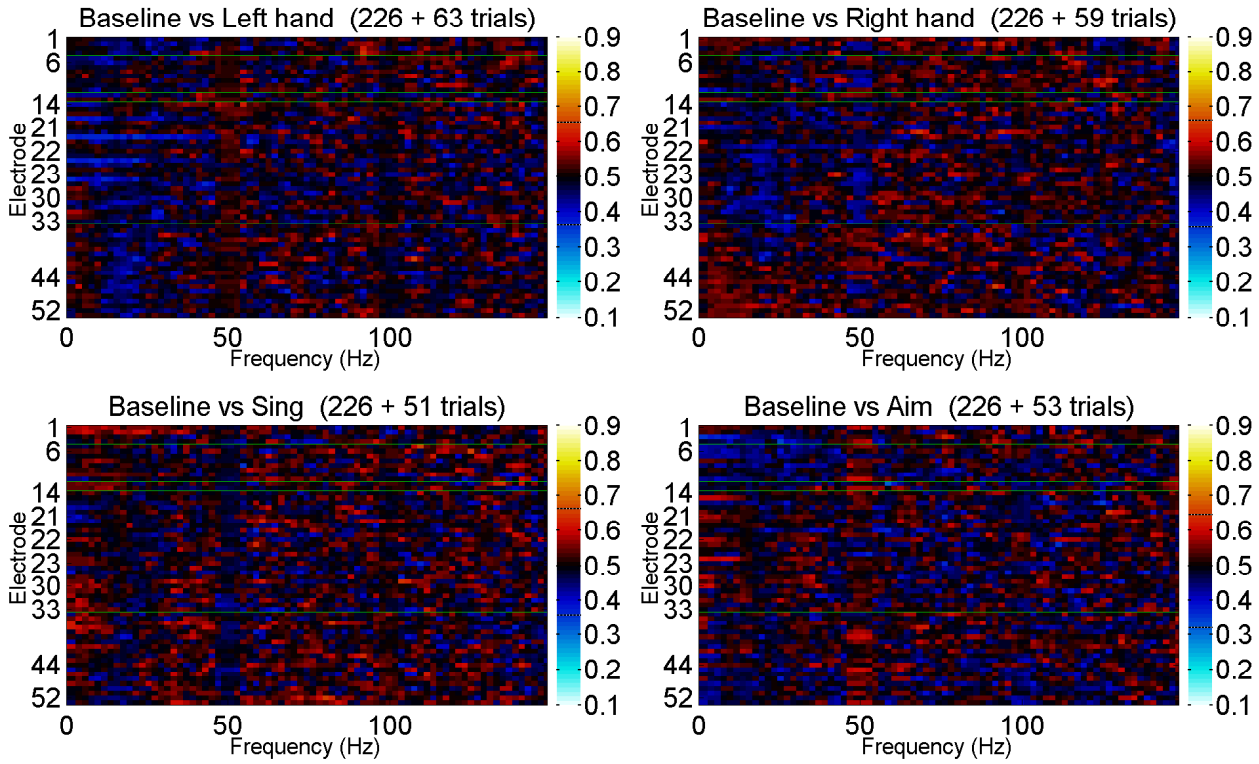


(b) Randomised class labels.

Figure 6.6.: Patient Pb_3. Comparison of four tasks (imagined left hand and right hand movement, singing, and aiming at a target). The colour represents discriminability (AUC score) between a task and the set of all ITIs. Horizontal lines indicate omitted electrodes.



(a) Actual class labels.



(b) Randomised class labels.

Figure 6.7.: Patient Pb_3. Comparison of the four tasks, sessions 6 and 7 only. The colour represents discriminability (AUC score) between a task and the set of all ITIs. Horizontal lines indicate omitted electrodes.

recording sessions 2–7 is shown in Figure 6.6a. In contrast, Figure 6.6b displays the result for randomised labels (see Section 2.5.2 for an explanation). PSD is only shown up to 150 Hz as there was no modulation visible in the range 100–250 Hz. The *left hand* and *sing* tasks show the highest modulation of PSD compared to the baseline state (left-hand panels, Figure 6.6a) in the range 10–40 Hz, especially on electrodes 44–52. A visual comparison with Figure 6.6b should convince any doubting reader that the vertical red structure is not a chance-level phenomenon. However, the AUC scores of 0.62 and 0.64 for *left hand* and *sing* are too low for single-trial classification. In fact, the AUC confidence range for this dataset is $[0.37, 0.63]$, which means the probability of an AUC value lying outside of this range by chance is below 5 %.

The second series of plots shown in Figure 6.7 are taken exclusively from sessions 6 and 7. It is clear that the patient either improved his strategy in the later sessions or was unable to concentrate properly on the task in the earlier sessions. Due to the lower number of 226 trials (sessions 6 and 7) compared to 519 (all sessions), the variance in the data is higher. Thus, the AUC confidence range increases to between $[0.28, 0.72]$ for *left hand* and $[0.27, 0.73]$ for the *sing* task. Once again, PSD modulation between 10–40 Hz is best visible for *left hand* and *sing* in Figure 6.7. The *sing* task is the only one to contain significant AUC scores outside of the confidence bound up to 0.76.

Even though the left hand and sing tasks would be promising candidates for a binary BCI, it was decided to first continue with a cognitively less demanding paradigm, namely the discrimination between singing and rest.

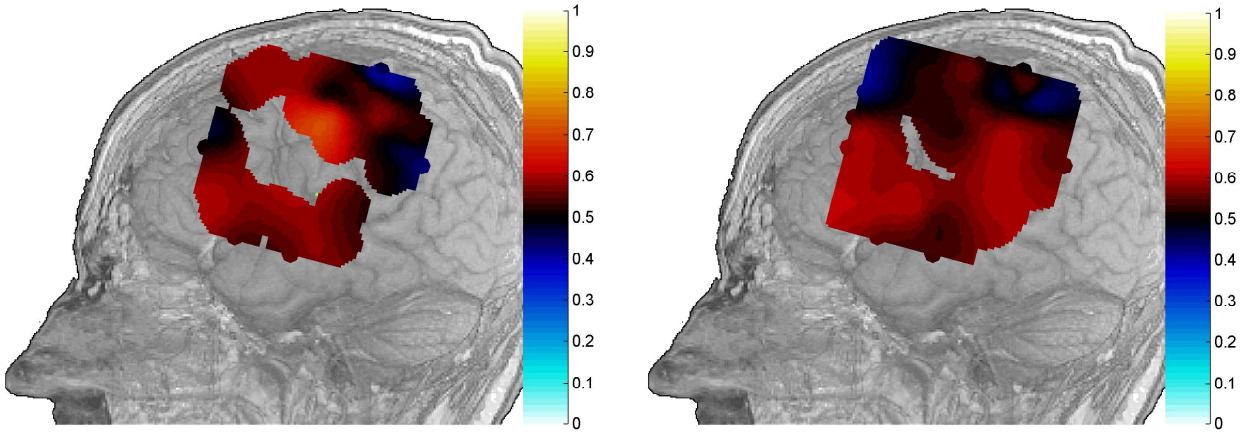
6.1.3.3. Singing Sessions (Electrocorticogram)

Two *sing-vs-rest* sessions were recorded with Pb_3 (see Table 4.1). The AUC scores using AR features are shown in Figure 6.8 for both sessions. The 5 % AUC confidence range for this data is $[0.29, 0.71]$ for sessions 1 and 2. AR coefficient 3 displays a high score on electrode 25 and surrounding electrodes, an area known as the inferior parietal lobe of the cortex (Figure 6.8a). This is in concordance with a study with healthy singers showing activation in functional magnetic resonance imaging (MRI) in the same area during imagined singing [63]. Overt singing in the MRI study activated primary sensorimotor cortex amongst others, which is slightly anterior to electrode 25, but close enough to also explain the modulation seen here. However, the highest AUC scores for sessions 1 and 2 are 0.67 and 0.61 respectively, well within the confidence range. This means that the modulation cannot be regarded as significant, which is why no classification results are presented for this data.

All in all, none of the ECoG paradigms achieved the restoration of communication in either Pa_3 or Pb_3. The following section investigates decreased cognitive ability in complete locked-in state (CLIS) as a possible reason for this.

6.2. Cognition Detection Study

Daily fluctuations in evoked response shape and amplitude were more noticeable than those caused by disease progression. Cognitive event related potentials revealed intact P3-like and N1/P2 responses to oddball target stimuli during the first three months of the investigation. However, the P3 response was missing in the final recording session (after three months in the completely locked-in state), which constitutes a major change. The spectrum of spontaneous ECoG activity showed peaks at 4–5 Hz in all sessions and at 15–20 Hz in the first three sessions. Alpha activity was particularly low in session 3, the session where the best P3 responses were recorded. Details follow.



(a) Session 1: AR coefficient 3 is displayed, 160 trials (78 *sing*, 82 *rest*) and 41 electrodes were analysed.

(b) Session 2: AR coefficient 4 is displayed, 160 trials (76 *sing*, 84 *rest*) and 39 electrodes were analysed.

Figure 6.8.: Patient Pb_3, *sing-vs-rest* measurement. Distribution of AUC score over the cortex, indicating discriminability between *sing* and *rest* tasks.

6.2.1. Significance of Evoked Response Potentials

Before calculating significance values of differences between evoked response potentials (ERPs), it is advisable to visualise the signal to ensure the electrodes were not measuring artefacts or spurious signals. The third recording session allows the best visual recognition of N1/P2 and P3 components (see Figure 6.9). Electrodes 46 and 47, which lie over the auditory cortex judging by the MRI scan, display the largest N1 (visible as a negative peak at 200 ms). Ten electrodes (five shown, bottom right) at the caudal end of the grid allow visual recognition of the P3 with 500–1000 ms latency.

A summation of the number of electrodes with significant differences between standards and deviants is shown in Tables 6.3 through 6.6. The total number of channels tested is 13 and 46 for Pa_3 and Pb_3, respectively. The reliability of the t-test for components reflecting cognitive processing was increased by (1) presenting the result for two different reference points, namely CAR and another quiet channel ideally displaying reduced electrical activity; (2) presenting the Monte Carlo result alongside the t-test as an estimate of how many electrodes would display significant differences by chance. This allows a summary of the results per patient as follows: For Pa_3, the only evidence of cognition is given by the Semantic Oddball test (see Table 6.4).

For Pb_3, the t-test finds significant ERPs in sessions 1–3 for Standard Oddball. Some ERPs due to Mismatch Negativity (MMN) are significant in sessions 1–3, yet session 4 has by far the highest number of significant electrodes. The responses to MMN and Oddball seem to be inversely related. The only other significant effect is found in the Priming test in session 1. A comparison with Table 4.6 leads to the conclusion that the active instruction to count the rare stimuli (runs annotated by (c)) did not increase the amount of significant components in those runs.

The colour-coded value of effect size E (per session, mean of runs) is plotted in Figure 6.10 for each session. Electrodes close to the auditory cortex display high effect size for MMN, best observed in sessions 2–3. This is in line with those who suggest the source of the N2a to be the superior temporal gyrus [3, 114]. In turn, the inferior parietal lobule has been suggested as a P3 generator [122], which fits well with the effect sizes found for the P3 in sessions 2–3, and to a lesser degree in session 1.

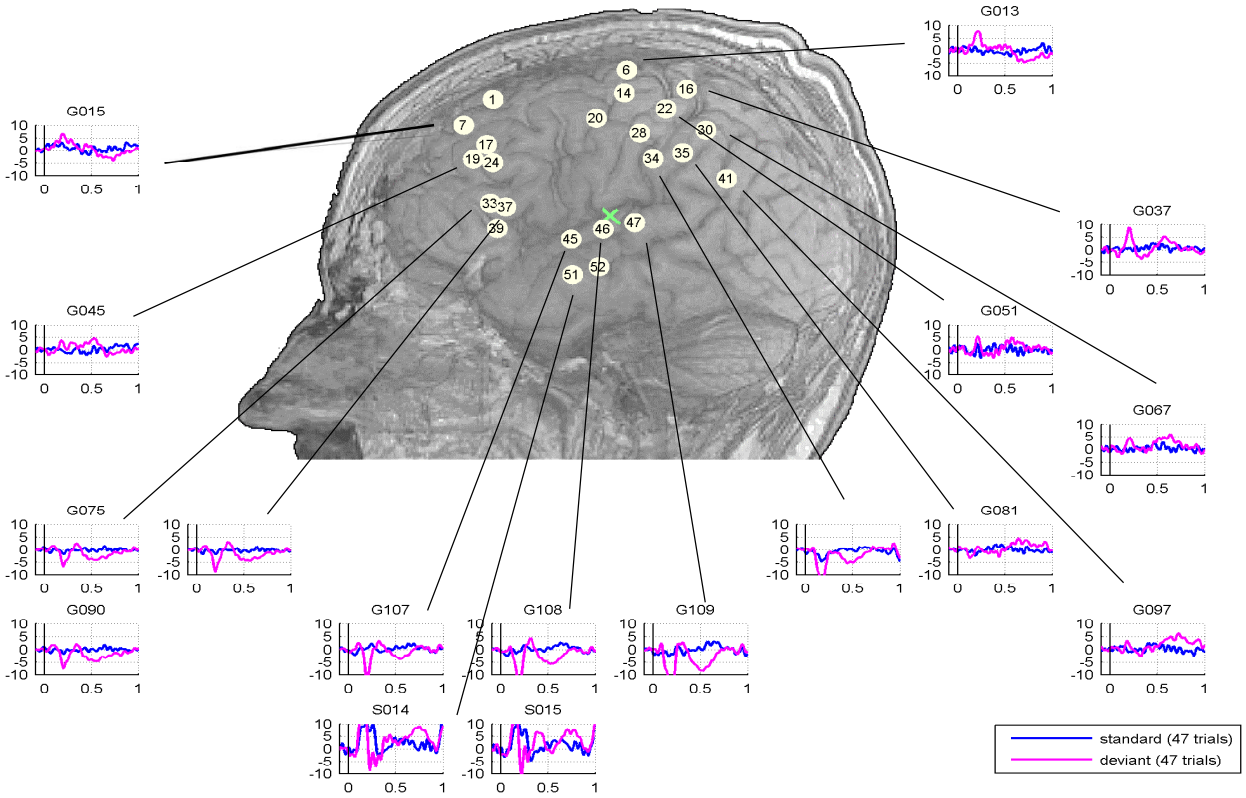


Figure 6.9.: Patient Pb_3. Time series of cortical response to standard and deviant tones of the Standard Oddball paradigm. For clarity, only a subset of electrodes is shown. Recording session 3, run 5, CAR reference.

Table 6.3.: Patient Pa_3. Number of electrodes showing a significant difference of means between standards and deviants (t-test, $\alpha = 0.05$, M=12) for MMN and Standard Oddball. Numbers in square brackets denote results for the Monte Carlo test. Numbers are shown for two references: CAR and channel G01.

	Standard Oddball					Mismatch Negativity				
Session	Run	Ch G01 Ref		CAR		Run	Ch G01 Ref		CAR	
S57	R01	0	[.7±1.6]	0	[.5±1]	R04	0	[.6±1.9]	0	[.3±1]
	R06	0	[.8±2]	1	[.4±.9]					
S58	R01	0	[.4±1.5]	1	[.3±1]					
	R02	0	[.1±.4]	0	[.1±.5]					
	R03	0	[0±0]	0	[.0±.2]	R04	0	[0±0.1]	0	[.1±.3]

Table 6.4.: Patient Pa_3. Number of electrodes showing a significant difference of means (t-test, $\alpha = 0.05$, M=12) between congruent and incongruent word pairs (Priming) and between attended semantic class and four other semantic classes (Semantic Oddball). Numbers in square brackets denote results for the Monte Carlo test.

	Semantic Oddball					Priming				
Session	Run	Ch G01 Ref		CAR		Run	Ch G01 Ref		CAR	
S57	R02	9	[.7±1.6]	8	[.7±1.3]	R03	0	[1.1±2.5]	0	[1±1.9]
						R05	0	[.7±1.9]	0	[.6±1.5]

Table 6.5.: Patient Pb_3. Number of electrodes showing a significant difference of means between standards and deviants (t-test, $\alpha = 0.05$, M=12) for MMN and Standard Oddball. Numbers in square brackets denote results for the Monte Carlo test.

	Standard Oddball					Mismatch Negativity				
Session	Run	Ch 44 Ref		CAR		Run	Ch 44 Ref		CAR	
S01						R01	0	[.7±3.4]	0	[.3±.8]
	R08	16	[.1±.5]	6	[.2±.7]	R06	1	[.3±1.6]	8	[.2±.6]
S02	R01	2	[.4±1.7]	5	[.5±.9]					
	R05	1	[.2±.8]	9	[.6±1.6]	R09	12	[.4±1.9]	5	[.8±1.9]
S03	R01	37	[1.7±4]	28	[1.2±2.3]					
	R02	2	[1.2±2.7]	12	[1±1.6]					
	R04	30	[1±2.4]	18	[.9±1.8]					
	R05	37	[.9±2]	30	[.8±1.6]	R06	9	[.7±1.5]	5	[.9±1.9]
S04	R01	0	[.5±1.7]	0	[.6±1.5]	R02	33	[.6±2.7]	37	[.7±2.7]
	R03	0	[.5±2.3]	1	[.3±.7]					

Table 6.6.: Patient Pb_3. Number of electrodes showing a significant difference of means (t-test, $\alpha = 0.05$, M=12) between congruent and incongruent word pairs (Priming) and between attended semantic class and four other semantic classes (Semantic Oddball). Numbers in square brackets denote results for the Monte Carlo test. Numbers are shown for two references: CAR and channel 44.

