

TOWARD USER-ADAPTED BRAIN COMPUTER INTERFACES: ROBUST INTERACTION AND MACHINE LEARNING BASED ON RIEMANNIAN GEOMETRY

By

Emmanuel K. Kalunga

Submitted in partial fulfilment of the requirement for the degrees of

Doctor of Philosophy (Ph.D.): Computer Science
École Doctorale : Sciences et Technologies de l'Information et de la
Communication

Université Paris-Saclay, UVSQ

and

Doctor Technologiae (D-Tech): Electrical Engineering
Faculty of Engineering and the Built Environment
Tshwane University of Technology

Supervisors:

Karim Djouani - Professor

Eric Monacelli - Associate Professor

Co-supervisors:

Sylvain Chevallier - Associate Professor

Yskandar Hamam - Professor

I hereby declare that this dissertation submitted for the degree of Doctor of Philosophy: Computer Science, at the Université Paris-Saclay, UVSQ, and the Tshwane University of Technology, TUT, is my own original work and has not previously been submitted to any other institution of higher education. I further declare that all sources cited or quoted are indicated and acknowledged by means of a comprehensive list of references.

Emmanuel K. Kalunga

Date

Abstract

In the last two decades, interest in Brain-Computer Interfaces (BCI) has tremendously grown, with a number of research laboratories working on the topic. Since the *Brain-Computer Interface Project* of Vidal in 1973, where BCI was introduced for rehabilitative and assistive purposes, the use of BCI has been extended to more applications such as neurofeedback and entertainment. The credit of this progress should be granted to an improved understanding of electroencephalography (EEG), an improvement in its measurement techniques, and increased computational power.

Despite the opportunities and potential of Brain-Computer Interface, the technology has yet to reach maturity and be used out of laboratories. There are several challenges that need to be addressed before BCI systems can be used to their full potential. This work examines in depth some of these challenges, namely the specificity of BCI systems to users physical abilities, the robustness of EEG representation and machine learning, and the adequacy of training data. The aim is to provide a BCI system that can adapt to individual users in terms of their physical abilities/disabilities, and variability in recorded brain signals.

To this end, two main avenues are explored: the first, which can be regarded as a high-level adjustment, is a change in BCI paradigms. It is about creating new paradigms that increase their performance, ease the discomfort of using BCI systems, and adapt to the user's needs. The second avenue, regarded as a low-level solution, is the refinement of signal processing and machine learning

techniques to enhance the EEG signal quality, pattern recognition and classification.

On the one hand, a new methodology in the context of assistive robotics is defined: it is a hybrid approach where a physical interface is complemented by a Brain-Computer Interface (BCI) for human machine interaction. This hybrid system makes use of users residual motor abilities and offers BCI as an optional choice: the user can choose when to rely on BCI and could alternate between the muscular- and brain-mediated interface at the appropriate time.

On the other hand, for the refinement of signal processing and machine learning techniques, this work uses a Riemannian framework. A major limitation in this field is the EEG poor spatial resolution. This limitation is due to the volume conduction effect, as the skull bones act as a non-linear low pass filter, mixing the brain source signals and thus reducing the signal-to-noise ratio. Consequently, spatial filtering methods have been developed or adapted. Most of them (i.e. Common Spatial Pattern, xDAWN, and Canonical Correlation Analysis) are based on covariance matrix estimations. The covariance matrices are key in the representation of information contained in the EEG signal and constitute an important feature in their classification. In most of the existing machine learning algorithms, covariance matrices are treated as elements of the Euclidean space. However, being Symmetric and Positive-Definite (SPD), covariance matrices lie on a curved space that is identified as a Riemannian manifold. Using covariance matrices as features for classification of EEG signals and handling them with the tools provided by Riemannian geometry provide a robust framework for EEG representation and learning.

Table of Contents

Abstract	iii
Table of Contents	v
1 Introduction	1
1.1 Background	1
1.2 Research Problem	3
1.3 Research Objectives and Contributions	5
1.3.1 Objectives	5
1.3.2 Research Contributions	6
1.4 Thesis Outline	7
2 Neurobiological Aspects of Brain-Computer Interfaces	9
2.1 Introduction	9
2.2 Signal Acquisition	11
2.2.1 Local Field Potentials	12
2.2.2 Electrical Signal Acquisition	13
2.2.3 Magnetic Signal Acquisition	17
2.2.4 Hemodynamic Techniques	17
2.2.5 Discussion	20
2.3 Neurological Phenomena	22
2.3.1 Event-Related Synchronisation-based BCI	22
2.3.2 Event-Related Potential	27
2.3.3 Visual Evoked Potential	32
2.3.4 Discussion	37
3 Signal Processing and Machine Learning for BCI	39
3.1 Signal Processing for Cognitive Functions	39
3.1.1 Motor Imagery Processing	42

3.1.2	SSVEP Processing	48
3.1.3	P300 Processing	52
3.1.4	Discussion	54
3.2	Riemannian Approaches in Machine Learning	56
3.3	New Trend in BCI Systems	58
3.3.1	Hybrid BCI systems	59
3.3.2	Passive BCI	60
3.4	Proposed Approach	61
4	Hybrid Brain-Computer Interface	64
4.1	Introduction	64
4.2	Proposed Hybrid Interface	65
4.3	Touchless Interface	66
4.4	SSVEP-based BCI	68
4.4.1	Material and EEG Recording	68
4.4.2	Signal Processing	69
4.5	Applications	70
4.6	Experimental Results	71
4.6.1	Validation of Proposed SSVEP Algorithm	71
4.6.2	Experiments in Virtual Environment	72
4.6.3	Application to Exoskeleton Arm Control	73
4.7	Conclusion	76
5	Riemannian Geometry for Brain-Computer Interfaces	79
5.1	Riemannian Manifold of Symmetric Positive-Definite Matrices	79
5.2	Covariance Matrix Estimation	81
5.2.1	Sample Covariance Matrix Estimator	81
5.2.2	Shrinkage Covariance Matrix Estimators	82
5.2.3	Fixed-Point Covariance Matrix Estimator	82
5.3	Classification of SSVEP Covariance Matrices	83
5.3.1	Machine Learning Approach for Classification	83
5.3.2	Means of Covariance Matrices	84
5.3.3	Minimum Distance to Mean Classifier for SSVEP	93
5.4	Online Classification	95
5.4.1	Curve-Based Online Classification	95
5.4.2	Outliers Removal with Riemannian Potato	99
5.5	Experimental Validation	100
5.5.1	Covariance Estimators Comparison	100
5.5.2	Effect of Outliers on Centre Estimation	103

5.5.3	From Euclidean to Riemannian Centres of Class	103
5.5.4	Classification Results and Analysis	107
5.6	Conclusion	112
6	Perspectives for Riemannian Approaches	119
6.1	Introduction	119
6.2	Data Augmentation	122
6.2.1	Generating Artificial Points on Riemannian Manifold . .	122
6.2.2	Classification	123
6.2.3	Experimental Data Description	123
6.2.4	Results and Discussion	124
6.3	Transfer Learning	128
6.3.1	User Specificity as Domain in Transfer Learning	128
6.3.2	Category of Proposed Transfer Learning	129
6.3.3	Composite Riemanian Mean	131
6.3.4	Experimental Results	133
6.4	Conclusion	136
7	Conclusion	138
Bibliography		140

Chapter 1

Introduction

The only way of discovering the limits of the possible is to venture a little way past them into the impossible.

— Arthur C. Clarke, *Profiles of the Future: An Enquiry into the Limits of the Possible*

This chapter follows the structure of the Tshwane University of Technology's theses. It is an introductory chapter that lays out the background of the research field, states the research problem and objective, and summarises the contribution brought by this work to the research filed.

1.1 Background

Brain-computer interfaces (BCI) also called brain machine interfaces (BMI) are devices that translates measured brain activity into tangible actions, allowing humans and apes to interact with the physical environment without using their muscular system. In the last two decades, interest in brain-computer interfaces has tremendously grown, with a number of research laboratories working on the topic. Since the *Brain-Computer Interface Project* [Vidal, 1973], joint effort from researchers in electronics, neuroscience, electrical engineering, signal processing, and machine learning – to name but a few, has promoted the use of BCI in different applications such as neurofeedback, entertainment, and assistance. Better understanding, improved measurement and processing of electroencephalograms (EEG) are at the centre of the growth of non-invasive

EEG based BCI. These interfaces have brought a complete paradigm shift to assistive technologies. In fact, unlike traditional human machine interfaces, BCI do not rely on motor abilities. Bypassing the neuromuscular pathways, BCI constitutes a golden opportunity for people with limited neuromuscular abilities or serious brain injuries. Brain-computer interfaces can be used for control or communication in replacement of traditional assistive devices [Wolpaw et al., 2002], or for improved human-machine interaction as a passive user feedback to the machine [Zander and Kothe, 2011].

BCI systems rely on neurological phenomena that can be measured in the brain signal – in response to a stimulus or a mental task, then quantified and interpreted using signal processing and machine learning techniques. Currently, the most used phenomena are the Steady State Visual Evoked Potential (SSVEP), Motor Imagery (MI) and P300 event-related potential. They respectively define three types of BCIs, each with limitations and advantages that can be exploited to achieve reliable brain-computer communication. In each BCI type, an appropriate experimental protocol is designed to stimulate the neurological response.

Although there are nowadays various techniques used to measure neuronal electrical activities in BCI (*e.g.* electrocorticography (ECoG), spikes and local field potentials (LFP), magnetoencephalography (MEG), etc.), EEG is still the main technique in BCI research. Despite its vulnerability to noise and low spatial resolution, EEG is appreciated over other techniques for its high temporal and spectral resolutions, its affordable, mobile and non-invasive acquisition equipment. Various electrodes types, configurations, and mounting are being proposed to improve the quality of recorded EEG, improve comfort, and reduce the setup time [Looney et al., 2012; Badcock et al., 2013].

Advances have been made in the signal processing and machine learning to extract the signal of interest from the ongoing brain activity and noise recorded in EEG. Particularly spatial filters are reported to successfully extract the signal of interest related to the BCI task. They have been used to achieve the most successful performance in various BCI types [Ang et al., 2012; Rivet et al., 2009; Spüler et al., 2012; Kalunga et al., 2013; Nakanishi et al., 2014]. Spatial patterns learned in filters are well captured by the covariance matrices of the

multichannel EEG signal, which are key components in the computation of spatial filters. Once a spatial filter has been applied, a standard classification algorithm (e.g. LDA, SVM) can be used.

Both spatial filter and classifier parameters are optimised offline using a training sample of recorded EEG data. A bias-variance threshold can be achieved through a cross-validation process. To use a BCI system, depending on the BCI type, a user might be required to go through a training where he will be trained to control his brain signals (*i.e.* elicit appropriate phenomena). The user is also required to record multiple EEG trials to constitute a training sample for the machine learning algorithm. The training sample should be large enough to avoid the problems of overfitting and signal components should be carefully selected in order to alleviate the curse of dimensionality to which BCI is prone due to the high dimensional feature space of multichannel EEG data.

Initially designed for clinical as well as rehabilitative and assistive purposes, brain computer interfaces have gained more grounds with applications to neurofeedback, navigation, training and education, gaming and entertainment, etc. [Milln and Carmena, 2010; Van Erp et al., 2012; Lotte et al., 2015; Abdulkader et al., 2015; Mensia, 2016; melomind, 2016].

1.2 Research Problem

Despite the opportunities seen in BCI and the advances made in BCI research, particularly in brain signal acquisition techniques, signal processing and machine learning approaches, there have been only a few applications that have done well in the market [Mensia, 2016; melomind, 2016; g.Tec, 2012]. The technology has not matured enough for a broad usage by the public in delicate applications. There is a number of limitations that should be overcome before BCI applications could be taken outside laboratories. In the current work some of these problems are addressed.

Problem 1: User's Physical Specificity

Current BCI systems are built around their potential to bypass the neuromuscular system. This perspective results in interfaces that are the sole remedy for

completely locked-in patients as they cannot use any traditional assistive devices (*i.e.* muscle dependent). In this approach, all the effort is turned toward the BCI system and its capacity to classify users intentions. No much attention is paid to the specificity of the user. Problems with this approach emerge as users adapt differently to BCI and express different needs. They can be depicted in three facts. Fist, the problem of BCI inefficiency (or illiteracy). There is a reported 15 to 30% of people who cannot use brain computer interface [[Allison and Neuper, 2010a](#)]. An important fact, however, is that while they show illiteracy with one BCI type (e.g. SSVEP), they can still be efficient in using another type of BCI (e.g. motor imagery). Secondly, the locked-in patients constitute a minority of potential BCI users. For rehabilitation and assistive applications, other than locked-in patients, the majority of people with motor disabilities or severe brain injuries retain different residual motor skills. Therefore the extend to which they rely on BCI command might differ. Lastly, there is a high cognitive load that accompanies the command of BCI interface, and can affect users differently. These facts show that BCI should not be designed as a disruptive unique solution for all users. There is a need to adapt to each user's special skills and needs.

