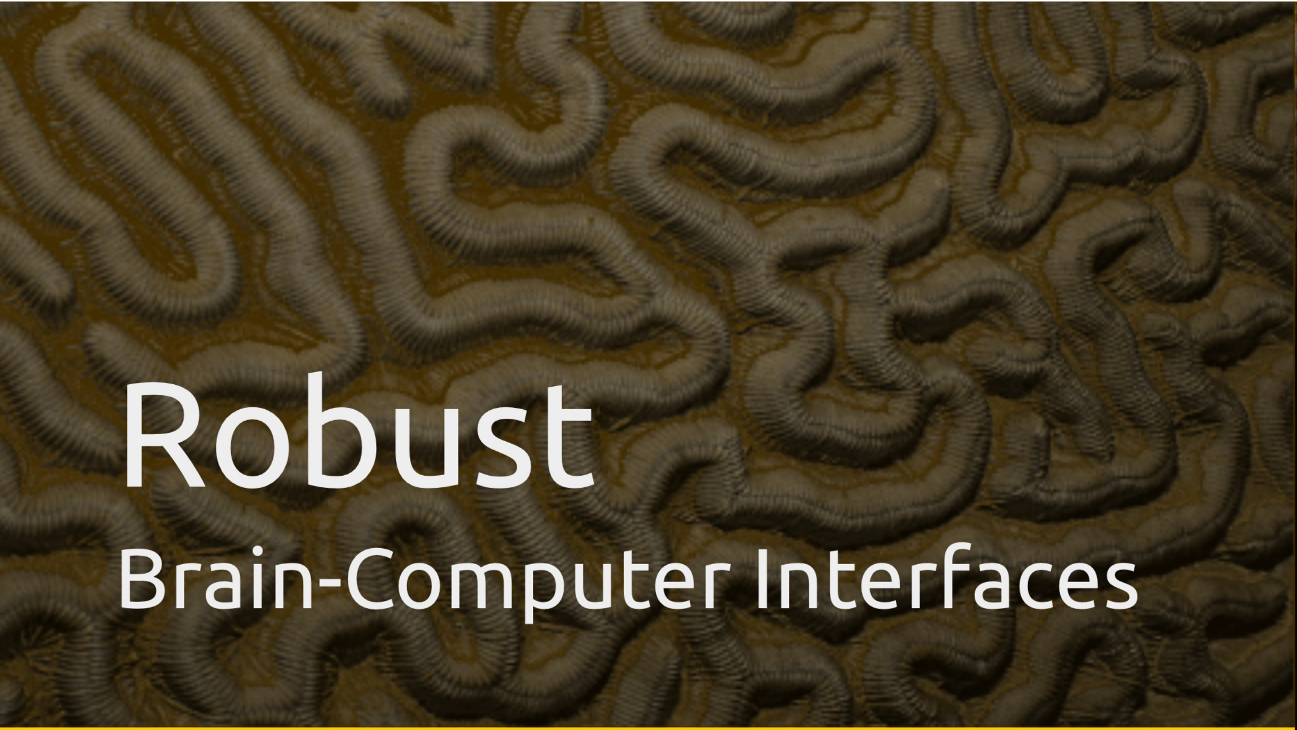

Robust
Brain-Computer Interfaces

Boris Reuderink

Robust Brain-Computer Interfaces

Boris Reuderink

Chairman and secretary:	
Prof. dr. ir. Ton J. Mouthaan	Universiteit Twente
Promotor:	
Prof. dr. ir. Anton Nijholt	Universiteit Twente
Assistant promotor:	
Dr. Mannes Poel	Universiteit Twente
Referee:	
Dr. Fabien Lotte	INRIA Bordeaux
Members:	
Dr. Dirk K.J. Heylen	Universiteit Twente
Prof. dr. Maja Pantic	Universiteit Twente
Prof. dr. Peter Desain	Radboud Universiteit Nijmegen
Prof. dr. Klaus-Robert Müller	Technische Universität Berlin
Paranymphs:	
Roald Dijkstra	
Luuk Peters	

The research reported in this thesis has been carried out at the Human Media Interaction (HMI) research group of the University of Twente.

The author gratefully acknowledges the support of the Brain-Gain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science.

CTIT dissertation series no. 11-209. Center for Telematics and Information Technology (CTIT). P.O. Box 217, 7500 AE Enschede, The Netherlands. ISSN: 1381-3617.

SIKS dissertation series no. 2011-44. The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

ROBUST BRAIN-COMPUTER INTERFACES

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Twente, op gezag van
de rector magnificus, prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen op
vrijdag 21 oktober 2011 om 16.45 uur

door

Boris Reuderink
geboren op 19 augustus 1981
te Wageningen.

Promotor: Prof. dr. ir. Anton Nijholt
Assistant promotor: Dr. Mannes Poel

ISBN: 978-94-6191-058-5

Into the distance, a ribbon of black
Stretched to the point of no turning back
A flight of fancy on a windswept field
Standing alone my senses reeled
A fatal attraction is holding me fast,
How can I escape this irresistible grasp?

Can't keep my eyes from the circling skies
Tongue-tied and twisted, just an earth-bound misfit, I

Ice is forming on the tips of my wings
Unheeded warnings, I thought, I thought of everything
No navigator to find my way home
Unladened, empty and turned to stone
A soul in tension — that's learning to fly
Condition grounded but determined to try

Can't keep my eyes from the circling skies
Tongue-tied and twisted just an earth-bound misfit, I

Above the planet on a wing and a prayer,
My grubby halo, a vapour trail in the empty air,
Across the clouds I see my shadow fly
Out of the corner of my watering eye
A dream unthreatened by the morning light
Could blow this soul right through the roof of the night

There's no sensation to compare with this
Suspended animation, a state of bliss

Can't keep my mind from the circling skies
Tongue-tied and twisted just an earth-bound misfit, I

— Pink Floyd, "Learning to Fly"

Preface

I have had a long-standing interest in how the human brain works. When at the end of my master's education the option arose to conduct research on brain-computer interfaces (BCIs), I had no choice but to pursue. The very intense past four years that followed culminated in this little book. This period was marked by great freedom to explore and follow intellectual curiosity, but also required perseverance, reflection on my strengths and weaknesses — and hard work. Looking back, I think of this period as a very valuable, and defining period of my life.

When I started four years ago, the BCI research was just starting at our Human Media Interaction (HMI) group, funded by the national BrainGain project. Through collaboration with Peter Desain's Cognitive Artificial Intelligence (CAI) group in Nijmegen, we quickly found our way in the field of BCIs. Over time, we developed our own best practises, and performed BCI research from the human-computer interaction (HCI) perspective. Simultaneously, I developed open source packages for signal processing and machine learning that powered many of our BCI demos.

Lots of people contributed to this great experience. First of all, I would like to thank the people from the HMI group. Specifically, my promotor Anton Nijholt for providing a place within the HMI group and allowing me to deviate from the sharp boundaries of the project assignment, and my daily supervisor Mannes Poel, who was always available for work-related reflections and kept an eye not only on the progress of my work, but was also interested in my personal well-being. During my travel to the university I often spoke with Dirk Heylen, who often stimulated me to reflect more deeply on my statements. I would also like to thank the secretaries Charlotte and Alice for their support, and thank Lynn for proofreading all my papers. Ronald shared my view on machine learning, and provided a listening ear for my doubts regarding the field. Dennis provided the occasional odd thought, folk music and general serendipity, and pushed me to think of the wider implications of my work.

I would like to thank the BrainMedia subgroup for the pleasant discussions and reflections on BCIs, and the experiences we shared. In particular, I would like to thank my roommate Christian with whom I shared the doubts and worries of obtaining a PhD, but also the successes, and Danny for our shared development of a vision on practical BCIs, her optimistic view, and our collaborations.

This work would not be the same without my weekly visits to the CAI group. Talking to Jason taught me that a simple mathematical proof can often save literally months of empirical research, and I learned a great variety of elegant (machine learning) tricks. When visiting, Rutger always had time for a creative brainstorm, and was as motivated as I am to improve the practise, and not only the theory of BCIs. I greatly enjoyed my collaboration with this group, and I would like to thank Peter Desain for creating many opportunities for me.

I know that the last years have been hard for the people around me. I would like to thank my friends and family for their support and acceptance of my sometimes lacking focus. My special thanks and admiration go to my wife Sanne and my daughter Lauren for enabling me to perform this research.

Contents

Samenvatting

EEN brein-computer interface (BCI) maakt directe communicatie tussen het brein en computers mogelijk, en omzeilt daarmee de traditionele route via de zenuwen en spieren. BCI's worden meestal ontworpen voor patiënten, die op geen enkele andere manier kunnen communiceren (door bijvoorbeeld verlamming). Het BCI-onderzoek van onze groep richt zich echter op gezonde gebruikers, specifiek op speltoepassingen. Deze ongebruikelijke focus leid tot andere eisen die we aan een BCI stellen.

Voor een algemene acceptatie van BCI-technologie moeten twee kernproblemen worden opgelost: 1) de investering die gedaan moet worden om een BCI te kunnen gebruiken moet klein zijn voor de gebruiker (zowel financiële investeringen, als investering van tijd), en 2) op de BCI moet vertrouwd kunnen worden; de BCI moet dus voorspelbaar reageren, met een constante nauwkeurigheid. Daarnaast moet een BCI natuurlijk zo worden toegepast dat het een meerwaarde biedt omdat de huidige generatie BCI's zich nog niet kan meten met de snelheid en betrouwbaarheid van invoerapparaten voor gezonde gebruikers.

Het eerste kernprobleem sluit de mogelijkheid uit dat hersensignalen worden gemeten met dure medische apparatuur. Relatief goedkope consumentenhardware is gelukkig al commercieel verkrijgbaar, wat het probleem van financiële investeringen grotendeels oplost. Wat overblijft is het deelprobleem dat (gezonde) gebruikers waarschijnlijk niet bereid zijn om weken of zelfs maanden te investeren — hetgeen gangbaar is voor BCI's gericht op patiënten — om de vaardigheid te leren die ze in staat stelt vrijwillig hun hersensignalen te sturen.

BCI's gebaseerd op *machine learning* reduceren het probleem van de te investeren tijd van maanden naar minuten. Dit doen ze door de persoonsafhankelijke patronen in spontane hersenactiviteit te herkennen. Maar zelfs met deze geavanceerde BCI's is een veeleisende en foutgevoelige kalibratiesessie, waarin de gebruiker geforceerd mentale taken moet uitvoeren, steevast nodig voor de BCI gebruik kan worden. Deze kalibratie is nodig omdat

de hersensignalen variëren van sessie tot sessie. De toepasbaarheid en de laagdrempeligheid van een BCI zouden sterk worden vergroot als deze herhaaldelijke kalibratiesessie vermeden kan worden.

Het tweede kernprobleem is gerelateerd aan het fundamentele probleem dat binnen het BCI-veld bekend staat als niet-stationaire signalen. De inherent variabele natuur van spontane hersenactiviteit leidt tot variaties in de signaaleigenschappen die de BCI gebruikt om hersenactiviteit te herkennen. Deze variabiliteit schendt de basisaanname van de machine learning methodes die gebruikt worden om een herkenner voor de BCI te trainen, en leidt tot grote fluctuaties in de nauwkeurigheid van de herkenning van de hersenactiviteit. Deze fluctuaties maken de huidige generatie BCI's onbetrouwbaar.

Beide kernproblemen zijn gerelateerd; zowel de inter-sessie variabiliteit als de onbetrouwbaarheid stammen af van niet volledig begrepen variaties van de hersensignalen over tijd, sessies en personen. In dit proefschrift onderzoeken we de aard van deze variaties, en ontwikkelen we twee nieuwe, complementaire technieken om deze kernproblemen op te lossen.

Om het probleem van niet-stationaire hersensignalen te bevestigen, laten we eerst zien dat een BCI, gebaseerd op veelgebruikte signaaleigenschappen, gevoelig is voor veranderingen in de gemoedstoestand van de gebruiker. Vervolgens presenteren we een methode gericht op het verwijderen van deze signaalveranderingen. Uitgaande van het inzicht dat een grote groep BCI's gebaseerd is op *relatieve* veranderingen in spectrale energie, maar de absolute energie gebruikt, ontwikkelen we een methode die een tweede-orde referentiepunt (SOB, second-order baseline) gebruikt om de relatieve veranderingen in het vuren van neuronen te isoleren. Voor zover wij weten is dit de eerste BCI die, zonder kalibratiesessie, spontane hersenactiviteit kan herkennen bij nieuwe gebruikers, zonder dat dit tot prestatieverlies leidt. Met deze SOB-methode hebben we het probleem van langzaam veranderende signaaldistributies omzeild. Maar nog steeds gaat de aanname, dat voorbeelden waarop de herkenner gebaseerd wordt onafhankelijk zijn van elkaar, niet op; hersenactiviteit lijkt op hersenactiviteit die kort daarvoor is waargenomen. In het bijzonder tijdens het trainen van een herkenner is het schenden van deze aanname problematisch, omdat de hoeveelheid informatie die de voorbeelden bevatten wordt overschat. Dit leidt tot *overfitting*; het model werkt alleen goed tijdens de kalibratiesessie. Daarom hebben we een generalisatie van de bekende support vector machine (SVM) herkenner afgeleid, die de chronologische structuur van herkenfouten mee kan nemen in de optimalisatie. Zowel op kunstmatige als echte BCI-data wordt overfitting verminderd, en leidt dit tot een verhoogde informatiedoorvoer.

Met de SOB-methode hebben we het eerste kernprobleem (investering van tijd) opgelost voor BCI's gebaseerd op het inbeelden van bewegingen. Het is waarschijnlijk dat deze aanpak ook voor andere mentale taken de kalibratiesessie overbodig maakt. Het tweede kernprobleem is deels opgelost met de generalisatie van de dependent-samples support vector machine (dSVM): de methode demonstreert dat het mogelijk is om de robuustheid te verbeteren door het modelleren van de structuur van de herkenfouten. Echter, het niet aannemen van onafhankelijke observaties roept nieuwe vragen op met betrekking tot de interpretatie van prestatiematen die gebruikt worden om BCI's te evalueren. Met de twee methodes gepresenteerd in deze thesis, hebben we een weg gebaand voor een nieuwe generatie BCI's — BCI's die betrouwbaar werken, zonder dat een kalibratiesessie nodig is.

Summary

A BCI enables direct communication from the brain to devices, bypassing the traditional pathway of peripheral nerves and muscles. Usually, BCIs are targeted at paralysed patients who have no other means of communication. The BCI research at our Human Media Interaction group focusses on BCIs for healthy users, especially on gaming applications, which poses some additional requirements on the BCI.

For successful BCI adoption in general, two key issues need to be addressed: 1) using a BCI should be easy, and require only small investments (of either time or money), and 2) the BCI should be dependable, that is to say it should function predictably with a known accuracy. In addition, a BCI should be applied such that it provides something unique (e.g. a covert measure of attention), since BCIs cannot yet compete on reliability and speed with existing input devices for non-patients.

The first key issue excludes the possibility that brain signals are recorded with expensive medical brain-imaging equipment. Fortunately, relatively cheap consumer hardware with semi-dry electroencephalography (EEG) sensors is already commercially available, solving the problem of monetary investments to a large extent. What remains is that non-paralysed users are probably not willing to invest weeks or even months to learn the skill to intentionally modify their brain signals, which is common practice with traditional BCIs aimed at patients. BCIs based on machine learning already reduce the problem of time investment from weeks to minutes, through automatic recognition of the user's naturally occurring brain signals. But still, a demanding and error-prone calibration session (in which the user is forced to demonstrate mental tasks) is required before each use of the BCI. This calibration session is needed because the electrical brain signals vary from session to session. Removing this need of a repeated training session would greatly expand the applicability of BCIs, and lower the barrier to entry.

The second key issue is related to the fundamental problem that is known in the BCI field as non-stationary signals. The inherent variable nature of

spontaneous EEG causes changes in the features that the BCI uses to detect and classify brain signals. This variability violates basic assumptions made by the machine learning (ML) methods used to train the BCI classifier, and causes the classification accuracy to fluctuate unpredictably. These fluctuations make the current generation of BCIs unreliable.

Both key issues are related; both the inter-session variability and the unreliability stem from not fully understood properties of fluctuation in the neuronal signal's feature distributions over time, sessions, and subjects. In this dissertation, we will investigate the nature of these variations in the EEG distributions, and introduce two new, complementary methods that we have devised to overcome these two key issues.

To confirm the problem of non-stationary brain signals, we first show that BCIs based on commonly used signal features are sensitive to changes in the mental state of the user. We proceed by describing a method aimed at removing these changes in signal feature distributions. Based on the insight that a large class of BCIs is based on *relative changes* in spectral power, but uses absolute power for classification, we have devised a method that uses a second-order baseline (SOB) to specifically isolate these relative changes in neuronal firing synchrony. To the best of our knowledge this is the first BCI classifier that works on out-of-sample subjects without any loss of performance. With these SOB features, we have effectively bypassed the problem of slowly changing non-stationary distributions. Still, the assumption made by ML methods, that the training data contains samples that are independent and identically distributed (iid), is violated, because EEG samples nearby in time are highly correlated. This chronological structure is especially troublesome during the training of the classifier, since it may lead to overfitting due to an overestimation of the amount of independent information present in the calibration session. We derived a generalization of the well-known SVM classifier, that takes the chronological structure of classification errors into account. Both on artificial data and real BCI data, overfitting is reduced with this dSVM, leading to BCIs with an increased information throughput.

With the SOB features we have addressed the first key issue (of investment of time) for a motor imagery task. It is likely that this approach also allows for cross-subject generalization of classifiers based on other neuronal signatures. The second key issue is partially addressed with the dSVM. The method demonstrates the feasibility of modeling the relatedness of brain signals recorded nearby in time, which is necessary to prevent overfitting in high-dimensional feature spaces derived from EEG. But, not assuming iid feature distributions raises new questions regarding the interpretation of BCI performance. With the two methods presented in this dissertation, we have paved the way for a new generation of BCIs — BCIs that work dependably, and without the need of recalibration.

Chapter 1

Introduction

A brain-computer interface (BCI) enables direct communication from the brain to devices, bypassing the traditional pathway of peripheral nerves and muscles. In the past, BCIs have been targeted mainly at paralysed patients or patients with motor disabilities who have hardly any other means of communication [21, 2, 73]. But the unique possibilities of BCI technology are by no means limited to those in need; BCI technology enables the use of signals related to attention, intentions and mental state, without relying on indirect measures based on overt behaviour or other physiological signals [75, 67, 35].

The traditional approach to BCIs is to provide the user with a device that is controlled through a fixed function of the brain signals, and let the users learn to voluntarily modify their brain signals, which takes weeks or even months [73]. An alternative, more user friendly approach is to adapt the BCI to the user's naturally occurring brain signals with machine learning (ML) methods (e.g. [69, 7]), which reduces the investment of time needed for the first use of a BCI from weeks to minutes. Due to inter-subject variations in the measured neuronal activity linked to specific mental tasks (the neuronal signature), a BCI session typically starts with a training session in which the subject performs a known series of mental tasks. After this session, these examples of brain activity are used to guide the optimization of a classification model that decodes neuronal activity based on a few seconds of the electroencephalography (EEG) signal. In this dissertation, we take this latter approach. Note that the two approaches mentioned above are not mutually exclusive.

The level of invasiveness is another major determining factor in BCI research. In this dissertation, we focus on the use of the electrical signals that are emitted by large ensembles of neurons, and are *measured on the outside of the scalp* (i.e. EEG). Most BCI research in Europe is based on these external EEG measures. Research groups in the United States generally take a more invasive approach, and place sensors under the scull, or even deep in the brain. This has the advantage that the measurements are less contaminated with external noise, and suffer less from spatial smearing (blurring) by the tissues between the neural sources and the sensors. The disadvantage is that one needs to undergo surgery, and that the implants usually work for a limited time. Less frequently used for BCIs are non-invasive measures of brain activity based on magnetic fields (i.e. magnetoencephalogram, MEG), or on depleting oxygen concentrations in the blood that indicate brain activity, such as functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS) [74].

Most of the current BCI research focusses on better neuronal signatures (i.e. neuroscience), better decoding of these neural signatures (i.e. signal processing, machine learning) and on developing specific applications for patients (e.g. speller grids). The traditionally clinical background of BCI practitioners is reflected in the focus on trial-based discrimination between a limited and tightly controlled set of tasks. Furthermore, the online evaluation is often performed in an environment similar to these controlled off-line experiments, with a classifier that is transplanted from off-line calibration data and used to classify batches of EEG *after* the task has been performed as indicated by a cue.

1.1 BCIs for healthy users

Compared to other groups, the recently started BCI research within our Human Media Interaction (HMI) group takes a more holistic approach, and attempts to create user friendly BCIs based on established neuronal signatures, and evaluate these in unconstrained environments both on efficacy, user experience and ethical considerations. HMI aims at BCIs for healthy users, specifically applied in gaming contexts.