	Semantic Oddball					Priming				
Session	Run	Ch 44 Ref		CAR		Run	Ch 44 Ref		CAR	
S01	R02	0	[.3±2.6]	0	[.3±.9]	R04	9	[.2±.8]	0	[.1±.7]
	R07	0	[.3±1.5]	1	[.2±.8]	R09	0	[.2±.9]	0	[.2±.7]
S02	R02	0	[.6±1.8]	1	[.8±1.3]	R03	0	[.2±.9]	0	[.4±1.1]
	R06	4	[1±2.8]	0	[.7±1.1]	R07	0	[.2±1]	0	[.3±1]
S03	R03	2	[1.5±3.2]	0	[1.5±2.4]					
S04	R04	0	[1.3±3.8]	0	[.8±1.7]					

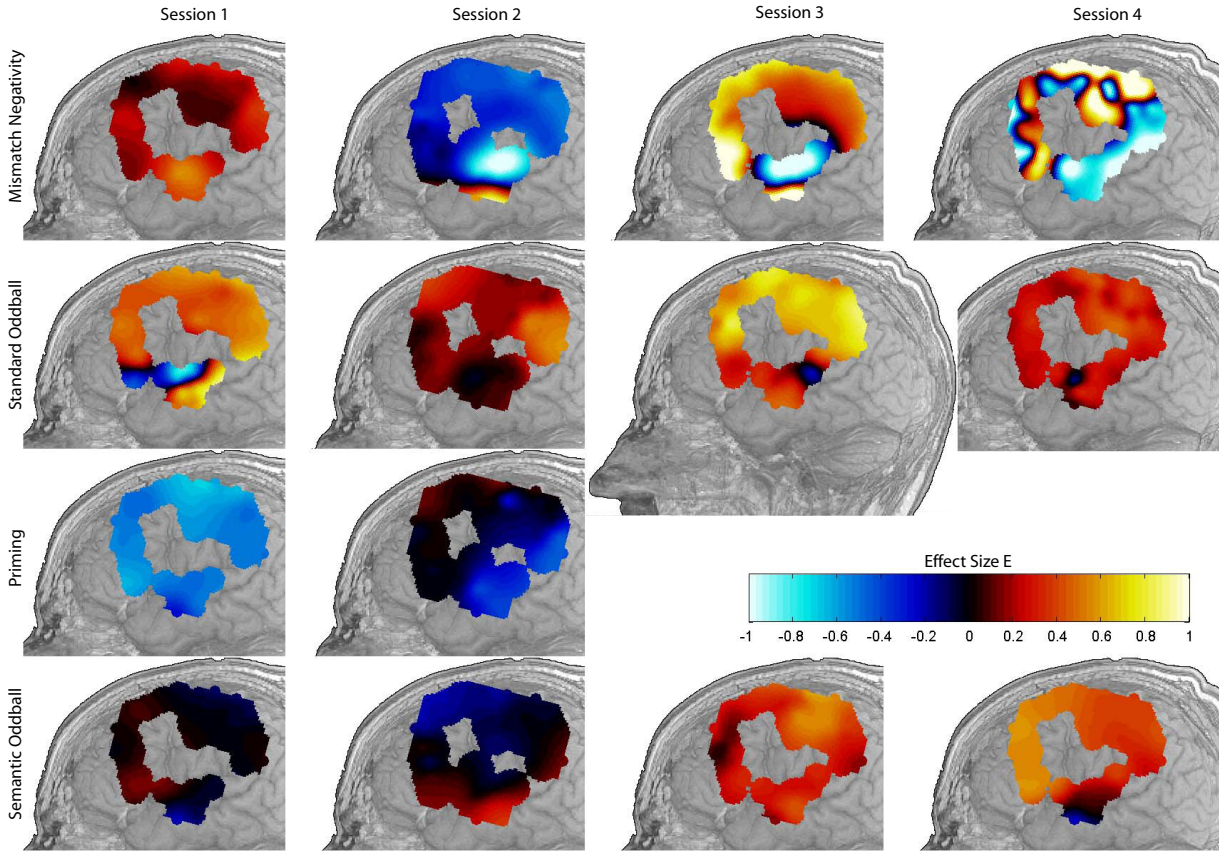
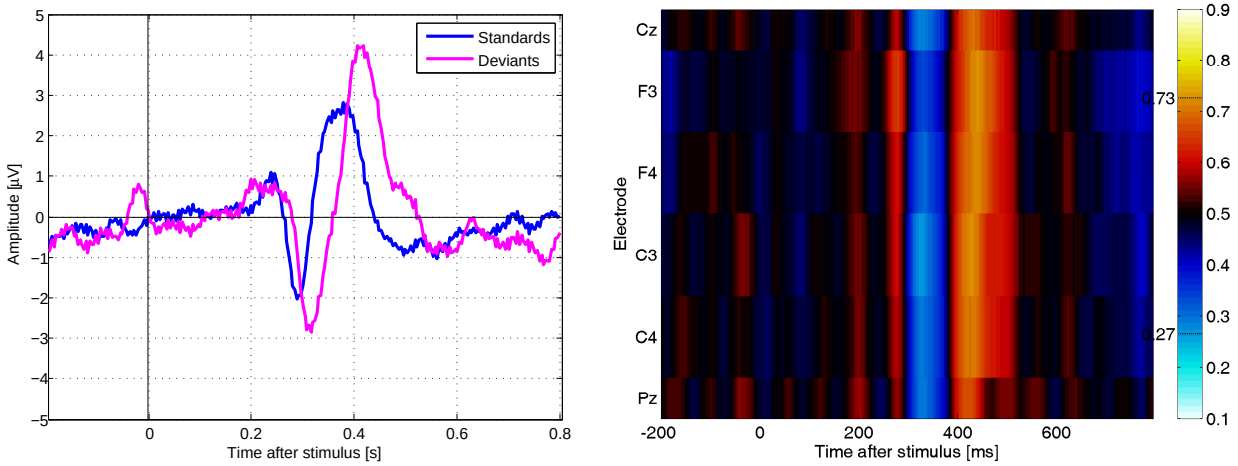


Figure 6.10.: Patient Pb_3. Effect sizes for Mismatch Negativity (top row), Standard Oddball (second row), Priming (third row) and Semantic Oddball (fourth row). Each subfigure represents the mean result for one session (one session per column). There were no Priming measurements in sessions 3 and 4. Bright regions depict high effect size (positive and negative), whereas red (positive) and blue (negative) depict lower effect size. The signal was rereferenced to channel 44. One subfigure shows the full head to give the reader a better impression of the grid location.



(a) Time series of electrode Cz.

(b) ROC plot.

Figure 6.11.: Patient Pa_3. Time series and corresponding ROC plot of Standard Oddball EEG recordings, sessions 1 and 3.

6.2.2. Latency of Evoked Response Potentials

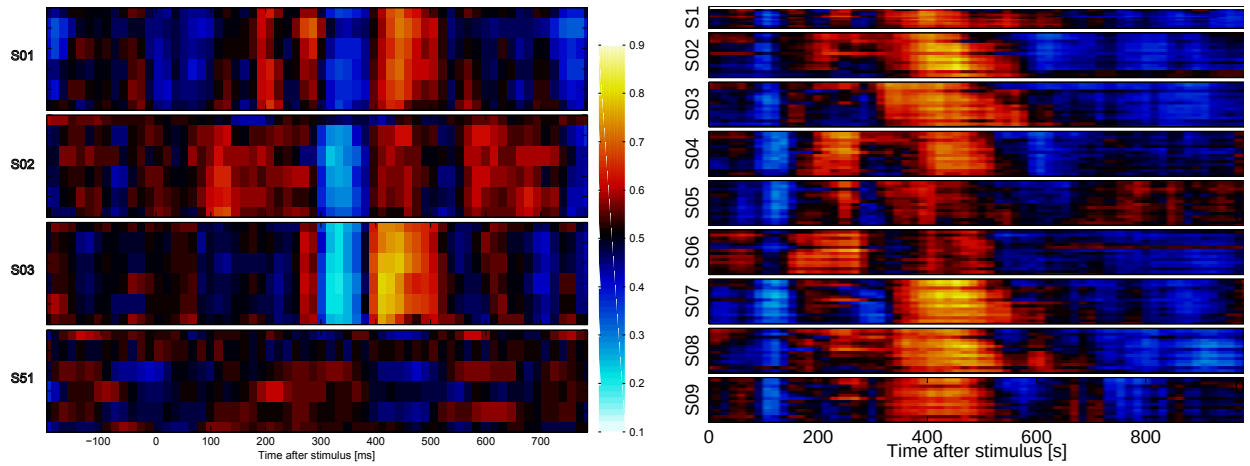
The information gathered on Pa_3's auditory oddball response is shown in Figure 6.11. It is probable that the ERP is a combination of P3a and P3b components, as a positive peak is visible in both standards and deviants. The P3b, which displays conscious recognition of the deviant tone, is evident as the peak is one µV higher and the peak latency is at 400 ms. Figure 6.11b highlights the discriminable time window as 400–500 ms — the effect weakens on the parietal electrode.

The remaining tests such as Priming, Pseudowords and Sentence Endings did not exhibit any significant effects (by visual analysis of time series and ROC plots, results not shown for brevity). As three of the cognition tests took place before the ECoG grid implantation, yet after onset of CLIS, one can speculate that the negative test results are a symptom of the CLIS rather than the invasive operation.

An inspection of Pb_3's response to auditory oddballs in the EEG, which had been recorded five months prior to the first ECoG examination, revealed no change in latency of either N1/P2 or P3 components. The N1 latency is 125 ms and the P2 latency is 250 ms. Peak latency of the P3 is between 500 ms (central electrodes) and 600 ms (parietal and occipital electrodes). The AUC scores for a number of EEG sessions are shown in Figure 6.12. From session 4 onwards, the N1 component is delayed slightly, but the P3 component's latency remains constant.

The peak latencies of components for the four ECoG paradigms are displayed in Figure 6.13 (channels 6, 7, 30, 36, 41, 45–47, 49, 50, 52). Most peaks are found in the 500–700 ms range. The N1/P2 latency in MMN and Standard Oddball stays constant over time. The significant differences between normals and deviants increase over time for MMN, while for the Standard Oddball they are visible best in session 3. No N1/P2 component is visible in the semantic tasks. The only session showing a significant effect is session 1 during the Priming task.

The same result is shown in Figure 6.14, yet sorted by channels. Of the displayed channels, those on the temporal lobe (45–47, 49, 50, 52) differ in many respects: they display an N1/P2 component of inverted polarity, probably due to their position inferior to the auditory cortex, for MMN. In the Standard Oddball paradigm, that group of electrodes displays a higher latency than the electrodes superior to the auditory cortex. In the Priming Test, they form a cluster of peaks at 340 ms and in the Semantic Oddball a cluster forms at



(a) Patient Pa_3. Four Standard Oddball sessions are shown. Sessions 01–03 are pre-implantation, session 51 is post-implantation.

(b) Patient Pb_3. Nine Standard Oddball sessions are shown, all pre-implantation.

Figure 6.12.: ROC plots of discriminability between standard and deviant tones in EEG. The ordinate is subdivided into sessions in chronological order from top to bottom, with each row representing one EEG channel. Time after stimulus is shown on the abscissa.

600 ms. Even though the effect was too weak to reject the null hypothesis, the latencies correspond roughly with the expected latencies of 400 ms and 600 ms, respectively (see Table 4.7). Most electrodes on the caudal side of the grid (30,36,41 in the chosen subset) reveal a delayed P3 (500–900 ms after stimulus).

In view of these latencies and the EEG latencies (Figure 6.12), one can conclude that no increase of N1/P2 latency took place within the eleven-month observation period.

The previous plots and statistical results exhibit hard to explain inter-session differences within the same paradigm. To promote a better understanding of these differences, a single electrode (41) in the caudal corner of the ECoG grid was chosen to visualise the averaged time series (see Figure 6.15). The N1/P2 component is present in the tonal paradigms, indicating an intact auditory system. The MMN recordings show a positive deflection at 400–600 ms, very weak in session 1 and inverted in session 4. The P3 component is at 500–1000 ms, but only visible in sessions 1 and 3 here. In session 2, the P3 component is visible on other channels (not shown) and a less stringent t-test (M=7) does find a significant difference at 500–600 ms on electrode 41. Amongst the semantic tests, only Priming displayed an effect in session 1. Session 4 exhibits generally increased signal amplitude.

6.2.3. Spectral Analysis

Spectral power of five frequency bands is shown for three electrodes positioned diagonally across the ECoG grid from a medial frontal to a lateral parietal position. For Pa_3, two groups of bars per subplot are shown (Figure 6.16) representing the two ECoG recordings. For Pb_3, four groups of bars are shown (Figure 6.17). Spectral power is displayed relative to the total power, which is represented by the subplot on the bottom right.

For Pa_3, the only obvious difference is decreased delta and increased theta PSD in session 58. For Pb_3, delta power increases monotonically from session to session, with theta, alpha and beta bands showing the opposite trend. Interestingly, there is a sharp drop in gamma power from sessions 3 to 4.

Both ECoG grids exhibit higher overall PSD for the outer grid electrodes than for the inner electrode. The relationship is inverted for frequencies above 13 Hz.

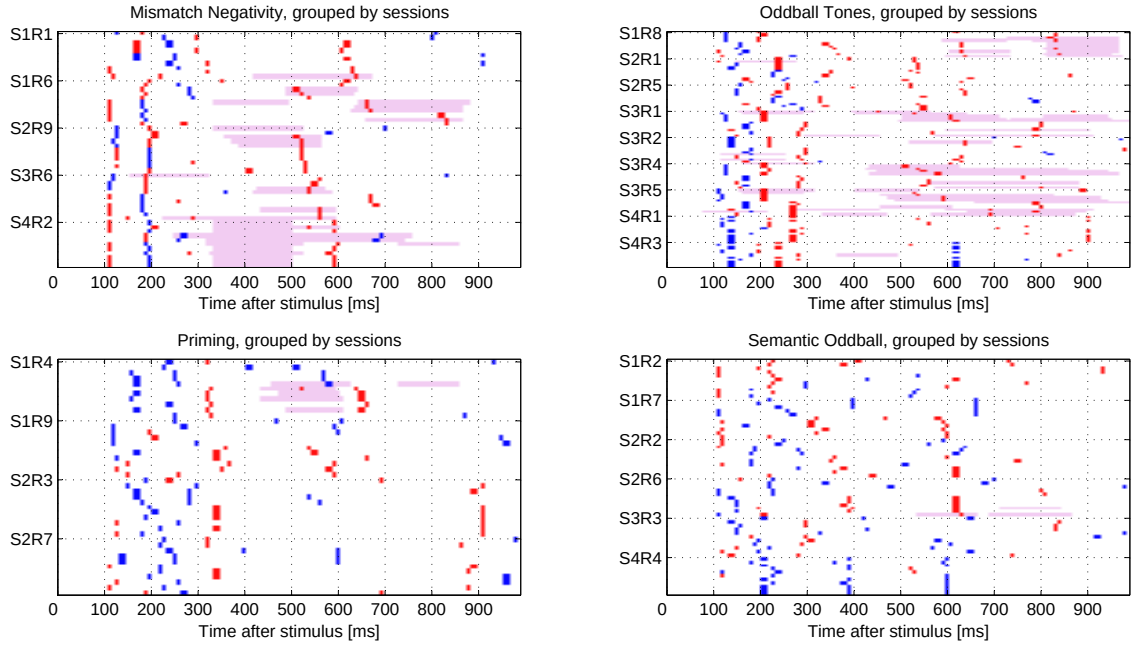


Figure 6.13.: Patient Pb_3. Peaks and significant time windows are shown for MMN, Standard Oddball (top row) and semantic tests (bottom row). Each horizontal row represents a recording run, with negative (blue dots) and positive (red dots) peaks represented along the abscissa. SsRr on the ordinate denotes Session s and Run r, below which one electrode is plotted per row. Significant differences (t-test, α=0.05, M=12) between standard and deviant tones are shaded pink. The signal was rereferenced to channel 44.

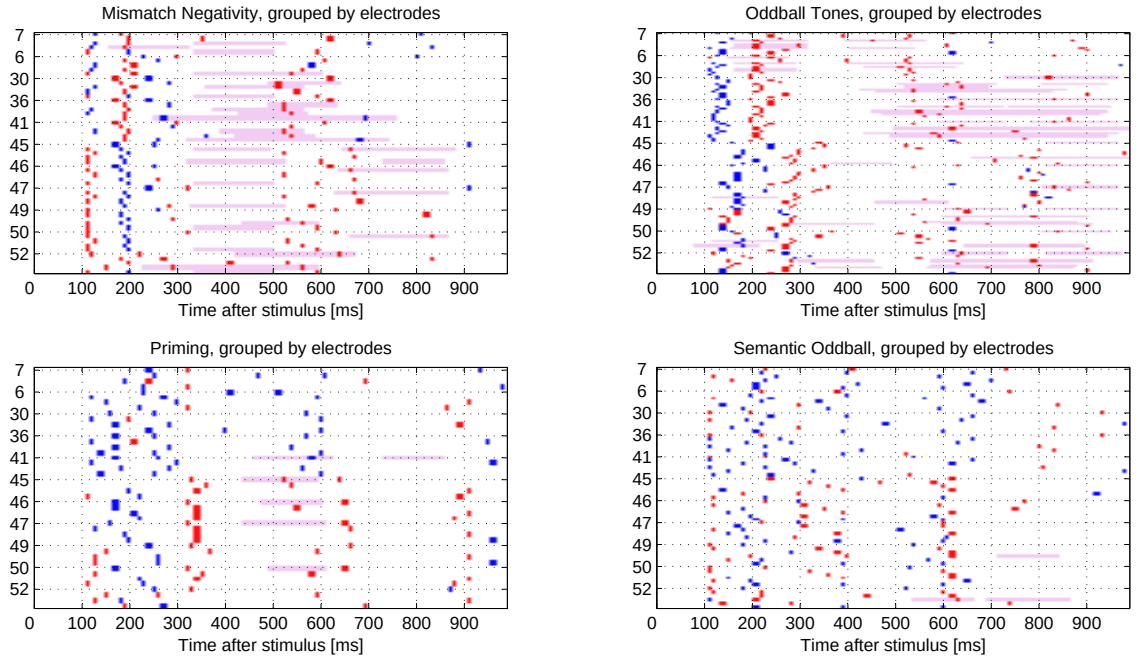


Figure 6.14.: As above, yet grouped by electrodes. Below each electrode (displayed on the ordinate) one recording run is plotted per row.