Problem 2: Robustness of EEG Representation and Machine Learning

The main limitations in EEG based BCI are related to signal quality of EEG, namely the poor spatial resolution of EEG and its vulnerability to artefacts. To avoid the influence of noise in the EEG, experiments are conducted in laboratories where the ambient noise is controlled, and tight experimental settings are used to restrict users' movements and avoid muscular noise. Environmental and muscular noise are not the only artefacts; ongoing brain activities that are not related to the neurological phenomenon used in the BCI task also reduce the signal-to-noise ratio. To alleviate these challenges, spatial filters are commonly used to reconstruct the most informative sources and separate signal from artefacts. However, spatial filters are fitted to the training data and the artefact therein. They perform well as long as the conditions in which the training data

were recorded are kept. In reality however, variations in EEG structure are observed along a recording due both to internal and external factors. Internally, there are evidences of intra-subject variabilities due to the changing state of mind and fatigue in users. Externally, environmental noise cannot be controlled out of laboratories. In such conditions, it becomes crucial to have a feature representation and learning algorithms that are robust to changing conditions and artefacts.

Problem 3: Scarcity of Training Samples

The algorithms used in the machine learning pipeline (*i.e.* spatial filters and classifiers) require sufficient training data to achieve a sound statistical learning. The sample of EEG to be classified should be drawn from the same distribution as the training sample used for the optimisation of machine learning parameters. This is guaranteed by using training and testing samples recorded from a single subject in similar experimental conditions. The training sample size is proportional to the dimension of the EEG feature space which is usually high due to multichannel recording, high temporal and spectral resolutions of recorded EEG. In BCI, it is difficult to constitute such large training samples for all subjects, as it requires a rigorous and long recording of EEG trials. It is a burden for BCI users and it is not always possible to record a sufficient and well labelled training sample due to different reasons (e.g. fatigue, lack of concentration). For user convenience, such a process should be kept short, or better, not required at all. When the training sample is not large enough, statistical learning is not possible, constrained by the curse of dimensionality, or over-fitting will be inevitable.

1.3 Research Objectives and Contributions

1.3.1 Objectives

Considering the problems that will be addressed, the objective of this research is to propose ways of achieving a brain-computer interface that is adapted to the needs and environment of the user, through leverage of user's special skills and robust machine learning.

1.3.2 Research Contributions

This research contributes to the maturation of brain computer interfaces on two levels: BCI methodology and machine learning.

On the level of BCI methodology, a new BCI approach in the context of rehabilitation and assistive technology that takes into account users' specificities is proposed. It consists of a hybrid BCI system where cerebral commands are combined with muscular commands to achieve an adapted human machine interaction. The muscular interface is designed to fit user's residual motor abilities, while the BCI type is selected based on the user's experience. The concept is demonstrated for patients with degenerative diseases that affect large muscles but spare the wrists and hands motor capacities. For such patients, an adapted 3D touchless interface is used for continuous control and a BCI based on steady-state visually evoked potential (SSVEP) – i.e a synchronisation of the brain electrical wave at the frequency of an oscillating visual stimulus, is used for discrete control (e.g. triggering specific actions). While the touchless interface allows the subject to use their residual motor abilities, the SSVEP-based BCI with state-of-the-art signal processing and machine learning [Kalunga et al., 2013] is able to provide timely intervention for a better control in a multimodal setup. Experimentally, the concept is evaluated for navigation in a virtual environment and in the control of a robotic arm exoskeleton designed to compensate for muscular dystrophy in the shoulder and elbow muscles occurring in our subjects of interest [Kalunga et al., 2014].

On the machine learning aspect, after establishing the key role played by covariance matrices of multivariate time series in statistical learning, the study gives an evaluation of different covariance matrix estimation techniques in terms of quality of estimation and impact on the classification accuracy yield by the learning algorithm. Instead of going through estimates of covariance matrices to compute spatial filters, the current study proposes a new approach that operates directly on the space on covariance matrices (*i.e.* a *Riemannian manifold*) and classifies them based on their distances from class centres to achieve a learning that is less prone to overfitting and robust to environmental changes and noise. It demonstrates that in this framework, it is indeed important to use Riemannian metrics as they describe the geometry of covariance matrices better

than the Euclidean ones [Kalunga et al., 2015c]. Metrics that are invariant to affine transformations are used to measure the distance between covariance matrices.

An online implementation of the described approach is subsequently proposed for classification in SSVEP based BCI. The algorithm is capable of identifying epochs where the user is focusing on SSVEP stimulus from epochs where the user is not, and eventually classify SSVEP epochs with state-of-the-art accuracy [Kalunga et al., 2016].

Finally the last part of the work presented in this PhD contributes to alleviating the problem of insufficient training sample in machine learning for BCI. It contributes with a data augmentation technique where, given a small training sample, tools from Riemannian geometry are used to generate artificial data within the convex hull of the original sample, thus enlarging the training sample [Kalunga et al., 2015a]. It also explores possibilities of transfer learning on covariance matrices such that training samples from previous BCI users are used to train a classifier for a new BCI user.

1.4 Thesis Outline

The rest of this thesis is organised as follows: Chapter 2 presents the advances in brain computer interfaces through a review of literature. It particularly discusses the state-of-the art in neuroimaging, describing the techniques used for brain signal measurement. The chapter also presents the main neurological phenomena captured in brain signals for BCI purposes.

In Chapter 3, signal processing methods as well as machine learning approaches that have been commonly used in various BCI types are presented. The newly introduced Riemannian approach to machine learning is presented in section 3.2. In section 3.3, new trends in BCI applications are presented. Major advances in BCI being laid down, the BCI approach proposed in this thesis is presented in section 3.4. Key choices and positions taken along the research are explained.

Chapter 4 presents the first contribution of this PhD, *i.e.* the hybrid BCI, its motivation and design. It presents the methods and techniques used in its multiple modalities. The motor modality is described in section 4.3, and the BCI

modality in section 4.4. Full description of experimental protocol for recording of EEG data used in subsequent chapters are given here. The experimental results are presented in section 4.6.

Chapter 5 presents the second contribution, the Riemannian framework used for EEG representation and learning. It analyses methods of covariance matrix estimation in section 5.2. The Riemannian classification framework is presented in sections 5.3 and 5.4. An experimental validation of the proposed approach is given in section 5.5.

Chapter 6 discusses perspectives of Riemannian approaches in BCI machine learning. It presents a data augmentation technique in section 6.2, and a transfer learning technique in section 6.3. They both use Riemannian tools to address the problem of data insufficiency in BCI. Chapter 7 concludes the work with a summary of contributions and future perspectives.

Chapter 2

Neurobiological Aspects of Brain-Computer Interfaces

Whenever you remove any fence, always pause long enough to ask yourself, ‘Why was it put there in the first place?’

— G.K. Chesterton

2.1 Introduction

Brain-computer interfaces also called brain-machine interfaces (BMI) are devices that translates measured brain activity into tangible actions, allowing humans and other animals to interact with the physical environment without using their muscular system. From the 1980s this technology has received growing attention. Researchers from various fields including neurology, neuroscience, computer science and electrical engineering have multiplied their effort to move brain-computer interfaces from proof of concept to working prototypes.

Ideas of reading into the human brain were steered up for the first time in 1929 when Hans Berger [Berger, 1929] published his work on the recoding of brain electrical activities, *electroencephalograms* (EEG). For decades that follow this breakthrough, EEG was used for the diagnosis of neurological diseases and the study of brain functions [Wolpaw et al., 2002; Daly and Wolpaw, 2008]. A further step was taken when EEG was explored for therapeutic possibilities. People could learn to intentionally control their EEG to limit frequency

of seizures in epilepsy, to treat hyperactivity and other disorders [Daly and Wolpaw, 2008].

Despite the ability of recording and analysing brain signal, no dive was taken into deciphering brain signals for interaction purposes. The idea of reading human thought from brain was contemplated more in fiction than in science. Relying on brain signals to interact would require detecting human intention from the recorded EEG that was not possible with the early understanding of EEG and its quality. It was impossible to recognise a brain activity induced by a specific intention from the vast electrical activity of neurons. Moreover a detection of intention would require a real-time analysis of EEG which was not foreseeable with the technology at hand.

The first attempt of using measured brain signals as carriers of information in man-computer communication or for the purpose of controlling external devices came in the 1970s, with the *Brain-Computer Interface project* [Vidal, 1973]. The project benefited from the advances made in EEG studies providing the evidence that beside the continuous ongoing activity, EEG waves contained time-locked disturbance in response to brief stimuli, and could also be altered by conscious decision [Donchin, 1969; Vidal, 1973]. With very limited computational power at the time, the project constituted a proof of concept that the authors believed would be achievable in the future given considerable advances in neurophysiology, in signal analysis techniques, and in computer science.

After the *Brain-Computer Interface project*, there were four factors that triggered advances toward brain-computer interface [Wolpaw et al., 2002]: the first factor is the advances made in neurophysiology particularly the progress in EEG measurement techniques, the understanding of how EEG was affected by conscious as well as unconscious experience, and better understanding of brain functions. The second factor is the development in computing technology and computational power allowing complex and online treatment of EEG. The third factor is the increasing social need of assisting people with severe motor disabilities especially locked-in people who could not use tradition assistive devices that rely on muscular functions. The final factor is the finding that EEG could be used to affect activity-dependent plasticity and contribute to the recovery of motor functions.

Indeed, BCI can be used as a replacement for deficient muscular functions in people with severe motor disabilities who cannot use conventional assistive methods that depend on voluntary muscle control [Wolpaw et al., 2002]. The targeted population include people suffering from neuromuscular disorders such as ALS, severe cerebral palsy, brainstem strokes, severe muscular dystrophy or peripheral neuropathy, and other acute disorders causing extensive paralysis [Daly and Wolpaw, 2008]. BCI can also be useful for the rehabilitation of functions that have been lost after accidents that damage the nervous system [Silvoni et al., 2011].

Brain-computer interfaces translate measured brain signals into tangible actions for a specific application. The functionality of such a system requires at least 3 components: a signal acquisition component that measures brain activity, a signal processing component that decipher the measured signal, and an application interface where the deciphered brain activity is used as command.

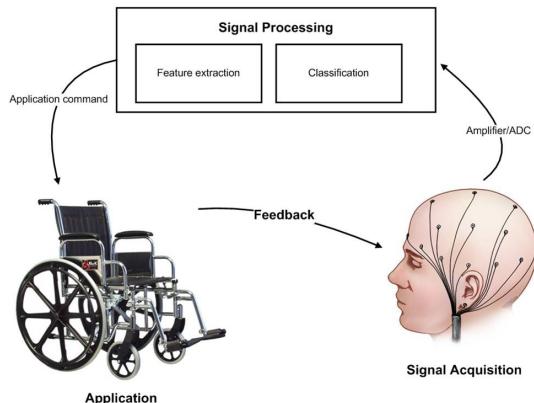


Figure 2.1: A standard BCI system with signal acquisition, signal processing and application components. The system provides feedback to the user.

In this chapter, signals used in BCI are presented, along with their measurement techniques, as well as their underlying neurological phenomena.

2.2 Signal Acquisition

To measure brain activity, BCI relies on brain imaging techniques used in neurophysiology. The existing methods for brain imaging used in BCI can be grouped

in three: electric signals, magnetic signals and hemodynamic signals. Electric and magnetic signals are two sides of the same coin and can be grouped under the term electromagnetic signals.

2.2.1 Local Field Potentials

The brain is made of billions¹ of interacting neurons constituting a neural network. A neuron is made of three major parts: a cell body, an axon, and dendrites [Purves, 2008]. Each neuron can be connected up to thousands other neurons. The connection between neurons is made at a junction called *synapse*. These junctions are often between an axon of one neuron and dendrites of the next neuron, and are referred to as axon-dendrite synaptic junctions (various other connections exist, e.g. axon-axon, dendrite-dendrite, dendrite-axon). The *presynaptic neuron* is passing information to the *postsynaptic neuron* [Herculano-Houzel, 2009; Purves et al., 2001].