BCI gaming applications are interesting for several reasons. First, the target population is huge, and gamers are known to be early adopters of new technology. Second, a less than stellar BCI efficacy might not be problematic in the context of a game, and might even contribute an immersive challenge (i.e. learn to control a magical in-game construct) [46]. If gamers embrace BCI technology, this will provides scales to mass-produce hardware and fund more research, which could eventually lead to feasible applications outside gaming. However, premature commercial exploitation of BCI technology is feared by the field, as claims of interpreting brain signals are often exaggerated by commercial parties — which could lead to a disappointed public.

The implied transition from a controlled lab to an unconstrained gaming environment poses some new challenges. During gaming, signals produced by facial expressions, speech and eye movement heavily contaminate the, in comparison weak, EEG signals. As such, some of the research at HMI explores the challenges and drawbacks of BCI combined with for example speech recognition [27]. Most of our studies allow the user to behave naturally. The drawback is that this implies careful interpretation of what measures are based on neuronal signals, and to what extent this is of importance for the target audience.

These challenges of unconstrained environments are balanced by some rather unique possibilities offered by the BCI. For example, the game can use indicators of imminent movement to anticipate future actions of the user, thereby blurring the boundary between the user's intentions and explicit behaviour in interaction. More fundamentally, measures of the user experience (e.g. workload, attention, or emotional states) can — if reliable measures are found — be used to adapt the game to keep the user in a state of being fully focused and immersed in the game. This mental state is known as "flow" [16]. For this reason, some of our work focusses on automatic recognition of mental states [56].

Another unique property of BCIs is that most conventional neural signatures are related to some form of attention. For example, imaginary movement produces changes in the sensory motor rhythm (SMR), and is modulated by attention [37]. Similarly, spelling applications for patients frequently use the P300 response that is strongly linked to attention [26]. Other examples include the steady-state visually evoked potential (SSVEP) response to flickering lights, and direct measures of visual attention [67]. This attentional aspect of these neural signatures is what makes them viable for active BCI control.

By measuring attention, a BCI can offer valuable information that measures of behaviour never can: it can provide an indication of the *context* that disambiguates the users behaviour. For example, a user making a phone call could direct speech commands to the computer if we could detect that the computer is the object of the user's (covert) attention.

These unique applications depend on a BCI that can reliably detect naturally occurring brain activity. Two key issues in the decoding of brain signals complicate the development of these applications.

1.2 Key challenges for BCI adoption

We identify two key issues that need to be addressed for wide BCI adoption in general: 1) using a BCI should be easy, and not require big investments by the user (e.g. money and time), and 2) the BCI should be dependable, that is to say it should function predictably, with a known accuracy [17, 68, 30][1]. In addition, the BCI should be applied such that it provides something unique for non-patients, since it cannot (yet) compete on reliability and speed with existing input devices.

Due to the availability of relatively cheap, semi-dry commercial EEG headsets for gaming that show promise for proper BCI use [9], the first key issue mainly revolves around investments of time. The investment of time can be separated into user training and setup time; both should be kept very short. The maximally acceptable setup time for amyotrophic lateral sclerosis (ALS) patients is around 30 minutes, with the maximum of 2–5 sessions in total for user training [30]. While current BCIs based on machine learning (ML) methods can achieve competitive performance without any user training [69, 4], the time needed to set up the recording equipment and to record a calibration session typically exceeds this acceptable setup time. Non-patients are probably even less willing to accept long investments of time. Ideally, we would like to reduce the setup time to one or two minutes, and remove the calibration and user training time altogether.

An example of the second key issue is described in the review of Wolpaw et al. [73], where ALS patients in the experiment by Kübler et al. [39] report to prefer a slower character-based speller over a faster word-based speller. They felt more independent since the former was completely under their control. Being in control implies that the BCI should not behave unexpectedly, but not necessarily that the BCI operates without errors. Lack of errors is not strictly needed since there is a fundamental trade-off between the number of errors and the speed of a BCI — the error rate can be reduced by integrating predictions over a longer period of time. But this trade-off only holds if the BCI makes mistakes with a constant probability. Therefore, a BCI with a constant error rate should be preferred over a BCI with a variable, but lower error rate, since the former can be relied upon, if slowed down to acceptable error rates.

Unexpected, erratic BCI behaviour can be caused by non-stationary signals. This is known to be a fundamental problem in the BCI field. The inherent variable nature of spontaneous EEG causes changes in the feature distributions used by the BCI to detect and classify brain signals. One source of this variability in the EEG is related to changes in the user state. For ex-

[1] Note that a known accuracy does not imply that the accuracy has to be high.

ample, differences in levels of alertness, fatigue, frustration and workload level can alter the characteristics of the EEG. This variability violates basic assumptions made by ML methods used to train the BCI classifier, and can result in a loss of performance during the application of the BCI [62, 5, 36]. Most of the published work on non-stationary signals in BCIs focusses on the changing distribution of the EEG, and explicitly attempts to reduce feature variability over time [28, 66, 5, 72, 44], or alternatively, to adapt the classifiers parameters to the changing distribution [62, 70]. An unsolved problem is that it is unclear how variability in the feature distributions influences the BCI performance, since commonly used spatial filtering methods can already remove task-irrelevant fluctuations to some degree. In this case, attempting to remove the variability could introduce new problems that are caused by difficulties in estimating the unrelated variability in EEG features.

Both key issues are interrelated, since they are based on not fully understood properties of fluctuation in feature distributions over time, sessions and subjects. In this dissertation, we will investigate the nature of these variations in the EEG distributions, and present two new, complementary methods that we have devised to overcome the key issues we have described.

1.3 Contributions

To ground the problem, we will show in Chapter 2 that even changes in the mental state of the user can induce changes in the EEG signals, and that BCIs based on commonly used signal features are sensitive to these changes. We will proceed in Chapter 3 by describing a method aimed at removing these changes in signal feature distributions. Based on the insight that a large class of BCIs are based on *relative changes* in spectral power, but uses absolute power for classification, we will describe the new second-order baseline (SOB) features that specifically isolate these changes in neural firing synchrony, thereby removing long-term and subject-specific deviations. Still, the assumption made by ML methods that the training data contains samples that are independent and identically distributed (iid) is violated, since samples nearby in time are highly correlated. This temporal dependence is especially troublesome during the training of the classifier: due to the overestimation of the amount of independent information contained in the training set it leads to overfitting. In Chapter 4 we will present a generalization of the well known support vector machine (SVM) classifier, that takes the temporal dependence of features (and hence the dependence of classification errors) into account. Both on artificial data and real BCI data, overfitting is reduced with this dependent-samples support vector machine (dSVM), leading to an increased information throughput.

Chapter 2

Influence of loss of control

IT is widely believed that BCI performance fluctuates over time due to non-stationary feature distributions [45, 5, 72, 69, 62, 36, 65, 73, 31]. These non-stationary feature distributions violate the basic assumption made by the classification models that the evaluation data (i.e. the online session) is distributed identically to the data the model was trained on. This problem is known as *covariate shift*, and leads to decreases in the classification performance. Among the hypothesised causes for these non-stationary feature distributions are changes in the mental state (e.g. fatigue, workload, loss of control (LOC)) and artifacts [5, 31]. Although it seems plausible that mental state changes that are detectable in the EEG can interfere with BCI operation, there is not much experimental evidence for this effect.

In this Chapter, we therefore describe an experiment[1] we performed to investigate the influence of a feeling of LOC on the detectability of movements with the left and right index finger through the EEG. Changes in the EEG related to users experiencing a state of LOC might lead to a decrease in BCI performance due to the aforementioned covariate shift. In turn, the user state is again influenced by the decreased performance of the BCI; for example, the non-working BCI could cause increased frustration, anger and reduced alertness. This interaction between the user state and the BCI performance might result in a positive feedback loop, leading to a BCI that spins out of control. Given the huge drawbacks of a BCI that can stop working depending on the mental state of a user, understanding the influence of changes in the mental state on the BCI is of great importance to develop reliable BCIs.

[1] The work described in chapter was accepted for publication in the journal IEEE Transactions of Neural Systems and Rehabilitation Engineering [55].

In the following sections we will describe previous work on the relation between mental states and BCI performance, the methods we used, our results, and a discussion of our experiment, followed by conclusions and directions for future research.

2.1 Previous work

The influence of frustration associated with LOC on a BCI is of great interest since it might cause the previously described feedback loop. This influence was previously investigated in [31, 76]. In this study, users were instructed to use real movement with their left or right hand to rotate respectively L or R-shaped objects to a target position in order to study the effect of loss of control on the BCI performance. The color of the letter indicated the angle of rotation, and the user could press a key to rotate the object in the direction indicated by the shape of the object every second. After performing a calibration block with cued left/right hand movement and two practise blocks with this so-called RLR paradigm, a LOC was simulated in the third block by occasionally using a wrong angle of rotation in the application. Both an event-related desynchronization (ERD) and an event-related potential (ERP) based classifier were trained on the first block, and applied to the other blocks in an off-line analysis. A significant difference between the training block and the LOC block was found for the distribution of ERD based features, but for ERP features no such difference was found. This seems to indicate that there is variability in ERD features related to loss of control.

However, the study described in [31, 76] is lacking on a few aspects. Most notable is the limitation that changes in BCI performance due to the induction of LOC, the progression of time, differences in stimulation and user behaviour cannot be distinguished. We were interested in the influence of LOC on the BCI performance independent of these other factors. Therefore we used 1) an interleaved block design to control for effects that manifest spontaneously over time, such as increasing fatigue, changing temperature, drying gel on the electrodes etc., 2) we used the same environment for training and evaluating the BCI classifiers to minimize environmental differences not related to LOC, 3) we used self-reported emotional ratings to validate the effect of loss of control on the mental state and 4) we tested and corrected for confounding behavioural changes, such as changes in the force, speed or order of the finger movements, and eye movements.

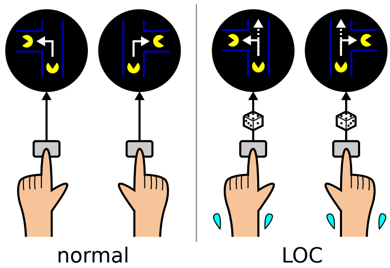

Figure 2.1: *In the normal condition, the left button rotated the player* 90° *counterclockwise; the right button rotated the player* 90° *clockwise. In the LOC condition, 15% of the keyboard input was ignored, and a visual lag was induced (not shown).*

2.2 Methodology

To study the effect of loss of control (LOC), we designed a Pacman game that periodically reduced the amount of control the user has over his avatar. The EEG was passively recorded during game play, and afterwards the offline performance of an ERD and an ERP based classifier was used to assess the influence of LOC on the BCI performance. In the rest of this section, we describe the data collection, the preprocessing and classification of the EEG, and the evaluation method in more detail.

2.2.1 Data collection

A game was designed to induce a state of LOC, with game play similar to the original Pacman game [52]. The major differences with other Pacman games is that our game periodically tried to induce a state of LOC in the user by responding unreliably to the keyboard commands. Since unreliable input is a proven method for frustration induction [60, 32, 19], we expected this method to induce mental state changes that were naturally associated with LOC. To simplify (simulated) BCI control, the user input was reduced to a button for the left index finger that turned Pacman 90 degrees counterclockwise, and a button for the right index finger that turned Pacman clockwise.

Experiment design

The LOC was induced in a randomized interleaved block design with experimental blocks of two minutes. In one third of these two-minute blocks LOC was induced, in the other blocks the game play remained unmodified. The LOC blocks were evenly distributed over the session by building a series of shuffled sequences of three blocks (one LOC and two normal blocks). In the LOC condition, the game randomly ignored 15% of the keystrokes, resulting in a barely playable game. In addition, the display occasionally lagged in the LOC condition. After each block, the user was asked to rate his mental state in terms of valence (pleasure), arousal and dominance (subjective feelings of control) on a Likert-scale presented under the self-assessment manikin (SAM) [10].

Experimental procedure

Subjects were asked to read and sign a form of consent, and were subsequently wired with the EEG and physiological sensors. The experimenter briefly explained the game and the self-assessment procedure. The subject was allowed to practise the controls for two minutes before the experiment was started. If users mentioned that the game was unresponsive during the experiment, the experimenter asked them to continue playing and promised to find the cause later. After 30 minutes, the experimenter stopped the experiment and the users were debriefed.

Sensors and recording

A BioSemi ActiveTwo EEG system was used to record the EEG and physiological signals at a sample rate of 512 Hz. EEG was recorded with 32 Ag/AgCl active electrodes placed at locations of the Extended International 10-20 system. To measure the influence of ocular and muscle artifacts, we recorded the EOG (horizontal and vertical pairs) and two pairs of EMG signals over the left and right flexor digitorum profundus (the muscles used to press with the index finger). Additional physiological sensors, such as temperature, respiration, the galvanic skin response and the blood volume pulse were recorded as well, but not used in the present study[2].

2.2.2 Preprocessing

The following preprocessing procedure was applied to reduce the influence of noise and artifacts caused by eye movements and muscle tension: first the recording was downsampled to 128 Hz to speed up processing. After downsampling, the data was high-pass filtered using a 4th-order Butterworth filter to remove frequencies below 0.2 Hz, and notch-filtered using a

[2]The recordings are available at `http://borisreuderink.nl/perm/affpac/`.

4th-order Butterworth filter from 49–51 Hz to remove power line noise. The EEG was then corrected for eye movements using a regression based subtraction method [61]. To prevent noise from spreading to other channels, we performed channel-level preprocessing before we applied the electrooculography (EOG) correction and re-referenced the signals to the common average reference (CAR).

2.2.3 Key press classification from EEG

Most motor imagery based BCIs are based on sensory-motor rhythms, specifically the event-related desynchronization (ERD) that occurs during both real and imaginary movement. As the ERD of real and imagined hand movement is similar [43], we used real movement to train BCIs that predict the movement from the EEG signal, since it provides a clear ground truth and allows for a tighter controlled experiment. In this section we will outline the classifiers used for detection of the ERD and the ERP associated with the movements executed to play the game.

ERD features

The ERD classification was based on the decrease in the Rolandic mu rhythm (8–12 Hz) and Rolandic beta frequencies (peak around 20 Hz) on the contra-lateral motor cortices that occurs when movement is initiated [48]. After preprocessing, we applied a 6th-order Butterworth band-pass filter to extract the frequencies from 8–30 Hz, which includes both the mu and beta rhythms. From this filtered EEG we extracted windows of one second, centered on the moment that keystroke was registered. Visual inspection confirmed that an ERD did indeed occur within this period. For these segments we trained subject-specific spatial filters with the common spatial patterns (CSP) algorithm.

The CSP algorithm [34] finds a matrix $\tilde{W}$ with spatial filters that map the EEG into a new space with basis vectors that have a high variance for the first class and a low variance for the second, and vice versa. Given the number of sensors s, and the number of samples n, W is an $s \times s$ transformation matrix with the following property:

$$\Sigma_{WX_1} = D \quad \text{and} \quad \Sigma_{WX_1} + \Sigma_{WX_2} = I, \tag{2.1}$$

where D is a diagonal matrix with elements in descending order, I is the identity matrix and Σ_{X_i} is the channel covariance matrix of the $s \times n$ EEG measurements matrix X for given class i. Rows of W that correspond to a high value in D have a high variance (power) for the first class and a low variance for the second, and vice versa. Because of this discriminatory property, the $\frac{m}{2}$ first and the $\frac{m}{2}$ last rows were picked to construct the final matrix $\tilde{W}$ with $m = 6$ spatial filters.

After applying the CSP algorithm, we calculated the variance (which corresponded to the band power in the mu and beta band) for each transformed channel, which resulted in m spatial band-power features.

ERP features

A less frequently used paradigm for classification of EEG related to movement is based on the Bereitschaftspotential (BP), a negative ERP related to movement initiation. The BP consists of an early phase beginning about 2 seconds before the movement onset, and a late phase with a steeper slope 400 ms before the onset [63]. We used the asymmetric distribution of the late BP over the scalp for classification of the laterality of the hand movements, which is known as the lateralized readiness potential (LRP).

For the ERP classification, we used the same preprocessing pipeline as used with the CSP classification up to the band-pass filter. Then we applied a (4th-order Butterworth) low-pass filter at 10 Hz, and again extracted windows of one second centered on the moment of registration of keyboard input. These trials were then transformed with a whitening transform P which has the property that the transformed signals are uncorrelated, and have unit variance:

$$\Sigma_{PX} \approx P\Sigma_X P^T = I. \tag{2.2}$$

With the eigenvalue decomposition $\Sigma_X = U\Lambda U^T$, we find that

$$P = \Lambda^{-\frac{1}{2}} U^T. \tag{2.3}$$

After whitening with P, we downsampled the signal by taking every fourth sample point, resulting in a $s \times \frac{e}{4}$ feature vector where $e = 128$ is the number of samples in a classification window. Despite superficial differences, this method for LRP classification is conceptually similar to the conventional approaches for ERP detection, such as [3, 8], but does not rely on time segment or channel picking.

Classification

The ERD and LRP features were used to train a final linear SVM classifier. The SVM's regularization parameter c was selected with a separate cross-validation loop on the two-minute blocks in training set.

2.2.4 Loss of control analysis

To analyze the influence of user LOC on the performance of a BCI, we trained a BCI on normal blocks, and measured the difference in performance of the same classifier between unseen, normal blocks and unseen blocks from the LOC condition. As we could not assume that the distribution of the EEG signal would remain stationary, training and evaluating a BCI on samples uniformly spread over the sessions might lead to an overestimation of the performance. In order to have a more reliable measure of how an online BCI would perform, we therefore used a special evaluation scheme where complete experimental blocks were left out for evaluation: The session consisted of a series of permutations of three experimental blocks; two normal blocks, and a block with LOC simulation. For every three blocks, we added the first normal block to the training set for the BCI classifier. The remaining normal and LOC block were used for evaluation. This way, the training data was spread over time, but we still have independent blocks for evaluation. Note that this evaluation is not symmetrical, since the classifiers were trained only on normal blocks, but tested on blocks of both the normal and LOC condition.

If there were difference between the normal and LOC conditions, we expected to find a lower performance on LOC blocks compared to block with normal control, since the model was optimized for a different distribution than the observations it was evaluated on had.

2.2.5 Performance measure

BCI classifiers are often evaluated by comparing their accuracy on out-of-sample trials. The choice for the accuracy measure is problematic, as accuracies (or equivalently, error rates) are hard to interpret when the prior probabilities of the classes are unequal and/or variable. Furthermore, the statistic does not take the time needed to perform a trial (key press) into account: due to our short inter-trial intervals (ITIs), lower accuracies were to be expected for our BCIs than the accuracies reported for more traditional BCI environments, where multiple seconds are used to detect an imagined movement. Despite these drawbacks, we will provide accuracy measures because it is commonly used.

A more informative measure is the information transfer rate (ITR), which conveniently captures the amount of information a user can communicate through a (noisy) channel with an optimal encoding strategy. It does so by combining the quality of and the time needed for the predictions. As such, the ITR is a better measure to evaluate BCI performance. Note that different formulas to calculate the ITR are used in the BCI literature, for example Wolpaw's definition in [73] is often used. The drawback of this definition is that it has a number of assumptions that are often violated in practice, most notably the assumption that all classes have the same prior probability. The ITR based on mutual information (MI) does not rely on these assumptions, hence we use MI to measure the information contained in the prediction of a single trial. Note that the labels of the trials still need to be independent of each other for a correct estimate of the ITR.

The MI expresses the decrease in uncertainty of a discrete variable $\mathscr{Y}$ (the true labels), given a discrete variable $\mathscr{Z}$ (the predictions of the classifier):

$$I(\mathscr{Z};\mathscr{Y}) = \sum_{\substack{y\in\mathscr{Y}\\ z\in\mathscr{Z}}} p(z,y)\log_2 \frac{p(z,y)}{p_1(z)\,p_2(y)}, \tag{2.4}$$

where $p(z,y)$ is the joint probability distribution and $p_1(z)$ and $p_2(y)$ are the marginal probability distribution functions of $\mathscr{Z}$ and $\mathscr{Y}$. With the base-2 logarithm the reduction in uncertainty is expressed in bits. We use the MI between the classifiers predictions and the ground truth as a second performance measure. The joint and marginal probabilities in (2.4) were estimated by their relative frequency of occurrence in the confusion matrix.

Finally, we calculate the third measure ITR R, in bits per minute, based on the MI (2.4), and the median[3] inter-trial interval $\text{med}(\Delta t)$:

$$R = 60\frac{I}{\text{med}(\Delta t)}. \tag{2.5}$$

As a fourth, and last performance measure, we use the area under the curve (AUC) of the receiver operating characteristic (ROC) [22] to express the ranking performance of the classifier. The AUC is equal to the probability that a randomly chosen instance of the first class is ranked above a randomly chosen instance of the second class; in other words, an AUC of 0.5 indicates random performance, and an AUC of 0 or 1 indicates perfect ranking ability. Like the MI, the AUC does not assume equal prior probabilities.