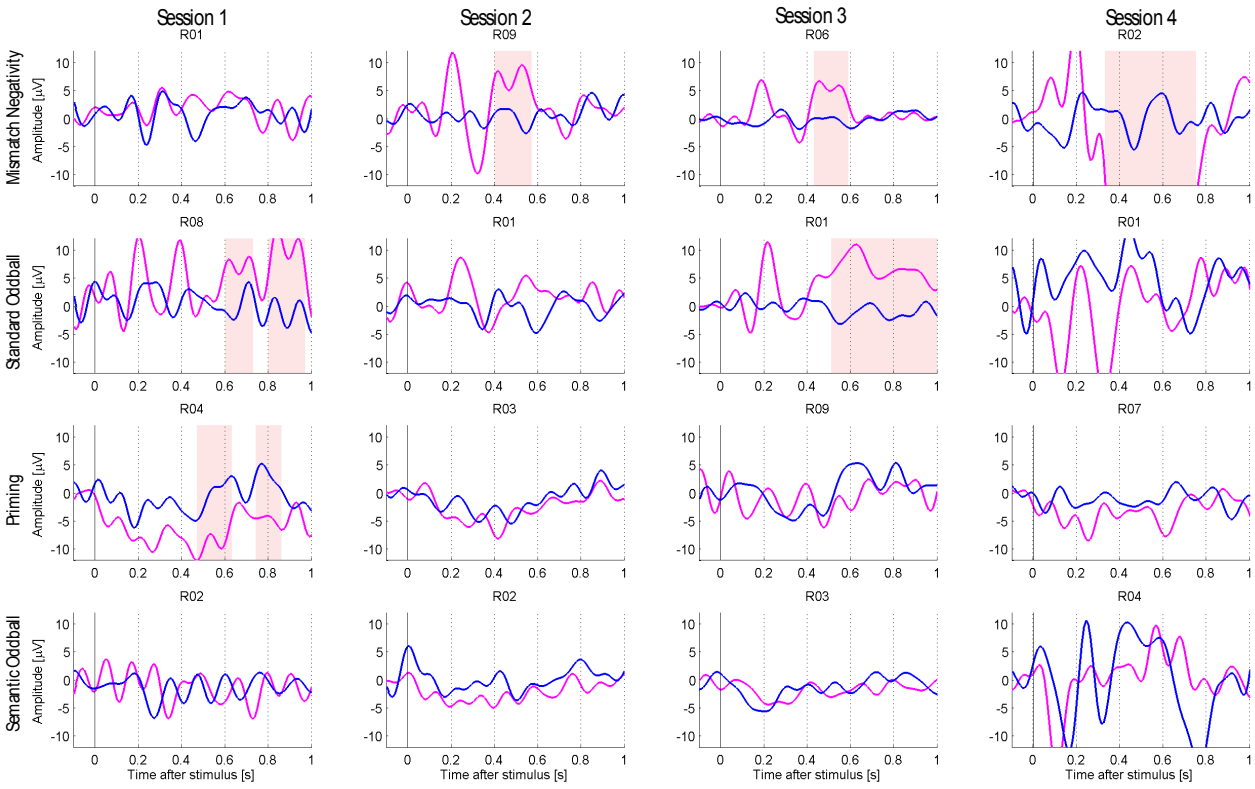


Figure 6.15.: Patient Pb_3. Mean time series for standards (blue line) and deviants (magenta line) of electrode 41 for MMN (top row), Standard Oddball (second row). Priming test (third row) shows second word of matching word pair (blue) and non-matching pair (magenta). Semantic Oddball (fourth row) shows the attended semantic group in magenta. Each column represents one recording session. Pink shaded background represents positive t-tests ($\alpha = 0.05$, M=12). M=12 indicates that the test is only regarded as significant if 12 successive samples test positive.

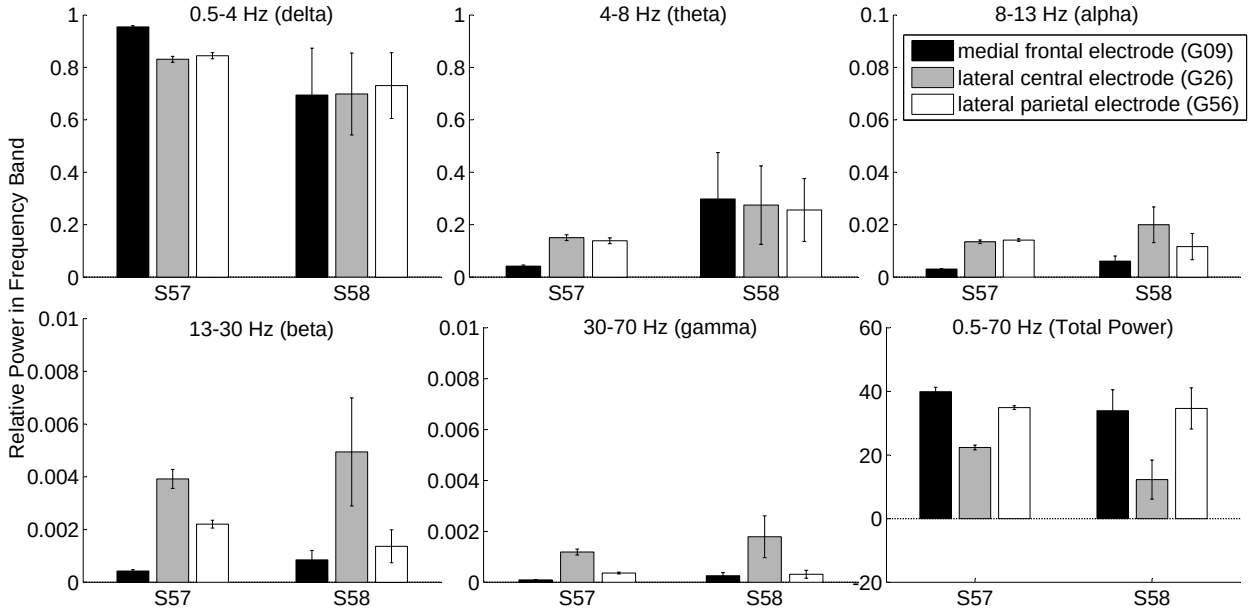


Figure 6.16.: Patient Pa_3. Relative spectral power estimate. Each subplot shows two bar groups for recording sessions 57–58. Note the changed scales on the ordinate for alpha, beta and gamma bands.

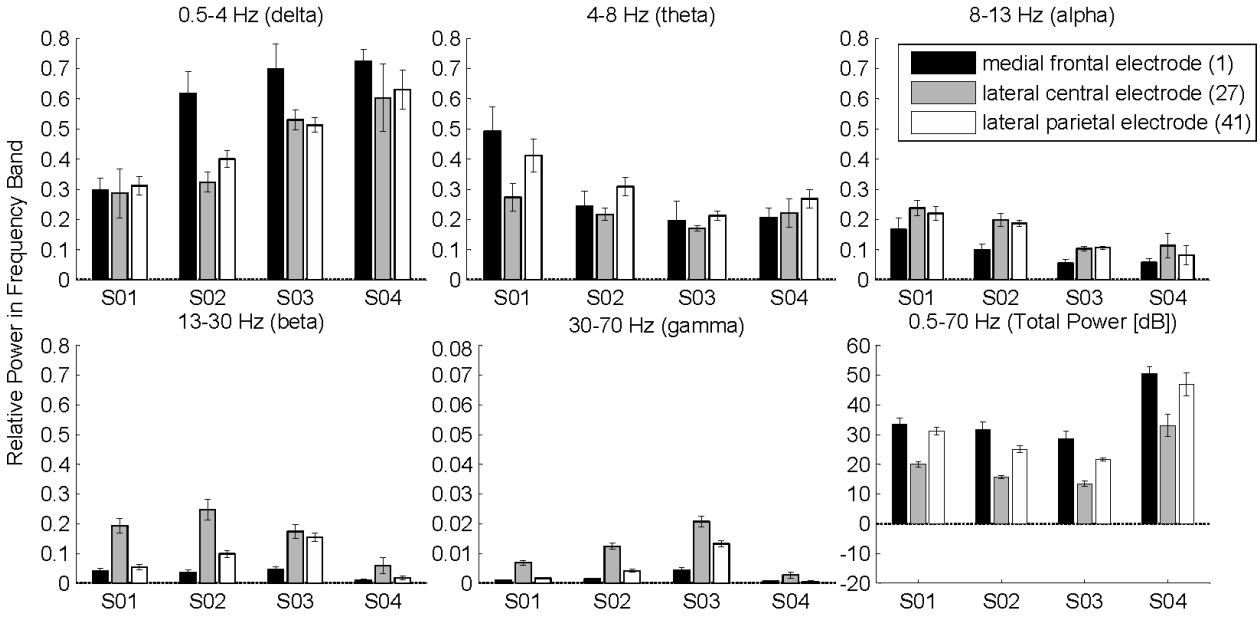


Figure 6.17.: Patient Pb_3. Relative spectral power estimate. Each subplot shows four bar groups for recording sessions 1–4. Note the changed scale on the ordinate for the gamma band.

Table 6.7.: Offline results for prerecorded data of six ALS patients (Group P1). r_C is specificity, r_E is sensitivity. b_{ECS} is the offline estimate of bit rate increase using the ECS. Bit rate decrease is indicated as negative bit rate increase in column 6.

Patient	r_C	r_E	acc_{ErrP}	Est. b_{ECS}	Est. Δ
Pa_4	0.96	0.21	0.87	3.83	-0.02
Pb_4	0.70	0.75	0.73	1.26	1.06
Pc_4	0.92	0.86	0.89	2.02	1.97
Pd_4	0.95	0.54	0.85	3.25	0.39
$Pd2_4$	0.97	0.38	0.89	3.82	0.16
$Pd3_4$	0.95	0.37	0.81	2.83	0.30
Pe_4	0.87	0.64	0.79	2.25	0.72
$Pe2_4$	0.89	0.26	0.71	2.21	-0.01
$Pe3_4$	0.87	0.33	0.70	1.96	0.09
Pf_4	0.88	0.38	0.70	1.84	0.25
$Pf2_4$	0.95	0.13	0.82	3.43	-0.11
Mean	0.90	0.44	0.80	2.61	0.44

6.3. Error-Related Potentials

This chapter presents for the first time a study in which healthy subjects as well as patients were shown to benefit from the effects of an error correction system (ECS) based upon error-related potentials (ErrPs). Prior to the study, data of 6 amyotrophic lateral sclerosis (ALS) patients was evaluated (ALS Group P1). Following that, 18 healthy subjects and 7 ALS patients participated in the newly designed study. The healthy subjects can be divided into young (Group H1) and older (Group H2), whereby the idea was to have Group H2 age-matched to the patients (ALS Group P2). It is uncommon for BCI studies to be undertaken with older people, but this was seen as a requirement here to investigate the effect of age on results. Details about this are kept for the discussion.

When studying the results, it is helpful to know that if and only if TP>FP holds, then the ECS increases bit rate. Tables of results will thus contain TP and FP along with bit rate information.

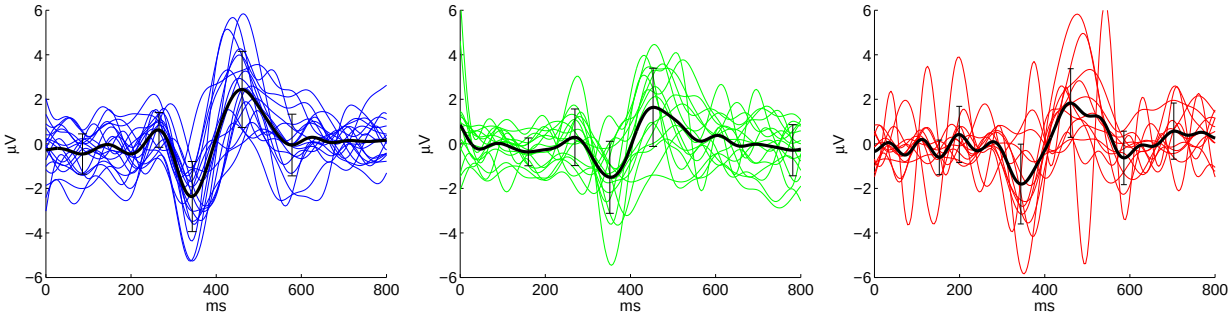
6.3.1. Offline Results

As mentioned above, offline CV estimates for the existing patient data were calculated to estimate whether the current study would stand any chance of success. Those results are presented in the next section, followed by results from the current study.

6.3.1.1. Prerecorded Data

The prerecorded data is taken from P3 speller recordings where no explicit instruction had been given to watch the feedback letter as it appeared (see Section 5.1). The results are shown in Table 6.7. Parameters used to obtain these results are: SVM (RBF kernel) with C=100, 100–800 ms CI after stimulus, 3–16 Hz bandpass filter (Fast Fourier Transform (FFT), removal of unwanted frequency bands, followed by inverse Fast Fourier Transform (iFFT)).

The CV estimate of bit rate increase shown in Table 6.7 served as preliminary evidence to show that implementing an online ECS for patients could be advantageous. Thus, the online study was conceived. The



(a) Group H1 (b) Group H2 (c) ALS Group P2

Figure 6.18.: Morphology of the ErrP found in each group. The ErrP is the hit-minus-miss difference potential. The signal is taken from the Cz electrode, lowpass filtered below 12 Hz and detrended.

first round of measurements was performed with healthy subjects. Patients followed after fine-tuning of the online classification. Study results are separated into two areas: offline analysis and introspection of the ErrP found in the recordings (next section), and online results (Section 6.3.2).

6.3.1.2. Current Study

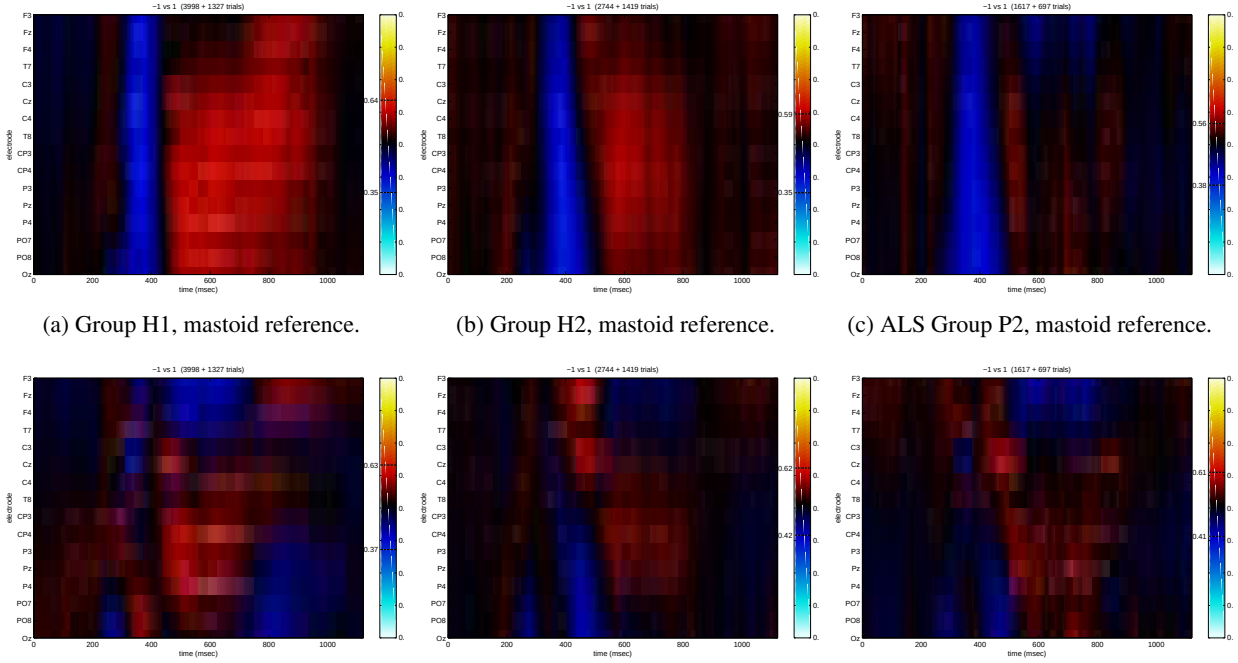
The analysis of the error-related potential signal serves to better understand the effect and improve future generations of error correction systems. More importantly for the current study, it is considered a necessary step to ensure that the correct potential (instead of an artefact) is classified. An overview in the time domain is shown in Figure 6.18. On average, the negativity appears at $t = 400$ ms followed by positivity at $t = 500$ ms. The timing of the ErrP shows great similarity between the three subject groups.

The ROC values describing discriminability between erroneous trials and correct trials are shown in Figure 6.19. The highest absolute AUC values were 0.65, 0.65, and 0.62 for the three groups, whereas the 99 % confidence intervals supporting the null-hypothesis that the ErrP does not lead to a discriminable signal were 0.552, 0.553, and 0.574, respectively (mastoid reference). As before, these plots can be used to check for electrooculogram (EOG) artefacts. The frontal channels in all groups do not show elevated AUC values and thus did not overly contribute to discriminability. Interestingly, the blue band around $t = 400$ ms (slightly earlier in Group H1, i.e., the younger subjects) is visible for all subject groups, but the ensuing broad band of positivity diminishes in length from Group H1 through to the ALS Group. In the online test, it will become evident whether classification of the ErrP is indeed harder to accomplish in patients.