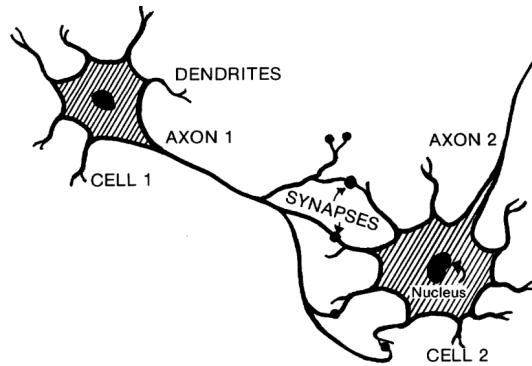


Figure 2.2: Neuron structure: showing main components of a neuron and its axon-dendrite synaptic connection to a neighbouring neuron [Purves et al., 2001]

The information sent between two neurons is mediated by a transient modification of voltage potential called action potential or *spike*. An activated neuron fires an action potential that is sent through its synapses to its postsynaptic partners. The excitatory and inhibitory postsynaptic potentials (EPSPs and IPSPs) cause a flow of charged ions between points at different potentials within and outside the neurons producing an electrical current, called *Local Field Potential* (LFP). Inside the neuron, positive ions propagate from the subsynaptic

¹The human brain contains about 85 billion neurons [Herculano-Houzel, 2012].

area to the rest of the neuron. Outside the neuron, the ions that have entered the cell are replaced by an ion flow directed toward the synaptic region along the extracellular space.

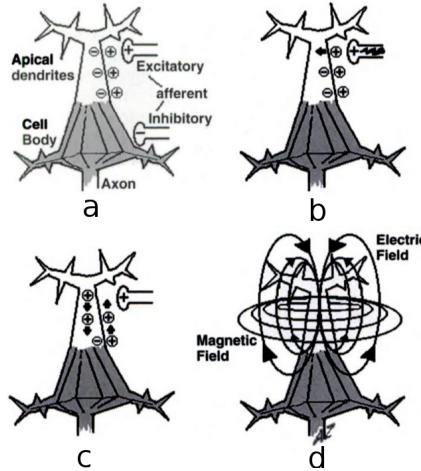


Figure 2.3: Electromagnetic activity in pyramidal neuron: (a) EPSPs converge to apical dendrites of the neuron. (b-c) Positive ions enter the neuron and propagate from synapses to the rest of the neuron. (d) the flow of current perpendicular to the apical dendrite is accompanied by a magnetic field that propagates orthogonally. [Proverbio and Zani, 2003]

2.2.2 Electrical Signal Acquisition

There are nowadays various techniques used to measure neuron electrical activities. They can be measured on the scalp, on the cortical surface, or within the cortex.

Electroencephalography is the measurement of the brain electric activity on the scalp. It is the most used measurement technique in brain-computer interfaces, and is at the origin of the expansion of BCI technology. Electrodes are spread over the scalp to cover all regions of the cortex.

The changes in electrical potentials – *electroencephalogram* (EEG), recorded at an electrode are the sum of electric field of neurons that are perpendicular to the scalp beneath the electrode. The recorded EEG is in the order of microvolts. EEG is recorded at an acquisition rate that can go as high as 1000 samples per second, providing a good time resolution. However, its spatial resolution is very

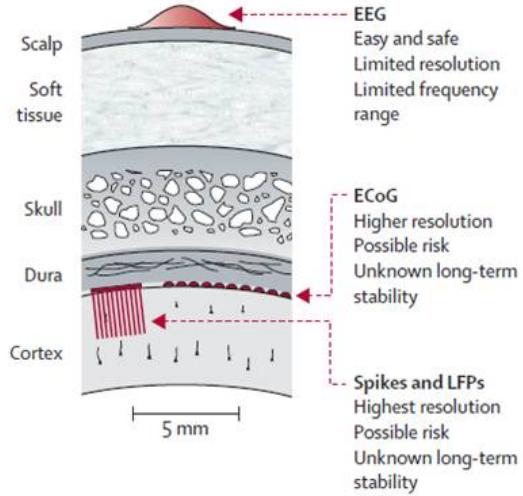


Figure 2.4: A cut through cortical layers. Electric activities can be measured from different layers. [Reproduced from [Daly and Wolpaw, 2008](#)]

low. It contains no depth information about the source. Moreover due to the dipole-like propagation of the electric potential of the source, the maximum of the distribution does not coincide with the source localisation [[Proverbio and Zani, 2003](#)]. The volume conduction affects the potential field as different biological layers do not have the same electrical properties and are inhomogeneous. The spatial resolution of EEG is affected by this, as the measured electrical field travels through different layers of the skull. EEG can measure brain electrical activities in spectral bands from 0 to 100 Hz. In BCI, it usually measures activities of up to 40 Hz, *i.e.* lower gamma band [[Schalk and Leuthardt, 2011](#)]. Activities in the upper gamma band, *i.e.* from 35 Hz to 100 Hz, have been measured mostly in emotion analysis [[Li and Lu, 2009](#); [Müller et al., 1999](#)].

Electrocorticography measures the same activity as EEG, but the electrodes that measure *electrocorticogram* (ECoG) are placed directly on the exposed surface of the cortex. For this reason it is also referred to as *intracranial electroencephalography* (iEEG). ECoG were recorded in humans and animals since the late 19th century [[Caton, 1875](#)]. Since then, ECoG has been used more in animals due to the fact that the placement of electrodes requires a skull surgery. The study of ECoG in humans is mostly done in epileptic subjects who await surgery. ECoG electrodes are temporarily placed to monitor

epileptic seizures and locate their focus zone [Ritaccio et al., 2012]. It is only very recently that ECoG has been considered for BCI [Huggins et al., 1999; Pfurtscheller et al., 2003]. The first BCI using ECoG in humans was done by Leuthardt et al. [2004]. Most BCI research is done on epilepsy patients and should coincide with the time ECoG electrodes are implanted for surgical purposes. This limit the number of ECoG-based BCI. There are few rare cases where ECoG electrodes have been implanted exclusively for research purposes [Wang et al., 2013; Sutter, 1992].

ECoG electrodes are usually in the form of electrodes array on a grid (Figure 2.5(b)) placed above (epidural) or below (subdural) the dura mater, *i.e.* the tough layer between the skull and the cortex [Schalk and Leuthardt, 2011]. The location of electrodes array is determined by the clinical need in epilepsy patients [Bundy et al., 2016]. For patients who have the ECoG electrodes implanted exclusively for research purposes, a fMRI is done prior to the placement to determine the cortical zone of interest for the BCI task [Wang et al., 2013].

ECoG is recorded at rates higher than 1 kHz, giving it a very high time resolution. The fact that the electrodes are placed directly on the surface of the cortex gives ECoG higher spatial resolution (*i.e.* 1.25 mm for subdural recording and 1.4 mm for epidural recording [Schalk and Leuthardt, 2011]), wider frequency bandwidth (0 to 500 Hz), and higher amplitude (*i.e.* 50 to 100 μ V) than EEG [Schalk and Leuthardt, 2011; Leuthardt et al., 2004; Spüler et al., 2014]. With a wider bandwidth, ECoG can capture neural electrical activity in the γ -band – which ranges from 30 Hz up, with higher precision than EEG. ECoG BCI research has shown that activity in the γ -band provide deeper information about movement and movement imagery such as direction and velocity both 2-dimensional and 3-dimensional [Bundy et al., 2016; Leuthardt et al., 2004].

Spikes and Local Field Potentials are intracortical measures of neural activities. The purpose is to measure the activity of a single neuron via its spikes, or the sum of activities of a small population of neurons local to a region – the local field potentials. Recording of neuronal spikes done approximately 50 years ago has shown that movement intent modulates spike timing from neurons of the motor cortex.

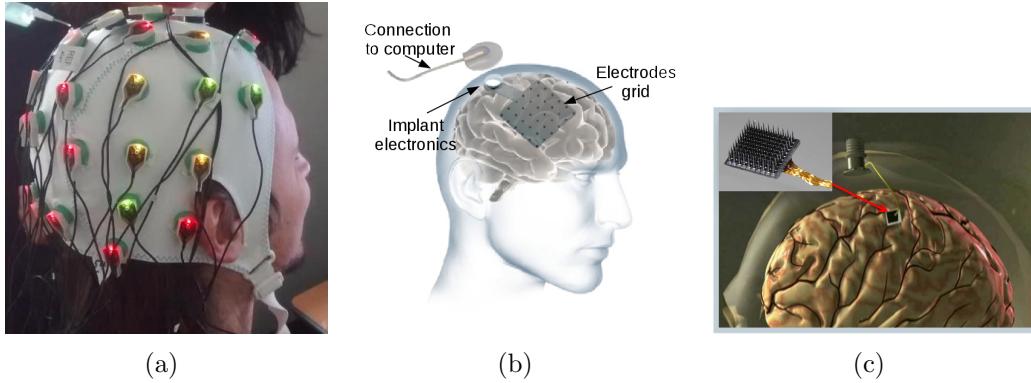


Figure 2.5: electric signals electrodes: (a) EEG electrodes cap being fitted on a subject’s head for recording. (b) ECoG Electrodes. Courtesy of Ripple, Inc. (c) A silicon-based cortical MEA (inset); implanted for intracortical neural recording via a percutaneous connection to a skull mounted pedestal connector [Homer et al., 2013].

The first clinical application using intracortical recording – also called stereoelectroencephalography, was achieved in 1998 [Kennedy and Bakay, 1998]. The electrodes are placed in the motor cortex. A fMRI or MRI is performed before the implant to locate the precise location. From the year 2000, substantial research has been done on these recording techniques, mostly on non-human primates [Serruya et al., 2002; Taylor et al., 2002; Musallam et al., 2004; Santhanam et al., 2006; Golub et al., 2014]. It has shown that it is possible to control devices such as prosthetics and achieve complex tasks such as 3D control, and reaching and grasping action with prosthetic devices by using motor cortex spiking pattern [Homer et al., 2013]. In humans, intracortical BCI has been studied on patients with spinal cord injuries [Hochberg et al., 2006, 2012], amyotrophic lateral sclerosis [Kennedy and Bakay, 1998], brainstem stroke [Kennedy et al., 2000], and mitochondrial myopathy [Kennedy et al., 2004]. By recording activities of individual neurons, recorded signals are not affected by activities happening in different brain regions. With electrodes implanted in different regions of the cortex, intracortical BCI could take parallel commands and offer many degrees of freedom.

2.2.3 Magnetic Signal Acquisition

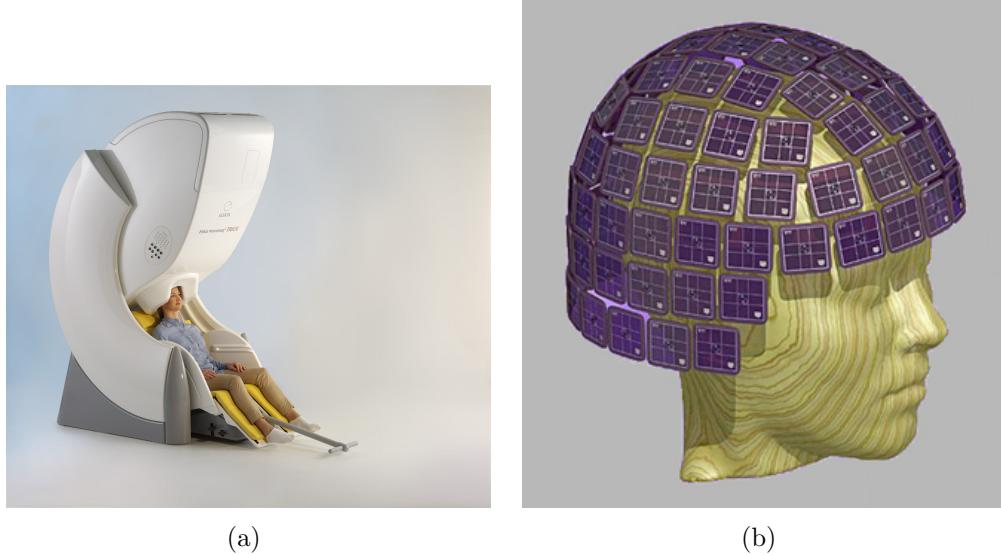
Electric and magnetic signals are two side of the same coin. They are both created by the same synaptic exchange between neurons 2.2.1. The magnetic effect of electric currents in neurons generates a magnetic field that propagates orthogonally to the flow of current [Gazzaniga et al., 2013; Proverbio and Zani, 2003].

Magnetoencephalography (MEG) is the measure of neurons magnetic field on the scalp. It is a neuroimaging technique used in neuroscience and clinical applications. It is very related to EEG as both are measured on the scalp. Both magnetic and electric fields propagate through different cranial layers before being measured on the scalp. Nonetheless MEG has an advantage over EEG; the magnetic field is not as influenced by the medium as is the electrical field. A drawback in measuring the magnetic activity of brain is that it is 8 orders of magnitude below the earth's magnetic field (in the order of 10^{-15} Tesla). Due to this, it cannot be measured in "open air". Electromagnetic isolation chambers are needed, making the MEG acquisition equipment bulky and expensive. MEG sensors are usually made of a magnetometer and two orthogonal planar gradiometers. Ranging from 64 to more than 300, MEG sensors are immersed in liquid helium and attached on a concave bottom of a container, where they typically lie at a distance of 3 - 4 cm from the cortex. The weak extracranial magnetic fields are amplified and transformed into a voltage [Paetau, 2002].