Originally we planned to use the Kullback-Leibler divergence (KLD) as a measure of change in the feature distributions as in [62], but the assumption that the features are normally distributed was violated heavily by both our ERD features (even after log-transforming) and our LRP features. This made the estimation of the KLD unfeasible due to the need for high-dimensional density estimation.

[3]We use the median instead of the mean because Δt appeared to follow a Poisson distribution.

2.2.6 Confounding factors

To induce mental state change associated with LOC, we intentionally degraded the quality of control. Behavioural changes (e.g. repetitive and force-full keystrokes, and more frequent gazing at the hands) might have occurred as a result of the method of induction. Therefore, potential differences in BCI performance might have been caused solely by these changes in behaviour. In this context, behavioural changes are confounding factors, and need to be corrected for.

However, we cannot discern behavior caused by the method of induction and behaviour caused by induced changes in the mental state. Correcting for confounding factors could therefore reduce the variation in mental state, and lead to an underestimation of the effect. Therefore, we performed our analysis both with and without correction for confounding behaviour.

Behavioural changes that we identified as confounding factors were the inter-trial interval (ITI), the repetition of keystrokes with the same hand, the fraction of keystrokes per hand, the force used to press a key, and eye movements. The ITI can be confounding because the EEG is analyzed over a short period of time; keystrokes that follow each other quickly could lead to masking of relevant EEG features, or worse, to the leaking of label information from one keystroke to the next. Repetition of strokes with the same hand might lead to increased performance for the same reason. Force is a confounding factor because force has an influence on the ERD [64]. Artifacts related to eye movements are known to have a profound influence on EEG analyses, but these were (greatly) attenuated by the EOG regression method during preprocessing.

To correct for the confounding factors, we used multi-variate frequency matching. Frequency matching involves stratifying the distribution of the confounding variable, and drawing samples such that the number of samples within each stratum is the same per condition [1]. In our case, the multivariate distributions of the previously described confounding variables in the normal and LOC conditions were matched.

These confounding factors were quantified as follows. To quantify the ITI, we used the logarithm of the difference in seconds between consecutive trials. The keystroke patterns were modeled with a discrete bivariate distribution of the label of the current and previous trial. For force, a bivariate (i.e. left and right arm's EMG power) distribution of the log-transformed electromyography (EMG) power was used. To calculate the EMG power, the procedure outlined in [29] was used: 1) apply a high-pass filter with a cut-off of 30 Hz, 2) apply the Hilbert transform to extract the envelope of the signal and 3) apply a low-pass filter with a cut-off of 40 Hz to smooth the signal.

A multi-dimensional histogram with regularly spaced bins was used to extract strata for frequency matching: 4 bins were used for log ITI, 2×2 bins were used for label patterns, and 5×5 bins were used for log EMG power for index fingers.

2.2.7 Statistical tests

Comparisons over subjects were performed using Wilcoxon signed-rank tests, on pairs of per-condition averages for each subject. This test is a non-parametric alternative to the commonly used paired Student's t-test, which could not be applied because the t-test's assumptions that the measurements are normally distributed and have equal variances do not hold for classification performance [18].

In addition to this over-subjects analysis, we performed a more sensitive meta analysis that combines the within-subject p-values to test for individual differences (as opposed to group differences). It combines the p-values of different subjects to reject the combined null hypothesis H_0, that states that each of the individual null hypotheses is true. The combined alternative, H_A, is that at least one is not true. For this purpose, Fisher's method was suggested in [42] for combining p-values:

$$X^2 = -2 \sum_{i=1}^{k} \log_e p_i, \tag{2.6}$$

where p_i is the p-value for subject i. When the null hypotheses are all true, and the p_i's are independent, X^2 follows a χ^2 distribution with $2k$ degrees of freedom. Note that opposing effects might be combined in a significant outcome with the Fisher's method if two-sided tests are used.

We used a significance level $\alpha = 0.05$ for all tests presented in this chapter.

2.3 Results

2.3.1 Subjects

Twelve healthy users (age 27±3.9) participated in the experiment. All participants had normal, or corrected to normal vision, and reported not to use medication. Only three of our subjects were female, and all subjects were right-handed. Most participants had some video game experience, and four subjects had previous experience with BCIs.

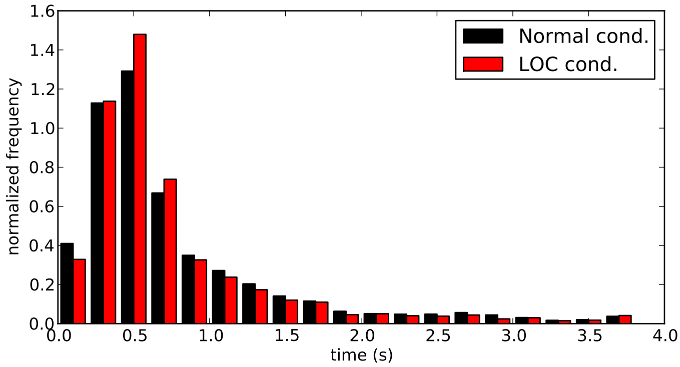

Figure 2.2: *The histogram for the time between key presses during game play for all subjects. The intervals for the normal condition are displayed in black, the LOC condition is displayed in red (gray). The histogram is dominated by short Δt's between key presses. The histogram of the LOC condition seems to be slightly more pinched around half a second.*

2.3.2 Self assessments

To verify the induction of changes in the mental state, we analysed the self reported emotional ratings of the SAM. Most subjects rated the LOC condition more negatively than the normal condition, over subjects this difference was significant (T=3, p<0.01). While we expected to find a trend towards more arousal in the LOC condition, there was no significant difference (T=23, p=0.26). The dominance dimension, which measures the amount of dominance, or control they have on their environment, indicate that people seemed to be significantly (T=3.5, p<0.01) more in control in the normal condition.

2.3.3 Confounding behavioural differences

In this section we describe the analysis of the characteristics of the user's behaviour, as it might have had a confounding influence on the BCI performance. Both differences in the ITI, and the pattern of consecutive keystrokes can indicate a confounding behavioural change. The per-subject statistics for these confounding factors are presented in Table 2.1. For the log ITI, we see an insignificant tendency to shorter intervals between key presses in the LOC condition. The probability that a key press was made with the same hand is significantly higher in the LOC condition. This may have been caused by increased repetition, by increased imbalance of the class ratios, or

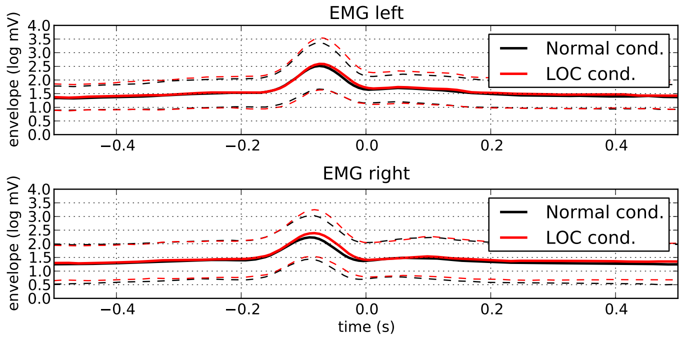

Figure 2.3: *The mean log EMG power and its standard deviation (dashed lines) is displayed for both the normal and LOC condition, time-locked to the key press at $t = 0$. The top plot shows the EMG power of the left bipolar channel for left index-finger presses, the bottom plot show the EMG power for right hand movement measured with the right bipolar channel.*

Table 2.1: *Statistics of confounding variables. The repetitiveness (second column) and log-EMG power (last column) differ significantly between the conditions. The EMG power was quantified as the maximum power in the interval [-0.2, 0] s. The start and end of the arrow signify the mean value in the normal and LOC condition respectively.*

	$\log \Delta t$	$p(y_t = y_{t-1})$	log pow. EMG L	log pow. EMG R
S0	-0.48 → -0.36	0.51 → 0.59	1.31 → 1.21	1.08 → 1.15
S1	-0.51 → -0.53	0.54 → 0.59	2.70 → 2.60	2.04 → 2.17
S2	-0.54 → -0.59	0.50 → 0.52	2.65 → 2.58	2.51 → 2.46
S3	-0.47 → -0.61	0.56 → 0.54	1.99 → 1.88	2.15 → 2.31
S4	-0.46 → -0.55	0.53 → 0.58	3.14 → 3.35	1.76 → 1.86
S5	-0.46 → -0.44	0.56 → 0.60	2.70 → 2.56	4.06 → 3.91
S6	-0.46 → -0.48	0.55 → 0.55	2.11 → 2.55	2.14 → 2.47
S7	-0.40 → -0.55	0.58 → 0.56	1.70 → 1.71	2.13 → 2.14
S8	-0.37 → -0.44	0.47 → 0.56	2.89 → 2.95	2.94 → 3.02
S9	-0.58 → -0.55	0.45 → 0.54	3.02 → 3.08	2.01 → 2.06
S10	-0.58 → -0.44	0.47 → 0.53	2.43 → 2.42	2.52 → 2.55
S11	-0.45 → -0.42	0.52 → 0.58	3.01 → 2.86	2.97 → 3.13
mean	-0.48 → -0.50	0.52 → 0.56	2.47 → 2.48	2.36 → 2.44
Wilc.	T=32, p=0.583	**T=6, p=0.010**	T=31, p=0.530	**T=12, p=0.034**

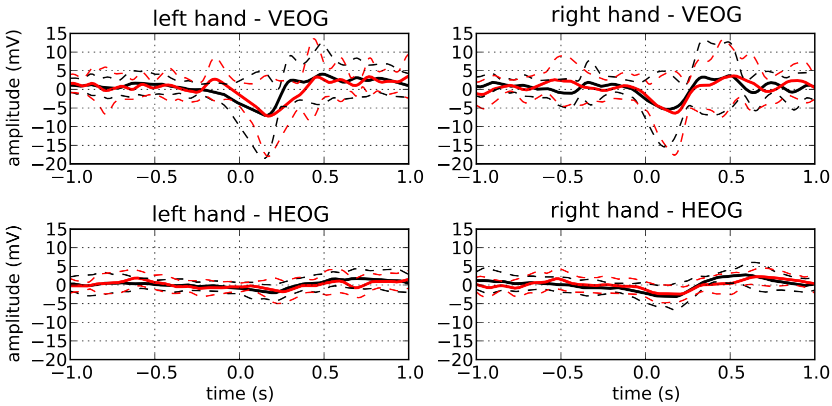

Figure 2.4: *The vertical eye movement (first row) show downward eye-gaze just after a keystroke at $t = 0$. The horizontal bipolar EOG channel is rather uneventful, only for right hand movement (last column) does there appear to be a delayed reaction to the key press, first negative (looking left) then positive (looking right). The dashed lines indicate the standard deviation. There is no significant difference (16 point Bonferroni corrected Wilcoxon signed-rank test over subjects) between the normal (black) and LOC condition (red).*

a combination thereof. Nevertheless, it indicates a significant behavioural change.

The temporal development of the EMG signal is displayed in Fig. 2.3. An increase in the EMG power is visible just before the stroke is registered, and a much weaker increase is visible when the key is released. Most of the activity is registered in the interval [-0.2, 0] s relative to the registration of the key press. We used the maximum EMG power in this interval to estimate the force used to press a key (see Table 2.1). Movements with the right index finger produce significantly more EMG power in the LOC condition.

Although we removed (most) of the influence of the EOG signal from the EEG, it is interesting to look at the user's gaze and blink behaviour during a key press (Fig. 2.4). We can see that users tend to look at their hands 200 ms after a key press, which is most visible in the vertical EOG, and at 300 ms, the variability of the vertical EOG signal seems to increase. This might be caused by eye-blinks, or an adjustment to the new movement direction of the avatar in the game.

In summary, our behaviour analysis has shown that the normal and LOC-conditions are very similar in the timing, the predictability of the key-strokes, the amount of force used to press the keys, and in eye movements.

Table 2.2: *The influence of LOC on a CSP classifier is shown below, without correction for confounding factors. The start and end of the arrow indicate the median performance for the normal and LOC condition respectively. The p-value of a Mann-Whitney U test on the per-block performance is displayed above the arrow. The row denoted with "Wilc." signifies the over-subject comparison with the Wilcoxon signed rank test. The row denoted by "Fish." presents the results of combining one-sided p-values for an increase in performance.*

	Accuracy	AUC	MI	ITR
S0	$0.679 \xrightarrow{p=0.35} 0.736$	$0.759 \xrightarrow{p=0.35} 0.843$	$0.110 \xrightarrow{p=0.35} 0.193$	$11.680 \xrightarrow{p=0.48} 13.763$
S1	$0.599 \xrightarrow{p=0.48} 0.618$	$0.616 \xrightarrow{p=0.64} 0.652$	$0.021 \xrightarrow{p=0.64} 0.041$	$2.601 \xrightarrow{p=0.64} 4.525$
S2	$0.619 \xrightarrow{p=0.92} 0.555$	$0.543 \xrightarrow{p=0.62} 0.578$	$0.001 \xrightarrow{p=0.77} 0.000$	$0.088 \xrightarrow{p=0.92} 0.025$
S3	$0.474 \xrightarrow{p=0.13} 0.519$	$\mathbf{0.466} \xrightarrow{p=0.05} \mathbf{0.518}$	$0.004 \xrightarrow{p=0.62} 0.003$	$0.452 \xrightarrow{p=0.77} 0.338$
S4	$0.519 \xrightarrow{p=0.92} 0.507$	$0.549 \xrightarrow{p=0.92} 0.526$	$0.002 \xrightarrow{p=0.77} 0.003$	$0.224 \xrightarrow{p=0.77} 0.351$
S5	$0.741 \xrightarrow{p=0.92} 0.750$	$0.828 \xrightarrow{p=0.92} 0.832$	$0.167 \xrightarrow{p=0.92} 0.188$	$15.616 \xrightarrow{p=0.92} 20.297$
S6	$0.538 \xrightarrow{p=0.65} 0.529$	$0.543 \xrightarrow{p=0.86} 0.522$	$0.003 \xrightarrow{p=0.59} 0.006$	$0.302 \xrightarrow{p=0.59} 0.664$
S7	$0.544 \xrightarrow{p=0.10} 0.600$	$0.565 \xrightarrow{p=0.27} 0.657$	$0.005 \xrightarrow{p=0.19} 0.027$	$0.532 \xrightarrow{p=0.08} 3.082$
S8	$0.542 \xrightarrow{p=0.34} 0.581$	$0.556 \xrightarrow{p=0.34} 0.589$	$0.004 \xrightarrow{p=0.48} 0.010$	$0.366 \xrightarrow{p=0.48} 1.035$
S9	$0.735 \xrightarrow{p=0.15} 0.768$	$0.797 \xrightarrow{p=0.15} 0.839$	$0.160 \xrightarrow{p=0.15} 0.218$	$18.879 \xrightarrow{p=0.15} 24.494$
S10	$0.612 \xrightarrow{p=0.35} 0.582$	$0.670 \xrightarrow{p=0.48} 0.651$	$0.046 \xrightarrow{p=0.35} 0.018$	$5.611 \xrightarrow{p=0.35} 2.108$
S11	$0.608 \xrightarrow{p=0.34} 0.553$	$0.594 \xrightarrow{p=0.95} 0.599$	$0.020 \xrightarrow{p=0.95} 0.023$	$2.371 \xrightarrow{p=0.64} 2.059$
mean	$0.601 \rightarrow 0.608$	$0.624 \rightarrow 0.651$	$0.045 \rightarrow 0.061$	$4.893 \rightarrow 6.062$
Wilc.	T=31.0, p=0.530	**T=12.0, p=0.034**	**T=14.0, p=0.050**	T=17.0, p=0.084
Fish.	p=0.123	p=0.078	p=0.259	p=0.222

However, there was a small but significant increase in repetition of the same movement, and a small significant increase in the force used with the right hand. After balancing the confounding variables and their interactions per subject, on average 25% of the original trials were removed.

2.3.4 Impact of loss of control on the BCI

To investigate the influence of LOC on the BCI performance, we trained the ERD based and the ERP-based classifier on blocks from the normal condition, and compared the performance on unseen normal blocks with the performance on blocks from the LOC condition. Please refer to Section 2.2.4 for more information on this procedure.

The performance of the CSP based features classifier on the normal and LOC blocks without correction for confounds is displayed in Table 2.2. The single trial detection accuracy may seem rather low (60%), but this is sim-

Table 2.3: *The influence of LOC on a CSP classifier is shown below, with correction for confounding factors enabled. Please refer to Table 2.2 for an explanation.*

	Accuracy	AUC	MI	ITR
S0	0.678 $\xrightarrow{p=0.64}$ 0.667	0.718 $\xrightarrow{p=0.35}$ 0.785	0.094 $\xrightarrow{p=0.82}$ 0.056	7.370 $\xrightarrow{p=0.95}$ 4.896
S1	**0.543** $\xrightarrow{p=0.05}$ **0.646**	0.572 $\xrightarrow{p=0.09}$ 0.686	**0.007** $\xrightarrow{p=0.05}$ **0.063**	0.734 $\xrightarrow{p=0.09}$ 6.751
S2	0.640 $\xrightarrow{p=1.00}$ 0.626	0.558 $\xrightarrow{p=0.92}$ 0.548	0.009 $\xrightarrow{p=0.19}$ 0.002	0.825 $\xrightarrow{p=0.27}$ 0.165
S3	0.541 $\xrightarrow{p=0.62}$ 0.553	0.483 $\xrightarrow{p=0.62}$ 0.514	0.007 $\xrightarrow{p=0.77}$ 0.005	0.482 $\xrightarrow{p=0.62}$ 0.450
S4	0.569 $\xrightarrow{p=0.19}$ 0.532	0.572 $\xrightarrow{p=0.77}$ 0.536	0.013 $\xrightarrow{p=0.92}$ 0.008	1.416 $\xrightarrow{p=0.92}$ 0.769
S5	0.753 $\xrightarrow{p=0.77}$ 0.743	0.857 $\xrightarrow{p=0.62}$ 0.821	0.198 $\xrightarrow{p=0.77}$ 0.180	14.086 $\xrightarrow{p=0.77}$ 15.423
S6	0.532 $\xrightarrow{p=0.21}$ 0.558	**0.498** $\xrightarrow{p=0.01}$ **0.579**	0.004 $\xrightarrow{p=0.10}$ 0.011	0.308 $\xrightarrow{p=0.06}$ 0.874
S7	0.549 $\xrightarrow{p=0.37}$ 0.577	0.613 $\xrightarrow{p=0.92}$ 0.623	0.010 $\xrightarrow{p=0.49}$ 0.015	0.964 $\xrightarrow{p=0.37}$ 1.534
S8	0.543 $\xrightarrow{p=0.48}$ 0.605	0.549 $\xrightarrow{p=0.23}$ 0.583	**0.002** $\xrightarrow{p=0.02}$ **0.017**	**0.166** $\xrightarrow{p=0.02}$ **1.394**
S9	0.687 $\xrightarrow{p=0.23}$ 0.740	0.783 $\xrightarrow{p=0.23}$ 0.833	0.086 $\xrightarrow{p=0.15}$ 0.162	8.770 $\xrightarrow{p=0.15}$ 16.045
S10	**0.658** $\xrightarrow{p=0.05}$ **0.617**	0.676 $\xrightarrow{p=0.82}$ 0.673	0.054 $\xrightarrow{p=0.24}$ 0.035	4.888 $\xrightarrow{p=0.35}$ 3.147
S11	0.585 $\xrightarrow{p=0.23}$ 0.551	**0.615** $\xrightarrow{p=0.01}$ **0.544**	0.021 $\xrightarrow{p=0.23}$ 0.005	1.329 $\xrightarrow{p=0.15}$ 0.366
mean	0.606 → 0.618	0.625 → 0.644	0.042 → 0.047	3.445 → 4.318
Wilc.	T=31.0, p=0.530	T=27.0, p=0.347	T=35.0, p=0.754	T=35.0, p=0.754
Fish.	p=0.237	**p=0.050**	p=0.063	**p=0.047**

ilar to the accuracies obtained in other studies that use short ITIs, such as [31]. This was also reflected in the mean ITR of 5.5 bits per minute, which is comparable to the ITRs obtained by naive users with motor-imagery based ERD BCIs. Despite this low recognition rate, the ERD BCIs performance did significantly *increase* in the LOC condition for the AUC and MI measures.