As described in Section 5.4, the effect of artefacts was not only analysed visually but also by classification of simulated test sessions with (1) artefact correction and (2) use of the frontal and EOG channels exclusively. This could only be done for the healthy participants as it was too problematic to record EOG in most patients. In case (1), the ECS bit rate did not change significantly (paired t-test, $p = 0.177$) but in case (2), the ECS bit rate worsened significantly (paired, two-tailed t-test, $p = 1.47 * 10^{-5}$). This evidence suggests that the ECS does not rely on ocular artefacts.

To better estimate the source of the ErrP, the amplitude of its three distinctive peaks is plotted topographically (see Figure 6.20). The global scale of the figure helps illustrate that the largest negative peak appears in Group H1 (Figure 6.20b).

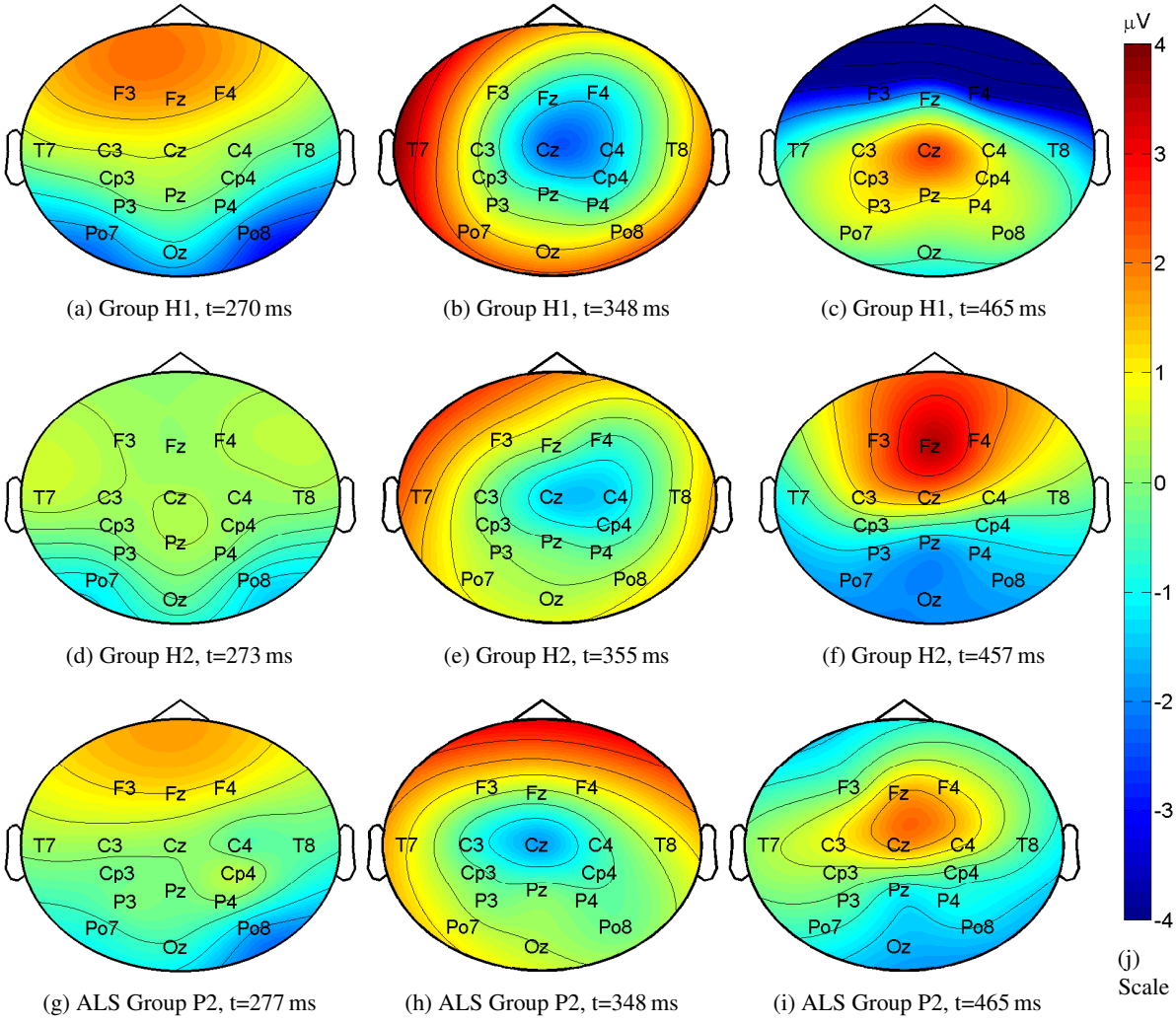
The source localisation method low resolution electromagnetic tomography (LORETA) [100] was chosen to localise the electrical source of the difference potential measured by the scalp electrodes. All data from healthy subjects was pooled and aligned to the stimulus, i.e., the timepoint when the feedback letter appeared.



(a) Group H1, mastoid reference. (b) Group H2, mastoid reference. (c) ALS Group P2, mastoid reference.

(d) Group H1, CAR reference. (e) Group H2, CAR reference. (f) ALS Group P2, CAR reference.

Figure 6.19.: AUC values representing discriminability between erroneous and correct trials. The abscissa represents time [ms] from stimulus presentation, the ordinate depicts the 16 electrodes. Frontal electrodes are located at the top while occipital electrodes are placed at the bottom. No filtering or eye movement correction was undertaken.



(a) Group H1, t=270 ms (b) Group H1, t=348 ms (c) Group H1, t=465 ms

(d) Group H2, t=273 ms (e) Group H2, t=355 ms (f) Group H2, t=457 ms

(g) ALS Group P2, t=277 ms (h) ALS Group P2, t=348 ms (i) ALS Group P2, t=465 ms

Figure 6.20.: Amplitude of the miss-hit difference (μV) at three distinctive peaks of the difference potential at electrode Cz. Each row represents one group of users.

The localisation was done on this average. The result is shown in Figure 6.21. The first peak was localised in the middle occipital gyrus (BA 19, MNI coordinates (40,-90,5)), indicating visual processing. The second peak (b) was localised in the postcentral gyrus (BA 5, MNI coordinates (5,-50,70)). The time points in between showed a gradual move of the activity from the occipital region to the parietal lobe. The third and strongest peak in the sense of global field potential is located in the superior frontal gyrus (BA 8, (5,40,55)). Subfigures (d) and (e) display the mean field potential over 116 ms following the third peak, as this time frame was the broadest discriminative band in the ROC plots.

Incidentally, the offline leave-one-out cross-validation (LOOCV) of session 1 data showed $acc_{ErrP} > acc_{P3}$ for most users. Even if the CV estimate is close to the actual accuracy in session 2, the inequality can be affected by an increase or decrease in acc_{P3} in session 2. To be able to check whether different acc_{P3} values skew acc_{ErrP} in the ErrP Test phase, all relevant values are shown in Table 6.8. It is worth noting that acc_{P3} varies strongly between the training and testing phase in some healthy subjects, which makes the task more difficult for the ErrP classifier. Interestingly, the patients' P3 performance remains stable.

The evaluation of mood and motivation brought to light two significant correlations: between mood and P3 amplitude ($\rho = 0.38$, $p = 0.009$, N=23) and between ErrP amplitude and P3 amplitude ($\rho = 0.331$, $p = 0.025$, N=23).

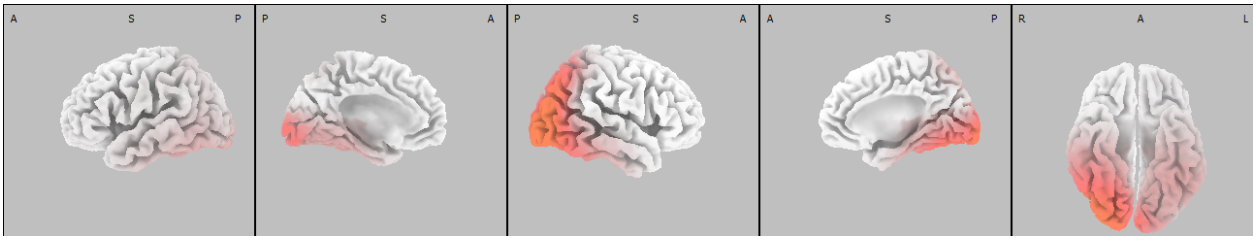
Lengthy training sessions can reduce motivation of users to work with a BCI. If possible, it is advisable to shorten or skip this phase. Some ideas to this end are presented in the following.

To estimate how quickly a new user can start using the ECS, by shortening or skipping the "ErrP Training" phase, a subject-wise CV was performed separately for each group: C_{ErrP} was trained with data of $n-1$ subjects and validated on data of the n^{th} subject. The CV projects a minor bit rate increase for healthy users, but not for patients (Figure 6.22a). But successive re-training with own data as it becomes available improves error detection.

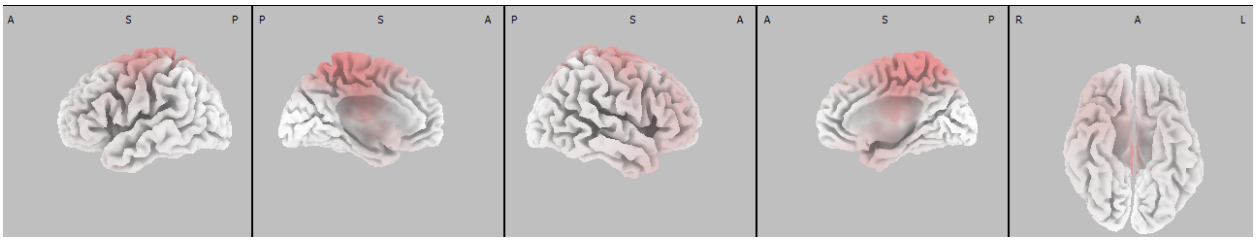
The decision whether to use foreign training data can be made under consideration of the efficacy estimate. This means that if the estimated bit rate gain is low for a user, it might be prudent to begin with a training phase and not use foreign training data. One way of estimating the bit rate gain is from P3 accuracy. Figure 6.22b clearly shows a negative correlation between acc_{P3} and bit rate increase, i.e., the worst performing users achieved the highest bit rate gains. This correlation is useful to estimate the bit rate gain for future users.

A second option to shorten the training phase is to reduce the amount of ErrP training data. This was tested with the healthy participants, as these consistently had enough training data to choose from. Reduction of the amount of ErrP trials used to train the classifier had a negative effect on acc_{ErrP}. Therefore, data containing at least 50 positives should be used if a combination with foreign training data is not an option.

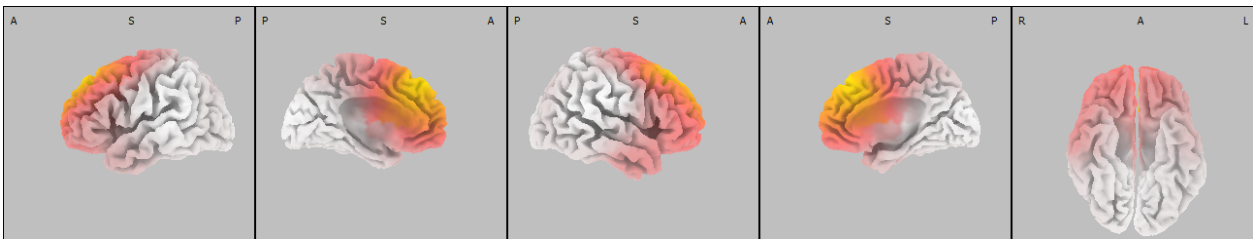
Lastly, preparation and training time depend on the number of electrodes necessary to pick up the error-related potential. The aim was to find out whether the ErrP could be recognised just as well with a subset of electrodes commonly thought to contain most information on the ErrP. Reduction of channels detrimentally affected classification accuracy acc_{ErrP} (see Figure 6.23). The montages chosen for this test were, in order from left to right in the figure:
(1) All electrodes;
(2) Fz, Cz, Pz, Oz and CP3, CP4, C3, C4 and P3, P4, PO7, PO8;
(3) Fz, Cz, Pz, Oz and CP3, CP4, C3, C4 ;
(4) Fz, Cz, Pz, Oz.



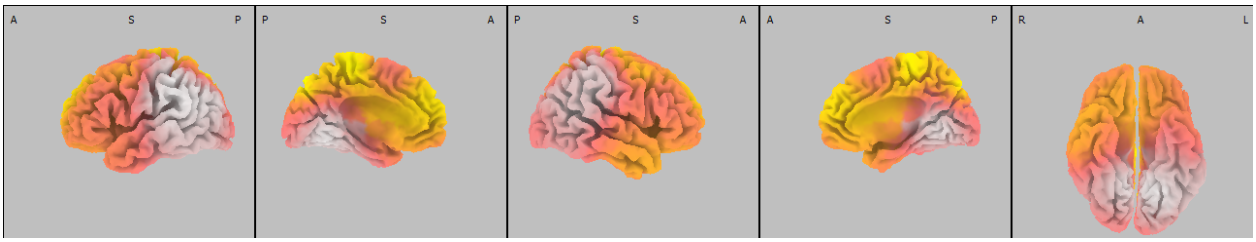
(a) Healthy subjects (N_h = 17), t = 285 ms (first peak, positive)



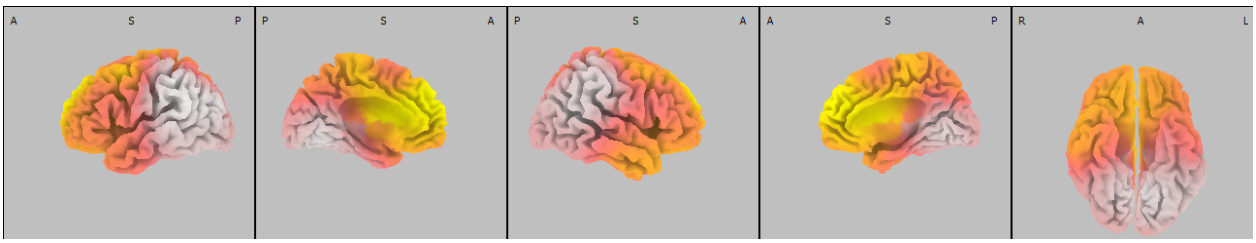
(b) Healthy subjects (N_h = 17), t = 359 ms (second peak, negative)



(c) Healthy subjects (N_h = 17), t = 484 ms (third peak, positive)



(d) Healthy subjects (N_h = 17), t = 484–600 ms

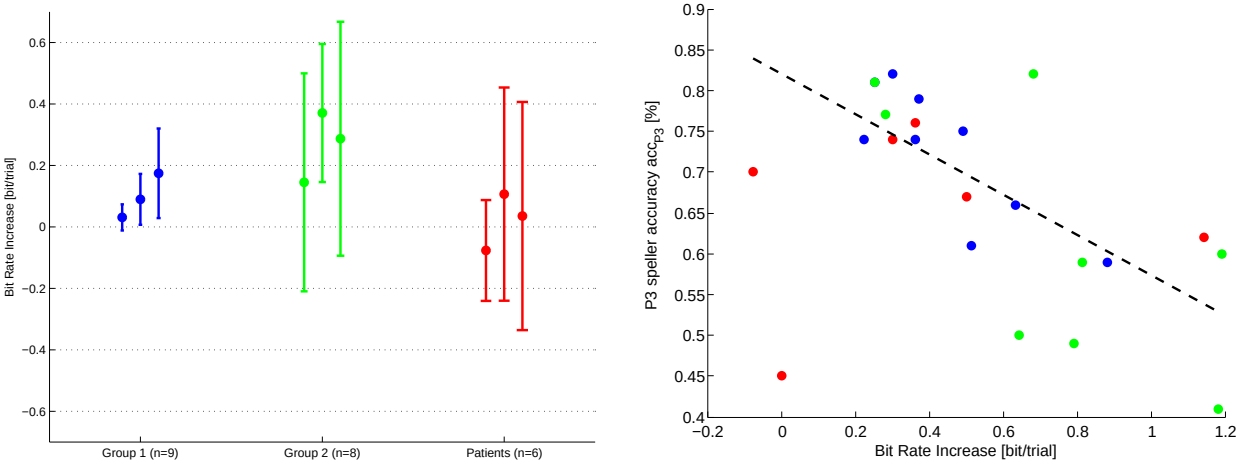


(e) All subjects (n=23), t = 484–600 ms

Figure 6.21.: Distribution of global field potential in the brain as determined by LORETA software. A global scale was used, indicating the prominence of the positive third peak in source space.

Table 6.8.: Healthy subjects' and patients' offline accuracy estimate and actual online result in relation to their baseline P3 accuracy for training and testing phases. The second recording session had to be cancelled for Hr_5 (no time) and patient Pg_5 (not enough data collected due to low signal quality caused by artefacts from ventilation equipment).

	P3 accuracy acc_{P3}		ErrP accuracy acc_{ErrP}	
User	ErrP Training	ErrP Test	CV estimate	ErrP Test
Ha_5	0.76	0.79	0.88	0.86
Hb_5	0.75	0.82	0.82	0.88
Hc_5	0.78	0.81	0.89	0.86
Hd_5	0.83	0.59	0.91	0.76
He_5	0.68	0.66	0.86	0.78
Hf_5	0.84	0.75	0.91	0.84
Hg_5	0.79	0.74	0.85	0.81
Hh_5	0.78	0.61	0.89	0.71
Hi_5	0.68	0.74	0.81	0.78
Hj_5	0.60	0.49	0.85	0.66
Hk_5	0.74	0.59	0.81	0.75
Hl_5	0.71	0.82	0.87	0.95
Hm_5	0.65	0.50	0.73	0.63
Hn_5	0.85	0.81	0.90	0.86
Ho_5	0.50	0.41	0.78	0.82
Hp_5	0.74	0.60	0.84	0.83
Hq_5	0.77	0.77	0.86	0.82
Hr_5	0.32	-	0.87	-
Pa_5	0.74	0.62	0.86	0.84
Pb_5	0.74	0.76	0.86	0.83
Pc_5	0.76	0.67	0.81	0.77
Pd_5	0.71	0.70	0.59	0.69
Pe_5	0.80	0.74	0.81	0.80
Pf_5	0.47	0.45	0.69	0.53
Pg_5	0.52	-	0.57	-



(a) Training and testing C_{ErrP} in a subject-wise cross-validation. Mean bit rate increase for each group of users. Three weight settings for w_{-1} are shown per user group: 0.15, 0.3 and 0.45 (see (5.5)).