MEG has similar temporal resolution to EEG, but has a higher spatial, and can better capture modulations in brain signals, thus improving control and information transfer rates in BCI. EEG and MEG can also be co-recorded in BCI tasks and used in to improve BCI performances [Mellinger et al., 2007; Henson et al., 2011; Foldes et al., 2015].

2.2.4 Hemodynamic Techniques

While EEG, ECoG, LFP, spikes, and MEG measure the direct electromagnetic activities of neurons, there are other neuroimaging techniques that measure the metabolic effect of neurons electrical activities. In fact there is a relationship, *i.e. neurovascular coupling*, between neuronal activity and subsequent regional



(a)

(b)

Figure 2.6: Elekta MEG acquisition system. (a) A MEG shield chamber for electromagnetic isolation (b) MEG sensors configuration. Each sensor location is equipped with three sensors: a magnetometer that measures normal field component, and two orthogonal planar gradiometers that measure gradient components [Team, 2016].

blood volume and flow. This coupling is explained by the fact that firing neurons involved in a neurological task requires more energy and oxygen, resulting in an increase of blood flow and oxygenation. The active neurons in the region do not use the totality of the provided oxygen. This results in a change in the ratio between oxygenated (oxyHb) and deoxygenated (deoxyHb) hemoglobin. This metabolic response to neuron activities is called the *hemodynamic response* and can be measured using different techniques such as Magnetic Resonance Imaging (MRI), Functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), Near Infrared Spectroscopy (NIRS), and Functional Near Infrared Spectroscopy (fNIRS). fMRI, NIRS, and fNIRS have a possibility of real time recording required for brain-computer interfaces.

fMRI or Blood-oxygen-level-dependent (BOLD) fMRI uses magnetic resonance to measure the concentration of oxyHb and deoxyHb making use of the difference in their magnetic properties [Matthews and Jezzard, 2004; Gosseries et al., 2008; Huettel et al., 2004; Sitaram et al., 2008]. fMRI consist of multiple

scans of MRI to capture brain activity. A transmitter coil covering the head is needed to generate a magnetic field responsible for the resonance and relaxation in oxyHb and deoxyHb. fMRI have high spatial resolution (*i.e.* few millimetres) and low temporal resolution (*i.e.* few seconds) compared to electromagnetic brain signals. fMRI-BCI capitalises on the ability of fMRI to locate brain activity to the millimetre, to characterise different spatial distribution of brain functions as BCI commands [Sitaram et al., 2007; Yoo et al., 2004]. Though fMRI-BCI can achieve high classification accuracy, they are held back by the low temporal resolution that limit the speed of MRI scans and the information transfer rate of the interface. Furthermore, the size and setup of fMRI acquisition equipment limit the mobility of users.

NIRS and **fNIRS** are recent hemodynamic techniques introduced in the late 1980s. They measure the intensity of light propagated through brain tissues. Since the concentrations of oxyHb and deoxyHb in brain tissues are indicators of neural activity, f/NIRS use the relationship between transmitted light and the concentration of the medium (*i.e.* chromophores such as oxyHb and deoxyHb) to calculate these concentrations by shining near-infrared light on the head and measuring the intensity of the exiting light as shown in Figure 2.7.

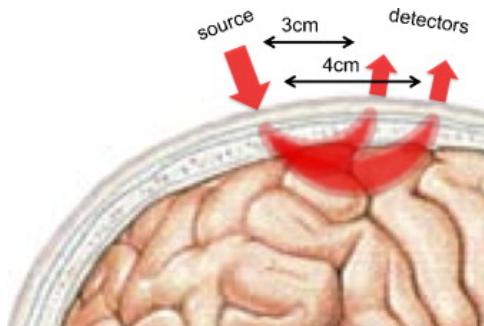


Figure 2.7: Trajectory of near-infrared light in the human brain. [Reproduced from Gervain et al., 2011].

f/NIRS has a good spatial resolution (few millimetres), better compared to electromagnetic signals measured on the scalp (*i.e.* EEG and MEG), but lower than fMRI. The acquisition equipment is lighter, easy to use, and enhances mobility of the subject. NIRS is tolerant to movement whereas other signals' recording techniques either does not allow user movement (*e.g.* fMRI, MEG)

or the signal quality is distorted by movement (e.g. EEG). As with fMRI-BCI, f/NIRS-BCI also relies on the ability to separate brain function based on their spatial distribution. Although the spatial resolution of f/NIRS is lower than the resolution of fMRI, f/NIRS-BCIs are easier to use than fMRI-BCIs. They have good classification accuracy, but are still slower than BCI that use electromagnetic signals. The BCI tasks are also limited to brain functions that can be measured close to the scalp. In fact f/NIRS cannot measure deep brain activity, because of light penetration which is limited to 15 mm and 5 mm into the cortex, for infants and adults respectively.

2.2.5 Discussion

Several neuroimaging techniques have the potential of being used in BCI. Each has characteristics that can be used to successfully classify brain functions. One requirement that all should meet to be considered as an input signal in BCI applications is the ability of near real time recording/scanning.

Neuroimaging techniques such as fMRI and MEG limit the mobility of the user and limit brain-computer interface to few applications. Moreover fMRI is not real time. Despite their good spatial resolution and temporal resolution (for MEG), they are not adapted for daily life interaction. f/NIRS is a good trade-off between user mobility and spatial resolution of the acquired signal. However, the fact that, as fMRI, it relies on hemodynamic response implies a low temporal resolution that might not be enough to capture transient brain responses to stimuli used in BCI, and exposes it to physiological noise such as cardiac cycle and respiratory effect that alter blood oxygenation more than other measurement techniques. In terms of user mobility and ability to capture fast brain response, electrical signals (EEG, ECoG and intracortical signals) remain so far the best options. It is proven that intracortical BCIs offer high interaction performances by decoding complex brain activities. However the risk and uncertainties surrounding the intracortical implantation of electrodes are still an issue for its recognition by the public. It is judged to be too invasive. The intracranial version of EEG – ECoG, alleviates many problems of EEG-based BCI, namely its vulnerability to noise (*i.e.* ocular, muscular and environmental). Although it is less invasive compared to intracortical measurement, ECoG still

requires surgery for electrodes placement.

Despite its relatively low spatial resolution and vulnerability to noise, EEG remains the sole technique that offers fast tracking of neural activities, affordable and light recording equipment allowing BCI users' mobility, safety and ease of use. It is vastly adopted as the input signal for BCI. However, some researchers believe that the future of BCI lies in invasive techniques. They argue that non-invasive techniques can only represent a limited number of brain responses – thus limited degrees of freedom, and that EEG weaknesses are hardly overcome. Another argument is that, for the same neurological phenomenon, non-invasive BCI requires longer training periods for the users to learn to produce a particular brain response voluntarily, and despite the training non-invasive BCI still have high error rates. A further argument is that although non-invasive medically, EEG measurement technique can also be seen as invasive in terms of human machine interaction: the gel, the tight electrode cap, the restriction to blink eyes during recording, etc. might be seen as invasive. These arguments have not stopped the EEG-based BCI community from pushing the limits. Wolpaw and McFarland disapproved the argument against non-invasive BCI by showing that with a comprehensive user training and good learning algorithms, EEG-based BCI could provide multidimensional point-to-point movement control that falls within the range of invasive BCI performances [Wolpaw and McFarland, 2004]. Research on improving EEG-based BCI performance has increased all the more, with better tools for the processing of EEG signals [Gramfort et al., 2014], and encouraging results [Mattout et al., 2013; Kalunga et al., 2016].

In conclusion, while other neuroimaging might be adequate for some BCI applications (e.g. neurofeedback), EEG constitutes a reasonable choice for signal input in BCIs for assistive purposes (e.g. communication and mobility). EEG-based BCIs are well tolerated with their limitations by patients with the need to communicate without their muscular systems [Kübler et al., 2005; Grbler et al., 2014]. With effort from different fields involved in BCI research, it is possible to reach better performance and tend toward those reported with invasive BCIs.

2.3 Neurological Phenomena

In deciphering brain signals, brain-computer interfaces identify a specific feature – a *neurological phenomenon*, from the signal that is associated with a given user’s intention. Neurological phenomena are variations in the brain signals associated with a cognitive activity (*i.e.* cognitive conscious information processing), or in response to a physical stimulus. Neurological phenomena induced by cognitive activities are said to be *endogenous*, while those triggered by external physical stimulus are said to be *exogenous*. Respectively, BCIs that rely on exogenous neurological phenomena are classified as exogenous or dependent BCIs as they dependent on an external stimulus, and their counterparts that rely on endogenous phenomena are classified as endogenous or independent BCIs as no external stimulus is needed. BCI research has mainly focused on the following phenomena: Event Related Desynchronisation (ERD) and Event Related Synchronisation (ERS), Event Related Potential (ERP), and visually Evoked Potential (VEP). There are discussed in the details in the next lines.

2.3.1 Event-Related Synchronisation-based BCI

Event-Related Desynchronisation and Synchronisation

Event related (de)synchronisation are either a decrease –*event-related desynchronisation* (ERD), or an increase – *event-related synchronisation* (ERS), of power in a given frequency band during a cognitive activity [Pfurtscheller, 1977; Pfurtscheller and Aranibar, 1977; Pfurtscheller and Neuper, 1994]; they are endogenous phenomena.

Pfurtscheller and Lopes da Silva [1999] interpret ERD as an electrophysiological correlate of activated cortical areas involved in processing of sensory or cognitive information or production of motor behaviour. ERD/ERS is mainly observed in the α rhythm, the μ rhythm also referred to as the upper α rhythm, the β rhythm and the γ rhythm. The α rhythm ranges from 8 to 12 Hz, the μ between 10 and 12 Hz, the β rhythm between 12 and 30 Hz, and the γ rhythm between 30 and 60 Hz. The low frequencies of oscillations in the brain signal are caused by synchronous neural activities that involve a large number of neurons. Hence slow oscillations are measurable in a large area of the brain. On the other

side, assemblies of only small numbers of neurons in synchrony oscillate at high frequencies [Singer, 1993]. The amplitudes of oscillations being proportional to the number of synchronous neurons, low frequencies have higher amplitude and high frequencies smaller ones. Therefore ERD/ERS in α -rhythm are more visible than in any other frequency bands.

Though easily measurable, lower α wave ERDs cannot be used to discriminate between tasks due to their wide topographical distribution. Moreover they might be obtained in response to any task. μ rhythm ERDs, however, are topographically restricted to some brain areas and happen only in response to specific activities. μ -rhythm ERD provoked by a given task will be observed mainly in the brain cortex in charge of the task. This specificity to tasks offers a possibility of discrimination amongst them [Pfurtscheller and Lopes da Silva, 1999].

Motor Imagery BCI Systems

The cognitive tasks used in current ERD/ERS-based BCI systems include motor imagery [Pfurtscheller and Neuper, 2001], mental tasks, e.g. sitting idle, doing a multiplication, composing a song [Kumar et al., 2010], composing letters, counting, rotating objects [Faradji et al., 2009]), or a combination of mental tasks and motor imagery tasks [Penny et al., 2000; Ozmen and Ktu, 2011].

A user performs a cognitive task while his brain signals are being recorded, for further processing and classification. The majority of studies conducted in ERD/ERS-based BCI are carried out on synchronous systems. Figure 2.8 illustrates a synchronous ERD/ERS-based BCI paradigm. The mental task is performed from the cue onset for a specific period of time. The trial starts with a beep. The user looks at the screen – where a fixation cross is displayed – waiting for the cue that will indicate the mental task to be performed.