When correction for confounding factors was performed, the results were different (Table 2.3); the over-subject differences disappeared, but there were more significant within-subject differences in sometimes opposing directions. Combined with Fisher's method, the one-sided p-values for a within-subject increase in performance was significant for both AUC and ITR. This indicates at least one individual increase in performance was significant at the $\alpha = 0.05$ level.

The spatial distribution of the movement related ERD is shown in Fig. 2.5. Subjects S0, S1, S5, S9 and S10 do display the prototypical ERD on the motor cortices. Remarkably, these activations are more pronounced in the LOC condition (second row), which supports the observed increase in performance. Note that the CSP classification is based on covariance of the EEG channels, while in this figure only the variance is shown.

In contrast to the ERD based classifiers, the ERP classifiers had a constant high performance with a minimum ITR of 11.6 bits per minute. Fur-

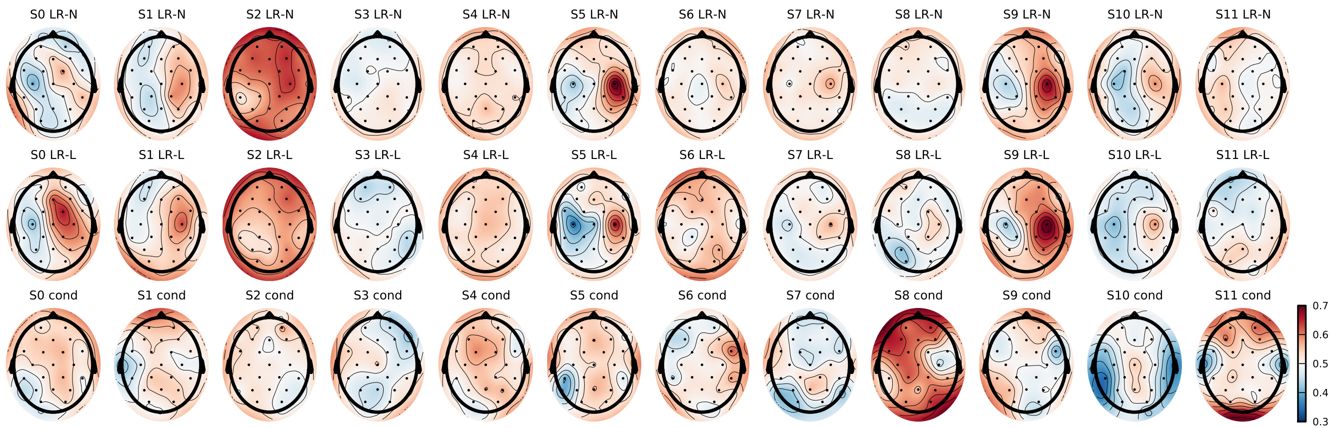

Figure 2.5: *These scalp plots display the difference between left and right hand movement in the normal (first row) and LOC condition (second row), and the difference between the normal and LOC in the last row. The color encodes the AUC-ROC ranking performance of the 8–30 Hz band power at the specified location; red indicates a positive rank correlation with the target class (right hand for the first two rows, or LOC in the last), blue a negative correlation. The conditions were corrected for confounding factors with frequency matching. Most subjects display a more pronounced spatial activation in the LOC condition.*

Table 2.4: *The influence of LOC on a ERP classifier is shown below, without correction for confounding factors. Please refer to Table 2.2 for an explanation.*

	Accuracy	AUC	MI	ITR
S0	$0.765 \xrightarrow{p=0.73} 0.747$	$0.824 \xrightarrow{p=0.95} 0.812$	$0.223 \xrightarrow{p=0.64} 0.179$	$24.349 \xrightarrow{p=0.48} 17.575$
S1	$0.802 \xrightarrow{p=0.34} 0.758$	$0.866 \xrightarrow{p=0.64} 0.823$	$0.282 \xrightarrow{p=0.23} 0.195$	$32.838 \xrightarrow{p=0.34} 22.092$
S2	$0.726 \xrightarrow{p=0.27} 0.750$	$0.790 \xrightarrow{p=0.49} 0.809$	$0.142 \xrightarrow{p=0.62} 0.154$	$16.269 \xrightarrow{p=0.37} 18.677$
S3	$0.711 \xrightarrow{p=0.77} 0.724$	$0.779 \xrightarrow{p=0.62} 0.788$	$0.133 \xrightarrow{p=0.77} 0.139$	$11.719 \xrightarrow{p=0.77} 16.195$
S4	$0.692 \xrightarrow{p=0.92} 0.714$	$0.767 \xrightarrow{p=0.37} 0.803$	$0.096 \xrightarrow{p=0.92} 0.110$	$11.672 \xrightarrow{p=0.77} 11.882$
S5	$0.720 \xrightarrow{p=0.77} 0.715$	$0.800 \xrightarrow{p=0.92} 0.790$	$0.142 \xrightarrow{p=0.92} 0.136$	$14.584 \xrightarrow{p=0.62} 15.041$
S6	$0.703 \xrightarrow{p=0.15} 0.726$	$0.766 \xrightarrow{p=0.15} 0.789$	$0.116 \xrightarrow{p=0.10} 0.155$	$13.460 \xrightarrow{p=0.10} 16.812$
S7	$0.772 \xrightarrow{p=0.77} 0.778$	$0.833 \xrightarrow{p=0.62} 0.857$	$0.225 \xrightarrow{p=0.77} 0.234$	$23.515 \xrightarrow{p=0.62} 26.844$
S8	$0.702 \xrightarrow{p=0.64} 0.742$	$\mathbf{0.778} \xrightarrow{p=0.05} \mathbf{0.830}$	$0.121 \xrightarrow{p=0.64} 0.175$	$10.529 \xrightarrow{p=0.23} 18.768$
S9	$0.831 \xrightarrow{p=0.48} 0.785$	$0.886 \xrightarrow{p=0.34} 0.873$	$0.328 \xrightarrow{p=0.48} 0.246$	$38.768 \xrightarrow{p=0.34} 29.488$
S10	$0.704 \xrightarrow{p=0.82} 0.701$	$0.790 \xrightarrow{p=0.95} 0.783$	$0.129 \xrightarrow{p=0.82} 0.120$	$14.146 \xrightarrow{p=0.95} 14.663$
S11	$0.835 \xrightarrow{p=0.81} 0.843$	$0.931 \xrightarrow{p=0.81} 0.921$	$0.344 \xrightarrow{p=0.64} 0.375$	$41.528 \xrightarrow{p=0.81} 38.285$
mean	0.747 → 0.749	0.818 → 0.823	0.190 → 0.185	21.115 → 20.5
Wilc.	T=32.0, p=0.583	T=30.0, p=0.480	T=36.0, p=0.814	T=37.0, p=0.875
Fish.	p=0.546	p=0.197	p=0.570	p=0.350

Table 2.5: *The influence of LOC on a ERP classifier is shown below, with correction for confounding factors enabled. Please refer to Table 2.2 for an explanation.*

	Accuracy	AUC	MI	ITR
S0	$0.710 \xrightarrow{p=0.35} 0.732$	$0.800 \xrightarrow{p=0.64} 0.820$	$0.134 \xrightarrow{p=0.48} 0.164$	$11.145 \xrightarrow{p=0.95} 9.779$
S1	$0.810 \xrightarrow{p=0.15} 0.777$	$0.870 \xrightarrow{p=0.48} 0.854$	$0.295 \xrightarrow{p=0.15} 0.233$	$28.990 \xrightarrow{p=0.23} 19.258$
S2	$0.720 \xrightarrow{p=0.92} 0.746$	$0.792 \xrightarrow{p=0.77} 0.801$	$0.121 \xrightarrow{p=0.62} 0.127$	$12.031 \xrightarrow{p=0.77} 12.806$
S3	$0.716 \xrightarrow{p=0.77} 0.702$	$0.763 \xrightarrow{p=0.62} 0.766$	$0.133 \xrightarrow{p=0.92} 0.134$	$9.135 \xrightarrow{p=0.27} 11.983$
S4	$0.696 \xrightarrow{p=0.62} 0.707$	$0.759 \xrightarrow{p=0.77} 0.787$	$0.110 \xrightarrow{p=0.77} 0.133$	$12.860 \xrightarrow{p=0.37} 13.445$
S5	$0.640 \xrightarrow{p=0.77} 0.649$	$0.729 \xrightarrow{p=0.92} 0.735$	$0.094 \xrightarrow{p=0.27} 0.071$	$7.364 \xrightarrow{p=0.92} 6.853$
S6	$0.671 \xrightarrow{p=0.47} 0.695$	$0.769 \xrightarrow{p=0.86} 0.760$	$0.086 \xrightarrow{p=0.47} 0.124$	$6.197 \xrightarrow{p=0.72} 7.992$
S7	$0.739 \xrightarrow{p=0.92} 0.733$	$0.824 \xrightarrow{p=0.77} 0.811$	$0.174 \xrightarrow{p=0.92} 0.164$	$15.149 \xrightarrow{p=0.92} 15.587$
S8	$\mathbf{0.709} \xrightarrow{p=0.05} \mathbf{0.799}$	$\mathbf{0.777} \xrightarrow{p=0.02} \mathbf{0.847}$	$\mathbf{0.106} \xrightarrow{p=0.05} \mathbf{0.265}$	$\mathbf{9.679} \xrightarrow{p=0.02} \mathbf{21.099}$
S9	$0.845 \xrightarrow{p=0.34} 0.803$	$0.925 \xrightarrow{p=0.23} 0.874$	$0.379 \xrightarrow{p=0.48} 0.278$	$35.293 \xrightarrow{p=0.34} 29.589$
S10	$0.671 \xrightarrow{p=0.82} 0.688$	$0.764 \xrightarrow{p=0.95} 0.765$	$0.089 \xrightarrow{p=0.82} 0.099$	$7.261 \xrightarrow{p=0.35} 9.269$
S11	$0.835 \xrightarrow{p=0.81} 0.822$	$0.921 \xrightarrow{p=0.34} 0.883$	$0.350 \xrightarrow{p=0.81} 0.331$	$23.343 \xrightarrow{p=0.64} 26.231$
mean	0.730 → 0.738	0.808 → 0.809	0.173 → 0.177	14.871 → 15.3
Wilc.	T=31.0, p=0.530	T=38.0, p=0.937	T=36.0, p=0.814	T=28.0, p=0.388
Fish.	p=0.392	p=0.484	p=0.456	p=0.162

thermore, they did not seem to behave differently in the LOC condition, not with (Table 2.5) and not without correction for confounding factors (Table 2.4). Visual inspection of the classifier's weights confirmed that the most discriminative features were located on the motor cortices, that is to say, the BCI was based on brain activity. The increase in performance for S8 is probably a false positive, since the combination over subjects with Fisher's method is not significant.

2.4 Conclusions and future work

In this chapter, we have presented an experiment in which a simulated non-responding brain-computer interface (BCI) controller was used to study whether changes in the user's mental state have an influence on the BCI performance. The self-reported emotional ratings confirmed that the loss of control (LOC) condition induced a more negative, and less dominant mental state. These different mental states were accompanied with minor behavioural changes for which we corrected the analysis.

Contrary to our expectations, we observed a significant performance *increase* during the LOC condition for the event-related desynchronization (ERD) based BCIs. For the event-related potential (ERP) based BCIs, we found no change in performance. The image of a BCI spiralling completely out of control that we sketched in the introduction appears to be an illusion. However, the difference in performance demonstrates that variabilities in the feature distributions related to LOC do in fact exist, and could be more dire under different circumstances.

For future work in this direction, a logical next step would be to investigate the origin of the increase of performance for ERD classifiers. We suspect that it might be related to a shift in attention from the game context during normal play to the movement of the hands in the LOC condition. Since the strength of the beta band event-related synchronization (ERS) is related to attention in constant isometric force motor tasks [37], an increase of attention on the motor task in the LOC condition could result in more pronounced beta ERD/ERS, and indirectly lead to better classification results. This would form an interesting hypothesis for a follow-up study. Related is also the study presented in [33] that shows a pronounced beta rebound when the observed movement does not match the movement the user was supposed to execute.

The recordings from our current experiments could also be analyzed for correlates with emotions, as we have user-reported ratings of emotions for every two-minute block in the experiment. The first steps in this direction have been taken in [56]. The recognition of emotions from electroencephalography (EEG) would be immensely valuable to both locked-in patients — who would otherwise have to verbalise their mood using other means, such as the P300 speller — and to healthy users.

Chapter 3

Cross-subject generalization

IN Chapter 2 we demonstrated that BCIs based on oscillatory (ERD) features are sensitive to changes in the mental state — changes that are unrelated to the mental task used to control the BCI. While the ERP based classifiers were proven to result in higher and more reliable performance than their ERD counterparts, here we focus on improving the ERD based classifiers that exhibit the problematic drifts in performance well known in the BCI field[1]. The reason we focus on a ERD based BCI is that it is well studied, and known to work when the user only imagines movement, and does not actually perform movement as was the case in Chapter 2; for lateralized readiness potential (LRP) classification we are not aware of BCI studies with fully imagined movement. Furthermore, improvement obtained with movement-related ERD might prove to be beneficial for other oscillatory brain signatures, such as those related to covert attention [e.g. 67].

Changes in BCI performance might be caused by changes in the generating processes (e.g. due to changes in attention, or different processes for motor control), or more generally, by changes in the joint distribution of task-relevant and task-irrelevant EEG features. This means that not only changes in the task-relevant neural activity, but also unrelated changes (e.g. LOC in the previous chapter) might affect the BCI. Changes or drifts in the feature distributions pose a problem for automatic discrimination, since the ability to generalize to new data of a static BCI classifier depends on the appropriateness of the learned decision function for the online data. When there is a change in the feature distribution compared to the calibration data (covariate shift), it is likely that the quality of the predictions will degrade; the BCI classifier does not generalize well to this data. This diffi-

[1] This chapter is based on [54], accepted for oral presentation at EMBC'2011.

culty of generalization holds for application of a trained BCI to future data within the same session, to new sessions of the same subject, and to other subjects, in increasing order of difficulty.

Being able to successfully apply a BCI to new sessions and subjects is relevant for various reasons. Practically, a BCI that does not require a calibration session before each use is much more user friendly. Secondly, if a BCI generalizes to new users this indicates that the mechanism used to classify is a generic one. This aids interpretation of the neural signatures used for control. And finally, if a BCI is known to operate robustly during changing contexts (i.e. mental states, recording environment), this means its operation is not depending on these contexts.

To achieve generalization under covariate-shift, either the classifier can be adapted to the changing feature distributions [71, 62], or the spatial filters themselves can be adapted [72, 44, 66, 28, 57, 5, 62]. The advantage of adapting the spatial filters is that a wide variety (e.g. disconnected sensors) of distribution changes can be tracked when they are estimated correctly. While adaptation of the classifier cannot handle all these changes, due to its simplicity it is easier to estimate, and can therefore be more robust. A further distinction can be made between methods that perform adaptation only during the (off-line) training phase [e.g. 72, 44, 5], and methods that perform online adaptation [71, 66, 28, 57, 62]. The off-line adaptation methods implicitly assume that there is a spatial subspace that is stationary, and retains most of the discriminatory information. It is unclear whether such a subspace does indeed exist. Therefore, we took the most flexible approach, that combines online adaptation with adaptation in feature space (i.e. adaptation of the spatial filters).

In this chapter, we will present a second-order baseline (SOB) procedure that reduces long-term, non task-related variations. This procedure enables the creation of a BCI that can be applied to new subjects *without a calibration session*. Since non-stationary distributions of specifically ERD based features have been reported in [62, 5, 31] and are supported by the results presented in Chapter 2, we evaluate our method with ERD based features — although it is in no way limited to this type of feature extraction.

In the next sections, we will describe prior work on subject-independent classification, outline our baselining method and describe an off-line experiment used to assess the performance on unseen subjects. Then we will describe and discuss the results, and end with conclusions and remarks for future work.

3.1 Previous Work

Progress has been made to make machine learning (ML) based BCIs generalize to new sessions and new users. The zero training method described in [36] is one of the first attempts to extend the applicability of the popular common spatial patterns (CSP) algorithm to generalize from one session to another session. The method attempts to find prototypical spatial filters from past sessions of a specific subject, and uses a small number of trials of the current session to update the BCI classifier. Using these prototypical filters a performance similar to CSP performance was obtained. Although this result is a promising step towards zero training, historical EEG data and a minimal calibration session are still required.

To overcome these limitations, an ensemble method [23] was developed that selects a sparse set of subject-dependent (SD) spatio-spectral filters derived from a large database with the recordings of 45 subjects. With a wide-band frequency filter (as used in our study), a subject-independent (SI) classifier performed almost as well (68% correct) as the average SD CSP classifier (70% correct). However, the SI classifier's predictions were post-processed with a non-causal bias-correction, which prevents online application. Without post-processing the best SI classifier still scores much lower than the SD classifiers with 63% of the trials correctly classified.

Combinations of different feature extraction methods and different classifiers were compared on their ability to discriminate between classes of imaginary movement in unseen subjects in [41]. Of all tested combinations, a filter-bank CSP classifier that used frequency filters with different bandwidths had the best SI performance (71% correct). This is slightly above the SI performance of naive log band-power features (68%), and far below the best SD classifier (82%).

These three studies indicate that constructing an SI BCI classifier that generalizes to new subjects is quite challenging. With complex feature extraction as done in [41] and spatial filter matching as done in [36, 23], the performance can approach the SD CSP performance.

3.2 Methods

During the initiation of imaginary movement, an ERD in the μ-band is often observed in the motor cortices. For BCIs based on ERD, the spatially filtered EEG is used to compute band-power related features which are then classified with a linear classifier.

Unrelated changes in the EEG signal that manifest over time pose a problem for this classification scheme, because the power, and not the change in power is used to classify the individual trials. To counter this problem, a pre-trial baseline is often used in neuroscientific experiments; for example, the trial power spectrum is often divided by the power spectrum obtained from a pre-trial baseline to study the ERD. Surprisingly, baselining is rarely used for BCIs, one exception being [38] where it did indeed result in a performance improvement. In this chapter we propose a new feature domain for ERD classification that uses a multi-variate pre-trial baseline to reduce the covariance between the EEG channels. Before we describe our baselining approach in more detail, we will re-introduce the CSP pipeline that serves as a control in this work.

3.2.1 CSP classification

The CSP algorithm [34, 51] was designed to find a set of m spatial filters that have a maximally different mean variance (band power) for two classes:

$$\Sigma_{WX} = I \tag{3.1}$$

$$\Sigma_{WX_+} = D, \tag{3.2}$$

where Σ_{WX} is the channel covariance of the band-pass filtered EEG signal X multiplied by the spatial filter matrix W, X_+ is the EEG signal generated during one specific task, I is the identity matrix and D is a diagonal matrix. The CSP transform can be decomposed into an unsupervised whitening transform to satisfy (3.1), and a class-specific (supervised) linear transformation to satisfy (3.2). Usually the $m = 6$ filters corresponding to the $\frac{m}{2}$ smallest and largest eigenvalues in D are used for classification, as extreme eigenvalues represent projections with the greatest mean difference in variance.

After projecting a trial to this $m \times n$ space, the logarithm of the variance of these m projections is typically used as a feature to automatically train a linear classifier. The combination of the feature extraction and a trained classifier results in the follow classification function:

$$f(X(i), \vec{w}, W) = \sum_m w_m \log\left(\left(\Sigma_{WX(i)}\right)_{m,m}\right) + w_0 \tag{3.3}$$

with bias w_0, feature weights $\vec{w}$, the m spatial filters W, and trial i's band-pass filtered signals $X(i)$. The variance of the projections is expressed with the diagonal of the covariance matrix diag(Σ). Traditionally, the logarithm is used to convert the band-power features to an approximately normal distribution as assumed by linear discriminant analysis (LDA) classifiers, but the logarithm is not strictly necessary for classification.

3.2.2 Direct covariance classification

If we reformulate (3.3) to work in the channel covariance ($\Sigma_{X(i)}$) space and drop the logarithm:

$$f(\Sigma_{X(i)}, \vec{w}, W) = \sum_m w_m \left(W \Sigma_{X(i)} W^T \right)_{m,m} + w_0, \tag{3.4}$$

we can see that $f(\Sigma_{X(i)})$ is just a linear transformation of the vectorized (flattened) trial covariance matrix denoted by $\text{vec}\left(\Sigma_{X(i)}\right)$:

$$f(\Sigma_{X(i)}, \vec{w}, W) = \sum_m \vec{w}_m \left(W_{m,\cdot} \Sigma_{X(i)} (W_{m,\cdot})^T \right) + w_0 \tag{3.5}$$

$$= \sum \left(W^T \operatorname{diag}(\vec{w})\, W \right) \circ \Sigma_{X(i)} + w_0 \tag{3.6}$$

$$= \vec{u}^T \operatorname{vec}\left(\Sigma_{X(i)}\right) + w_0, \tag{3.7}$$

where $\circ$ is the Hadamard (element wise) product, $W_{m,\cdot}$ is the m-th row of W, and $\operatorname{diag}(\vec{w})$ is a diagonal matrix containing the values of $\vec{w}$ on its diagonal. The combination of the spatial filters and the band-power feature weights

$$\vec{u} = \operatorname{vec}\left(W^T \operatorname{diag}(\vec{w})\, W \right) \tag{3.8}$$

can thus be modeled directly with a regularized linear classifier [20]. This simplification enables the classifier to learn spatial filters simultaneously with the projection's variance weighting in a single, supervised learning step.