(b) Correlation for all user groups of P3 speller accuracy with bit rate increase (Group H1 is blue, Group H2 is green, ALS Group P2 is red).

Figure 6.22.: Methods to reduce training time of the ECS (a) and correlation for all user groups of P3 speller accuracy with bit rate increase (b).

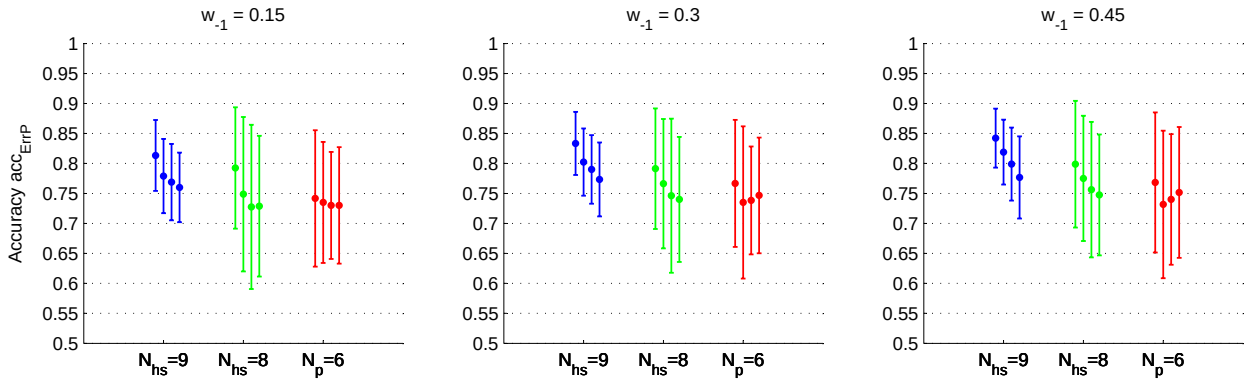


Figure 6.23.: Effect of the successive reduction of channels a priori known to be furthest away from the area where the ErrP displays its peak amplitude. Each subfigure displays a different weight w_{-1}. There are four error bars per user group, depicting mean and standard deviation for each of four electrode configurations in descending order from left to right, containing 16, 12, 8 and 4 electrodes, respectively.

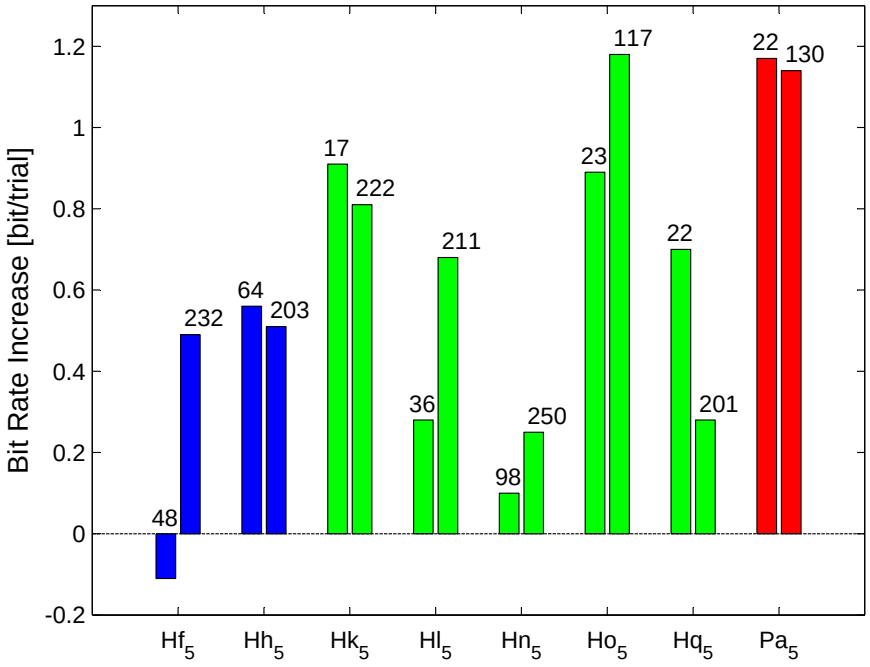


Figure 6.24.: Bit rate increase (first bar) during free spelling as compared to bit rate increase (second bar) during copy spelling for eight participants (Group H1 is blue, Group H2 is green, patient is red).

6.3.2. Online Results

The online results are shown in Table 6.9. The column b_{ECS} displays online bit rate with the ECS in operation, and the final column displays the increase that was achieved. The ALS Group P2 displayed less increase in bit rate — reasons for this are discussed later. There is only a 2 % difference in mean acc_{ErrP} between Group H1 and Group H2.

Due to an operator error, the ECS was inactive during Patient Pe_5's session. The result was achieved in an a posteriori reconstruction of the online session. As the ECS only produced one false positive (FP), it is unlikely that the patient's behaviour would have been affected significantly, had the ECS been active online. The subject would not have been distracted by incorrectly deleted letters. In other words, this can be regarded as an online result.

To determine whether belonging to a particular group has an effect on variables such as bit rate improvement, accuracy and amplitude of P3 and ErrP, an analysis of variance (ANOVA) was performed. Where possible, a repeated measures ANOVA was used to account for both recording sessions. No significant differences between the groups H1, H2 and P2 were found, indicating that the ECS worked equally well for all participants (Table 6.10).

Free spelling is a very realistic scenario, as it includes the cognitive load of thinking of what to spell during the BCI task. It was possible to test free spelling with eight users (amongst them only one patient). The bit rate increase achieved by free spelling is shown in Figure 6.24. Due to the low number of participants and spelled words, the significance of this result cannot be evaluated. It leads us to believe, however, that bit rate improvement is possible on the same level with free spelling as with copy spelling.

Table 6.9.: Online results for healthy subjects (top two groups) and patients (bottom group), with group means. r_C is specificity, r_E is sensitivity. b_{ECS} is b_S with the ECS active, Δ is the absolute increase or decrease in b_S compared to without ECS.

	TN	FP	TP	FN	r_C	r_E	b_{ECS}	Δ	acc_{P3}	acc_{ErrP}
Ha_5	252	10	34	36	0.96	0.49	3.34	**0.37**	0.79	0.86
Hb_5	179	1	14	26	0.99	0.35	3.57	**0.30**	0.82	0.88
Hc_5	215	0	13	38	1.00	0.25	3.41	**0.25**	0.81	0.86
Hd_5	132	5	45	52	0.96	0.46	1.75	**0.88**	0.59	0.76
He_5	185	50	94	28	0.79	0.77	2.26	**0.63**	0.66	0.78
Hf_5	170	3	25	34	0.98	0.42	3.01	**0.49**	0.75	0.84
Hg_5	284	0	27	73	1.00	0.27	2.82	**0.36**	0.74	0.81
Hh_5	122	2	22	57	0.98	0.28	1.64	**0.51**	0.61	0.71
Hi_5	131	9	17	33	0.94	0.34	2.65	**0.22**	0.74	0.78
Mean					.96 $\pm$ .1	.40 $\pm$ .2	2.72 $\pm$ .7	**.44** $\pm$ **.2**	.72 $\pm$ .1	.81 $\pm$ .1
Hj_5	108	11	51	71	0.91	0.42	0.79	**0.79**	0.49	0.66
Hk_5	128	4	39	51	0.97	0.43	1.78	**0.81**	0.59	0.75
Hl_5	170	2	30	9	0.99	0.77	3.91	**0.68**	0.82	0.95
Hm_5	127	31	70	86	0.80	0.45	0.67	**0.64**	0.50	0.63
Hn_5	202	0	12	36	1.00	0.25	3.41	**0.25**	0.81	0.86
Ho_5	43	5	53	16	0.90	0.77	1.18	**1.18**	0.41	0.82
Hp_5	116	2	48	32	0.98	0.60	2.18	**1.19**	0.60	0.83
Hq_5	148	6	17	30	0.96	0.36	3.01	**0.28**	0.77	0.82
Mean					.94 $\pm$ .1	.51 $\pm$ .2	2.12 $\pm$ 1.2	**.73** $\pm$ **.4**	.62 $\pm$ .2	.79 $\pm$ .1
Pa_5	71	9	38	12	0.89	0.76	2.33	**1.14**	0.62	0.84
Pb_5	98	0	9	22	1.00	0.29	3.02	**0.36**	0.76	0.83
Pc_5	135	3	23	44	0.98	0.34	2.28	**0.50**	0.67	0.77
Pd_5	90	3	1	38	0.97	0.03	2.02	**-0.08**	0.70	0.69
Pe_5	87	1	8	23	0.99	0.26	2.76	**0.3**	0.74	0.80
Pf_5	29	15	22	31	0.66	0.42	-0.11	**-0.11**	0.45	0.53
Mean					.91 $\pm$ .1	.35 $\pm$ .2	2.05 $\pm$ 1.1	**.35** $\pm$ **.5**	.7 $\pm$ .1	.74 $\pm$ .1

Table 6.10.: Top row: One-way ANOVA with p-values describing significant differences between Groups H1, H2 and P2. Bottom rows: Repeated ANOVA with p-values describing significant with-in group differences between sessions and between-group differences.

	Δb_S	P3		ErrP		
		Acc	Ampl	Acc	Ampl	Lat
p	0.11	0.25	0.18	0.42	0.81	0.37
Sessions p	-	**0.01**	0.09	-	0.069	0.97
Groups p	-	0.21	0.54	-	0.79	0.93

7. Summary

The previous chapter's results are discussed here. Thereafter, the thesis is evaluated. To conclude, an outlook on some promising topics for future work is given.

7.1. Discussion

The offline results propose that amongst the synchronisation features tested, PLV could become a valuable addition to MEG BCIs. Cross-validation results suggest that some subjects would benefit from using the PLV+AR combination instead of AR alone (Table 6.1).

7.1.1. Multiclass Experiments

The error estimates using feature selection are higher than without feature selection in the two-class as well as the three-class setting (Table 6.2). However, feature selection is needed for the online BCI to reduce computational load, especially if combinations of various feature types are to be used. This is the reason for focussing on the result with reduced features. Overall, the result for ternary classification is significantly better than that for binary classification (paired t-test, $p = 0.002$).

The task combinations shown in Table 6.2 are the best on a per-subject basis. It is interesting to see that for most subjects, a motor task combined with the two nonmotor tasks *subtraction* and *navigation* is the best combination (even though subjects reported difficulties with the nonmotor mental tasks (Table 4.3)). This might have to do with the fact that cortical activation for these tasks is spatially disjunct. The best two-class pairs are *subtraction* together with a motor task in eight of ten subjects. This is in contrast to [29], where the most discriminable pair consists of two nonmotor tasks. Reasons for the difference could be (1) the *auditory* task was not included in the study presented here; (2) functional motor tasks were allowed here (subjects Ha_2, Hb_2, Hf_2, Hh_2, Hj_2 used functional motor tasks for hand imagery). From this result, it stands to reason that some class combinations generally lead to higher discriminability, such as the combination of a motor with a nonmotor task.

The analysis of artefacts shows that, with the exception of subject Hd_2, the electromyogram (EMG) hand movement artefacts are evenly distributed across all tasks. This implies that hand movements do not unfairly bias the classifier during the offline analysis. However, the possiblity of an influence of eye or head movement artefacts on the result cannot be ruled out. Furthermore, there is no guarantee that the subjects did not move their tongue or feet during the recording.

As the multiclass study progressed, it became increasingly evident that the process of measuring patients in the MEG would become too time-consuming and costly (the complexity of one such measurement, which took place at a later stage, is described in Section 4.2.5.1). Given that patients could reach similar two-class accuracy levels as the healthy subjects, and that the additional cognitive load required to choose from three tasks does not overwhelm patients, it is reasonable to assume that patients can also increase their ITR using three tasks. However, both points are problematic, as patients' results are generally inferior, and cognitive deterioration is disputed (Section 2.3.2.1).

A comparison with the state-of-the-art result in EEG in [34] (Table 3.2, page 26, N_h =8, 0.66±0.26 b_W/trial, range = 0.32–1.14) shows that these results in MEG (N_h =10, 0.5±0.25 b_W/trial, range = 0.09–0.95) do not reach the state-of-the-art. However, one should appreciate that the direct comparison with the study in [34] is not entirely fair. Firstly, data from three different recording paradigms with three, four and six classes

was used, and secondly, results for two subjects were not given. Without feature selection, the MEG result presented here is 0.6±0.29 b_W/trial, and the omission of the worst two performers further boosts this result to 0.69±0.22, range 0.36–1.07. Additionally, due to the exploratory nature of this study, trial length is 8.8 s as compared to 4.5 s, resulting in a lower value of b_W/min.

A comparison with existing three-class BCIs based on motor imagery alone should be undertaken to see whether there is an advantage to using nonmotor tasks combined with motor tasks, as suggested by the results presented here. It was found that it is possible to select a subset from a larger group of mental tasks individually to obtain a higher bit rate for each BCI user. In the offline analysis, eight subjects benefit by using a three-task combination instead of binary decision tasks (Table 6.2). The relative improvement in ITR for these eight subjects is 32 %. This is a motivation to continue the online measurements where the subjects spell a word by selecting letters from a ternary decision tree. The use of four or five classes does not increase bit rate further, which is in line with previous studies.

The MEG measurements with patient Pb_3 bring to light a willful μ-rhythm modulation in the left motor cortex during the *navigation* and *foot* tasks. For the *foot* task, this modulation is more lateralised than would be expected, as the part of the motor cortex representing the feet lies within the interhemispheric fissure (Figure 6.5). Possibly Pb_3 was using mental imagery different to the instruction. More importantly, even though a central frontal activation is visible for the aim task, discriminability is too low to implement a single-trial communication system with ternary classification for Pb_3.

After implantation of the ECoG grid, four-class screening with different imagery tasks was performed. Singing and left hand imagery seem promising (Figure 6.6a), but discriminative power is again not sufficient for single-trial classification, which prompted another series of measurements with singing as the only task. As in the previous results, averaging over all trials shows that Pb_3 modulated his brain activity insufficiently for single-trial communication (Figure 6.8). The ECoG implants were thus unsuccessful in restoring communication in Pa_3 and Pb_3. Attempts to explain the negative outcome of the ECoG measurements with CLIS patients are discussed in the following section.

7.1.2. Cognition Detection in CLIS Patients

The cognition detection tests with Pa_3 and Pb_3 commenced when it became evident that standard BCI paradigms had been unsuccessful. At the behavioural level, it was often impossible to determine whether the patients were vigilant and able to follow instructions, vigilant and unwilling to participate, or too drowsy to cooperate during a particular BCI experiment. The aim was to determine, at a lower level, the intactness of the patient's auditory and cognitive system, because vision is progressively impeded in ALS . If it can be shown that these two requirements for BCI communication in locked-in syndrome (LIS) are affected on an hourly, daily, or progressive basis, better decisions on the type and time of suitable BCI training can be made. Additionally, the tests were seen as a measure to document the patients' transition from LIS into CLIS using chronic ECoG recordings, a procedure which has not been reported on previously.

One of the difficulties with chronic ECoG recordings in humans is the continually increasing number of noisy channels. It is plausible that the rigid grid loses contact with the cortex or dura, thus allowing 50 Hz mains interference to overpower the brain signal. Another possibility is tissue growth under the electrodes. The changes in recording reference unavoidably lead to inaccuracies when comparing sessions, despite the rereferencing that was done. The reader will also notice that no age-matched control group was available. The only possibility would have been to ask epilepsy patients with temporary ECoG grids, but this would have been too time-consuming due to far travel distances and grid locations vary considerably in epilepsy patients.

No continuous deterioration of the patients' cognitive state is evident as reflected in the tonal and semantic tests (Tables 6.3–6.6). The LIS and CLIS sessions will now be discussed separately.

For Pb_3 in LIS, task-related ERPs were detected, yet fluctuate from session to session. P3 components are most evident in sessions 1 and 3, but also present in session 2. Strikingly, the number of Mismatch Negativity components in any session is inversely proportional to the P3 ERP count (Table 6.5). The inverse relationship is also visible in the effect sizes (Figure 6.10), although the Mismatch Negativity effect is strong in session 3 also. Generally, the effect sizes support the t-test results. Mismatch Negativity is expressed especially well in the auditory cortex. The Standard Oddball response is strongest in a caudal location on the grid, whereas Mismatch Negativity is located further anterior and medial. ERPs were not found in the semantic paradigms requiring a higher level of attention and cognitive processing, except for run 4 in session 1 of the Priming test. Although this might indicate the impairment of the respective semantic processing functions, one should be cautious in jumping to conclusions: the Priming N400 response can even be lacking in some healthy individuals [32]. Furthermore, even if not statistically significant, Figure 6.13 does reveal a cluster of peaks at 600 ms during Semantic Oddball and at 350 ms during Priming. The latency and cortical area (Priming: electrodes 45–47, Semantic Oddball: electrodes 45–52, see Figure 6.13) correspond roughly with the expectation (Table 4.7), although the P600 component location is further temporal than expected.

In the CLIS sessions, there is no evidence of the input being processed at all for Pa_3. As Pa_3 had already been in CLIS for a long time, no similar data was available from the LIS phase for comparison. With Pb_3, session 4 was recorded in CLIS (more than three months after the patient's last communication with muscular assistance). No Standard Oddball responses, but many Mismatch Negativity responses were found, indicating low cognitive ability with intact auditory pathways. This evidence is no proof for cognitive deterioration due to ALS. For instance, medical complications in the weeks before the CLIS measurement and administration of antibiotics due to an acute medical condition (see Table 4.6, high blood pressure and heart rate) might just as well have lead to a low P3 count. Alternatively, the lack of P3 responses could serve as evidence that ERP components changed drastically in CLIS.