Motor imagery provides the most intuitive and affordable cognitive task for the large population of users and has therefore dominated BCI research. It also has a record of best classification accuracy. The tasks that induces the most separable features in the EEG are the imagery of right-hand movement, the imagery of left-hand movement, the imagery of foot movement and the imagery

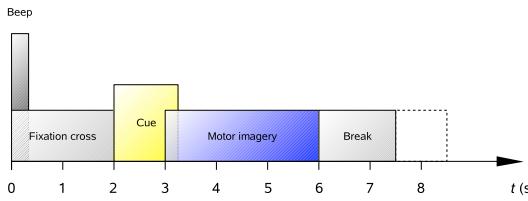


Figure 2.8: Standard ERD/ERS-based BCI system paradigm. The break before the next trial should last at least a second to allow the changes in the ongoing EEG/MEG to recover.

of tongue movement [Ang et al., 2012]. When a person is at rest (*i.e.* not involved any motor activity), there is a high activity in the 8-12 Hz band (*i.e.* μ rhythm) and the 18-26 Hz band (*i.e.* β rhythm) in the motor cortex. This activity is also known as *sensory motor rhythm* (SMR). It has been shown that for right hand movement there is a decrease, ERD, of SMR in the left hemisphere of the sensory-motor cortex, and the ERD occurs prior to the actual movement, during the preparation phase preceding the movement [Pfurtscheller and Lopes da Silva, 1999]. It has also been established that motor imagery (mental imagination of movements) activates similar brain areas (functions) to those activated during the preparation phase of actual movement [Jeannerod, 1995; Roland et al., 1980]. In general, voluntary hand movement results in bilateral ERD in the hand area and ERS in the foot area (see homunculus in 2.9); while a simple mental imagination of the same movement results in the contralateral β ERD and ipsilateral β ERS, both in the hand area [Pfurtscheller et al., 1997; Pfurtscheller and Neuper, 1994; Toro, C. and Deuschl, G and Thatcher, R and Sato, S. and Kufta, C and Hallett, M., 1994]. The fact that in mental imagination of one-sided hand movements the ERD remains mostly limited to the contralateral hemisphere is of key value in the classification of motor imagery-based BCI. Several studies [Lotte et al., 2007] have focused on the imagery of right hand and left hand movement – since these two tasks present the most discriminative characteristics because of their asymmetrical electrocortical responses – to build a 2-class BCI and have the best classification accuracy achieved in ERD/ERS-based BCI [Zhang et al., 2012]. It is to be mentioned that ERD elicited by motor imagery of different parts of an upper limb cannot be discriminated [Pfurtscheller and Lopes da Silva, 1999]. For instance, the imagery of left wrist movement and any left finger movement will activate the same

brain region (contralateral ERD and ipsilateral ERS in the hand region) as the one activated during the imagery of the left hand. For lower limbs, imagination of either foot movement results in a μ or β ERD in the foot area between both hemispheres such that it becomes impossible to discriminate between imagery of the left foot and of the right foot [Pfurtscheller and Lopes da Silva, 1999]. It is expected that the imagery of foot movement activates the foot area and the imagery of the tongue, the tongue area. However it is generally observed that the area activated by these two tasks are mixed up and not easily interpretable.

BCI systems must therefore use some complex algorithms to extract the most discriminative features and achieve a multiclass discrimination, e.g. 4-class: right hand, left hand, feet, and tongue [Dornhege et al., 2004b; Brunner et al., 2007].

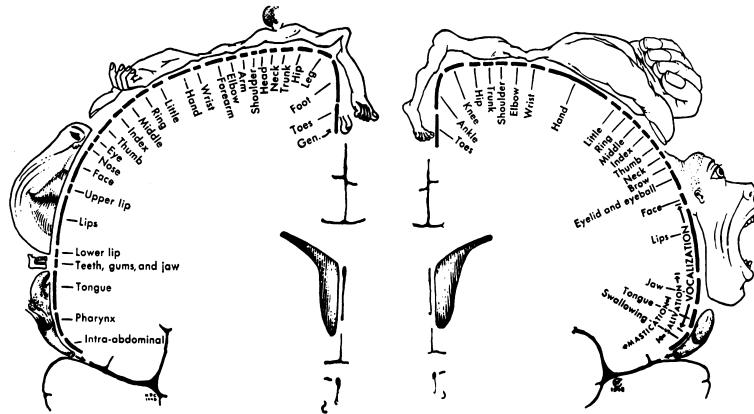


Figure 2.9: Sensory homunculus on the left and Motor homunculus on the right. The cortical homunculus initially developed by Dr. Wilder Penfield shows a disproportionate human body laid on the cortex from the prefrontal cortex (top) to the cerebellum (bottom). The size of a given body part of the homunculus is descriptive of the amount of cerebral tissue or cortex devoted to the specific body region which is proportional to how richly innervated that region is. This image is taken from [Schoot, 1993]

Challenges in ERD/ERS-based BCI systems

BCIs relying on ERD/ERS face a number of challenges: First is the low signal-to-noise ratio. ERD/ERS phenomena are submerged in a much larger brain

activity. It is thus difficult to distinguish a synchronisation or desynchronisation in a single trial.

Secondly, classification into classes representing different intentions is based on the topographical distribution (activated brain regions) and the ERD/ERS frequency range. However, as noted, EEG has poor topographical resolution and relatively narrow bandwidth. ERD/ERS in ECoG and other signals with better spatial resolution are reported to possess better signal characteristics for classification [Wilson et al., 2006; Schalk and Leuthardt, 2011; Power et al., 2012; Naseer and Hong, 2013; Wang et al., 2013]. Spatial filters are needed to alleviate the poor topographical resolution. The impact of spatial filtering on can be seen in [Hill et al., 2006].

Moreover, ERD and ERS areas are not always the same in different subjects due to physiological differences between them. The (de)synchronised frequency band is also not the same amongst subjects. Besides, the time where (de)synchronisation happens with reference to a cue is not the same either. This forces the BCI systems to identify the relevant brain area (e.g. spatial filter), the frequency band, and the time interval of significant (de)synchronisation [Yang et al., 2014]. The last task becomes even more complex in asynchronous BCI systems where there is no cue, therefore no reference (baseline). The term ERD implies that a baseline measured some seconds before the event represents a larger synchronisation [Pfurtscheller and Lopes da Silva, 1999].

A major problem in ERD/ERS-based BCI systems is the user training required. Users need to learn how to perform the cognitive tasks such that they can modulate their brain signals in a way that is detectable by the BCI system. Even after training, some users still cannot produce signals that are classifiable by the system. This phenomenon is known as *BCI illiteracy* and affects an estimate of 15 to 20% of BCI users [Allison and Neuper, 2010a]. Though the problem of BCI is not exclusive to ERD/ERS BCI, it is more prominent here [Hammer et al., 2012]. There are many attempts to explain the causes of BCI illiteracy and possible ways of alleviating the problem [Allison and Neuper, 2010a; Hammer et al., 2012; Jeunet et al., 2016].

2.3.2 Event-Related Potential

Event Related Potentials (ERP) are time-locked deflections in the EEG voltage (or electrical activity of a population of neurons) in response to a sensory stimulus. A commonly accepted hypothesis is that they are the result of a reorganisation of the phases or changes in specific frequency bands in the ongoing brain signals [Pfurtscheller and Lopes da Silva, 1999]. Having a very small amplitude compared to the ongoing brain activity, they are extracted through an averaging of aligned signal segments of repeated trials. After averaging, only the time-locked phenomena will remain, and all unrelated EEG will cancel out.

The resulting ERP consists of several positive and negative deflections called components of the ERP. They are designated with a ‘N’ (for negative components) or ‘P’ (for positive components) followed by a number indicating the time when they happened after the stimulus. Each component reflects a neural process involved in the response to the stimulus. The first components are usually sensory processes (*i.e.* P120 is the first positive component observed in response to a visual stimulus). They are then followed by more complex processes such as decision, recognition, and emotion related processes. N250 reflects the neural processing of a person’s own face, P300 reflects the processing of an odd event, N450 marks a processing of conflict, Error Related Potential is negative component observed after an error committed in a selection task [Luck, 2014].

Only the P300 [Polich, 2007; Donchin, 1981] and the error related potential (ErrP) [Miltner et al., 1997] components have been explored in BCI applications. ERPs are mere responses to sensory stimuli. A user would not have a voluntary control of the ERPs and cannot use them as input to a BCI. The *oddball paradigm* has allowed a “pseudo” voluntary control of P300 components, hence its usability in brain-computer interfaces [Ritter and Vaughan, 1969]. On the other hand, the ErrP has been used in BCI not as a control input, rather as a feedback channel. It allows the detection of errors (from the human and the machine) in human-machine interactions [Perrin et al., 2012].

P300

P300 is a positive deflection in the ERP, typically 300 ms after the perception of an odd event that creates a surprise effect for the subject [Donchin, 1981].

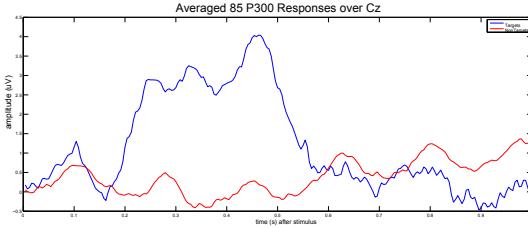


Figure 2.10: 2550 P300 trials have been averaged to obtain the enhanced P300 (blue line). The enhanced P300 is compared to 12750 trials not containing the P300. The data used are subject A's recorded signals from the BCI competition III data set II. A visual oddball paradigm as described in [Donchin et al., 2000] is used to elicit the P300.

Contrary to the intuition that P300 might be an exogenous phenomenon, Sutton et al. established that it was endogenous. The subject must have perceived the event, analysed it and established its oddity for a P300 to be elicited. It is related to the psychological reaction of the subject to the stimulus rather than to the physical characteristics of this stimulus [Sutton et al., 1965, 1967]. P300 amplitude is proportional to the temporal probability of the stimulus (e.g. sequential probability) which can roughly be defined as (1/total number of stimuli). It is also, to a lesser extent, related to the stimulus probability: (stimulus time/total trial time) [Fitzgerald and Picton, 1967].

P300 has a latency that varies with the difficulty of discriminating the improbable stimulus from the standard ones. The 300 ms latency is typical in young adults. Older subjects and those with decreased cognitive analogies have smaller P300 with a longer latency. Subjects with a greater ability to solve simple problems will generally have shorter latency. Within the adult population, the latency of P300 increases with age. Three positive waves overlap during the P300 latency: the P3a near 250 ms, the P3b near 350 ms, and a positive slow wave.

P300-based BCI systems

In BCI, the *oddball* paradigm is used in a scenario where the subject has a “pseudo” voluntary control of P300 generation [Ritter and Vaughan, 1969]. In this paradigm the subject is presented with a sequence of events that can be classified into two categories, this is the traditional two-stimulus oddball. A three-stimulus variation of the oddball paradigm can also be used [Polich, 2007].

In the traditional two-stimulus oddball, events in one of the two categories are rarely presented, thus eliciting a P300.

Auditory and visual stimuli are used to elicit the P300 with only a few studies focusing on auditory stimuli [Elshout, 2009]. Subjects in the complete locked-in state lose all voluntary control and cannot use visual stimuli. For such subjects, auditory P300-based BCI could be of great importance. Different sounds are played (e.g. notes, words) and for a given task the subject is asked to focus on a particular sound. When that sound is played a P300 is elicited around 300 ms later. Despite the opportunity they represent for people in a complete locked-in state, auditory P300-based BCI have low information transfer rate and have been explored by only a few studies [Sellers et al., 2006; Elshout, 2009; Käthner et al., 2013; Kaufmann et al., 2013b].

The most popular application of P300-based BCI systems is the P300 speller [Farwell and Donchin, 1988]. The subject is presented with a screen, containing a metric of characters. Rows and columns of the matrix are flashed one after the other in a randomised order. The selected character is at the intersection of the row and column which, when flashed, were followed by a P300. The flashes are repeated several times to enhance the detection of P300 through averaging. It was pioneered by Farwell and Donchin when for the first time they used the oddball paradigm and the flashing matrix to spell words conveyed to a voice synthesiser. They achieved a communication rate of 12 bits or 2.3 characters per minute [Farwell and Donchin, 1988]. Since then, several improvements have been made.

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	0

Figure 2.11: A P300-speller screen

A considerable amount of work has been devoted to improving the machine

learning algorithms for better detection of P300 [Hoffmann et al., 2005; Rakotomamonjy et al., 2005; Zhang et al., 2007; Krusienski et al., 2008; Rivet et al., 2009; Verschore et al., 2012; Lenhardt et al., 2008; Panicker et al., 2010]. They have significantly contributed to the development of P300-based BCI.

Other stimulation paradigms, different from the row-column flashing matrix have been proposed, and yield good performances: for example using *chequerboard paradigm* where individual characters are flashed randomly. In the chequerboard paradigm flashing objects or human faces improves ERP-based BCI [Hoffmann et al., 2008; Kaufmann et al., 2011, 2013b; Chen et al., 2015].

As stated earlier, the detection of P300 is done through averaging of repeated trials. This slows down the communication rate of the BCI. An important trend in P300 BCI is the detection of P300 in a single trial. This is being achieved by experimental paradigms and signal processing technique that enhance the evoked P300 in a single trial and with adequate machine learning algorithms [Bayliss and Ballard, 1998; Yin et al., 2013; Ishita et al., 2007; GüclüTürk et al., 2010; Kaufmann et al., 2013a]

Most of P300 BCI systems are synchronous; the timing is dictated by the stimulation system. Few implementations of asynchronous P300 BCI have been made. It is an effort to discriminate between control state (*i.e.* P300 being elicited) and rest state (*i.e.* the subject does not aim at any target), and to dynamically determine the number of trials needed for P300 detection [Lenhardt et al., 2008; Zhang et al., 2008; Schettini et al., 2014].