For direct covariance classification, we have chosen to explicitly decorrelate the channels with a symmetric whitening transform P that is estimated on the training trials to satisfy (3.1):

$$P = \Sigma_X^{-\frac{1}{2}} = U \Lambda^{-\frac{1}{2}} U^T \quad \text{with} \quad U \Lambda U^T = \Sigma_X, \tag{3.9}$$

and only learn the rotational spatial filters and their associated weights implicitly as in (3.7) with a linear support vector machine (SVM). Without this whitening transform, the SVM's ℓ_2 regulariser strongly biasses the classifier to focus on high-powered sources that are probably not originating form the brain.

In summary, we used the covariance of whitened trials as features to directly train a linear classifier that is, except for the log transform, almost equivalent to the commonly used CSP pipeline.

3.2.3 Covariance classification with a second order-baseline

Likewise, the proposed second-order baseline (SOB) method learns the spatial filters implicitly, but instead of the static whitening transform (3.9) a pre-trial baseline is used to *adaptively* normalize ongoing second order covariance statistics.

We estimated a whitening transform $P(i)$ for each trial i based on past pre-trial baselines, and applied this transform to the data of the trial itself. Without a change in brain activity, the covariance during normalized trial $P(i)X(i)$ would approximate the identity matrix, even during (slow) sensor covariance changes. But when the coupling between or the power in certain brain regions changes, a perturbation appears in the sensor covariance during the normalized trial. This perturbation can be used for classification. The specific whitening transform (3.9) has the crucial property that it removes correlations, but at the same time minimizes the distorting to preserve the relation between projections and locations on the scalp. This preserves the task-relevant topography, which is needed to have consistent features over time and over subjects. A similar normalization procedure was outlined in [66] to adapt the session covariance matrix in order to reduce the influence of non-stationarities. The main differences between our method and [66] is that in our method each trial is normalized differently based on the pre-trial baseline, and that this baseline period is used to estimate the resting state covariance instead of the global session covariance.

Estimating the symmetrical whitening transform from the pre-trial baseline covariance $\Sigma_{B(i)}$ for trial i is difficult, since there is a large number of parameters to estimate from a few independent samples for the s sensors. To improve the robustness, we used the regularized Ledoit-Wolf covariance estimator [40], and used an exponentially weighted moving average (EMWA) to combine the baseline covariance of past trials into a covariance estimate $\Sigma_B(i)$ for the baseline of trial i:

$$\hat{\Sigma}_{B(i)} = \alpha S^*_{B(i)} + (1-\alpha)\hat{\Sigma}_{B(i-1)}, \tag{3.10}$$

where $S^*_{B(i)}$ is the Ledoit-Wolf covariance estimate of the baseline before trial i, and α is known as the forgetting factor that determines the rate of adaptation. With

$$\alpha = 1 - \sqrt[n]{\tfrac{1}{2}} \tag{3.11}$$

the forgetting factor α is then associated with a decay that halves in n trials.

Specifically, we calculate a whitening transform $P(i)$ for each of the trials $X(i)$ based on $\hat{\Sigma}_{B(i)}$, and apply this $P(i)$ to the Ledoit-Wolf covariance estimate of the current trial $S^*_{X(i)}$:

$$\tilde{X(i)} = P(i) S^*_{X(i)} P(i)^T \quad \text{with} \quad P(i) = \Sigma^{-\frac{1}{2}}_{B(i)}. \tag{3.12}$$

The new features $\mathrm{vec}\left(\tilde{X(i)}\right)$ are more robust against time and subject related variations, but are still sensitive to task-related (co)variance changes.

3.2.4 Dataset

To evaluate the performance of the invariant features we used the movement imagery dataset[2] from the experiment described in [59] contributed to Physiobank [25]. This dataset contains sessions of 109 different subjects with trials for actual and imagined movement [58]. The EEG was recorded with 64 electrodes positioned spread over the scalp according to the international 10-10 system, sampled at 160 Hz.

We have chosen to use the blocks where the subjects had to imagine either movement of both feet or movement with both hands, as a preliminary experiment indicated that BCI classification performance above chance level can be obtained with small training sets. There are about 22 trial per class for each subject.

3.2.5 Preprocessing

To preprocess the data, we applied a 6th-order Butterworth notch filter at 60 Hz, applied a 6th-order Butterworth filter between 8–30 Hz and extracted trials for movement imagery of both hands or of both feet in the interval from $[-2, 4]$ s after the stimulus. All evaluated methods used the same interval $[0.5, 4]$ s after the stimulus presentation for classification. For the SOB method, the interval $[-2, 0]$ s was used to estimate the pre-trial baseline. The same preprocessing was used for all BCI classifiers.

3.2.6 Evaluation

We used two CSP-based pipelines and a log band-power (logBP) based pipeline as a comparison method in both an SD and an SI BCI classification scheme. One CSP pipeline was based on CSP projected log band-power features classified with an LDA, the other CSP classifiers used band-power features without the log transform, classified with a linear soft-margin SVM [15]. The logBP classifiers simply used the variance of each band-pass filtered channel as a feature for a linear SVM. The whitened covariance features and SOB normalized covariance features were also classified with a linear SVM. The SVM's c-parameter was always estimated using a sequential (chronological) 5-fold cross-validation procedure with a logarithmic step size for the c-values, on the training set.

To evaluate these classifiers in an SD context, the first half of the session was used for training, and the second half was used for evaluation. This

[2] `http://www.physionet.org/pn4/eegmmidb/`

Table 3.1: *The subject-dependent accuracy of the different pipelines on the last 22 trials of each session for the 51 test subjects.*

Pipeline	Half life	Mean (std.)	Median
logBP SVM		62.2 (11.5)	63.6
CSP logvar LDA		**69.5** (14.6)	68.2
CSP var SVM		68.6 (15.0)	68.2
whcov SVM		68.9 (15.2)	**72.7**
SOB cov SVM (best)	13.8 trials	67.1 (13.3)	68.2
SOB cov SVM (worst)	1 trial	64.8 (13.1)	63.6

simulates a calibration and application session, respectively. Chronological separation of training and test set is needed since random splits lead to overly optimistic performance estimates. As the dataset also contains blocks with other mental tasks, we have only about 22 trials in total for training and 22 trials for evaluation per subject.

The performance of SI application was assessed by training an SI classifier on the first 50 users, and then applying the classifier to the test set formed from the remaining 51 subjects (we removed 8 subjects from the test pool because they had fewer trials). The final SI performance was calculated on the second half of the predictions for these 51 test subjects to allow for a paired comparison with the SD classifiers.

3.3 Results

3.3.1 Subject-dependent classification

The performance of the various control features is shown in Table 3.1, as well as the performance on the newly proposed SOB covariance features. For subject-dependent (SD) classification, the LDA classification of CSP log-variance features had the highest mean accuracy. However, there was no significant difference between the performance of this CSP pipeline and direct covariance classification (whcov SVM), and the latter had a higher median performance (72.7% accuracy). This demonstrates the feasibility of direct covariance classification. The logBP features performed much worse than the spatially filtered alternatives. The SOB features worked almost as well as the CSP features when low α-values were used; with faster adaption rates the SOB did not perform as well with SD application.

Table 3.2: *The subject-independent accuracy of the different pipelines on the last 22 trials of each session for the 51 test subjects.*

Pipeline	Half life	Mean (std.)	Median
logBP SVM		58.1 (11.1)	54.5
CSP logvar LDA		59.3 (11.8)	54.5
CSP var SVM		56.4 (9.7)	54.5
whcov SVM		59.1 (10.8)	54.5
SOB cov SVM (best)	4.0 trials	**67.3** (13.4)	**68.2**
SOB cov SVM (worst)	18.9 trials	64.9 (13.0)	63.6

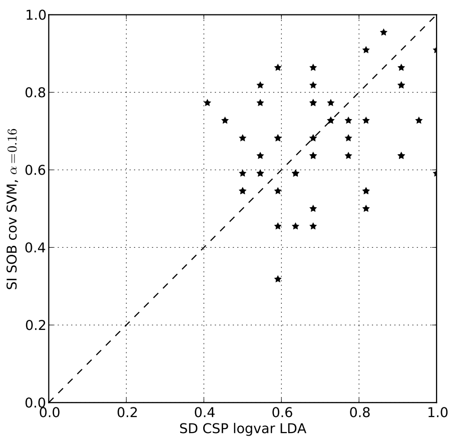

Figure 3.1: *The accuracy of a subject-dependent CSP pipeline versus the performance of a subject-independent SOB pipeline. There is no significant difference between the classifiers, despite the fact that the SOB pipeline was not trained on the subject.*

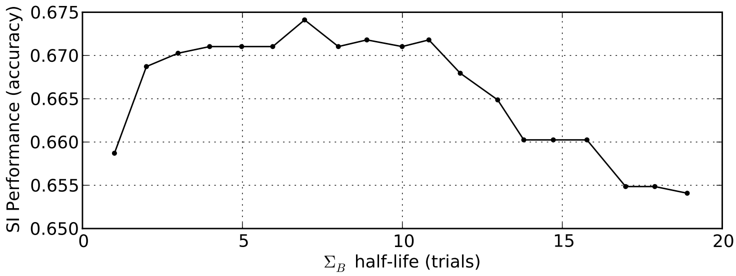

Figure 3.2: *The mean SI accuracy of the SOB as a function of the half-life of the baseline estimate $\Sigma_{B(i)}$ of the predictions on the last 22 trials of all the 59 test subjects.*

3.3.2 Subject-independent classification

The performance of the various control methods severely degraded due to inter-subject variability (Table 3.2) when these classifiers were applied in a subject-independent (SI) fashion. The CSP based classifiers, which did outperform the naive logBP classifier with subject-dependent training, now performed at the same level as logBP classifier with SI application — the advantage that spatial filtering provided in the SD training disappeared with SI application. The performance of the SOB based predictions however, was not affected at all. The accuracy of the best subject-independent SOB-based predictions was not even significantly different from the best subject-dependent (CSP log-variance LDA based) predictions (p=0.16 with a Wilcoxon signed-rank test on 51 paired observations, see Fig. 3.1).

The best SI performance was obtained with a volatile baseline with a half life of 4 trials, while with SD application the best results were obtained with long half lives (low α's). Note that even the worst performing SOB classifier outperformed all of the control classifiers. Fig. 3.2 displays the fraction of correctly classified trials as a function of the amount of smoothing of the pre-trial baseline covariance. The best performance was obtained with a baseline half-life between 2–11 trials.

The most contributing spatial filters that were learned implicitly by the SI SOB covariance classifier are shown in Fig. 3.3. These filters can be easily extracted with an eigenvalue decomposition of the covariance matrix, see (3.6). The most relevant features originated from the motor cortex region, but also occipital and central parietal features contributed to the classifiers predictions. There was no apparent contribution of muscle or eye move-

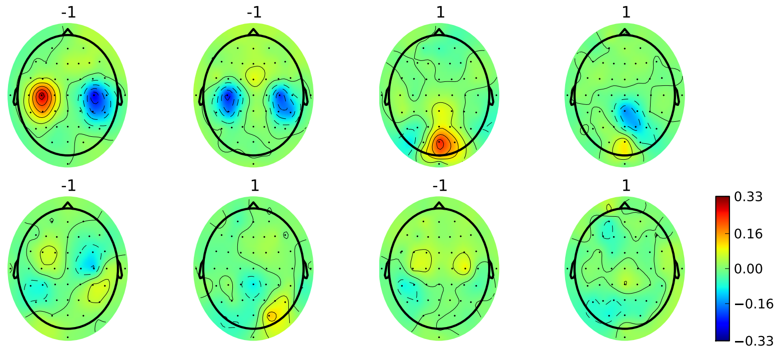

Figure 3.3: *The most influential spatial filters $W_{i,\cdot}$ for the best SI SOB classifier scaled by the magnitude of their weight w_i. The number above the plot is the weight w_i (i.e. a positive sign indicates filters with a response that corresponds to imagery of foot movement, a negative sign indicates imagined hand movement). Most of the contribution seems to originate from the motor areas, the central parietal regions and the visual cortex.*

ment artifacts to the classification.

3.4 Discussion

The results indicate that the new second-order baseline covariance features provide a robust alternative to CSP features for classification of motor imagery, and generalize to new, unseen subjects without additional calibration or training. Apparently, the normalization performed with the SOB removes enough inter-subject variability to generalize to new subjects. However, the dataset used in this research contains rather few trials, hence the SD CSP performance might have suffered from insufficient training data. Nevertheless, recording more trials is not always an option, and the SD performance obtained in our study is similar to scores presented in [23] where much longer sessions were used. Further, when a similar SD CSP pipeline with a broadband spectral filter was applied to naive (i.e. users that have not used a BCI before) users in [6], the results were barely above chance although 280 trials were available for evaluation. The performance increased dramatically though when subject-specific *spectral* filters were used.

The SOB's α parameter seems of some importance for generalization over subjects. While for SD classification a long half-life was preferred, α's with a short half-life were preferred for SI classification. Presumably, slow adaptation is preferred for SD classification because it allows the classifier to model and exploit session-specific variations, such as for example bad channels and EEG artifacts. For SI classifications modeling these variations is generally not helpful as they are not consistent over subjects. Shorter half-lives reduce these variations, and are thus preferred for SI classification.

It is noteworthy to mention that the best α-value for SOB-covariance features was selected based on the performance on the test subjects. This might slightly overestimate the true performance. Usually these hyper-parameters (e.g. the SVM's c-parameter) are set based on performance estimates obtained with cross-validation. The half-life constant α could be chosen with cross-validation, but since the SOB is a preprocessing method this is often computationally impractical. Since current BCI pipelines have several preprocessing hyper-parameters that are fixed a priori (e.g. the cut-off values in the band-pass filter, or the $m = 6$ spatial filters), we can imagine that a fixed α can be used as well. Given that even the worst α performs better than the alternatives in SI classification, the performance gain seems fairly robust for a wide range of α-values (Fig. 3.2). Therefore, we expect that choosing $\alpha = 0.16$ (a half-life of 4 trials) a priori will be adequate in practice. This value is probably independent of the mental task used.

3.5 Conclusions and future work

We have presented an SOB procedure that reduces inter-subject and inter-session variability, and demonstrated that SOB-covariance features allow for cross-subject motor imagery classification without a loss of performance compared to within-subject classification with the popular CSP based BCI classifier. The advantage of the SOB based covariance features is that they are robust against inter-session and inter-subject variation, and that standard classifiers such as the SVM can be used without the need of any adaptation or post-processing of the outputs (e.g. bias-adaptation). Furthermore, the online processing is simplified as it can be implemented as a stateless, fixed pipeline that does not handle the incoming data differently during a calibration session.

In addition to the practical advantages of removing the need for the laborious calibration sessions, changing from subject-dependent to subject-independent BCIs also simplifies multi-disciplinary BCI research. It allows researchers to work with validated BCI classifiers that are known to work with a certain probability on the target population, and focusses on the intended brain regions. The development of subject-independent BCIs can facilitate new applications for which collecting enough subject-specific training data before each session is not feasible, such as for example fatigue detection, screening of neurological disorders or classification of emotional states.

The method described uses a single, broad frequency band. For future work, the features can be extended to include multiple frequency bands as in [41, 20]. Work with naive BCI has shown that subject-specific frequency bands can dramatically improve performance [6].

Another interesting research direction is to generalize to recordings with different electrode layouts. As the learned covariance weights were quite sparse, the correspondence of a few key sensor locations might be enough to generalize to new sensor configurations. This sparsity in covariance space seems to suggest that it is more natural to think in covariance (or coherence) between brain regions, than to think of power in spatially filtered sources. The second row in Figure 3.3 shows spatial filters with a dipolar structure on the motor cortices. The combination of these and many more filters is needed to isolate a specific covariance pair — that is, the intriguing dipolar structure is probably residue caused by the decomposition of the weight matrix.

Finally, although the method presented works causally, it should be validated in an online experiment with a user in the loop.

Chapter 4

An SVM for structured errors

NON-STATIONARY feature distributions violate the basic assumption made by most ML methods that the data is independent and identically distributed (iid). In Chapter 3, we demonstrated for the first time that cross-subject EEG classification is possible by making the training set's and test set's distribution similar with a normalization procedure. But this normalization procedure does not result in a performance gain with subject-dependent (SD) application (i.e. the usual scenario where a calibration session is used before application). This is consistent with prior work on non-stationary signals in BCIs that attempts to reduce feature variability over time [28, 66, 5, 72, 44, 53] or alternatively to adapt the classifier's parameters to the changing distribution [62, 70], where at best modest improvements are reported.

It is unclear to what degree variability in the feature distributions influences the SD BCI, since the variable features might be irrelevant to the specific classification problem. The stationary subspace analysis (SSA) method [72, 44] hinges on this assumption. But even if this assumption holds, attempting to remove the irrelevant variability could introduce residual noise. As a consequence the performance could suffer, while a robust classifier might have been able to learn these variations in the data in the first place. This may be the reason why most of the aforementioned studies have been demonstrated on datasets that were artificially augmented with strong non-stationary distributions, or had to include additional data or labels to learn from.

All these proposed methods focus on assuring that the features are identically distributed during training and testing. Still, being identically distributed does not imply that the samples are independent of each other.

The dependence of trials over time, which can be caused for example by inertia of neural processes or by filtering artifacts (or even the SOB procedure), can result in prediction errors that are dependent in time. If the classifier observes a series of dependent errors that are caused by a single (hidden) event, and assumes that these errors are unrelated, this can severely bias the classifier and lead to overfitting. Incorporating a priori known dependencies in the model thus enables the classifier to reduce the penalty for related errors (e.g. caused by a period of distraction of the user) and to focus on structural errors present in the data.

In this Chapter, we focus on the violation of the independence aspect of the iid assumption. Instead of attempting to remove dependencies in the feature space as done in previous work, we modeled the dependencies in the objective function of the classifier. Specifically, we present a generalization of the SVM that uses latent slack variables to model the structure of misclassified training instances. By penalizing these latent slack variables, the dependent-samples support vector machine (dSVM) was able to diminish the influence of related errors and to reduce overfitting. The dSVM was evaluated on real BCI datasets, and showed an improvement in performance.

4.1 Previous work

The iid assumption is ubiquitous in machine learning and statistics. When using real data, the ubiquitous iid assumption of the training examples can be, and often is violated. For example, when samples are drawn from different subjects, it is to be expected that samples drawn from the same subject display similar (subject specific) variability, in other words, the uncertainty of the observations decreases if the subject is known. The implications of classifying such non-iid data has received surprisingly little attention from the ML community. Methods that deal with conditionally iid (i.e. for each label y_i, the corresponding instance $\vec{x}_i$ is generated according to some fixed probability distribution $P(\vec{x}_i|y_i)$) data do exist, such as for example hidden Markov models (HMMs) and conditional random fields (CRFs). These models allow for structure between the *labels.* Methods for non-stationary data with direct dependencies between the observations $\vec{x}_i$ have received considerably less attention. With BCI data, it is safe to assume that the labels are (relatively) independent. The features derived from the EEG however do display temporal structure, and sessions (or users) can be considered as groups with different feature distributions.

A dependence over time is common in time series. This implies that when doing prediction on a time series, knowing the error term of one data point is informative of the error of the next. With this knowledge, better models for time series can be built. This dependency is known as *serial correlation* in the field of econometrics, and is known to bias the standard error of uncorrected regression coefficients to be biased downwards (overfitting). Various methods do exist to estimate the first-order serial correlation coefficient ρ for consecutive error terms such as the iterative Cochrane-Orcutt procedure or a grid-search method that minimizes the standard error (Hildreth-Lu method) [49]. For linear regressors, a second set of regressors based on lagged error terms can be used to correct the problem (feasible generalized least squares).

In statistics, there are methods that model the dependencies within groups of measurements. These so-called mixed-effect models separately model the contribution of different users (known as fixed-effects) and the contribution of random-effects. An example of the application of mixed-effect models to inter-subject variability in BCIs can be found in [24]. Since these mixed-effect methods are aimed at explanatory analyses, they rely on estimating statistics for every group (user). While this improves the understanding of the classification problem, they generally do not allow for prediction of out-of-sample users since the fixed-effects of this user still need to be estimated. The SOB procedure presented in Chapter 3 can be regarded as mixed-effects model, since the adaptive whitener models and subtracts the fixed effects. The difference is that the SOB quickly re-estimates the fixed effects model for prediction.