A comparison with EEG recorded five months prior to the first ECoG cognition test can only be undertaken for Pb_3, as no components were found in the tonal tests for Pa_3. No changes over time in latency of the N1/P2 component was found. Regarding the P3 component, frontal electrodes in EEG and ECoG are in agreement at 500 ms, but parietal electrodes on the ECoG grid display a higher latency than the corresponding EEG electrodes (a range of 500–800 ms as compared to 600 ms at electrode Oz in EEG).

The P3 latency is also stable: 500 ms at Pz, 600 ms at Oz in EEG, while it appears after 500 ms on frontal electrodes and in a range from 500–800 ms on parietal electrodes of the ECoG grid in this study (Figure 6.9). The latency of the N1/P2 complex seems to vary with the recording site. The low latency of the electrode strip could be due to its orientation or position on the temporal lobe. Another group of electrodes (45–47,49–50,52) displays a high latency. These channels are all in the vicinity of the auditory cortex.

For Pb_3, spectral properties vary between electrodes, but on average an upward trend of delta power and a downward trend of alpha power was found (Figure 6.17). These values are relative to the total power, which was constant over the LIS sessions but increased in the CLIS session, possibly due to medication. The high gamma PSD in session 3 could be an indicator of heightened vigilance, explaining the high effect sizes in that session. However, a causal relation between spectral power and significant ERP components is debatable and should ideally be compared with BCI results from that time. Unfortunately, the BCI results were consistently weak so that a comparison is not meaningful. Interestingly, the delta and alpha power trend is reversed for Pa_3, possibly due to the fact that this patient had been in the CLIS stage much longer than Pb_3 (Figure 6.17). In any event, it is difficult to speak of a trend given only two sessions and the relatively high error bars in session 58.

Overall, the evidence of a dramatic decline of normal brain functioning in the transition from LIS to CLIS is not compelling, indicating — at least at an electrophysiological level — partially intact cognitive functioning. Unfortunately, not much can be concluded from the measurements with Pa_3, as there is no comparison to the LIS phase. Rather than a continuous decline, the variation of the ERPs between sessions

could mean that information processing abilities in Pb_3 fluctuated on a daily basis. To conclusively determine an increase in cognitive impairment independent of the current vigilance state of the patient, it has become evident that more measurements should have been performed with both patients. Additionally, because the patients' BCI accuracies are too close to chance level, and show very little variation, there is no information to correlate the cognition detection tests with. This would have been a requirement to investigate the viability of predicting BCI results with the cognition test.

7.1.3. Error-Related Potentials

Not only do the results show that an ECS based on ErrPs resulting from P3 speller feedback significantly increases bit rate in a large group of healthy subjects, this was also found to be true for a group of six ALS patients (Table 6.9). The result can be considered credible since the test was performed online in a typical P3 BCI speller setting and it was ensured that the basis of the classification was not eye movement artefacts.

While the communication speed of all healthy subjects improved with the ECS, two patients did not benefit from error correction. The latency of the ErrP is similar between the three groups, but the amplitude is highest in the young Group H1.

The online results of the ALS Group P2 highlight that the ECS presented here is clearly effective in increasing bit rate on average, but a minimal bit rate decrease was observed in one of six patients. This comes as no surprise — the LOOCV estimate of acc_{ErrP} predicts a bit rate decrease in the case of Pd_5. In such cases the bit rate has to be monitored and the ECS switched off if the trend is confirmed. Another patient's bit rate was unchanged at 0 ($acc_{P3} < 0.5$ is problematic for Pf_5). Additionally, even though the mean accuracy in Figure 6.23 seems to indicate superior results of Group H2 over the ALS Group P2, the low mean is mainly attributable to patient Pf_5. A comparison with the *median* accuracy, which is more robust against outliers, shows Group P2 on par with Group H2.

In contrast to patients, all healthy participants benefit from the ECS during online copy spelling. The increase for participants Hj_5, Hm_5, and Ho_5 is noteworthy. Their original p_{P3} is below 50 %, which means the ECS increased their bit rate on average from 0 to 0.87. However, as with all BCI experiments, it cannot be ruled out that the ability to move the eyes works in favour of the healthy subjects. These results could be improved further on a per-patient basis by optimising certain parameters, such as the time window used for ErrP classification or the SVM hyperparameter w_{-1}. The defaults used here are recommended as a starting point for all BCI users.

Table 6.8 shows that the a priori estimate of acc_{ErrP} is reasonable, deviating by more than 10 % in only five cases (often attributed to major deterioration of acc_{P3} in the second session). If this is anything to go by, then the CV estimate of bit rate improvement for Group P1 (patients Pa_4–$Pf2_4$) of 0.44 b_S per trial (Table 6.7) predicts that significant improvement in communication speed can be achieved in patients, even with the standard P3 speller feedback[1].

Impedances differ significantly between the two sessions for two subjects from Group H1, five subjects from Group H2, and four of six patients from Group P2 (t-test, $p = 0.05$), but this has no effect on bit rate increase. This demonstrates that the recognition of ErrPs is insensitive to changes in electrode impedances.

The argument could be put forward that acc_{P3} was reduced artificially by manual reduction of the number of flashes, thus making it easier for acc_{ErrP} to exceed acc_{P3}. It is clear that the ECS allows a far higher bit rate gain for people with low p_{P3}, which makes it less interesting to test on subjects whose baseline accuracy is above 80 %. Secondly, the aim was to achieve acc_{P3} similar to typical patients' p_{P3} to find out how the system works for the intended target population. The p_{P3} of Group P2 was 66 % (in line with the finding by [68, Tab. 3]), which confirms that p_{P3} was not manipulated to an undue degree.

[1] Influence of eye movements on the old results cannot be ruled out, as no EOG was collected.

Overall, it can be stated that the ECS presented here is a significant improvement over the current state-of-the-art: merely one of the studies presented in Section 3.4 has implemented an online ECS for healthy subjects (with bit rate improvements inferior to those presented here), and none have shown their system to work with patients. A strength of the method lies in its independence on the underlying BCI paradigm. There is no reason why it should fail with any other BCI paradigm which presents feedback and it is hoped that the ECS will be picked up by other researchers for the benefit of all interested patients.

7.2. Conclusion

To conclude, the thesis "Brain connectivity analysis, nonmotor imagery and the detection of error-related potentials can increase the communication speed of patients using a BCI" must be evaluated.

The experiments with healthy subjects provide evidence that the use of phase-locking values as expression of brain connectivity improves BCI accuracy. However, the improvement did not reach statistical significance and could not be tested with patients due to time constraints and the recording modality of that experiment (MEG) being unsuitable for many patients.

Nonmotor imagery was shown to be more effective in boosting bit rate. The investigation initially performed with healthy subjects had the side effect of demonstrating the advantages of multiclass BCI. The advantage of using nonmotor imagery was evident in one late-stage ALS patient. Using ECoG recording rechnology, higher brain modulation was found for imagined singing than for any other imagined motor task. The difficulty with this patient was that the strength of the modulation was not sufficient for single-trial classification, implying that communication speed could not actually be measured in that case.

The strongest point of the thesis is the claim that error-related potentials increase BCI communication speed in patients. This was shown to be true for a large cohort of patients (and subjects) in an EEG study. The bit rate increase was significant in the realistic online spelling sessions. The evidence presented leads one to believe that the thesis is true.

In addition to that result, this dissertation for the first time examines methods of measuring BCI performance and cognition detection in two CLIS patients with chronic ECoG implants.

7.3. Outlook

Some ideas for future work to further strengthen the justification of the thesis are given in the following.

Beginning with the connectivity and multiclass aspect, it is important to use the best binary combinations found in the study in an online two-class experiment. Thereafter, the best ternary combinations should be tested online to ensure that the accuracies projected here are achievable. If possible, the additional cognitive effort required should be gauged as well. Thereafter, the online three-class BCI should be tested with patients in EEG. ECoG would be too limited as it only covers part of the cortex, and the tasks investigated in this work are not limited to the motor cortex. MEG is an option for mobile patients, but they would have to switch to EEG later. However, future recording technology might enable high-density contactless recordings in portable systems, bringing the advantages of MEG to patients. This would also increase the relevance of the multiclass MEG results presented here.

The most important lesson learnt from the multiclass study is that a pairwise combination of imagined movement and nonmotor tasks promises higher accuracy than using imagined movement alone. The best of these task combinations should be tested with binary online classification in a new study, before the best ternary task combinations are tested online. Connectivity features failed to deliver on the promise of increased classification accuracy. However, the recommendation for future work would be to try further preprocessing methods such as wider frequency bands, spatial filters, and the extraction of more complex connectivity

features based on autoregressive coefficients, such as causality.

Concerning cognition detection in ALS patients, it would be important to repeat the measurement more often with future patients and to keep the measurements comparable, i.e., use the same reference electrode, even if various electrodes do not deliver signal. From an ethical standpoint it would be equally important to obtain informed consent from the patient as early as possible, as the presentation of the tones is repetitive and can be annoying for some.

An area bearing great promise for communication in patients is auditory BCIs [43, 71], although the newest concept has not yet been tested with patients [119]. These BCIs are also based on the P3 potential and thus operate on a similar basis as the cognition detection — it is conceivable that a cognition detection test could give the supervisor of such a BCI study with patients a more comprehensive view of the expected outcome, or help in finding reasons for low performance with auditory BCIs.

Concerning the recognition of error-related potentials, future work should address the question of implementation of error correction systems in other BCI paradigms. The current work centres around the visual P3 speller. The next step should be to implement this in one of the auditory paradigms cited above.

Further automation of the ErrP classifier training could be achieved by developing a measure which combines accuracy and specificity of ErrP recognition (e.g., $acc_{ErrP} \cdot r_C$), circumventing the need for manual selection of the hyperparameter w_{-1}. To elaborate on this point, more sessions with patients using the ECS would certainly help to elucidate the influence of a FP on succeeding trials further. This influence should be used as a weight factor for r_C in the above formula.

In some patients, the ECS did not improve performance. An ECS which monitors the accuracy of the primary classifier and only intervenes when the expected bit rate increase is above a predefined threshold, is worth investigating. There are two preconditions for such a system to work, though: Firstly, (visual) feedback of results is required to elicit the ErrP effect even when it is not used. Secondly, primary classifier accuracy can only be monitored during supervised learning (e.g., copy spelling instead of free spelling).

An effort is underway to make the newly developed ECS accessible to all BCI2000 users in the near future, in the hope that those people performing experiments with ALS patients will be able to integrate the ECS into their software to help patients communicate faster.

A. Nomenclature

General

ms	Milliseconds	
N_p	Number of patients in a study	
N_h	Number of healthy subjects in a study	
R	Run, a recording containing a series of trials	
S	Session, a recording containing a series of runs with breaks in-between	
t	Time, in milliseconds unless otherwise stated, after an important stimulus	

Electrophysiology

delta	Delta frequency band, 0.1–4 Hz	p. 15
theta	Theta frequency band, 4–8 Hz	p. 15
alpha	Alpha frequency band, 8–13 Hz	p. 15
beta	Beta frequency band, 14–30 Hz	p. 15
gamma	Gamma frequency band, above 30 Hz	p. 15
μ-rhythm	Resting-state rhythm in the alpha range (8–12 Hz), suppressed during (imagined) movement	p. 1
N1/P2	Component sensitive to auditory attention	p. 9
N2a	Involuntary and attention-independent response to an occasional deviation in a repetitive auditory stimulus sequence	p. 9
P3a	Elicited by infrequent shifts in tone pitch or intensity, task-irrelevant	p. 9
P3b	Elicited only by task-relevant infrequent shifts in tone pitch or intensity	p. 9
N400	Expected for word pairs with different semantic contexts	p. 9
hEOG	Horizontal component of electrooculogram	p. 14
vEOG	Vertical component of electrooculogram	p. 14

Signal Processing

a, b	Channels, electrodes or sensors	p. 16
N	Number of samples in a predefined time window	p. 16
n	Sample at time point n	p. 16
φ	Instantaneous phase	p. 16
f	Frequency	p. 16

Evaluation Criteria

p	Accuracy	p. 14
M	Number of classes	p. 17
b_W	Wolpaw bit rate	p. 17
b_S	Dal Seno bit rate	p. 17
κ	Cohen's kappa coefficient	p. 18
N	Number of test data vectors	p. 18
c	Trial duration in seconds	p. 17
r_C	Specificity, $\frac{TN}{TN+FP}$	p. 18
r_E	Sensitivity, $\frac{TP}{TP+FN}$	p. 18
α	Significance level for statistical tests, usually 5%	p. 19
p	If $p < \alpha$, the null hypothesis is rejected, i.e., the result is considered significant	p. 43
E	Effect size (Cohen's d)	p. 44
M	Minimum number of successive positive t-tests required for statistical significance	p. 44

Error Potentials, Classification

C	SVM hyperparameter, regularisation constant	p. 20
C_{P3}	P3 classifier	p. 45
C_{ErrP}	ErrP classifier	p. 45
acc_{P3}	Accuracy of P3 classifier	p. 45
acc_{ErrP}	Accuracy of ErrP classifier	p. 45
w_{-1}	SVM hyperparameter, weight for class -1, to achieve trade-off between r_C and r_E	p. 52

B. Abbreviations

ALS amyotrophic lateral sclerosis . . . 72
ALSFRS ALS functional rating scale . . . 12
ALSFRS-R revised ALSFRS . . . 12
ANOVA analysis of variance . . . 80
AR autoregressive . . . 15
AUC area under curve . . . 56
BCI brain-computer interface . . . 1
CAR common average reference . . . 10
CI classification interval . . . 53
CLIS complete locked-in state . . . 62
Coh coherence . . . 53
CV cross-validation . . . 53
DLL dynamic link library . . . 52
ECoG electrocorticogram . . . 2
ECS error correction system . . . 72
EEG electroencephalogram . . . 1
EMG electromyogram . . . 83
ERD event-related desynchronisation . . . 23
EOG electrooculogram . . . 73
ERP evoked response potential . . . 63
ErrP error-related potential . . . 72
FIR finite impulse response . . . 33
FFT Fast Fourier Transform . . . 72
FP false positive . . . 80
GUI graphical user interface . . . 52
ICA independent component analysis . . . 14
iFFT inverse Fast Fourier Transform . . . 72
ITI inter-trial interval . . . 57
ITR information transfer rate . . . 53
LDA linear discriminant analysis . . . 20
LIS locked-in syndrome . . . 84
LOOCV leave-one-out cross-validation . . . 76

LORETA	low resolution electromagnetic tomography	73
MEG	magnetoencephalogram	3
MMN	Mismatch Negativity	63
MRI	magnetic resonance imaging	62
OVO	one-versus-one	20
OVR	one-versus-the-rest	20
PLV	phase-locking value	16
PSD	power spectral density	57
RBF	radial basis function	20
RCE	recursive channel elimination	53
ROC	receiver operating characteristics	18
RFE	recursive feature elimination	53
SCP	slow cortical potential	1
SD	standard deviation	19
SMR	sensorimotor rhythm	13
SNR	signal-to-noise ratio	10
SSVEP	steady-state visual evoked potential	8
SQUID	superconducting quantum interference device	10
SVM	support-vector machine	53
SWLDA	step-wise linear discriminant analysis	20

Bibliography

Internet references are current in June 2009. The most recent modification date is given in brackets if known.

[1] S. Abrahams, L. Goldstein, J. Kew, D. Brooks, C. Lloyd, C. Frith, and P. Leigh. Frontal lobe dysfunction in amyotrophic lateral sclerosis: A PET study. *Brain*, 119:2105–2120, 1996.

[2] S. Abrahams, L. Goldstein, A. Al-Chalabi, A. Pickering, R. Morris, R. Passingham, D. Brooks, and P. Leigh. Relation between cognitive dysfunction and pseudobulbar palsy in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 62:464–472, 1997.

[3] K. Alho, I. Winkler, C. Escera, M. Huotilainen, J. Virtanen, I. P. Jääskeläinen, E. Pekkonen, and R. J. Ilmoniemi. Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology*, 35(2):211–224, Mar 1998.

[4] B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser. BCI demographics: how many (and what kinds of) people can use an SSVEP BCI? *IEEE Trans Neural Syst Rehabil Eng*, 18 (2):107–116, Apr 2010. doi: 10.1109/TNSRE.2009.2039495. URL `http://dx.doi.org/10.1109/TNSRE.2009.2039495`.

[5] C. Anderson and Z. Sijercic. Classification of EEG signals from four subjects during five mental tasks. In *Proceedings of the Conference on Engineering Applications in Neural Networks*, pages 407–414, 1996.

[6] M. Averbeck, M. Grote-Kusch, P. Leiberich, E. Olbrich, S. Schöbel, and A. Schröder. *Skalen zur Erfassung der Lebensqualität*. Swets, 1997.

[7] D. Bamber. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12(4):387–415, 1975. ISSN 0022-2496. doi: DOI:10.1016/0022-2496(75)90001-2. URL `http://www.sciencedirect.com/science/article/B6WK3-4D7JNKG-8D/2/752ed837f02a9523cda7e96258f5516c`.

[8] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *J Neural Eng*, 4(2):R32–R57, Jun 2007. doi: 10.1088/1741-2560/4/2/R03. URL `http://dx.doi.org/10.1088/1741-2560/4/2/R03`.

[9] G. Bauer, F. Gerstenbrand, and E. Rumpl. Varieties of the locked-in syndrome. *J Neurol*, 221(2): 77–91, Aug 1979.

[10] **M. Bensch**, A. A. Karim, J. Mellinger, T. Hinterberger, M. Tangermann, M. Bogdan, W. Rosenstiel, and N. Birbaumer. Nessi: An EEG-Controlled Web Browser for Severely Paralyzed Patients. *Comput Intell Neurosci*, page 71863, 2007. doi: 10.1155/2007/71863. URL `http://dx.doi.org/10.1155/2007/71863`.