Visual P300 stimulation presented thus far requires a gaze control from the subject, which is not achievable by locked-in patients. A new paradigm was therefore developed to allow the use of visual P300 BCI by locked-in patient. One or several characters are presented in a rapid sequence in the middle of a screen. In the stream of character, when the intended character is displayed (or magnified), a P300 should be elicited [Acqualagna et al., 2010; Treder et al., 2011; Aloise et al., 2012; Acqualagna and Blankertz, 2013]. Tactile P300 has also been investigated for locked-in patients [Kaufmann et al., 2013a]

Challenges in P300-based BCI systems

Amongst the limitations in P300-based BCI is the low information transfer rate, due to the repetition of trials required to enhance P300 extraction. Acceptable classification is obtained after more than 5 repetitions. In [Rakotomamonjy et al., 2005] for instance 15 repetitions are used and make a trial duration (*i.e.* character epoch) of 35.4 seconds. This makes P300-based BCIs very slow. Single-trial detection approaches are a solution to this problem, but are hardly achievable.

The amplitude of the P300 decreases with time. The first occurrences of the rare stimuli will elicit a larger P300 [Courchesne et al., 1975] than the later occurrences will. This decrease might be explained by the local versus global probability of the rare stimulus. While the local probability of the rare stimuli within an oddball paradigm trial is the same over the entire BCI experiment, their global probability increases, creating a sense of habituation. In a P300 speller paradigm, this is first observed at the character epoch level as the user gets used to the stimuli that are being repeated, and then over the sessions as the user becomes used to the nature of the rare stimulus. The amplitude of P300 will decrease with the habituation, thus deteriorating the BCI performance. A potential solution would be to consider the inter-trial variability of the P300 while training the classifier [Rakotomamonjy et al., 2005].

On the subject's side, although P300-based BCI systems do not require initial training, they require continuous attention from the user, who should pay close attention to stimuli and notice every time the rare stimulus occurs. Over the long run, this might be tiring for some subjects or just not achievable for those with attention disorders [Szurovi et al., 2010; Krusienski et al., 2008].

Error-Related Potential

The neural processing of incorrect response generate a negative going deflection (Ne) in the ERP. The Ne has been observed in experiences with multiple-choice selection with tasks. Once a person observes an erroneous response, the error-related potential is elicited [Gehring et al., 1993].

The error-related potential could thus improve the performance of non invasive BCI. The erroneous action can either be automatically corrected or simply undone, as proposed by [Perrin et al. \[2012\]](#). The erroneous command is automatically replaced by the second most probable output of a probabilistic classifier.

2.3.3 Visual Evoked Potential

A Visual Evoked Potential (VEP) is an electrophysiological potential in the primary visual cortex in response to a visual stimulus. In general, a VEP contains three components illustrated in figure 2.12: a negative deflection at around 75 ms from the stimulus referred to as N75, a positive deflection at 100 ms from the stimulus referred to as P100, and a second negative deflection 135 ms after the stimulus called N135.

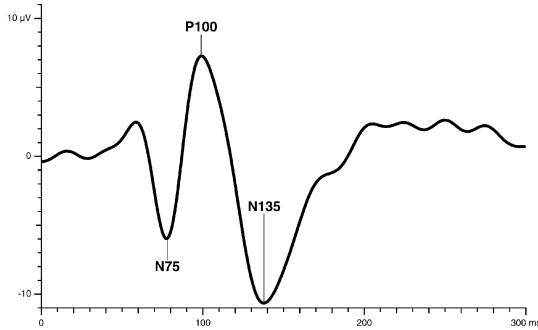


Figure 2.12: A Standard Visual Evoked Potential

VEP can be either transient or steady, *i.e.* *Steady-State Visually Evoked Potential* (SSVEP). Transient VEP can be defined as the response to an isolated or infrequent stimulus that provides enough time for the system to return to its initial state before onset of the next stimulus. The steady state response of SSVEP corresponds to a periodic succession of transient evoked potentials [[Capilla et al., 2011](#)]. Neuronal activity in the primary visual cortex is synchronised at the stimulation's fundamental frequency and its harmonics. This phenomenon is being increasingly used in brain-computer interfaces.

Figure 2.13 is very expressive with regards to the nature of SSVEP. While all three major voltage deflection (*i.e.* N75, P100, and N135) are observable

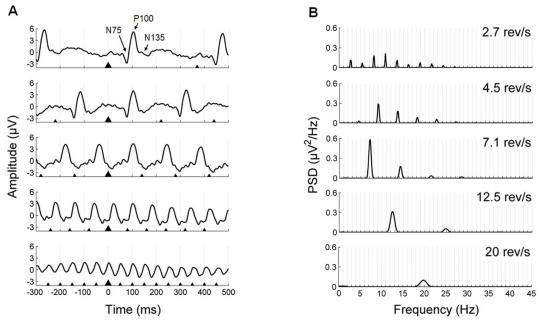


Figure 2.13: Visual evoked potentials at stimulation frequency of 2.7 Hz, 4.5 Hz, 7.1 Hz, 12.5 Hz, and 20 Hz. A illustrates the signal in time domain and B illustrates the frequency spectrum. [Reproduced from [Capilla et al., 2011](#)] .

in steady state responses at lower frequencies (e.g. 2.7 and 4.5 Hz) – making them very similar to transient evoked responses, they are less visible when the frequency of the stimuli train is increased. Only the P100 is still present at all stimulation frequencies. The amplitude of the steady state response appears to be attenuating as the stimulation frequency increases. This attenuation can be explained by latent inhibition, meaning that the transient excitation of the neural generators responding to the first stimulus in a sequence spreads to neurons that, in turn, feed back to them, attenuating the response to an incoming stimulus. High stimulation frequencies, with periods far shorter than the width of a P100, will suffer more from this inhibition.

At lower frequencies (≤ 7 Hz), the frequency component corresponding to the stimulation frequency is very weak. From 7 Hz up, this component is predominant. In both cases, harmonics of the stimuli frequency are visible especially for intermediate frequencies. The presence of harmonics is explained by the number of positive and negative voltage deflation found in a single VEP. At higher stimulation frequencies, the late deflation of preceding VEP cancel out (or overlap with) the early deflations of the ongoing VEP, leaving out a single VEP component per VEP. This explains the absence of harmonics in SSVEP from higher stimulation frequencies. With regards to amplitude of response, [Pastor \[2003\]](#) reached similar conclusions in their studies. They show that responses to low and high stimulation frequencies are less visible than responses to intermediate frequencies (see Figure 2.14). Another factor affecting the attenuation of responses to high simulation frequencies is the low pass filtering

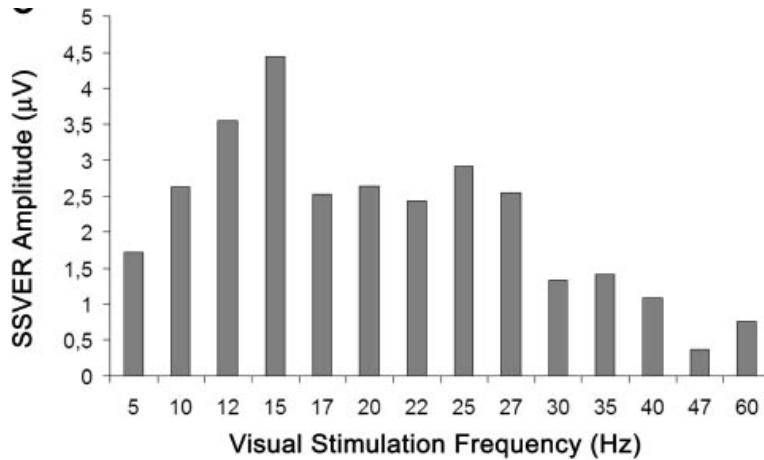


Figure 2.14: Average of the mean values of the amplitude of the FFT fundamental frequency of the SSVEP recorded at the three occipital leads (Oz, O1, O2) at the different stimulation frequencies. The amplitude of the occipital SSVEP, expressed in microvolts, reached a maximum at 15 Hz and then fell, with a plateau up to 27 Hz, declining at higher frequencies. [Reproduced from [Pastor, 2003](#)].

characteristics of the skull [[Nunez and Srinivasan, 2005](#); [Bédard et al., 2006](#)].

SSVEP-based BCI systems

There are various techniques to design stimulus for SSVEP in BCI. They are reported in [[Zhu et al., 2010](#)]. Different stimulation frequencies are used to build multiple BCI commands.

As in other BCI systems, offline applications of SSVEP-based BCI are used to investigate the parameters influencing the performance of the system. SSVEP-based BCI, especially synchronous systems, have the advantage of focusing on EEG activity that occurs at known frequencies. Making use of this feature, many studies have reduced the feature extraction methods to a simple frequency spectrum quantification, e.g. Fourier transforms-based methods. The target whose stimulation frequency has the largest amplitude in the frequency spectrum of the brain signal recorded in the occipital region is considered to be the one that the subject is gazing at [[Muller-Putz and Pfurtscheller, 2008](#); [Pfurtscheller et al., 2010](#)]. Due to inter-trial and inter-subject variability of the frequency spectrum features, parameters optimisation methods are introduced or classifiers such as *support vector machines* that can be trained and

used to classify the frequency spectrum features into classes [Kalunga et al., 2013]. Methods using canonical correlation analysis are very successful in the identification target’s stimulus frequency [Lin et al., 2006; Kalunga et al., 2013; Nakanishi et al., 2014].

Over the past years, interest in SSVEP-based BCI has increased due to the advantages it presents over other BCI systems. SSVEP have a higher signal-to-noise ratio, leading to higher classification accuracy, and a fast information transfer rate [Nakanishi et al., 2014]. Moreover, due to the fact that SSVEP is an inherent response of the brain, SSVEP-BCIs’ users do not need to go through intensive training.

It should be mentioned that the highest performances have been achieved in synchronous systems. Even in some asynchronous systems, the subjects are supposed to be continuously gazing at one target stimulus. This keeps the classification simpler as it avoids the complexity of discriminating between intentional control (IC) state and no-control (NC) state. To alleviate the complexity of having to discriminate continuously between NC and IC, some BCI systems activate the SSVEP target stimuli only when needed. Once the stimuli are activated, the system is invariably in the IC state, and when deactivated, it is in NC state [Cheng et al., 2002; Pfurtscheller et al., 2010].

SSVEP-based BCI is often employed as a dependent BCI [Wolpaw et al., 2000], that is, some residual muscular capabilities are required to move the eye toward the blinking stimulus as opposed to independent BCI, such as Motor Imagery (MI), where the communication does not rely on any motor capability. It has been shown that SSVEP could be used as an independent BCI [Morgan et al., 1996; Mller et al., 2006] as the brain oscillations are strongly related to the focus of attention. Using covert attention, *i.e.* shifting the focus of attention without moving the eyes, subjects can generate different SSVEP responses.

Visual stimulus plays a crucial role, affecting the BCI performance, and should be designed carefully. An in-depth review of the literature shows that LED stimuli provide better results than those obtained on computer screen [Zhu et al., 2010; Oralhan and Tokmaki, 2016]. A cognitive study indicates that any stimulation between 2 and 50 Hz induces visible oscillations in the visual cortex [Herrmann, 2001]. Common values employed in SSVEP studies are between 12

and 25 Hz, as they induce oscillations with higher amplitudes [Zhu et al., 2010]. One should note that safety of the subject should be taken into account as some frequency ranges of the stimulation could trigger epileptic seizure [Fisher et al., 2005].

The phase of the stimulation signal can also be modulated, enhancing the BCI performance by boosting the Information Transfer Rate (ITR) [Pan et al., 2011; Nakanishi et al., 2014]. An important constraint in that case is that experimental setup requires a synchronization between the display and the recording system, to ensure the correct estimation of the stimulus' phase. Better alternatives are available when considering systems with such constraints: code-modulated VEP (c-VEP) has yield the highest ITR in BCI [Spüler et al., 2012; Bin et al., 2011]. In c-VEP, the sole difference is that the stimulus flickering is based on pseudorandom sequences instead of the fixed frequencies of SSVEP.

Challenges in SSVEP-Based BCI Systems

Although SSVEP relies on the perception of the subject rather than eye movement, the majority of current SSVEP-based BCI paradigms requires eye movements for the perception of stimuli. To operate such systems, the subject must possess a functional visual system which should, moreover, entirely be devoted to the BCI application. Nonetheless, studies are investigating the possibility of an SSVEP-based BCI without the need of gazing [Lopez-Gordo et al., 2010].

This limitation is due to the fact that SSVEP can only be elicited within a limited frequency band. Also, due to the fact that harmonics of a stimulation frequency cannot be used in other target stimuli. Applications using computer monitors are faced with another limitation in usable frequencies due to the monitors refresh rates. The refresh rate must be a multiple of the stimulation frequency to avoid discrepancies in the generated frequency. Jia et al. [2011] proposed a stimuli coding method that combines frequency and phases. On a single frequency many stimuli can be coded using different phases, thus increasing the number of targets. c-VEP is an alternative to SSVEP that does not have this constraint [Spüler et al., 2012].