The structure of prediction errors implies that not only the sum of the errors, but also the distribution of the errors should be taken into account in the optimization of the classifier, since related errors might lead to overfitting [49]. One approach to bias the classifier towards certain distributions of errors is the quadratic-loss Y-SVM (QLYSVM) [50]. The QLYSVM uses a quadratic penalty for pairs for margin errors, and optimizes

$$\arg\min_{\vec{w},b,\xi} \quad \frac{1}{2}\|\vec{w}\|^2 + c\vec{\xi}^T YSY\vec{\xi} \tag{4.1}$$

$$\text{s.t.} \quad y_i\left(\vec{w}^T\vec{x}_i + b\right) + \vec{\xi}_i \geq 1 \tag{4.2}$$

$$\vec{\xi}_i \geq 0, \tag{4.3}$$

where $\vec{x}_i$ is the feature vector for instance i, Y is a matrix with the target labels $y_i \in \{-1, 1\}$ on its diagonal, $\vec{\xi}$ is a vector with a positive slack variable for each training instance that allow support vectors to penetrate the margin at a cost, and S is a symmetric positive-definite matrix that determines the cost of pairs of margin errors. The dual of this QLYSVM can be solved by adding $\frac{1}{2c}S^{-1}$ to the kernel matrix K and solving the hard-margin SVM's dual. When S is a diagonal matrix, the QLYSVM degenerates to the quadratic loss SVM.

Unfortunately, the QLYSVM cannot be used to exploit the decreasing dependence with time: to reduce the weight of a temporary disturbance, the cost of misclassifying two highly dependent instances should be less than the cost of misclassifying two more independent instances. The resulting cost matrix S is not positive definite, which prevents optimization of the QLYSVM's dual. Furthermore, the inversion of S is expensive, and might suffer from numerical inaccuracies depending on the kernel function used to construct S.

4.2 The dependent-samples SVM

Given training examples $\{\vec{x}_i\}$ and corresponding labels $\vec{y}_i \in \{-1, 1\}$, the dSVM can be formulated as the following constrained, convex optimization problem:

$$\arg\min_{\vec{w}, b, \xi} \quad \frac{1}{2}\|\vec{w}\|^2 + c \sum_i \vec{\xi}_i \tag{4.4}$$

$$\text{s.t.} \quad y_i \left(\vec{w}^T \phi(\vec{x}_i) + b\right) + D\vec{\xi}_i \geq 1 \tag{4.5}$$

$$\vec{\xi}_i \geq 0, \tag{4.6}$$

where $\vec{w}$ is the classifier's weight vector and b is the bias term. The function $\phi(\cdot)$ maps examples from input space to a feature space. The c parameter determines the cost of penetrating the SVM's margin (margin errors); $\vec{\xi}_i$ is a positive, latent slack-variable that accommodates for these margin errors, and is related to a specific example i through $(D\vec{\xi})_i$. When D is the identity matrix I, each slack variable $\vec{\xi}_i$ is linked to a single instance i of the training set, and the dSVM degenerates to the traditional soft-margin SVM [15]. The columns $D_{i,\cdot}$ contain the non-negative dispersion function for a specific latent error $\vec{\xi}_i$ that needs to be specified a priori.

For ease of notation, we write this in matrix form with $Y_{i,i} = \vec{y}_i$ and $X_{\cdot,i} = \phi(x_i)$:

$$\arg\min_{\vec{w}, b, \vec{\xi}} \quad \frac{1}{2}\|\vec{w}\|^2 + c\vec{\xi}^T\vec{1} \tag{4.7}$$

$$\text{s.t.} \quad Y\left(\vec{w}^T X + b\right)^T + D\vec{\xi} - 1 \succeq 0 \tag{4.8}$$

$$\vec{\xi} \succeq 0. \tag{4.9}$$

With the Lagrange multipliers $\vec{\alpha}$ and $\vec{\gamma}$, the constraints can be put in the primal objective function $\mathscr{L}_p$ that needs to be minimized for $\vec{w}$ and b, and maximized for the dual variables $\vec{\alpha}$ and $\vec{\gamma}$:

$$\mathcal{L}_p(\vec{w}, b, \vec{\xi}, \vec{\alpha}, \vec{\gamma}) = \frac{1}{2}\vec{w}^T\vec{w} + c\vec{\xi}^T\vec{1} - \vec{\alpha}^T\left(YX^T\vec{w} + b\vec{y} + D\vec{\xi} - \vec{1}\right) - \vec{\gamma}^T\vec{\xi} \quad (4.10)$$

$$\text{s.t.} \quad \vec{\alpha} \succeq 0 \quad \text{and} \quad \vec{\gamma} \succeq 0, \quad (4.11)$$

Setting the derivative of the primal variables $\vec{w}$, b and $\vec{\xi}$ to zero yields their minimum,

$$\frac{\delta}{\delta\vec{w}}\mathcal{L}_p = \vec{w}^T - \vec{\alpha}^T YX^X = 0 \qquad \Rightarrow \vec{w} = XY\vec{\alpha} \quad (4.12)$$

$$\frac{\delta}{\delta b}\mathcal{L}_p = -\vec{\alpha}^T\vec{y} = 0 \qquad \Rightarrow \vec{\alpha}^T\vec{y} = 0 \quad (4.13)$$

$$\frac{\delta}{\delta\vec{\xi}}\mathcal{L}_p = c\vec{1}^T - \vec{\alpha}^T D - \vec{\gamma}^T = 0 \qquad \Rightarrow c\vec{1}^T = \vec{\alpha}^T D + \vec{\gamma}^T, \quad (4.14)$$

that result in the dual $\mathcal{L}_d$ after substitution:

$$\mathcal{L}_d(\vec{\alpha}) = \vec{\alpha}^T\vec{1} - \frac{1}{2}\vec{\alpha}^T YKY\vec{\alpha}, \quad (4.15)$$

where the kernel matrix $K_{i,j} = \phi(\vec{x_i})^T\phi(\vec{x_j})$. The dual $\mathcal{L}_d$ needs to be maximized with respect to the dual variables $\vec{\alpha}$, subject to the Karush–Kuhn–Tucker (KKT) conditions (4.13), (4.14) and (4.11):

$$\arg\max_{\vec{\alpha}} \quad \mathcal{L}_d = \vec{\alpha}^T\vec{1} - \frac{1}{2}\vec{\alpha}^T YKY\vec{\alpha} \quad (4.16)$$

$$\text{s.t.} \quad 0 \preceq \vec{\alpha}^T D \preceq c, \quad \vec{\alpha}^T\vec{y} = 0. \quad (4.17)$$

The only difference of this dual with the soft-margin SVM's dual is the appearance of the dispersion matrix D in the constraint (4.17); with $D = I$ the dual degenerates to the soft-margin SVM's dual. The dSVM's dual (4.16) and constraints (4.17) form a quadratic programming (QP) problem, and an optimal $\vec{\alpha}$ can be found with a standard QP solver. With $\vec{\alpha}$ found, $\vec{w}$ is defined through (4.12).

The bias b is usually calculated from support vectors on the boundary of the margin using (4.8) in the traditional soft-margin SVM formulation, i.e. from the x_i where $\vec{\alpha}_i > 0$ and $(I\vec{\xi})_i = 0$. With the dSVM, this definition of unconstrained support vectors is troublesome, since the latent slack variables are associated with multiple (possibly all) support vectors through D. Therefore, we have chosen to use an implicit, regularized bias instead, which can be implemented by either adding a constant feature, or by using an inhomogeneous kernel. To remove the bias term, the constraint (4.13) has to be removed, resulting in

$$\arg\min_{\vec{\alpha}} \quad \mathscr{L}_d = \frac{1}{2}\vec{\alpha}^T Y K Y \vec{\alpha} - \vec{1}^T \vec{\alpha} \tag{4.18}$$

$$\text{s.t.} \quad 0 \preceq \vec{\alpha}^T D \preceq c, \tag{4.19}$$

which is still a standard QP problem.

Classification with the dSVM is done as with the SVM, using (4.12) and (4.8), with $b = 0$:

$$f(X_2) = \vec{w}^T X_2 + b = \vec{\alpha}^T Y X^T X_2 = \vec{\alpha}^T Y K_2. \tag{4.20}$$

The main difference between the dSVM and the traditional soft-margin SVM is that dSVM has groups of support vectors that share a "support budget", whereas the former has a budget per support vector. In addition, the dispersion matrix D has to be specified a priori based on domain knowledge (e.g. for time series classification with a sliding window, the dependence is apparent from the feature construction). The dispersion functions (columns of D) do not have to be the same function with an offset. For example, when combining the data of multiple subjects, the D matrix could model a dependency of all trials (instances) on a subject, and a dependency on trials of the same subject near in time. This would allow the dSVM to remove subjects that do not display the expected brain signal altogether from the training set, while still spending modeling power on an as diverse set of subjects as possible.

4.3 Validation

To validate the dSVM introduced in the previous section, we constructed an artificial dataset to demonstrate the dSVM's robustness against dependent disturbances. After demonstrating the dSVM on this artificial dataset, we will present the evaluation of the dSVM on a real BCI dataset.

4.3.1 Artificial data

The dataset was constructed by sampling from two Gaussian distributions with equal covariance but different means. A non-stationary perturbation in feature space was generated with the Laplace distribution function

$$f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}, \tag{4.21}$$

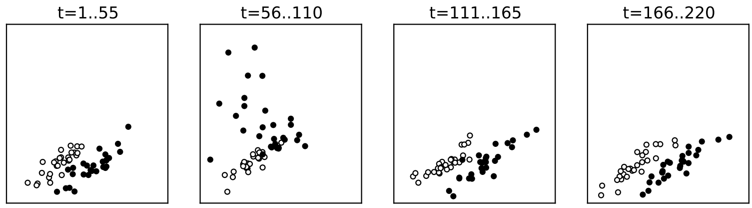

Figure 4.1: *Sequential snapshots of the artificial dataset used to evaluate the dSVM. The data is sampled form two Gaussian distributions with different means but equal covariance. In the second frame, we introduce a latent error, that leaks to a range of related of samples.*

with $b = 9$ and $\mu = 60$, where the argument x is the position in the dataset. This makes the instances non-iid (see Figure 4.1), with a known dispersion function.

A Laplace distribution was chosen based on its similarity to the empirical distribution found in a preliminary experiment, in which we modeled the interdependence of slack variables in the training set of a BCI experiment. As the strongest dependence in this preliminary experiment was found between slack variables of the same class, the perturbation was limited to one class only.

Both a soft-margin SVM and a dSVM were trained on this artificially constructed dataset. Model selection for the cost parameter c was performed with a line search on cross-validated accuracy with four sequential subsets corresponding to the frames in Figure 4.1. The dSVM's dispersion matrix D was constructed using (4.21):

$$\tilde{D}_{i,j} = \begin{cases} \vec{g}_i \cdot f(|\vec{t}_i - \vec{t}_j|) & \text{if } \vec{y}_i = \vec{y}_j, \\ 0 & \text{otherwise} \end{cases} \tag{4.22}$$

where $\vec{t}_i$ is the time offset of an instance, and $\vec{g}$ chosen such that the columns of D are normalized to one.

The resulting margins and hyperplanes are displayed in Figure 4.2. It can be seen that the soft-margin SVM was sensitive to the outliers produced by our non-stationary perturbation, with the result that the hyperplane was almost orthogonal to the hyperplane that would separate the two Gaussian distributions. In contrast, the dSVM was able to exploit the known dependence between the training instances, and separated the two Gaussian distributions correctly in the presence of the perturbation. This demonstrates

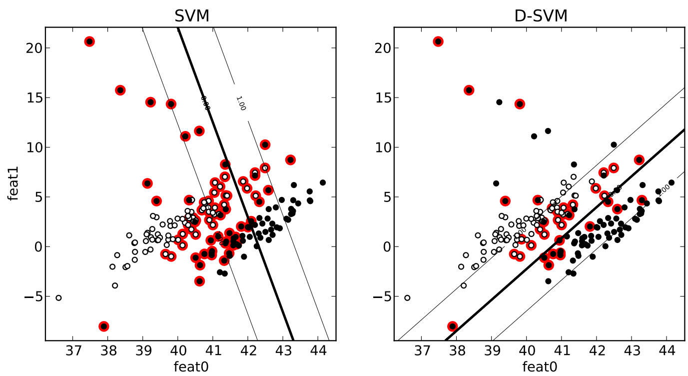

Figure 4.2: *The hyperplane (thick black line) and margin (between the thin black lines) for the standard soft-margin SVM (left) and the dSVM (right). Both classifiers were trained on the artificial data displayed in Figure 4.1. The two classes are indicated with white and black dots respectively, red edges indicate the support vectors. While the dependent outliers displace the hyperplane of the standard SVM, the dSVM is able to reduce the weight of these points based on their relatedness.*

how the dSVM can use dependency information to avoid overfitting if the training data is non-iid.

4.3.2 BCI data

Description of the datasets A suitable BCI corpus was selected based on a preliminary screening, based on its large interdependence of classification errors close in time. This corpus contains EEG signals for 12 subjects, recorded during 30 minutes of natural game play, synchronized with annotations of key presses. The game was designed to evoke changes in the mental state by periodically ignoring 15% of the keyboard input. Please refer to Chapter 2 for experimental details. Due to the unconstrained nature of this game, the key strokes follow each other rather quickly. While this dataset satisfies the non-iid assumption of the dSVM, the disadvantage of this dataset is that the short trials might make the task-related signals undetectable, as the recommended inter-trial interval (ITI) for the motor related ERD is at least 10 sec [48].

A BioSemi ActiveTwo EEG system was used to record 32 channels of EEG and physiological signals at a sampling rate of 512 Hz. The 32 Ag/AgCl active electrodes were placed at locations of the Extended International 10-20 system. To measure the influence of ocular and muscle artifacts, EOG (horizontal and vertical pairs) and two pairs of EMG signals over the left and right flexor digitorum profundus (the muscles used to press with the index finger) were recorded.

For this chapter, the recordings were preprocessed as follows: first the recording was downsampled to 128 Hz to speed up processing. After downsampling, the data was high-pass filtered using a 4th-order Butterworth filter to remove frequencies below 0.2 Hz, and notch-filtered using a 4th-order Butterworth filter from 49–51 Hz to remove power line noise. The EEG was then corrected for eye movements using a regression based subtraction method [61].

Feature extraction After preprocessing, the signals were filtered with a 6th-order Butterworth bandpass filter with a passband of 8–30 Hz, windows of 1 sec centered on the moment the key-press was registered were extracted, and the Ledoit-Wolf covariance estimator [40] was used to estimate the channel-covariance in the trial window. Finally, a symmetrical whitening transform $P = \Sigma^{-\frac{1}{2}}$ based on the mean channel covariance matrix Σ was calculated on the training set, and applied to all trials. The resulting 32×32 features were vectorized and used for classification. This pipeline is conceptually similar to the popular CSP based [34, 51] classification of ERD

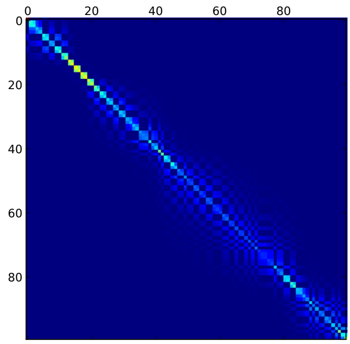

Figure 4.3: *An color-coded example dispersion matrix D for the first 100 trials of subject S0. The used dependence function is a Laplace distribution with width $b = 2$ for trials of the same class. The green diagonal entries around trial 20 indicate independent trials, while the pattern around trial 60 indicates a closely packed group of trials that are interdependent.*

related to movement imagery, but it delegates learning of the spatial filters to the classifier.

Performance evaluation For BCI data it is expected that the iid data assumption does not hold, as there are non-stationary perturbations in the data. This complicates the evaluation procedure as one cannot simply use cross-validation when the trials (instances) are inter-dependent. We opted for a simulation approach, in which we used the first 800 trials for training, used the performance on the following 400 trials for model selection, and used the remaining ≈ 800 trials as the test set. Nevertheless, non-stationary perturbations could be present in the validation and test set, and could skew the results, as the performance is measured based on the iid assumption. To improve the robustness of the performance measure, both the validation and test sets were split into five continuous parts, and the median of the performance on each part was used. As a measure of performance we chose to use the information transfer rate (ITR) based on mutual information (MI) between the predictions and the true class labels, which is an information-theoretic measure of the communication speed through an unreliable communication channel (the BCI). The advantage of ITR over, for example, accuracy is that it takes both the quality of the prediction and the speed of communication into account.

Table 4.1: *The ITR for the SVM and dSVM classifiers. ITRs below 1 bit/min have been removed for clarity.*

Subject	SVM bits/min	dSVM bits/min	b-param
S0	5.26	6.38	1.3
S1	–	–	0.01
S2	–	–	8.9
S3	–	–	0.48
S4	–	1.02	0.11
S5	11.9	11.2	0.11
S6	–	–	5.5
S7	–	–	38
S8	–	–	8.9
S9	13.8	18.0	0.3
S10	2.58	5.1	1.3
S11	–	–	2.1

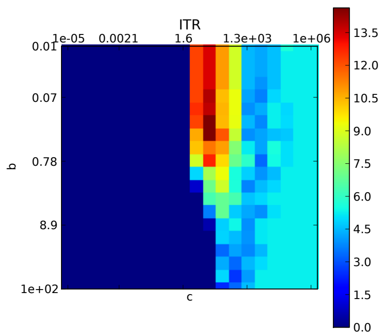

Figure 4.4: *The ITR on the test set of S9 for different combinations of the cost parameter c and dispersion parameter b. The top row has such a small b that the dSVM degenerates in the traditional SVM. The optimal spread appears to be $b = 0.3$, which indicates that modeling the inter-dependency of trials is beneficial for this subject.*

Results On the training and validation set, we performed a grid-search with the soft-margin SVMs and the dSVMs with different cost parameters c, and different width parameters b in (4.22) for the dSVM (see Figure 4.3 for an example of the resulting D) with a linear, inhomogeneous kernel. The offset parameter μ in (4.22) was set to zero. The performance of the selected SVM and dSVM classifiers are presented in Table 4.1. Unfortunately, for a number of subjects no well-performing classifier could be trained with either classifier. For the subjects for which a classifier could be trained, the dSVM seems to outperform the standard soft-margin SVM, with S5 being the only (minor) exception. The best subject displays a dramatic improvement in performance, from 13.8 to 18.0 bits per minute (see Figure 4.4).

4.4 Discussion

The results on the artificial dataset and on the real BCI datasets indicate that modeling the inter-dependency of training instances can improve the classification performance. On the BCI dataset, we have quite a few subjects for which the classifiers perform at chance level. This is, however, a property of the dataset, and not a limitation of the dSVM. For the subjects that had classifiable brain signals, the performance improved in general. Particularly promising is the observation that the best performing subject displayed the biggest improvement in performance. The question remains how this result would generalize to more traditional BCI datasets that are more easily classified, and generally have much smaller training sets.

We demonstrated that the modeling the interdependence of the training data can be helpful. Other methods that attempt to constrain the undesirable influence of non-stationary feature distributions have focused mainly on the changes in the distribution. The dSVM and these methods are complementary; to assess their efficacy they should be evaluated in combination, since features for BCI data probably violates both the independence and the identically distributed assumption.

4.5 Conclusions and future work

In this chapter, we presented the dSVM, which is a generalization of the soft-margin SVM. The dSVM is able to take the dependency of classification errors into account. Using a real BCI dataset, we have demonstrated the feasibility of improving the performance on an unseen test set by modeling the temporal interdependence between the training instances. Since the dSVM and SOB methods presented in Chapter 3 are complementary, they should be evaluated in combination. Correcting for the iid assumption might finally give a estimate of the magnitude of the problems caused by non-stationary feature distributions in BCI.

For continuous classification of EEG signals, features extracted with a sliding window might form a natural candidate for the dSVM. When the interdependence is less clear, it might be possible to learn the dependence function from the data. A naive method would be to iteratively re-estimate the covariance of margin errors. Including the estimation of this dependency function in the optimization criterion is an interesting direction for future research. We have taken the first steps in this direction in a logistic regression framework.

An open problem is how the performance on non-iid datasets should be assessed. We have chosen for the simple approach of using a robust statistic on multiple estimates. For future work, a well-founded, reliable method needs to be devised.

Chapter 5

Conclusions

In this chapter, we will summarize the contributions of the last three chapters. Drawbacks and merits of the proposed methods will be discussed, and embedded in a unified view. We will reflect on the progress we have made towards the creation of robust brain-computer interfaces (BCIs), and outline directions to continue improving BCIs.