[11] **M. Bensch**, J. Mellinger, M. Bogdan, and W. Rosenstiel. A Multiclass BCI using MEG. In *Proceedings of the 4th Int. Brain-Computer Interface Workshop*, pages 191–196, Graz, 09 2008.

[12] H. Berger. Über das Elektrenkephalogramm des Menschen. *Arch. Psychiat. Nervenkr.*, 87:527–570, 1929.

[13] M. Besserve, K. Jerbi, F. Laurent, S. Baillet, J. Martinerie, and L. Garnero. Classification methods for ongoing EEG and MEG signals. *Biol Res*, 40(4):415–437, 2007. doi: /S0716-97602007000500005. URL http://dx.doi.org//S0716-97602007000500005.

[14] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398(6725):297–298, Mar 1999. doi: 10.1038/18581. URL http://dx.doi.org/10.1038/18581.

[15] N. Birbaumer. Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43(6):517–532, Nov 2006. doi: 10.1111/j.1469-8986.2006.00456.x. URL http://dx.doi.org/10.1111/j.1469-8986.2006.00456.x.

[16] B. Blankertz, C. Schafer, G. Dornhege, and G. Curio. Single trial detection of EEG error potentials: A tool for increasing BCI transmission rates. *Lecture notes in computer science*, pages 1137–1143, 2002.

[17] B. Blankertz, G. Dornhege, C. Schäfer, R. Krepki, J. Kohlmorgen, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):127–131, Jun 2003. doi: 10.1109/TNSRE.2003.814456. URL http://dx.doi.org/10.1109/TNSRE.2003.814456.

[18] R. A. Boyajian, C. Amo, S. M. Otis, J. S. Romine, and R. A. Smith. Magnetic source imaging of cortical dysfunction in amyotrophic lateral sclerosis. *Am J Phys Med Rehabil*, 87(6):427–437, Jun 2008. doi: 10.1097/PHM.0b013e318174e7f1. URL http://dx.doi.org/10.1097/PHM.0b013e318174e7f1.

[19] V. Braitenberg and A. Schüz. *Cortex : statistics and geometry of neuronal connectivity*. Springer, Berlin, 2., thoroughly rev. ed. edition, 1998. ISBN 3-540-63816-4. URL http://swbplus.bsz-bw.de/bsz066222397cov.htm.

[20] C. Brunner, R. Scherer, B. Graimann, G. Supp, and G. Pfurtscheller. Online control of a brain-computer interface using phase synchronization. *IEEE Trans Biomed Eng*, 53(12 Pt 1):2501–2506, Dec 2006. doi: 10.1109/TBME.2006.881775. URL http://dx.doi.org/10.1109/TBME.2006.881775.

[21] A. Buttfield, P. W. Ferrez, and J. del R Millán. Towards a robust BCI: error potentials and online learning. *IEEE Trans Neural Syst Rehabil Eng*, 14(2):164–168, Jun 2006. doi: 10.1109/TNSRE.2006.875555. URL http://dx.doi.org/10.1109/TNSRE.2006.875555.

[22] R. T. Canolty, E. Edwards, S. S. Dalal, M. Soltani, S. S. Nagarajan, H. E. Kirsch, M. S. Berger, N. M. Barbaro, and R. T. Knight. High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793):1626–1628, Sep 2006. doi: 10.1126/science.1128115. URL http://dx.doi.org/10.1126/science.1128115.

[23] J. M. Cedarbaum, N. Stambler, E. Malta, C. Fuller, D. Hilt, B. Thurmond, and A. Nakanishi. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). *J Neurol Sci*, 169(1-2):13–21, Oct 1999. doi: http://dx.doi.org/10.1016/S0022-510X(99)00210-5.

[24] J. M. Cedarbaum and N. Stambler. Performance of the Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS) in multicenter clinical trials. *Journal of the Neurological Sciences*, 152(Supplement 1):s1–s9, 1997. ISSN 0022-510X. doi: http://dx.doi.org/10.1016/S0022-510X(97)00237-2. URL http://www.sciencedirect.com/science/article/B6T06-3RD1D51-1/2/a48437cb9c72df66c4fc10309ab81a67.

[25] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[26] J. Charcot and A. Joffroy. Deux cas d'atrophie musculaire progressive. *Arch. Physiol. Neurol. Pathol*, 2:354–367,745–760, 1869.

[27] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res. (JAIR)*, 16:321–357, 2002.

[28] J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[29] E. Curran, P. Sykacek, M. Stokes, S. Roberts, W. Penny, I. Johnsrude, and A. Owen. Cognitive tasks for driving a brain computer interfacing system: a pilot study. *IEEE Transactions on Rehabilitation Engineering*, 12(1), 2003. URL citeseer.ist.psu.edu/curran01cognitive.html.

[30] B. Dal Seno, M. Matteucci, and L. Mainardi. The Utility Metric: A Novel Method to Assess the Overall Performance of Discrete Brain-Computer Interfaces. *IEEE Trans Neural Syst Rehabil Eng*, Sep 2009. doi: 10.1109/TNSRE.2009.2032642. URL http://dx.doi.org/10.1109/TNSRE.2009.2032642.

[31] V. Dalbello-Haas, J. M. Florence, and L. S. Krivickas. Therapeutic exercise for people with amyotrophic lateral sclerosis or motor neuron disease. *Cochrane Database Syst Rev*, 2(2):CD005229, 2008. doi: 10.1002/14651858.CD005229.pub2. URL http://dx.doi.org/10.1002/14651858.CD005229.pub2.

[32] J. Daltrozzo, N. Wioland, V. Mutschler, P. Lutun, B. Calon, A. Meyer, T. Pottecher, S. Lang, A. Jaeger, and B. Kotchoubey. Cortical information processing in coma. *Cogn Behav Neurol*, 22(1):53–62, Mar 2009. doi: 10.1097/WNN.0b013e318192ccc8. URL http://dx.doi.org/10.1097/WNN.0b013e318192ccc8.

[33] F. P. de Lange, O. Jensen, M. Bauer, and I. Toni. Interactions between posterior gamma and frontal alpha/beta oscillations during imagined actions. *Front Hum Neurosci*, 2:7, 2008. doi: 10.3389/neuro.09.007.2008. URL http://dx.doi.org/10.3389/neuro.09.007.2008.

[34] G. Dornhege, B. Blankertz, G. Curio, and K. Müller. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Transactions on Biomedical Engineering*, 51(6):993–1002, 2004.

[35] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Trans Biomed Eng*, 53 (11):2274–2281, Nov 2006. doi: 10.1109/TBME.2006.883649. URL http://dx.doi.org/10.1109/TBME.2006.883649.

[36] T. Elbert, B. Rockstroh, W. Lutzenberger, and N. Birbaumer. Biofeedback of slow cortical potentials. I. *Electroencephalogr Clin Neurophysiol*, 48(3):293–301, Mar 1980.

[37] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein. ERP components on reaction errors and their functional significance: a tutorial. *Biol Psychol*, 51(2-3):87–107, Jan 2000.

[38] L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol*, 70(6):510–523, Dec 1988.

[39] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch. EMG and EOG artifacts in brain computer interface systems: A survey. *Clin Neurophysiol*, 118(3):480–494, Mar 2007. doi: 10.1016/j.clinph.2006.10.019. URL http://dx.doi.org/10.1016/j.clinph.2006.10.019.

[40] P. W. Ferrez and J. del R Millán. Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE Trans Biomed Eng*, 55(3):923–929, Mar 2008. doi: 10.1109/TBME.2007.908083. URL http://dx.doi.org/10.1109/TBME.2007.908083.

[41] P. Ferrez, J. del R Millán, and T. Test. You are wrong!-automatic detection of interaction errors from brain waves. In *International Joint Conference on Artificial Intelligence*, volume 19, pages 1413–1418. Citeseer, 2005.

[42] D. Friedman, V. Kazmerski, and M. Fabiani. An overview of age-related changes in the scalp distribution of P3b. *Electroencephalogr Clin Neurophysiol*, 104(6):498–513, Nov 1997. doi: http://dx.doi.org/10.1016/S0168-5597(97)00036-1.

[43] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, Jan 2009. doi: 10.1111/j.1469-8986.2008.00783.x. URL http://dx.doi.org/10.1111/j.1469-8986.2008.00783.x.

[44] T. Geng, J. Q. Gan, M. Dyson, C. S. Tsui, and F. Sepulveda. A novel design of 4-class BCI using two binary classifiers and parallel mental tasks. *Comput Intell Neurosci*, page 437306, 2008. doi: 10.1155/2008/437306. URL http://dx.doi.org/10.1155/2008/437306.

[45] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller. Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis. *IEEE Trans Biomed Eng*, 51(6):954–962, Jun 2004. doi: 10.1109/TBME.2004.826671. URL http://dx.doi.org/10.1109/TBME.2004.826671.

[46] J. Grosskreutz, J. Kaufmann, J. Frädrich, R. Dengler, H.-J. Heinze, and T. Peschel. Widespread sensorimotor and frontal cortical atrophy in Amyotrophic Lateral Sclerosis. *BMC Neurol*, 6:17, 2006. doi: 10.1186/1471-2377-6-17. URL http://dx.doi.org/10.1186/1471-2377-6-17.

[47] D. Guthrie and J. S. Buchwald. Significance testing of difference potentials. *Psychophysiology*, 28(2):240–244, Mar 1991.

[48] E. Gysels, P. Renevey, and P. Celka. Fast feature selection to compare broadband with narrowband phase synchronization in brain-computer interfaces. *Methods Inf Med*, 46(2):160–163, 2007. URL http://www.ncbi.nlm.nih.gov/pubmed/17347748.

[49] E. Gysels and P. Celka. Phase synchronization for the recognition of mental tasks in a brain-computer interface. *IEEE Trans Neural Syst Rehabil Eng*, 12(4):406–415, Dec 2004. doi: 10.1109/TNSRE.2004.838443. URL http://dx.doi.org/10.1109/TNSRE.2004.838443.

[50] S. Halder, **M. Bensch**, J. Mellinger, M. Bogdan, A. Kübler, N. Birbaumer, and W. Rosenstiel. Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Comput Intell Neurosci*, page 82069, 2007. doi: 10.1155/2007/82069. URL http://dx.doi.org/10.1155/2007/82069.

[51] P. S. Hammon and V. R. de Sa. Preprocessing and meta-classification for brain-computer interfaces. *IEEE Trans Biomed Eng*, 54(3):518–525, Mar 2007. doi: 10.1109/TBME.2006.888833. URL http://dx.doi.org/10.1109/TBME.2006.888833.

[52] H. A. Hanagasi, I. H. Gurvit, N. Ermutlu, G. Kaptanoglu, S. Karamursel, H. A. Idrisoglu, M. Emre, and T. Demiralp. Cognitive impairment in amyotrophic lateral sclerosis: evidence from neuropsychological investigation and event-related potentials. *Brain Res Cogn Brain Res*, 14(2):234–244, Aug 2002. doi: http://dx.doi.org/10.1016/S0926-6410(02)00110-6.

[53] P. Haselager, R. Vlek, J. Hill, and F. Nijboer. A note on ethical aspects of BCI. *Neural Netw*, 22 (9):1352–1357, Nov 2009. doi: 10.1016/j.neunet.2009.06.046. URL http://dx.doi.org/10.1016/j.neunet.2009.06.046.

[54] A. J. Haufler, T. W. Spalding, D. L. S. Maria, and B. D. Hatfield. Neuro-cognitive activity during a self-paced visuospatial task: comparative EEG profiles in marksmen and novice shooters. *Biol Psychol*, 53(2-3):131–160, Jul 2000. doi: http://dx.doi.org/10.1016/S0301-0511(00)00047-8.

[55] H. Hayashi and E. A. Oppenheimer. ALS patients on TPPV: totally locked-in state, neurologic findings and ethical implications. *Neurology*, 61(1):135–137, Jul 2003.

[56] N. J. Hill, T. N. Lal, M. Schröder, T. Hinterberger, B. Wilhelm, F. Nijboer, U. Mochty, G. Widman, C. Elger, B. Schölkopf, A. Kübler, and N. Birbaumer. Classifying EEG and ECoG signals without subject training for fast BCI implementation: comparison of nonparalyzed and completely paralyzed subjects. *IEEE Trans Neural Syst Rehabil Eng*, 14(2):183–186, Jun 2006. doi: 10.1109/TNSRE.2006.875548. URL http://dx.doi.org/10.1109/TNSRE.2006.875548.

[57] T. Hinterberger, B. Wilhelm, J. Mellinger, B. Kotchoubey, and N. Birbaumer. A device for the detection of cognitive brain functions in completely paralyzed or unresponsive patients. *IEEE Trans Biomed Eng*, 52(2):211–220, Feb 2005. doi: 10.1109/TBME.2004.840190. URL http://dx.doi.org/10.1109/TBME.2004.840190.

[58] T. Hinterberger, G. Widman, T. N. Lal, J. Hill, M. Tangermann, W. Rosenstiel, B. Schölkopf, C. Elger, and N. Birbaumer. Voluntary brain regulation and communication with electrocorticogram signals. *Epilepsy Behav*, 13(2):300–306, Aug 2008. doi: 10.1016/j.yebeh.2008.03.014. URL http://dx.doi.org/10.1016/j.yebeh.2008.03.014.

[59] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens. An efficient P300-based brain-computer interface for disabled subjects. *J Neurosci Methods*, 167(1):115–125, Jan 2008. doi: 10.1016/j.jneumeth.2007.03.005. URL http://dx.doi.org/10.1016/j.jneumeth.2007.03.005.

[60] J. Kalcher, D. Flotzinger, C. Neuper, S. Gölly, and G. Pfurtscheller. Graz brain-computer interface II: towards communication between humans and computers based on online classification of three different EEG patterns. *Med Biol Eng Comput*, 34(5):382–388, Sep 1996. URL http://www.springerlink.com/content/3jl8103813080316/fulltext.pdf.

[61] J. Kamiya. *Altered states of consciousness*, chapter Operant control of the EEG alpha rhythm and some of its effects on consciousness, pages 489–501. Wiley, New York, 1969.

[62] Z. A. Keirn and J. I. Aunon. A new mode of communication between man and his surroundings. *IEEE Trans Biomed Eng*, 37(12):1209–1214, Dec 1990. doi: 10.1109/10.64464. URL http://dx.doi.org/10.1109/10.64464.

[63] B. Kleber, N. Birbaumer, R. Veit, T. Trevorrow, and M. Lotze. Overt and imagined singing of an Italian aria. *Neuroimage*, 36(3):889–900, Jul 2007. doi: 10.1016/j.neuroimage.2007.02.053. URL http://dx.doi.org/10.1016/j.neuroimage.2007.02.053.

[64] B. Kotchoubey, H. Schleichert, W. Lutzenberger, and N. Birbaumer. A new method for self-regulation of slow cortical potentials in a timed paradigm. *Appl Psychophysiol Biofeedback*, 22(2):77–93, Jun 1997.

[65] B. Kotchoubey, S. Lang, S. Winter, and N. Birbaumer. Cognitive processing in completely paralyzed patients with amyotrophic lateral sclerosis. *Eur J Neurol*, 10(5):551–558, Sep 2003.

[66] A. Kreilinger, C. Neuper, G.Pfurtscheller, and G.-R. Mueller-Putz. Implementation of Error Detection into the Graz-Brain-Computer Interface, The Interaction Error Potential. In *AAATE 2009*, 2009.

[67] A. Kübler. Brain-computer interfaces for communication in paralysed patients and implications for disorders of consciousness. In S. Laureys and G. Tononi, editors, *The Neurology of Consciousness*, chapter 17, pages 217–233. Elsevier Ltd, 2009. URL http://www.elsevier.com/wps/find/bookdescription.cws_home/716351/description.

[68] A. Kübler and N. Birbaumer. Brain-computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients? *Clin Neurophysiol*, 119(11):2658–2666, Nov 2008. doi: 10.1016/j.clinph.2008.06.019. URL http://dx.doi.org/10.1016/j.clinph.2008.06.019.

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

[70] A. Kübler, 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):1775–1777, May 2005. doi: 10.1212/01.WNL.0000158616.43002.6D. URL http://dx.doi.org/10.1212/01.WNL.0000158616.43002.6D.

[71] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey. A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Ann N Y Acad Sci*, 1157:90–100, Mar 2009. doi: 10.1111/j.1749-6632.2008.04122.x. URL http://dx.doi.org/10.1111/j.1749-6632.2008.04122.x.

[72] J. Lachaux, E. Rodriguez, J. Martinerie, and F. Varela. Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208, 1999.

[73] J. Lakerveld, B. Kotchoubey, and A. Kübler. Cognitive function in patients with late stage amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*, 79(1):25–29, Jan 2008. doi: 10.1136/jnnp.2007.116178. URL http://dx.doi.org/10.1136/jnnp.2007.116178.

[74] T. N. Lal, M. Schröder, N. J. Hill, H. Preißl, T. Hinterberger, J. Mellinger, M. Bogdan, W. Rosenstiel, T. Hofmann, N. Birbaumer, and B. Schölkopf. A brain computer interface with online feedback based on magnetoencephalography. In *ICML*, pages 465–472, 2005.

[75] S. Laureys, F. Perrin, M.-E. Faymonville, C. Schnakers, M. Boly, V. Bartsch, S. Majerus, G. Moonen, and P. Maquet. Cerebral processing in the minimally conscious state. *Neurology*, 63(5):916–918, Sep 2004.

[76] S. Laureys, A. M. Owen, and N. D. Schiff. Brain function in coma, vegetative state, and related disorders. *Lancet Neurol*, 3(9):537–546, Sep 2004. doi: 10.1016/S1474-4422(04)00852-X. URL `http://dx.doi.org/10.1016/S1474-4422(04)00852-X`.

[77] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans Biomed Eng*, 52(9):1541–1548, Sep 2005. doi: 10.1109/TBME.2005.851521. URL `http://dx.doi.org/10.1109/TBME.2005.851521`.