The implementation of asynchronous systems that can discriminate between IC and NC with minimal false positive still poses a challenge. This is crucial

for real life applications, yet studies investigating the matter and evaluating the performance of such systems are very few in number when compared to the attention SSVEP-based BCI have drawn in recent years.

2.3.4 Discussion

The presented neurological phenomena have all been used with considerable success in BCI. They each have advantages and drawbacks. The choice of a neurological phenomenon will depend on the specific needs in the BCI application. An adequate threshold should be found between the efficiency of the system and the comfort of the BCI user. Indeed the neurological phenomenon used in the BCI impact both the system's performance and the comfort of users.

In general, there is a duality or complementarity between endogenous and exogenous BCIs, the strengths found in one are usually the weaknesses found in the other. While exogenous BCIs suffer from the fact that they depend on the muscular functions and on an external stimulus, endogenous BCI are free from this dependency; except visual P300 that might need gaze control. Endogenous BCIs require that the user be trained, while exogenous BCI can be used with no training. Endogenous BCI have very low signal amplitude, while their counterpart enjoy a relatively stronger signal amplitude. Endogenous BCI are flexible; the user can shift between several mental tasks in a single BCI application. That is usually not possible with exogenous BCI.

Hence, BCIs that rely on exogenous BCI cannot be used by patients in a complete locked-in state. With no muscular function left, they still retain sensory and cognitive abilities that can be leveraged in endogenous BCI. There is however a vast population of patients who do retain gaze control, for whom visual techniques can still be used. Moreover, SSVEP and P300 are related to attention and perception rather than to gaze control. An appropriate BCI paradigm leverage this characteristic for their application in locked-in patients. The dependence to perception and attention also marks the difference between evoked potential-based BCI (*i.e.* P300 and SSVEP) and muscular devices such as eye-trackers, devices that rely on eye-fixation and saccades.

BCI illiteracy can be observed with any type of BCI. 15 to 20% of users cannot generate neurological responses necessary to control a particular BCI.

[Allison and Neuper \[2010b\]](#) discussed the possible causes of illiteracy and proposed few potential solutions that mainly consist of improving BCI accuracy in general. An interesting observation is that a subject who is illiterate in one BCI modality (e.g. SSVEP) might effectively use another modality (e.g. ERD). Different neurological modalities could be combined in a hybrid interface, where their impact is weighted depending on the user's abilities. There are also questions being raised about changing current approaches to BCI altogether. For example, recently [Jeunet et al. \[2016\]](#) challenged the standard training protocol used in MI-based BCI. They used the protocol followed in BCI tasks to train users on non-BCI task. They found that about the same rate 17% of users could not perform the learnt task, which is about the rate of BCI illiteracy. Their findings suggest that the training protocol used in BCI is not optimal and might by a influential factor in BCI illiteracy. These results will surely prompt more digging in approaches used in BCI.

Chapter 3

Signal Processing and Machine Learning for Brain-Computer Interfaces

It is not by trying to improve the candle that we invented electricity.

— Niels Bohr

3.1 Signal Processing for Cognitive Functions

Brain-computer interfaces translate brain signals into control or communication signal; the signal processing and machine learning component is therefore fundamental. The usual steps in the translation of brain signals consist of signal preprocessing, feature extraction, and finally a feature classification (or regression).

Signal preprocessing is fundamentally a cleaning up of data. The operations involved vary depending on the authors. They are usually generic operations such as epoching and removal of non-EEG signal added during the recording. [Bashashati et al. \[2007\]](#) report on techniques used to this end.

Feature extraction aims at identifying the characteristics in the EEG signals that bear relevant information for the classification task. Thus, it plays an important role in the design and choice of appropriate classifiers. The performance of BCI depends as much on the features used as on the classifier. Feature extraction techniques are a set of operations or transformations applied on the raw

EEG such that the neurological phenomenon induced by the BCI task is either enhanced or adequately represented. From the EEG signal they can extract:

- time features e.g. EEG signal amplitudes, signal power [Rivet et al., 2009],
- frequency features e.g. band power, power spectral density [Bhattacharyya et al., 2010],
- parametric features e.g. Auto Regressive (AR) and adaptive AR (ARR) parameters [Zhang et al., 2015a],
- time-frequency features e.g. wavelet transforms [Kumar et al., 2010; Dingyin et al., 2011; Zhang et al., 2015b], short-time Fourier transform [Kumar et al., 2010], empirical mode decomposition [Gaur et al., 2015; Wang et al., 2011; Liu et al., 2011],
- spatial features e.g. Independent Component Analysis (ICA) [Wang and James, 2006; Wang and Jung, 2013; Brunner et al., 2007], Common Spatial Patterns (CSP) [Ang et al., 2012; Barachant et al., 2010a; Blankertz et al., 2008], Canonical Correlation Analysis (CCA) [Nakanishi et al., 2014; Kalunga et al., 2013], Principal Component Analysis (PCA) [Kottaimalai et al., 2013; Yu et al., 2014], xDAWN [Rivet et al., 2009].

Detailed reviews of classic feature extraction methods can be found in [Lotte et al., 2007; Nicolas-Alonso et al., 2012; Bashashati et al., 2007; Khorshidtalab and Salami, 2011; Krusienski et al., 2011]. The choice of feature extraction techniques is guided by the neurological phenomenon used in the BCI. For instance, in SSVEP, relevant information will be contained in the spectral features while in P300, temporal features will suffice. Fukunaga [1990] defines feature extraction for classification as a search, among all possible singular transformations, for the best subspace which preserves class separability as much as possible in the lowest possible dimensional space.

Spatial filters (from which spatial features are extracted) have proven to be successful tools in extracting (or enhancing) the neurological phenomenon from the EEG signal. The information of interest is hidden in the recorded EEG, a mixture of simultaneous active brain sources and noise sources in the recording

environment. The signal of interest could be overlapped in time, space, and frequency with multiple signals. Using *a priori* knowledge on the phenomenon of interest or knowledge deduced from pre-recorded data, spatial filters extract better signal features that make subsequent processing tractable. Supervised and semi-supervised approaches lead to highest classification performances. Once a spatial filter has been applied to EEG data, any feature extraction technique can further be used merely for feature representation.

With relevant features, standard classification algorithms such as Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) [Rivet et al., 2009; Pfurtscheller et al., 2010; Spüler et al., 2012], Artificial Neural Networks [Haselsteiner and Pfurtscheller, 2000; Sturm et al., 2016], Quadratic Discriminant Analysis [Friedman, 1989; Bhattacharyya et al., 2010], Hidden Markov Model [Obermaier et al., 2001; Lee and Choi, 2003; Yan et al., 2008], and nearest neighbour [Cincotti et al., 2003; Schlg et al., 2005; Bhattacharyya et al., 2010] can be applied.

The signal processing pipeline can be represented as in Figure 3.1. Removing the preprocessing and the feature representation phases that do not require training, the machine learning pipeline of most successful BCI systems [Ang et al., 2012; Spüler et al., 2012; Rivet et al., 2009] consists mainly of spatial filtering and classification/regression blocks.



Figure 3.1: BCI signal processing pipeline. The fundamental blocks for machine learning are in blue (*i.e.* spatial filtering and classification). Filters and classifiers are learned from training data. Gray blocks (*i.e.* preprocessing and feature representation) do not require learning.

Spatial filters and classifiers are trained offline on pre-recorded data. The best preprocessing and feature representation can be chosen based on the offline analysis through cross validation.

Hereafter, classical couplings of spatial filters and classifiers are illustrated as they are used in machine learning of various BCI types *i.e.* motor imagery, P300, and SSVEP.

Filtering is based on a modelling of EEG as a linear mixture of multiple

neural ensembles to which noise is added:

$$X = AS + R \quad (3.1.1)$$

where $X \in \mathbb{R}^{C \times N}$ is the EEG recorded on C channels over N samples, $S \in \mathbb{R}^{K \times N}$ are the brain sources, $A \in \mathbb{R}^{C \times K}$ is a matrix defining the mixture of the brain sources, and $R \in \mathbb{R}^{C \times N}$ is the additive noise, with $C \leq K$. Spatial filters are designed to extract signal of interest (3.1.1) based on *a priori* knowledge on the neurological phenomenon of interest or knowledge deduced from pre-recorded data.

3.1.1 Motor Imagery Processing

Common Spatial Patterns are spatial filters that has been particularly designed for motor imagery tasks classification [Koles et al., 1990]. CSP extracts EEG spatial components that are common to two imagery tasks, but maximising the variance of the signal recorded during one task while minimising the variance in the other task. The distribution of filtered samples belonging to one imagery task has maximal variance, while the distribution of samples in the other task has minimal variance, or vice-versa as illustrated in Figure 3.2. CSP should capture the contralateral effect of ERS versus ERD induced by motor imagery tasks.

CSP has been successfully coupled with linear discriminant analysis classifiers [Dornhege et al., 2004a; Popescu et al., 2007]. Indeed like CSP, LDA also relies on data class scatter. LDA projects samples into a space where the within-class covariance matrices are minimised, while the between-class covariance matrix is maximised. Here the covariance matrices are proportional to the class scatter.

Common Spatial Patterns

CSP model is given by:

$$S = WX \quad (3.1.2)$$

CSP finds the filter $W \in \mathbb{R}^{C \times C}$ that minimises the variance of the filtered signal S in one condition and maximises it in the other. Neglecting additive

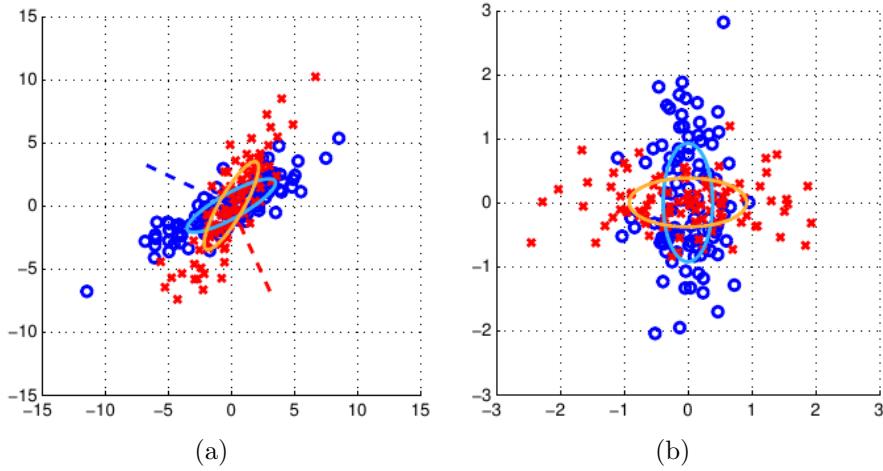


Figure 3.2: CSP effect on class distribution. CSP is applied on a 2D toy data set containing samples from two classes marked with red crosses and blue circles. (a) Samples distribution is shown before CSP filtering. The ellipses show estimates of each class covariance. It can be seen that the two classes are highly correlated. The dashed lines show the direction of the CSP projections \mathbf{w}_j ($j = 1, 2$). (b) Distributions after CSP projections. The two distributions are orthogonal, showing that the two classes are uncorrelated. Each axis gives the largest variance in one class and the smallest in the other. [Image from [Blankertz et al., 2008](#)].

noise in Equation (3.1.1), the CSP model is equivalent to finding the inverse of A : $W = A^{-1}$. In EEG modelling, A is called the *forward model* or the *mixing matrix* and W the *reverse model* or *de-mixing matrix*. A describes the spatial pattern.

Let $X_i \in \mathbb{R}^{C \times N}$ be a band-pass filtered signal of EEG recorded at epoch i . An estimate of its covariance matrix $\hat{\Sigma}_i \in \mathbb{R}^{C \times C}$ can be computed as:

$$\hat{\Sigma}_i = \frac{X_i X_i^T}{\text{trace}(X_i X_i^T)} \quad (3.1.3)$$

A class covariance matrix is obtained as:

$$\hat{\Sigma}_{(c)} = \frac{1}{N_c} \sum_{i=1}^{N_c} \hat{\Sigma}_i \quad (3.1.4)$$

In a two imagery task, $c \in \{+, -\}$, and N_c is the number of epochs in class c .