5.1 Summary of contributions

In Chapter 2 we showed that the feature distributions used for BCI classification are indeed non-stationary, and that these changes can be caused by changes in the mental state of the user. Changes in the mental state were induced with a modified Pacman game that caused episodes of frustration by reducing the amount of control the user had over the Pacman character. BCI classifiers based on spectral event-related desynchronization (ERD) features that were trained on the episodes with full control performed significantly better on the episodes with reduced control.

In Chapter 3 we described the second-order baseline (SOB) method that reduces inter-session and inter-subject variability in signal feature distributions. The procedure is based on the insight that a large class of BCIs is based on ERD which is observed by a *relative change* in spectral power, but use absolute power instead for classification. An off-line experiment with 109 subjects has shown that these new features are robust enough to train a subject-independent (SI) BCI classifier. The data of new, unseen users was classified with an accuracy as high as conventional BCIs that, unlike our method, do require a calibration session prior to BCI use. To the best of our knowledge this is the first causal SI BCI classifier that works as well as subject-dependent (SD) classifiers.

In Chapter 4 we linked the problem of non-stationary feature spaces to the violation of the basic independent and identically distributed (iid) assumption underlying most statistics. We derived a generalization of the well-known support vector machine (SVM) classifier, that takes the chronological dependence of features (and hence the dependence of classification errors) into account. Both on artificial data and real BCI data, overfitting was reduced with this dependent-samples support vector machine (dSVM), leading to an increased information throughput for the BCI.

In Chapter 1 we outlined two key issues for BCIs. We have fully addressed the first issue of investment of time by completely removing the need for a calibration session with the SOB based ERD features. It is likely that this approach also allows for subject-independent classification of other ERD based brain signatures, and even of event-related potential (ERP) based signatures. This still has to be validated experimentally. Future research should investigate whether the dramatic performance gains associated with subject specific spectral filters [6] can be incorporated into a generalized SOB method.

The second key issue of requiring predictable, dependable BCI performance was partially addressed with the work on the dSVM. The method demonstrates that it is feasible to model the relatedness of brain signals recorded nearby in time. Modeling the interdependence of trials is necessary to prevent overfitting when using high-dimensional features spaces based on electroencephalography (EEG). Not assuming iid feature distributions raises new questions regarding performance evaluations though. Nevertheless, we have taken the first steps in handling non-stationary feature distributions from a new angle, which has provided the valuable insight that structured errors can indeed lead to overfitting.

With the two new methods presented in this dissertation, we have paved the way for a next generation of BCIs — BCIs that work dependably, without the need of recalibration .

5.2 Discussion

The influence of non-stationary feature distributions on BCI performance is often described as severe, and one easily gets the impression that the effects of the ever varying EEG is the cause of the sometimes disappointing accuracy of BCI systems. While it is difficult to get a clear estimate of the impact these fluctuations in feature distributions on the classification performance (because most variability is irrelevant for classification, and an upper limit of the performance is unknown), the results in Chapter 2 show that the difference in performance caused by different mental states is very modest, despite the fact that changes in the mental state were strongly induced. Furthermore, the trials sampled from deviating ERD feature distribution were classified significantly better, contrary to what one would expect. These results, combined with fact that the SOB did not result in an expected performance increase for within subject classification, indicate that the *direct* influence of changing feature distributions on the performance is very weak.

This does not imply that problems caused by non-stationary feature distributions do not exist, but merely that it is not straightforward to assess the magnitude of their influence.

The common remedy to these ill-defined non-stationarities is to adapt the classifiers or feature spaces to track changes in the feature distributions. Our SOB normalization procedure falls into this category. Although the method is able to track feature changes over time, it only results in performance gains when applied across subjects. This seems to indicate that the feature distributions are often approximately stationary within a session. The independence part of the iid assumption is probably what is being violated, as shown in Chapter 4.

The unsupervised BCI feature extraction in the form of covariance matrices (introduced in tensor form in [20]) has some distinct advantages over the supervised ERD feature extraction with the common spatial patterns (CSP) algorithm. First and foremost, it prevents the stacking of supervised methods with potentially different objective functions which can result in the final supervised layer underestimating the noise in the training set due to overfitting in the first supervised method (i.e. CSP). Other advantages include that the spatial filter is learned implicitly, which simplifies multi-class classification, and the fact that the features are extracted without class labels opens up faster methods for cross-validation and performance measures. While the SOB procedure can be used with CSP feature extraction, the dSVM presented in Chapter 4 critically depends on implicit learning of the spatial filters.

An additional performance increase is to be expected when the SOB based normalization and the dSVM with implicit learning of the spatial filters are combined. Both methods attempt to solve different aspects of the violation of the iid assumption: the SOB reduces long-term distribution changes, and thereby ensures that future features are almost distributed identically to the features calculated on the training data, and the dSVM reduces overfitting caused by violating the independence aspect of the iid assumption. The results based on the SOB procedure presented in Chapter 3 might still suffer from overfitting due to dependence of trainings trials.

Similarly, the results presented in Chapter 4 could still suffer from changing distributions. Despite these suboptimal conditions, both methods demonstrate performance improvements on real BCI data.

5.3 The road ahead

With the work described in this dissertation we have made strides in the development of robust BCI methods, with applications both to advancing BCI research and improved reliability of BCIs applied in practice. Usually, though, research answers fewer questions that it raises. In this section, we will present directions to find new interesting answers based on the experience gained in the past four years.

5.3.1 A different kind of BCI research

Our subject-independent (SI) classifier presented in Chapter 3 can facilitate the simplification of BCI research. The current practice in the BCI field is that classifiers are trained for each subject individually. This has the drawback that within a single experiment a great variety of classifier models is used; all these classifiers have to be carefully trained and inspected in order to ensure that the right signals are used for classification. In Chapter 3 we showed that a single model can suffice for a large subject pool, without any loss of performance. As a result only a single set of spatial filters need to be inspected, and more data can be used. This reduces the chance of using subject-specific artifacts for classification.

As the SI BCI do no longer depend on a calibration session, the BCI can be designed and *fully trained before the actual experiment* is performed. Preparing a fully trained BCI has the advantage that it does not depend on an error-prone calibration (a single artifact can severely change the resulting classifier). For subjects, a pre-trained BCI has the advantage that it rewards a predetermined model of brain signals, and not some unrelated but class-relevant artifacts. User training might be improved by such validated, targeted feedback. The separation of the sensitive training phase and application phase could stimulate multi-disciplinary BCI research, as fully developed BCIs can be shared with researchers from other disciplines. For example, the BCI field would greatly benefit from contributions from human-computer interaction (HCI) field [17], for example on the topics of user experience, evaluation methodology and application design. The required HCI studies would become much easier with the availability of a pretrained, validated BCI that focusses on the intended brain regions.

We see the transition from subject-dependent to subject-independent models as a necessary step in the development of the field, parallel to the developments that have taken place in speech recognition research. We have demonstrated an SI model that discriminates types of motor imagery within a single experimental environment using the same recording hardware. The next phase is to perform discrimination independent of the recording hardware, and independent of the specifics of the experiment for a variety of brain signatures. Continuing the parallel with speech recognition, current BCI research resembles small-vocabulary speech recognition, since a minimal number of actions is being discriminated. Future research effort could therefore focus on working towards more open-vocabulary models. In practice, this means the integration of multiple SI models, such as for example the P300 response, the error potential, covert attention etcetera into a single BCI system. The combination of detectors for the different brain signatures might lead to novel BCI paradigms, and to disambiguation of variability naturally present in the brain signals.

5.3.2 Better feature spaces

The method for SI BCI classification is still largely based on features inherited from SD classification. The increase of training data and the increased diversity of the samples accompanying the transition from SD to SI classification creates the opportunity of (re)discovering appropriate feature spaces.

An obvious and already often practised improvement for the features used in Chapters 3 and 4 is the use of features based on multiple frequency bands instead of features based on a single broad frequency band [41, 20].

A different path is the unification of known BCI feature spaces. The features used to extract different BCI signatures (e.g. steady-state visually evoked potential (SSVEP), motor imagery related ERD, P300) are usually designed specifically for the task, despite that these methods are all based on either phase-locked amplitude (ERP) or oscillator power (ERD/event-related synchronization (ERS)). A combined feature space that can capture both the first order ERP features and the second order ERD/ERS features can result in an improvement in performance [12], and could provides a single unified feature space that can be applied to all known BCI tasks. The unification of the two main feature spaces would help to reduce the balkanization of the BCI methods (i.e. each group introduces a new variant of a BCI classification pipeline).

One way to implement unification would be through the definition of specific kernels for use in kernel machines. An example of such a kernel is a complex non-homogeneous polynomial kernel used on specific frequency bins of discrete Fourier transform (DFT) transformed trials. Working in the DFT domain opens another interesting possibility of shift-invariant ERP classification, in other words, a P300 could be detected at variable positions in the analysis window due to the circular wrapping of the DFT. Applications include uncued ERP classification and the detection of non phase-locked ERPs.

A strong neurological motivation for the inclusion of phase features in oscillatory analyses is provided with the recent discovery of cross-frequency coupling (CFC): sometimes the activity for a given frequency band cannot be detected on its own, but only emerges after phase-locking the high-frequency power to the phase of a low-frequency task-relevant activity [13, 47]. CFC takes these non-linear interactions between signals in different frequency bands into account. The discovery of these interactions have caused some excitement in the neuroscience community [11, 14] since they assess the information being carried in the precise timing of activity within, and across, physically separated brain areas. Although pestered by the huge increase in feature dimensionality, CFC based features could eventually culminate in an ultimate feature space that supports ERP and ERD based classification as well as more complex interactions.

Another route to improve the BCIs feature space is through unsupervised, semi-supervised or transfer learning. In general, the amount of *labeled* training data is limited. In contrast, unlabeled data is abundant, and increasingly so with each recorded session. With these methods this data might be utilized, although the methods' applicability to BCI datasets with their excessive and non-iid noise needs to be validated.

5.3.3 Towards uncued BCI classification

A limitation of most BCI research to date — including our own research — is that the pace of communication is completely determined by the system. That is, the system dictates a window in which the user can perform a predetermined mental task to control the system. This is often true even when no stimulation is necessary to perform the mental task (e.g. motor imagery). For most practical applications the continuous classification of brain signals is preferred, since the explicit trial based communication is unnecessarily restrictive for the user. The transition form cued trials to uncued classification implies that the user can decide to perform a function at any time. This has two implications for the design of the BCI: 1) the BCI has to perform robustly when the user is not engaged with the BCI and performs unrelated tasks, and 2) the BCI classifier should not depend on a specific alignment of the trial with the sliding classification window.

The first challenge of producing few false positives can be addressed by selecting mental tasks with big contrasts between the active tasks and the resting state. For identifying these tasks we depend on the field of neuroscience.

The second issue of alignment reshapes the definition of the learning problem used to train the BCI classifier. Since it is unknown what part of the brain signature is being processed, or even if the user is performing a task at all, the BCI has to discriminate reliably under a much larger variety of circumstances. Usually uncued classification is achieved by training a BCI classifier on cued trails, and subsequently modifying the classification pipeline to produce predictions at regular intervals. This method further violates the iid assumption, and lacks proper theoretical backing. A better approach would be to use an overlapping sliding classification window, and let the machine learning (ML) methods model the mapping from the space of the different (partial) observations to labels. But since features extracted with this sliding window share most of their values with the previous window they are by definition, strongly related. Therefore, classification problems based on a sliding window form a natural candidate for classification methods that relax the iid assumption, such as the dSVM presented in Chapter 4.

An additional complexity is added by uncued BCI operation, since a pretrial baseline cannot be easily defined. Since the SOB method (Chapter 3) depends on such a baseline, an alternative baseline method needs to be devised. One approach would be to use a fixed delay for the baseline; preliminary results are presented in [53].

5.3.4 The iid assumption

One of the elusive aspects of BCIs remains, and that is the problem of non-stationary feature distributions. While variations in the performance are frequently stated in the literature, hard evidence of these fluctuations is difficult to gather. We believe this is due to a bias induced in the past years by favouring robustly performing feature spaces and classifiers with a low capacity for overfitting (e.g. linear classifiers).

By definition, the common assumption that data is iid is violated when the feature distributions are produced by a non-stationary process. Since the iid assumption abounds both in the ML and the common evaluation methods, weakening the iid assumption requires the reevaluation of known best practises. The problem caused by incorrectly assuming iid is a form of overfitting. Combined with the small training sets resulting from per session re-training, this might explain why previous attempts to use non-linear classifiers, or more complex feature spaces have been unsuccessful to date.

In Chapter 4 we showed that modeling the dependence of classification errors is beneficial for the BCI performance. Still, the relation between the changes in the feature distributions and the iid assumption needs to be better understood. Based on this understanding better classifiers and evaluation guidelines that take the relatedness of measures into account need to be developed. The violated iid assumption is one of the most intriguing aspects of working with BCI data, and our struggle with non-iid data might lead to new insights and novel classification models.

Bibliography

[1] Sharon Anderson, Ariane Auquier, Walter W. Hauck, David Oakes, Walter Vandaele, and Herbert I. Weisberg. *Statistical Methods for Comparative Studies: Techniques for Bias Reduction*. Wiley series in probability and mathematical statistics. John Wiley & Sons, Inc., New York, 1980.

[2] Niels Birbaumer, Nimr Ghanayim, Thilo Hinterberger, Iver Iversen, Boris Kotchoubey, Andrea Kübler, Juri Perelmouter, Edward Taub, and Herta Flor. A spelling device for the paralysed. *Nature*, 398:297–298, 1999.

[3] Benjamin Blankertz, Guido Dornhege, Christin Schäfer, Roman Krepki, Jens Kohlmorgen, Klaus-Robert Müller, Volker Kunzmann, Florian Losch, and Gabriel Curi. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):127–131, 2003.

[4] Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Klaus-Robert Müller, Volker Kunzmann, Florian Losch, and Gabriel Curio. The Berlin Brain-Computer Interface: EEG-based communication without subject training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):147–152, 2006.

[5] Benjamin Blankertz, Motoaki Kawanabe, Ryota Tomioka, Friederieke U. Hohlefeld, Vadim Nikullin, and Klaus-Robert Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 113–120, 2007.

[6] Benjamin Blankertz, Florian Losch, Matthias Krauledat, Guido Dornhege, Gabriel Curio, and Klaus-Robert Müller. The Berlin Brain-Computer Interface: Accurate performance from first-session in BCI-

naïve subjects. *IEEE Transactions on Biomedical Engineering*, 55(10): 2452–2462, 2008.

[7] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Müller. Optimizing spatial filters for robust EEG single-trial analysis. *Signal Processing Magazine*, 25(1):41–56, 2008.

[8] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Müller. Single-trial analysis and classification of ERP components — a tutorial. *NeuroImage*, 56(2):814–825, May 2011.

[9] Pavel Bobrov, Alexander Frolov1, Charles Cantor, Irina Fedulova, Mikhail Bakhnyan, and Alexander Zhavoronkov. Brain-computer interface based on generation of visual images. *PLoS ONE*, 6(6):e20674, 2011.

[10] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, March 1994.

[11] Ryan T. Canolty, Erik Edwards, Sarang S. Dalal, Maryam Soltani, Srikantan S. Nagarajan, Heidi E. Kirsch, Mitchel S. Berger, Nicholas M. Barbaro, and Robert T. Knight. High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793):1626–1628, 2006.

[12] Christoforos Christoforou, Robert Haralick, Paul Sajda, and Lucas C. Parra. Second-order bilinear discriminant analysis. *Journal of Machine Learning Research*, 11:665–685, 2010.

[13] Michael X. Cohen. It's about time. *Frontiers in Human Neuroscience*, 5 (2), 2011.

[14] Laura L. Colgin, Tobias Denninger, Marianne Fyhn, Torkel Hafting, Tora Bonnevie, Ole Jensen, May-Britt Moser, and Edvard I. Moser. Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*, 462, November 2009.

[15] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[16] Mihaly Csikszentmihalyi. *Flow: the psychology of optimal experience.* Harper & Row, New York, 1990.

[17] José del R. Millán, Rüdiger Rupp, Gernot Müller-Putz, Roderick Murray-Smith, Claudio Giugliemma, Michael Tangermann, Carmen Vidaurre, Febo Cincott, Andrea Kübler, Robert Leeb, Christa Neuper, Klaus-Robert Müller, and Donatella Mattia. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in Neuroscience*, 4:161, 2010.

[18] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, December 2006.

[19] Holger Diener and Karina Oertel. Experimental approach to affective interaction in games. In *Edutainment*, volume 3942/2006, pages 507–518, 2006.

[20] Jason Farquhar. A linear feature space for simultaneous learning of spatio-spectral filters in BCI. *Neural Networks*, 22:1278–1285, 2009.

[21] Lawrence A. Farwell and Emanual Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.

[22] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2005.

[23] Siamac Fazli, Florin Popescu, Márton Danóczy, Benjamin Blankertz, Klaus-Robert Müller, and Cristian Grozea. Subject-independent mental state classification in single trials. *Neural Networks*, 22(6):1305–1312, 2009.

[24] Siamac Fazli, Márton Danóczy, Jürg Schelldorfer, and Klaus-Robert Müller. ℓ_1-penalized linear mixed-effects models for high dimensional data with application to BCI. *NeuroImage*, 56(4):2100–2108, 2011.

[25] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23):e215–e220, 2000.

[26] Heather M. Gray, Nalini Ambady, William T. Lowenthal, and Patricia Deldin. P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology*, 40:216–224, 2004.

[27] Hayrettin Gürkök, Mannes Poel, and Job Zwiers. Classifying motor imagery in presence of speech. In *Proceedings of the 2010 International*

Joint Conference on Neural Networks (IJCNN 2010), pages 1235–1242, 2010.

[28] Jeremy Hill, Jason Farquhar, Thomas N. Lal, and Bernhard Schölkopf. Time-dependent demixing of task-relevant EEG signals. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006*, 2006.

[29] Altinus Lucilus Hof. EMG and muscle force: An introduction. *Human Movement Science*, 3(1–2):119–153, 1984.

[30] Jane E. Huggins, Patricia A. Wren, and Kirsten L. Gruis. What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*, 2011. In press.

[31] Sabine Jatzev, Thorsten O. Zander, Monica De Filippis, Christian Kothe, Sebastian Welke, and Matthias Rötting. Examining causes for nonstationarities: the loss of controllability is a factor which induces nonstationarities. In *Proceedings of the 4th International BCI Workshop & Training Course*, pages 138–143. Verlag der Technischen Universität Graz, 2008.

[32] Jonathan Klein, Youngme Moon, and Rosalind W. Picard. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14:119–140, 2002.

[33] Thomas Koelewijn, Hein T. van Schie, Harold Bekkering, Robert Oostenveld, and Ole Jensen. Motor-cortical beta oscillations are modulated by correctness of observed action. *NeuroImage*, 40:767–775, 2008.

[34] Zoltan J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 79(6):440–447, December 1991.

[35] Matthias Krauledat, Guido Dornhege, Benjamin Blankertz, Florian Losch, Gabriel Curio, and Klaus-Robert Müller. Improving speed and accuracy of brain-computer interfaces using readiness potential features. In *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS 2004)*, pages 4511–4515, 2004.

[36] Matthias Krauledat, Michael Tangermann, Benjamin Blankertz, and Klaus-Robert Müller. Towards zero training for brain-computer interfacing. *PLoS ONE*, 3:e2967, 2008.

[37] Rumyana Kristeva-Feige, Christoph Fritsch, Jens Timmer, and Carl-Hermann Lücking. Effects of attention and precision of exerted force on beta range EEG-EMG synchronization during a maintained motor contraction task. *Clinical Neurophysiology*, 113:123–131, 2002.

[38] Julien Kronegg, Guillaume Chanel, Sviatoslav Voloshynovskiy, and Thierry Pun. EEG-based synchronized brain-computer interfaces: A model for optimizing the number of mental tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(1):50–58, 2007.

[39] Andrea Kübler, Boris Kotchoubey, Thilo Hinterberger, Nimr Ghanayim, Juri Perelmouter, Margarete Schauer, Christoph Fritsch, Edward Taub, and Niels Birbaumer. The thought translation device: a neurophysiological approach to communication in total motor paralysis. *Experimental Brain Research*, 124(2):223–232, 1999.

[40] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, February 2004.

[41] Fabien Lotte, Cuntai Guan, and Kai Keng Ang. Comparison of designs towards a subject-independent brain-computer interface based on motor imagery. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009)*, pages 4543–4546, 2009.

[42] Thomas M. Loughin. A systematic comparison of methods for combining p-values from independent tests. *Computational Statistics & Data Analysis*, 47:467–485, 2004.

[43] Dennis J. McFarland, Laurie A. Miner, Theresa M. Vaughan, and Jonathan R. Wolpaw. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topography*, 12(3):117–186, October 2000.