[78] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain-computer interfaces. *J Neural Eng*, 4(2):R1–R13, Jun 2007. doi: 10.1088/1741-2560/4/2/R01. URL `http://dx.doi.org/10.1088/1741-2560/4/2/R01`.

[79] S. J. Luck. *An introduction to the event-related potential technique*. Cognitive Neuroscience. MIT Press, 2005.

[80] A. C. Ludolph, K. J. Langen, M. Regard, H. Herzog, B. Kemper, T. Kuwert, I. G. Böttger, and L. Feinendegen. Frontal lobe function in amyotrophic lateral sclerosis: a neuropsychologic and positron emission tomography study. *Acta Neurol Scand*, 85(2):81–89, Feb 1992.

[81] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalogr Clin Neurophysiol*, 103(3):386–394, Sep 1997. doi: http://www.ece.uvic.ca/~btill/research/bci/chmaprefs/McFarland1997.pdf.

[82] J. Mellinger, T. Hinterberger, **M. Bensch**, M. Schröder, and N. Birbaumer. Surfing the Web with Electrical Brain Signals: The Brain Web Surfer (BWS) for the Completely Paralysed. In *In Proceedings of 2nd World Congress of the International Society of Physical and Rehabilitation Medicine ISPRM*, Prague, Czech Republic, 05 2003.

[83] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kübler. An MEG-based brain-computer interface (BCI). *Neuroimage*, 36(3):581–593, Jul 2007. doi: 10.1016/j.neuroimage.2007.03.019. URL `http://dx.doi.org/10.1016/j.neuroimage.2007.03.019`.

[84] R. G. Miller, C. E. Jackson, E. J. Kasarskis, J. D. England, D. Forshew, W. Johnston, S. Kalra, J. S. Katz, H. Mitsumoto, J. Rosenfeld, C. Shoesmith, M. J. Strong, S. C. Woolley, and Q. S. S. of the American Academy of Neurology. Practice parameter update: The care of the patient with amyotrophic lateral sclerosis: multidisciplinary care, symptom management, and cognitive/behavioral impairment (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, 73(15):1227–1233, Oct 2009. URL `http://www.neurology.org/cgi/content/full/73/15/1227`.

[85] E. Mugler, **M. Bensch**, S. Halder, W. Rosenstiel, M. Bogdan, N. Birbaumer, and A. Kübler. Control of an Internet Browser Using the P300 Event Related Potential. *International Journal of Bioelectromagnetism*, 10(1):56–63, 2008.

[86] E. M. Mugler, C. A. Ruf, S. Halder, **M. Bensch**, and A. Kübler. Design and Implementation of a P300-Based Brain-Computer Interface for Controlling an Internet Browser. *IEEE Trans Neural Syst Rehabil Eng*, Aug 2010. doi: 10.1109/TNSRE.2010.2068059. URL http://dx.doi.org/10.1109/TNSRE.2010.2068059.

[87] A. R. Murguialday, J. Hill, **M. Bensch**, S. Martens, S. Halder, F. Nijboer, B. Schölkopf, N. Birbaumer, and A. Gharabaghi. Transition from the locked in to the completely locked-in state: A physiological analysis. *Clin Neurophysiol*, Dec 2010. doi: 10.1016/j.clinph.2010.08.019. URL http://dx.doi.org/10.1016/j.clinph.2010.08.019.

[88] N. Neumann and B. Kotchoubey. Assessment of cognitive functions in severely paralysed and severely brain-damaged patients: neuropsychological and electrophysiological methods. *Brain Res Brain Res Protoc*, 14(1):25–36, Nov 2004. doi: 10.1016/j.brainresprot.2004.09.001. URL http://dx.doi.org/10.1016/j.brainresprot.2004.09.001.

[89] C. Neuper, G. R. Müller, A. Kübler, N. Birbaumer, and G. Pfurtscheller. Clinical application of an EEG-based brain-computer interface: a case study in a patient with severe motor impairment. *Clin Neurophysiol*, 114(3):399–409, Mar 2003. doi: http://dx.doi.org/10.1016/S1388-2457(02)00387-5.

[90] E. Niedermeyer and F. L. da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, November 2004. ISBN 0781751268. URL http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0781751268.

[91] F. Nijboer, E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N. Birbaumer, and A. Kübler. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clin Neurophysiol*, 119(8):1909–1916, Aug 2008. doi: 10.1016/j.clinph.2008.03.034. URL http://dx.doi.org/10.1016/j.clinph.2008.03.034.

[92] F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N. Birbaumer, and A. Kübler. An auditory brain-computer interface (BCI). *J Neurosci Methods*, 167(1):43–50, Jan 2008. doi: 10.1016/j.jneumeth.2007.02.009. URL http://dx.doi.org/10.1016/j.jneumeth.2007.02.009.

[93] F. H. Norris. Amyotrophic Lateral Sclerosis: The Clinical Disorder. In R. Smith, editor, *Handbook of Amyotrophic Lateral Sclerosis*, pages 3–38. Marcel Dekker, 1992.

[94] B. Obermaier, C. Neuper, C. Guger, and G. Pfurtscheller. Information transfer rate in a five-classes brain-computer interface. *IEEE Trans Neural Syst Rehabil Eng*, 9(3):283–288, Sep 2001. doi: 10.1109/7333.948456. URL http://dx.doi.org/10.1109/7333.948456.

[95] T. Ogawa, H. Tanaka, and K. Hirata. Cognitive deficits in amyotrophic lateral sclerosis evaluated by event-related potentials. *Clin Neurophysiol*, 120(4):659–664, Apr 2009. doi: 10.1016/j.clinph.2009.01.013. URL http://dx.doi.org/10.1016/j.clinph.2009.01.013.

[96] R. C. Oldfield. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1):97–113, Mar 1971.

[97] P. Olejniczak. Neurophysiologic basis of EEG. *J Clin Neurophysiol*, 23(3):186–189, Jun 2006. doi: 10.1097/01.wnp.0000220079.61973.6c. URL http://dx.doi.org/10.1097/01.wnp.0000220079.61973.6c.

[98] R. W. Orrell. Motor neuron disease: systematic reviews of treatment for ALS and SMA. *Br Med Bull*, 93:145–159, 2010. doi: 10.1093/bmb/ldp049. URL http://dx.doi.org/10.1093/bmb/ldp049.

[99] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda. Response error correction–a demonstration of improved human-machine performance using real-time EEG monitoring. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):173–177, Jun 2003. doi: 10.1109/TNSRE.2003.814446. URL http://dx.doi.org/10.1109/TNSRE.2003.814446.

[100] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int J Psychophysiol*, 18(1):49–65, Oct 1994.

[101] S. H. Patel and P. N. Azzam. Characterization of N200 and P300: selected studies of the Event-Related Potential. *Int J Med Sci*, 2(4):147–154, 2005.

[102] W. Penfield. Mechanisms of voluntary movement. *Brain*, 77(1):1–17, 1954.

[103] E. Pereda, R. Q. Quiroga, and J. Bhattacharya. Nonlinear multivariate analysis of neurophysiological signals. *Prog Neurobiol*, 77(1-2):1–37, 2005. doi: 10.1016/j.pneurobio.2005.10.003. URL http://dx.doi.org/10.1016/j.pneurobio.2005.10.003.

[104] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. da Silva. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159, May 2006. doi: http://dx.doi.org/10.1016/j.neuroimage.2005.12.003.

[105] F. Piccione, F. Giorgi, P. Tonin, K. Priftis, S. Giove, S. Silvoni, G. Palmas, and F. Beverina. P300-based brain computer interface: reliability and performance in healthy and paralysed participants. *Clin Neurophysiol*, 117(3):531–537, Mar 2006. doi: 10.1016/j.clinph.2005.07.024. URL http://dx.doi.org/10.1016/j.clinph.2005.07.024.

[106] M. L. V. Quyen, J. Foucher, J. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, and F. J. Varela. Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. *J Neurosci Methods*, 111(2):83–98, Oct 2001.

[107] R. Radtke, A. Erwin, and C. Erwin. Abnormal sensory evoked potentials in amyotrophic lateral sclerosis. *Neurology*, 36:796–801, 1986.

[108] A. B. Randolph, S. Karmakar, and M. M. Jackson. Toward predicting control of a brain-computer interface. In *Proceedings of the International Conference on Information Systems*, Milwaukee, 2006. URL http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1174&context=icis2006.

[109] E. Rodriguez, N. George, J. P. Lachaux, J. Martinerie, B. Renault, and F. J. Varela. Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397(6718):430–433, Feb 1999. doi: 10.1038/17120. URL http://dx.doi.org/10.1038/17120.

[110] G. Schalk, J. R. Wolpaw, D. J. McFarland, and G. Pfurtscheller. EEG-based communication: presence of an error potential. *Clin Neurophysiol*, 111(12):2138–2144, Dec 2000.

[111] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans Biomed Eng*, 51(6):1034–1043, Jun 2004. doi: 10.1109/TBME.2004.827072. URL http://dx.doi.org/10.1109/TBME.2004.827072.

[112] R. Scherer, G. R. Müller, C. Neuper, B. Graimann, and G. Pfurtscheller. An Asynchronously Controlled EEG-Based Virtual Keyboard: Improvement of the Spelling Rate. *IEEE Transactions on Biomedical Engineering*, 51(6), 2004.

[113] R. Scherer, A. Schlögl, F. Lee, H. Bischof, J. Jansa, and G. Pfurtscheller. The self-paced graz brain-computer interface: methods and applications. *Comput Intell Neurosci*, page 79826, 2007. doi: 10.1155/2007/79826. URL http://dx.doi.org/10.1155/2007/79826.

[114] M. Scherg, J. Vajsar, and T. W. Picton. A Source Analysis of the Late Human Auditory Evoked Potentials. *Journal of Cognitive Neuroscience*, 1(4):336–355, 1989. doi: 10.1162/jocn.1989.1.4.336. URL http://www.mitpressjournals.org/doi/abs/10.1162/jocn.1989.1.4.336.

[115] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller. A fully automated correction method of EOG artifacts in EEG recordings. *Clin Neurophysiol*, 118(1):98–104, Jan 2007. doi: 10.1016/j.clinph.2006.09.003. URL http://dx.doi.org/10.1016/j.clinph.2006.09.003.

[116] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller. Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *J Neural Eng*, 2(4):L14–L22, Dec 2005. doi: 10.1088/1741-2560/2/4/L02. URL http://dx.doi.org/10.1088/1741-2560/2/4/L02.

[117] A. Schlögl, J. Kronegg, J. E. Huggins, and S. G. Mason. Evaluation Criteria for BCI Research. In G. Dornhege, J. del R. Millan, T. Hinterberger, D. J. McFarland, and K.-R. Müller, editors, *Toward Brain-Computer Interfacing*, chapter 19, pages 327–342. MIT Press, 2007.

[118] T. Schreiner. Development and Application of a Python Scripting Framework for BCI2000. Master's thesis, Eberhard-Karls-Universität Tübingen, 1 2008. URL http://edoc.mpg.de/423656.

[119] M. Schreuder, B. Blankertz, and M. Tangermann. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. *PLoS One*, 5(4):e9813, 2010. doi: 10.1371/journal.pone.0009813. URL http://dx.doi.org/10.1371/journal.pone.0009813.

[120] E. W. Sellers and E. Donchin. A P300-based brain-computer interface: initial tests by ALS patients. *Clin Neurophysiol*, 117(3):538–548, Mar 2006. doi: 10.1016/j.clinph.2005.06.027. URL http://dx.doi.org/10.1016/j.clinph.2005.06.027.

[121] S. Silvoni, C. Volpato, M. Cavinato, M. Marchetti, K. Priftis, A. Merico, P. Tonin, K. Koutsikos, F. Beverina, and F. Piccione. P300-based brain-computer interface communication: evaluation and follow-up in amyotrophic lateral sclerosis. *Front. Neuroprosthetics*, 1:1–12, 2009. ISSN ISSN 1662-9957. doi: 10.3389/neuro.20/001.2009. URL www.frontiersin.org/neuroscience/neuroprosthetics/paper/10.3389/neuro.20/001.2009/.

[122] M. E. Smith, E. Halgren, M. Sokolik, P. Baudena, A. Musolino, C. Liegeois-Chauvel, and P. Chauvel. The intracranial topography of the P3 event-related potential elicited during auditory oddball. *Electroencephalogr Clin Neurophysiol*, 76(3):235–248, Sep 1990.

[123] M. E. Smith, L. K. McEvoy, and A. Gevins. Neurophysiological indices of strategy development and skill acquisition. *Cognitive Brain Research*, 7(3):389–404, 1999. ISSN 0926-6410. doi: DOI:10.1016/S0926-6410(98)00043-3. URL http://www.sciencedirect.com/science/article/B6SYV-3V91NR5-G/2/126832d3e142977a6d3f82b04a65122f.

[124] L. Song, E. Gysels, and E. Gordon. Phase synchrony rate for the recognition of motor imagery in BCI. In *Advances in Neural Info. Proc. Sys*, volume 18, 2006.

[125] L. Song and J. Epps. Classifying EEG for brain-computer interface: learning optimal filters for dynamical system features. *Intell. Neuroscience*, 2007:1–11, 2007. ISSN 1687-5265. doi: http://dx.doi.org/10.1155/2007/57180.

[126] M. B. Sterman. Sensorimotor EEG operant conditioning: experimental and clinical effects. *Pavlov J Biol Sci*, 12(2):63–92, 1977.

[127] G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. E. Frye, C. K. Hauser, N. E. Schwartz, T. M. Vaughan, J. R. Wolpaw, and E. W. Sellers. A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. *Clin Neurophysiol*, 121(7):1109–1120, Jul 2010. doi: 10.1016/j.clinph.2010.01.030. URL http://dx.doi.org/10.1016/j.clinph.2010.01.030.

[128] M. van Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, A. Nijholt, N. Ramsey, P. Haselager, L. Vuurpijl, S. Gielen, and P. Desain. The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4):041001, 2009. URL http://stacks.iop.org/1741-2552/6/i=4/a=041001.

[129] F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie. The brainweb: phase synchronization and large-scale integration. *Nat Rev Neurosci*, 2(4):229–239, Apr 2001. doi: 10.1038/35067550. URL http://dx.doi.org/10.1038/35067550.

[130] K. K. Vesco, R. C. Bone, J. C. Ryan, and J. Polich. P300 in young and elderly subjects: auditory frequency and intensity effects. *Electroencephalogr Clin Neurophysiol*, 88(4):302–308, 1993. doi: http://dx.doi.org/10.1016/0168-5597(93)90054-S.

[131] P. Vieregge, B. Wauschkuhn, I. Heberlein, J. Hagenah, and R. Verleger. Selective attention is impaired in amyotrophic lateral sclerosis–a study of event-related EEG potentials. *Brain Res Cogn Brain Res*, 8 (1):27–35, May 1999.

[132] G. Visconti, B. Dal Seno, M. Matteucci, and L. Mainardi. Automatic Recognition of Error Potentials in a P300-Based Brain-Computer Interface. In *Proceedings of the 4th International Brain-Computer Interface Workshop & Training Course*, pages 238–243, 2008.

[133] A. Walter, **M. Bensch**, D. Brugger, W. Rosenstiel, M. Bogdan, N. Birbaumer, and A. Gharabaghi. BCCI - A Bidirectional Cortical Communication Interface. In *In Proceedings of the International Joint Conference on Computational Intelligence*, pages 440–445, Funchal, 10 2009.

[134] Q. Wei, Y. Wang, X. Gao, and S. Gao. Amplitude and phase coupling measures for feature extraction in an EEG-based brain-computer interface. *J Neural Eng*, 4(2):120–129, Jun 2007. doi: 10.1088/1741-2560/4/2/012. URL http://dx.doi.org/10.1088/1741-2560/4/2/012.

[135] J. Weston, A. Elisseeff, G. BakIr, and F. Sinz. The spider machine learning toolbox. http://www.kyb.mpg.de/bs/people/spider/, 2005. URL http://www.kyb.mpg.de/bs/people/spider/.

[136] B. Wilhelm, M. Jordan, and N. Birbaumer. Communication in locked-in syndrome: effects of imagery on salivary pH. *Neurology*, 67(3):534–535, Aug 2006. doi: 10.1212/01.wnl.0000228226.86382.5f. URL http://dx.doi.org/10.1212/01.wnl.0000228226.86382.5f.

[137] J. A. Wilson, E. A. Felton, P. C. Garell, G. Schalk, and J. C. Williams. ECoG factors underlying multimodal control of a brain-computer interface. *IEEE Trans Neural Syst Rehabil Eng*, 14(2):246–250, Jun 2006. doi: 10.1109/TNSRE.2006.875570. URL http://dx.doi.org/10.1109/TNSRE.2006.875570.

[138] M. Witgert, A. R. Salamone, A. M. Strutt, A. Jawaid, P. J. Massman, M. Bradshaw, D. Mosnik, S. H. Appel, and P. E. Schulz. Frontal-lobe mediated behavioral dysfunction in amyotrophic lateral sclerosis. *Eur J Neurol*, Oct 2009. doi: 10.1111/j.1468-1331.2009.02801.x. URL http://dx.doi.org/10.1111/j.1468-1331.2009.02801.x.

[139] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris. An EEG-based brain-computer interface for cursor control. *Electroencephalogr Clin Neurophysiol*, 78(3):252–259, Mar 1991.

[140] 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 Transactions on Rehabilitation Engineering*, 8:164–173, 2000.

[141] 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):767–791, Jun 2002.

[142] J. R. Wolpaw, D. J. McFarland, T. M. Vaughan, and G. Schalk. The Wadsworth Center brain-computer interface (BCI) research and development program. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):204–207, Jun 2003. doi: 10.1109/TNSRE.2003.814442. URL http://dx.doi.org/10.1109/TNSRE.2003.814442.

[143] W. L. Woon and A. Cichocki. Novel features for brain-computer interfaces. *Comput Intell Neurosci*, page 82827, 2007. doi: 10.1155/2007/82827. URL http://dx.doi.org/10.1155/2007/82827.