[44] Frank C. Meinecke, Paul von Bünau, Motoaki Kawanabe, and Klaus-Robert Müller. Learning invariances with stationary subspace analysis. In *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 87–92, September 2009.

[45] Klaus-Robert Müller, Michael Tangermann, Guido Dornhege, Matthias Krauledat, Gabriel Curio, and Benjamin Blankertz. Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167: 82–90, 2008.

[46] Anton Nijholt, Danny Plass-Oude Bos, and Boris Reuderink. Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Computing*, 1(2):85–94, October 2009.

[47] Daria Osipova, Dora Hermes, and Ole Jensen. Gamma power is phase-locked to posterior alpha activity. *PLoS ONE*, 3(12):e3990, December 2008.

[48] Gert Pfurtscheller and Fernando Henrique Lopes da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110:1842–1857, 1999.

[49] Robert S. Pindyck and Daniel L. Rubinfeld. *Econometric models and economic forecasts*. McGraw-Hill, New York, 1991.

[50] Filippo Portera and Alessandro Sperduti. A generalized quadratic loss exploiting target information for support vector machines. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI2004)*, pages 628–632. IOS Press, 2004.

[51] Herbert Ramoser, Johannes Müller-Gerking, and Gert Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446, 2000.

[52] Boris Reuderink, Anton Nijholt, and Mannes Poel. Affective Pacman: A frustrating game for brain-computer interface experiments. In *Proceedings of the 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2009)*, volume 9 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 221–227, May 2009.

[53] Boris Reuderink, Jason Farquhar, and Mannes Poel. Slow sphering to suppress non-stationarities in the EEG. *International Journal of Bioelectromagnetism*, 2010. To appear.

[54] Boris Reuderink, Jason Farquhar, Mannes Poel, and Anton Nijholt. A subject-independent brain-computer interface based on smoothed, second-order baselining. In *Proceedings of the 33st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2011)*, 2011. To appear.

[55] Boris Reuderink, Mannes Poel, and Anton Nijholt. The impact of loss of control on movement BCIs. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2011. To appear.

[56] Boris Reuderink, Christian Mühl, and Mannes Poel. Valence, arousal and dominance in the EEG during game play. *International Journal of Autonomous and Adaptive Communication Systems*, 2012. To appear.

[57] Abdul Satti, Cuntai Guan, Damien Coyle, and Girijesh Prasad. A covariate shift minimisation method to alleviate non-stationarity effects for an adaptive brain-computer interface. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR 2010)*, pages 105–108, 2010.

[58] Gerwin Schalk, 2011. Personal correspondence.

[59] Gerwin Schalk, Dennis J. McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R. Wolpaw. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004.

[60] Jocelyn Scheirer, Raul Fernandez, Jonathan Klein, and Rosalind W. Picard. Frustrating the user on purpose: A step toward building an affective computer. *Interacting with Computers*, 14:93–118, February 2002.

[61] Alois Schlögl, Claudia Keinrath, Doris Zimmermann, Reinhold Scherer, Robert Leeb, and Gert Pfurtscheller. A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*, 118:98–104, 2007.

[62] Pradeep Shenoy, Matthias Krauledat, Benjamin Blankertz, Rajesh P. N. Rao, and Klaus-Robert Müller. Towards adaptive classification for BCI. *Journal of Neural Engineering*, 3(1):R13–R23, 2006.

[63] Hiroshi Shibasaki and Mark Hallett. What is the bereitschaftspotential? *Clinical Neurophysiology*, 117(11):2341–2356, 2006.

[64] Andrej Stančák Jr., Andrea Riml, and Gert Pfurtscheller. The effects of external load on movement-related changes of the sensorimotor EEG rhythms. *Electroencephalography and Clinical Neurophysiology*, 102: 495–504, 1997.

[65] Shiliang Sun and Changshui Zhang. Adaptive feature extraction for eeg signal classification. *Medical and Biological Engineering and Computing*, 44:931–935, 2006.

[66] Ryota Tomioka, Jeremy Hill, Benjamin Blankertz, and Kazuyuki Aihara. Adapting spatial filtering methods for nonstationary BCIs. In *2006 Workshop on Information-Based Induction Sciences (IBIS2006)*, 2006.

[67] Marcel van Gerven and Ole Jensen. Attention modulations of posterior alpha as a control signal for two-dimensional brain-computer interfaces. *Journal of Neuroscience Methods*, 179:78–84, 2009.

[68] Marcel van Gerven, Jason Farquhar, Rebecca Schaefer, Rutger Vlek, Jeroen Geuze, Anton Nijholt, Nick Ramsey, Pim Haselager, Louis Vuurpijl, Stan Gielen, and Peter Desain. The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4):041001 (10pp), 2009.

[69] Jacques J. Vidal. Real-time detection of brain events in EEG. *Proceedings of the IEEE*, 65(5):633–641, May 1977.

[70] Carmen Vidaurre, Alois Schlögl, Rafael Cabeza, Reinhold Scherer, and Gert Pfurtscheller. Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. *IEEE Transactions on Biomedical Engineering*, 54(3):550–556, March 2007.

[71] Carmen Vidaurre, Motoaki Kawanabe, Paul von Bünau, Benjamin Blankertz, and Klaus-Robert Müller. Toward unsupervised adaptation of LDA for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 58(3):587–597, March 2011.

[72] Paul von Bünau, Frank C. Meinecke, and Klaus-Robert Müller. Stationary subspace analysis. In *Independent Component Analysis and Signal Separation 8th International Conference (ICA 2009)*, volume 5441/2009 of *Lecture Notes in Computer Science*, pages 1–8, 2009.

[73] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.

[74] Jonathan R. Wolpaw, Gerald E. Loeb, Brendan Z. Allison, Emanuel Donchin, Omar Feix do Nascimento, William J. Heetderks, Femke Nijboer, William G. Shain, and James N. Turner. BCI meeting 2005—workshop on signals and recording methods. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):138–141, 2006.

[75] Thorsten O. Zander and Christian Kothe. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, 8 (2):025005 (5pp), 2011.

[76] Thorsten O. Zander, Christian Kothe, Sebastian Welke, and Matthias Rötting. Utilizing secondary input from passive brain-computer interfaces for enhancing human-machine interaction. In *Lecture Notes in Computer Science*, volume 5638/2009, pages 759–771, 2009.

Notation

c	Scalar c
A	Matrix A
A^T	A transposed
$\vec{x}$	Column vector x
$\vec{1}$	A vector $[1\ 1\ \cdots\ 1]^T$ of matching shape
$E[x]$	Expected value of x
$A \circ B$	Hadamard (element-wise) product of matrix A and B
$A_{i,\cdot}$	The i-th row of A
$A_{\cdot,j}$	The j-th column of A
$\mathrm{vec}(A)$	Column vector based on the stacked columns of A
$\mathrm{diag}(A)$	Column vector based on elements on the diagonal of A
$\prec, \preceq, \succ, \succeq$	Element wise vector/matrix inequalities

Publication list

[77] Gido Hakvoort, Boris Reuderink, and Michel Obbink. Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system. Technical Report TR-CTIT-11-03, Human Media Interaction, Faculty of EEMCS, University of Twente, Enschede, February 2011.

[46] Anton Nijholt, Danny Plass-Oude Bos, and Boris Reuderink. Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Computing*, 1(2):85–94, October 2009.

[79] Anton Nijholt, Boris Reuderink, and Danny Oude Bos. Turning shortcomings into challenges: Brain-computer interfaces for games. In *Proceedings of the 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2009)*, volume 9 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 153–168, May 2009.

[80] Danny Oude Bos and Boris Reuderink. BrainBasher: a BCI game. In *Extended Abstracts of the International Conference on Fun and Games 2008*, pages 36–39, Eindhoven, The Netherlands, October 2008. Eindhoven University of Technology.

[81] Danny Plass-Oude Bos, Boris Reuderink, Bram van de Laar, Hayrettin Gürkök, Christian Mühl, Mannes Poel, Dirk K. J. Heylen, and Anton Nijholt. Human-computer interaction for BCI games: Usability and user experience. In *Proceedings of the International Conference on CYBERWORLDS 2010*, pages 277–281, 2010.

[82] Danny Plass-Oude Bos, Boris Reuderink, Bram L. A. van de Laar, Hayrettin Gürkök, Christian Mühl, Mannes Poel, Anton Nijholt, and Dirk K. J. Heylen. *Brain-Computer Interfacing and Games*, pages 149–178. Human-Computer Interaction Series. Springer, London, June 2010.

[83] Danny Plass-Oude Bos, Matthieu Duvinage, Oytun Oktay, Jaime Delgado Saa, Huseyin Gürüler, Ayhan Istanbullu, Marijn van Vliet, Bram van de Laar, Mannes Poel, Linsey Roijendijk, Luca Tonin, Ali Bahramisharif, and Boris Reuderink. Looking around with your brain in a virtual world. In *IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain*, 2011.

[84] Boris Reuderink. Games and brain-computer interfaces: The state of the art. Technical Report TR-CTIT-08-81, Human Media Interaction, Faculty of EEMCS, University of Twente, September 2008.

[85] Boris Reuderink and Mannes Poel. Robustness of the common spatial patterns algorithm in the BCI-pipeline. Technical Report TR-CTIT-08-52, Human Media Interaction, Faculty of EEMCS, University of Twente, Enschede, July 2008.

[52] Boris Reuderink, Anton Nijholt, and Mannes Poel. Affective Pacman: A frustrating game for brain-computer interface experiments. In *Proceedings of the 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2009)*, volume 9 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 221–227, May 2009.

[53] Boris Reuderink, Jason Farquhar, and Mannes Poel. Slow sphering to suppress non-stationarities in the EEG. *International Journal of Bioelectromagnetism*, 2010. To appear.

[54] Boris Reuderink, Jason Farquhar, Mannes Poel, and Anton Nijholt. A subject-independent brain-computer interface based on smoothed, second-order baselining. In *Proceedings of the 33st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2011)*, 2011. To appear.

[55] Boris Reuderink, Mannes Poel, and Anton Nijholt. The impact of loss of control on movement BCIs. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2011. To appear.

[56] Boris Reuderink, Christian Mühl, and Mannes Poel. Valence, arousal and dominance in the EEG during game play. *International Journal of Autonomous and Adaptive Communication Systems*, 2012. To appear.

[91] Bram van de Laar, Danny Plass-Oude Bos, Boris Reuderink, Mannes Poel, and Anton Nijholt. How much control is enough? optimizing fun with unreliable input. *International Journal of Ergonomics*, 2011. Submitted.

[92] Bram L. A. van de Laar, Danny Oude Bos, Boris Reuderink, and Dirk K. J. Heylen. Actual and imagined movement in BCI gaming. In *Proceedings of the AI and Games Symposium at the Artificial Intelligence and Simulation of Behaviour 2009 Convention (AISB 2009)*, pages 59–63, Heriot-Watt University, Edinburgh, Scotland, April 2009. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

[93] Bram L. A. van de Laar, Danny Oude Bos, Boris Reuderink, and Dirk K. J. Heylen. Evaluating user experience of actual and imagined movement in BCI gaming. *International Journal of Gaming and Computer-Mediated Simulations*, 2(4):33–47, 2010.

[94] Marijn van Vliet, Christian Mühl, Boris Reuderink, and Mannes Poel. Guessing what's on your mind: Using the N400 in brain computer interfaces. In *Proceedings of International Conference on Brain Informatics (BI 2010)*, volume 6334/2010 of *Lecture Notes in Computer Science*, pages 180–191, 2010.

SIKS Dissertation Series

Since 1998, all dissertations written by Ph.D.-students who have conducted their research under auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series.

2011-36 Erik van der Spek (UU), *Experiments in serious game design: a cognitive approach.*

2011-35 Maaike Harbers (UU), *Explaining Agent Behavior in Virtual Training.*

2011-34 Paolo Turrini (UU), *Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations.*

2011-33 Tom van der Weide (UU), *Arguing to Motivate Decisions.*

2011-32 Nees-Jan van Eck (EUR), *Methodological Advances in Bibliometric Mapping of Science.*

2011-31 Ludo Waltman (EUR), *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality.*

2011-30 Egon van den Broek (UT), *Affective Signal Processing (ASP): Unraveling the mystery of emotions.*

2011-29 Faisal Kamiran (TUE), *Discrimination-aware Classification.*

2011-28 Rianne Kaptein(UVA), *Effective Focused Retrieval by Exploiting Query Context and Document Structure.*

2011-27 Aniel Bhulai (VU), *Dynamic website optimization through autonomous management of design patterns.*

2011-26 Matthijs Aart Pontier (VU), *Virtual Agents for Human Communication — Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots.*

2011-25 Syed Waqar ul Qounain Jaffry (VU)), *Analysis and Validation of Models for Trust Dynamics.*

2011-24 Herwin van Welbergen (UT), *Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior.*

2011-23 Wouter Weerkamp (UVA), *Finding People and their Utterances in Social Media.*

2011-22 Junte Zhang (UVA), *System Evaluation of Archival Description and Access.*

2011-21 Linda Terlouw (TUD), *Modularization and Specification of Service-Oriented Systems.*

2011-20 Qing Gu (VU), *Guiding service-oriented software engineering — A view-based approach.*

2011-19 Ellen Rusman (OU), *The Mind's Eye on Personal Profiles.*

2011-18 Mark Ponsen (UM), *Strategic Decision-Making in complex games.*

2011-17 Jiyin He (UVA), *Exploring Topic Structure: Coherence, Diversity and Relatedness.*

2011-16 Maarten Schadd (UM), *Selective Search in Games of Different Complexity.*

2011-15 Marijn Koolen (UvA), *The Meaning of Structure: the Value of Link Evidence for Information Retrieval.*

2011-14 Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets.*

2011-13 Xiaoyu Mao (UvT), *Airport under Control. Multiagent Scheduling for Airport Ground Handling.*

2011-12 Carmen Bratosin (TUE), *Grid Architecture for Distributed Process Mining.*

2011-11 Dhaval Vyas (UT), *Designing for Awareness: An Experience-focused HCI Perspective.*

2011-10 Bart Bogaert (UvT), *Cloud Content Contention.*

2011-09 Tim de Jong (OU), *Contextualised Mobile Media for Learning.*

2011-08 Nieske Vergunst (UU), *BDI-based Generation of Robust Task-Oriented Dialogues.*

2011-07 Yujia Cao (UT), *Multimodal Information Presentation for High Load Human Computer Interaction.*

2011-06 Yiwen Wang (TUE), *Semantically-Enhanced Recommendations in Cultural Heritage.*

2011-05 Base van der Raadt (VU), *Enterprise Architecture Coming of Age — Increasing the Performance of an Emerging Discipline..*
2011-04 Hado van Hasselt (UU), *Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference.*
2011-03 Jan Martijn van der Werf (TUE), *Compositional Design and Verification of Component-Based Information Systems.*
2011-02 Nick Tinnemeier(UU), *Work flows in Life Science.*
2011-01 Botond Cseke (RUN), *Variational Algorithms for Bayesian Inference in Latent Gaussian Models.*
2010-53 Edgar Meij (UVA), *Combining Concepts and Language Models for Information Access.*
2010-52 Peter-Paul van Maanen (VU), *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention.*
2010-51 Alia Khairia Amin (CWI), *Understanding and supporting information seeking tasks in multiple sources.*
2010-50 Bouke Huurnink (UVA), *Search in Audiovisual Broadcast Archives.*
2010-49 Jahn-Takeshi Saito (UM), *Solving difficult game positions.*
2010-48 Withdrawn, .
2010-47 Chen Li (UT), *Mining Process Model Variants: Challenges, Techniques, Examples.*
2010-46 Vincent Pijpers (VU), *e3alignment: Exploring Inter-Organizational Business-ICT Alignment.*
2010-45 Vasilios Andrikopoulos (UvT), *A theory and model for the evolution of software services.*
2010-44 Pieter Bellekens (TUE), *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain.*
2010-43 Peter van Kranenburg (UU), *A Computational Approach to Content-Based Retrieval of Folk Song Melodies.*
2010-42 Sybren de Kinderen (VU), *Needs-driven service bundling in a multi-supplier setting — the computational e3-service approach.*
2010-41 Guillaume Chaslot (UM), *Monte-Carlo Tree Search.*
2010-40 Mark van Assem (VU), *Converting and Integrating Vocabularies for the Semantic Web.*
2010-39 Ghazanfar Farooq Siddiqui (VU), *Integrative modeling of emotions in virtual agents.*
2010-38 Dirk Fahland (TUE), *From Scenarios to components.*
2010-37 Niels Lohmann (TUE), *Correctness of services and their composition.*
2010-36 Jose Janssen (OU), *Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification.*
2010-35 Dolf Trieschnigg (UT), *Proof of Concept: Concept-based Biomedical Information Retrieval.*
2010-34 Teduh Dirgahayu (UT), *Interaction Design in Service Compositions.*
2010-33 Robin Aly (UT), *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval.*
2010-32 Marcel Hiel (UvT), *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems.*
2010-31 Victor de Boer (UVA), *Ontology Enrichment from Heterogeneous Sources on the Web.*
2010-30 Marieke van Erp (UvT), *Accessing Natural History — Discoveries in data cleaning, structuring, and retrieval.*
2010-29 Stratos Idreos(CWI), *Database Cracking: Towards Auto-tuning Database Kernels.*
2010-28 Arne Koopman (UU), *Characteristic Relational Patterns.*
2010-27 Marten Voulon (UL), *Automatisch contracteren.*
2010-26 Ying Zhang (CWI), *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines.*
2010-25 Zulfiqar Ali Memon (VU), *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective.*
2010-24 Dmytro Tykhonov, *Designing Generic and Efficient Negotiation Strategies.*
2010-23 Bas Steunebrink (UU), *The Logical Structure of Emotions.*
2010-22 Michiel Hildebrand (CWI), *End-user Support for Access to*
Heterogeneous Linked Data.
2010-21 Harold van Heerde (UT), *Privacy-aware data management by means of data degradation.*
2010-20 Ivo Swartjes (UT), *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative.*
2010-19 Henriette Cramer (UvA), *People's Responses to Autonomous and Adaptive Systems.*
2010-18 Charlotte Gerritsen (VU), *Caught in the Act: Investigating Crime by Agent-Based Simulation.*
2010-17 Spyros Kotoulas (VU), *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications.*
2010-16 Sicco Verwer (TUD), *Efficient Identification of Timed Automata, theory and practice.*
2010-15 Lianne Bodenstaff (UT), *Managing Dependency Relations in Inter-Organizational Models.*
2010-14 Sander van Splunter (VU), *Automated Web Service Reconfiguration.*
2010-13 Gianluigi Folino (RUN), *High Performance Data Mining using Bio-inspired techniques.*
2010-12 Susan van den Braak (UU), *Sensemaking software for crime analysis.*

2010-11 Adriaan Ter Mors (TUD), *The world according to MARP: Multi-Agent Route Planning.*

2010-10 Rebecca Ong (UL), *Mobile Communication and Protection of Children.*

2010-09 Hugo Kielman (UL), *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging.*

2010-08 Krzysztof Siewicz (UL), *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments.*

2010-07 Wim Fikkert (UT), *Gesture interaction at a Distance.*

2010-06 Sander Bakkes (UvT), *Rapid Adaptation of Video Game AI.*

2010-05 Claudia Hauff (UT), *Predicting the Effectiveness of Queries and Retrieval Systems.*

2010-04 Olga Kulyk (UT), *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments.*

2010-03 Joost Geurts (CWI), *A Document Engineering Model and Processing Framework for Multimedia documents.*

2010-02 Ingo Wassink (UT), *Work flows in Life Science.*

2010-01 Matthijs van Leeuwen (UU), *Patterns that Matter.*

Propositions

Boris Reuderink

October 7, 2011

1. The assumption in machine learning and statistics that our observations are *independently and identically distributed* is a lie that we need in order to do our research (Chapter 4).
2. While the BCI community is generally opposed to hyping the technology in the media, the field often contributes to unrealistic expectations by using improper evaluation methodologies.
3. Multiple-test correction is used to control false positives when a series of statistical tests is performed for a single hypothesis. This is only sensible if one corrects also for all the fruitless tests performed on the same dataset.
4. Much of the observed "non-stationary" BCI classification performance within sessions is caused by the researchers interpreting structure in random noise.
5. It would be beneficial if the obligatory peer review for articles was extended to review software and scripts developed for the article as well.
6. The Kolmogorov complexity of a PhD thesis in a technical field is inversely related to its quality.
7. Many insights from machine learning apply to our daily lives as well; for example, most researchers have issues with the non-differentiable cost associated with missing a deadline.
8. Hoewel een computer is een uitstekend medium is om ideeën op uit te werken, ontstaan deze ideeën meestal niet achter een beeldscherm.
9. Ouderschap vormt gek genoeg geen belemmering voor het werken aan een proefschrift; blijkbaar is slaaptekort niet cumulatief.