CSP solves the following problem:

$$\begin{aligned} & \underset{W}{\text{maximise}} \quad \text{trace}(W^T \hat{\Sigma}_{(+)} W) \\ & \text{subject to} \quad W^T (\hat{\Sigma}_{(+)} + \hat{\Sigma}_{(-)}) W = \mathbf{I}. \end{aligned} \quad (3.1.5)$$

The constraint in (3.1.5), where \mathbf{I} is the identity matrix, forces $W^T \hat{\Sigma}_{(-)} W$ to

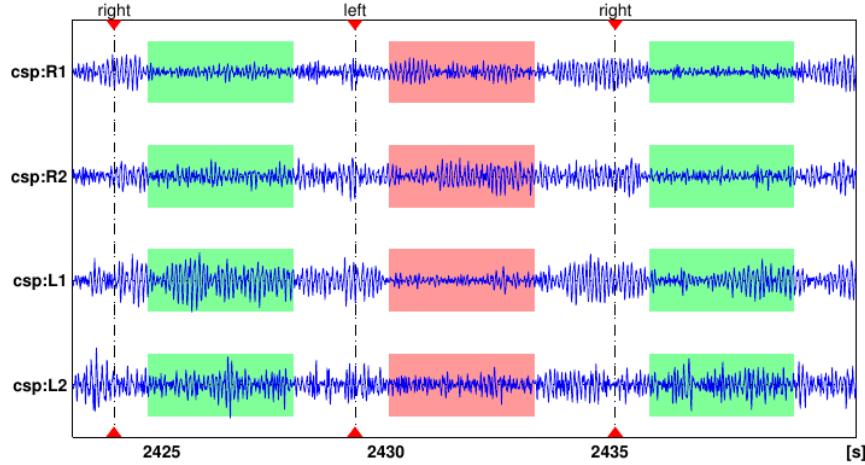


Figure 3.3: Effect of CSP filtering. A continuous EEG signal containing two right-hand imagery epochs and one left-hand imagery is filtered using four CSP filters (\mathbf{w}_j): csp:R1, csp:R2, csp:L1, and csp:L2. The resulting signals from csp:R1 and csp:R2 have large variance during left hand imagery, while signals from csp:L1 and csp:L2 have large variance during right hand imagery. [Image from [Blankertz et al., 2008](#)].

be minimal when $W^T \hat{\Sigma}_{(+)} W$ is maximal. There are many ways of solving this problem. A simple way is to solve it as a generalised eigenvalue problem [[Koles et al., 1990](#); [Blankertz et al., 2008](#)]:

$$\hat{\Sigma}_{(+)} \mathbf{w} = \lambda \hat{\Sigma}_{(-)} \mathbf{w}, \quad (3.1.6)$$

where \mathbf{w}_j ($j = 1, \dots, C$) are the generalised eigenvectors that constitute the vectors of the matrix W , and $\lambda_j = \lambda_j^{(+)} / \lambda_j^{(-)}$, with $\lambda_j^{(c)} = \mathbf{w}_j^T \hat{\Sigma}_{(c)} \mathbf{w}_j$ where $0 \leq \lambda_j^{(c)} \leq 1$ and $\lambda_j^{(+)} + \lambda_j^{(-)} = 1$. $\lambda_j^{(c)}$ is the variance in the filtered signal $\mathbf{s}_j = \mathbf{w}_j^T X$ in the epochs corresponding to class (c) . Thus filtering (*i.e.* projecting) the EEG signal X_i with \mathbf{w}_j that corresponds to the largest $\lambda_j^{(+)} \rightarrow 1$, will maximise the variance in class $(+)$ while minimising it in class $(-)$ as illustrated in Figures 3.3 and 3.2.

This discussion of CSP is limited to two-class motor imagery. The application of CSP has been extended to multi-class cases [Dornhege et al., 2004a; Grosse-Wentrup and Buss, 2008].

Linear Discriminant Analysis

Given a set of n C -dimensional samples x_1, x_2, \dots, x_n ($x_i \in \mathbb{R}^C$), LDA finds a lower dimensional space (typically $C-1$) where the data are the most separable. In a two-class case ($c \in \{+, -\}$), $n_{(+)}$ samples belong to the positive subset $\mathcal{X}^{(+)}$, and $n_{(-)}$ samples belong to the negative subset $\mathcal{X}^{(-)}$. LDA finds a projection vector \mathbf{w} that will achieve the mapping

$$y = \mathbf{w}^T x \quad (3.1.7)$$

that obtains the n $(C-1)$ -dimensional samples y_1, \dots, y_n ($y_i \in \mathbb{R}^{d-1}$), where the positive subset of the projected data $\mathcal{Y}^{(+)}$ is separable from the negative subset $\mathcal{Y}^{(-)}$. If the original C -dimensional samples are highly overlapped, not even the best \mathbf{w} could separate them in a lower dimension. A successful application of CSP will avoid such problems.

For separability brings the idea of distance between subsets, LDA uses the difference of projected samples means:

$$|\tilde{\mathbf{m}}_{(+)} - \tilde{\mathbf{m}}_{(-)}|$$

If $\mathbf{m}_{(c)}$ is the C -dimensional class mean,

$$\mathbf{m}_c = \frac{1}{n_{(c)}} \sum_{x \in \mathcal{X}^{(c)}} x, \quad (3.1.8)$$

$\tilde{\mathbf{m}}_{(c)}$ is the mean of projected samples belonging to $\mathcal{Y}^{(c)}$ and is given by

$$\begin{aligned} \tilde{\mathbf{m}}_{(c)} &= \frac{1}{n_{(c)}} \sum_{y \in \mathcal{Y}^{(c)}} y \\ &= \frac{1}{n_{(c)}} \sum_{y \in \mathcal{Y}^{(c)}} \mathbf{w}^T x = \mathbf{w}^T \mathbf{m}_c, \end{aligned} \quad (3.1.9)$$

Maximising the distance between class means does not ensure class separability. If both subset are very scattered, their samples can still be highly overlapped

even when their respective means are very distant. This introduces a second criterion in LDA: the distance between class means should be large relative to the *within-class scatter* given by:

$$\tilde{s}_{(c)}^2 = \sum_{y \in \mathcal{Y}^{(c)}} (y - \tilde{\mathbf{m}}_{(c)})^2 \quad (3.1.10)$$

The total within-class scatter is $\tilde{s}_{(+)}^2 + \tilde{s}_{(-)}^2$. LDA uses the function $\mathbf{w}^T x$ with the vector \mathbf{w} that maximises the criteria

$$J(\mathbf{w}) = \frac{|\tilde{\mathbf{m}}_{(+)} - \tilde{\mathbf{m}}_{(-)}|^2}{\tilde{s}_{(+)}^2 + \tilde{s}_{(-)}^2} \quad (3.1.11)$$

Figure 3.4 illustrates a projection of data verifying this criterion. To express

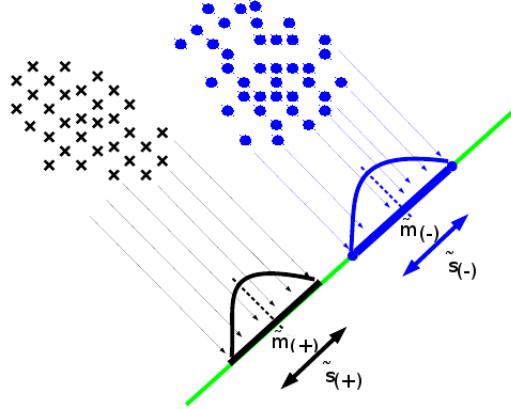


Figure 3.4: LDA mapping. The distance $|\tilde{\mathbf{m}}_{(+)} - \tilde{\mathbf{m}}_{(-)}|$ is maximised relative to class variance such as the separability is maximised ($J(\mathbf{w})$)

$J(\mathbf{w})$ in terms of \mathbf{w} and the d-dimensional observed samples, (3.1.9) and (3.1.10) are inserted into (3.1.11), which yields

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \Sigma_B \mathbf{w}}{\mathbf{w}^T \Sigma_W \mathbf{w}}, \quad (3.1.12)$$

where

$$\Sigma_B = (\mathbf{m}_{(+)} - \mathbf{m}_{(-)})(\mathbf{m}_{(+)} - \mathbf{m}_{(-)})^T \quad (3.1.13)$$

is the *between-class scatter matrix*, and

$$\Sigma_W = \sum_{x \in \mathcal{X}^{(c)}} (x - \mathbf{m}_{(c)})(x - \mathbf{m}_{(c)})^T \quad (3.1.14)$$

is the *within-class scatter matrix*.

Equation 3.1.12 is a generalised Rayleigh quotient and can be solved for \mathbf{w} as a generalised eigenvalue problem [Duda et al., 2001].

Quadratic Discriminant Analysis (QDA) is a generalisation of LDA to non-linear boundaries (*i.e.* conic section). Unlike LDA, QDA allows different class covariance in the projected space. QDA has also been used for classification of motor imagery tasks [Bhattacharyya et al., 2010]. Figure 3.5 illustrates the significance of QDA over LDA.

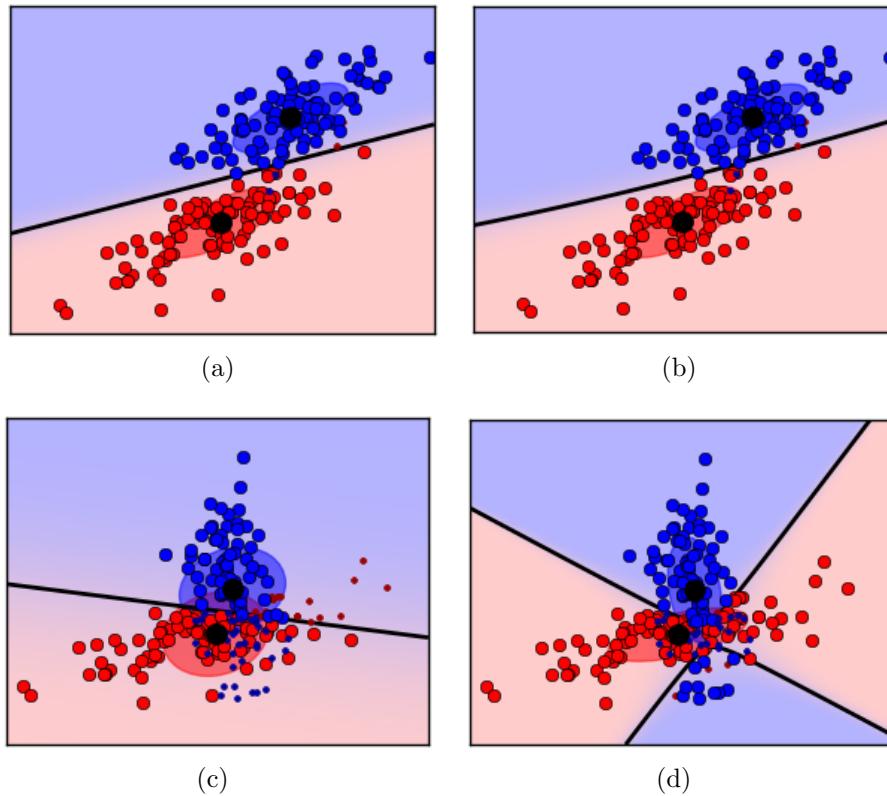


Figure 3.5: QDA versus LDA. Data from two subsets shows with red and blue circles are classified with either LDA (in (a) and (c)) or QDA (in (b) and (d)). The black lines show the classifier separation line. Wrongly classified data are shown with squares instead of circles. Class covariances are shown with ellipses of corresponding colours. In (a) and (b), the two subsets have similar covariance, while in (c) and (d) they have different covariance [Pedregosa et al., 2011]

3.1.2 SSVEP Processing

Canonical Correlation Analysis (CCA) was recently introduced to classification of SSVEP signal [Lin et al., 2006]. It has since been used in the most successful SSVEP-based BCI [Bin et al., 2009; Nakanishi et al., 2014]. It is known that flickering visual stimuli induced an SSVEP that is correlated to the stimuli, *i.e.* the phase and frequency of the signal of interest are known. Using this information about the signal of interest, CCA will extract the EEG spatial components that correlate the most with the SSVEP stimuli. When used as spatial filters, CCA works well when coupled with Support Vector Machine (SVM) classifiers [Spüler et al., 2012; Kalunga et al., 2013].

Canonical Correlation Analysis

Let $Y \in \mathbb{R}^{H \times N}$ be a multivariate signal representing the stimulation signal used in recording the EEG signal X . Per SSVEP principle, X is expected to be correlated to Y . Thus CCA will find two projection directions \mathbf{w}_X and \mathbf{w}_Y such that $\mathbf{w}_X^T X$ and $\mathbf{w}_Y^T Y$ have maximal correlation. \mathbf{w}_X and \mathbf{w}_Y maximises the correlation function $\rho(\mathbf{w}_X, \mathbf{w}_Y)$:

$$\begin{aligned} \rho(\mathbf{w}_X, \mathbf{w}_Y) &= \text{corr}(\mathbf{w}_X^T X, \mathbf{w}_Y^T Y) \\ &= \frac{\mathbf{w}_X^T \Sigma_{XY} \mathbf{w}_Y}{\sqrt{\mathbf{w}_X^T \Sigma_X \mathbf{w}_X \mathbf{w}_Y^T \Sigma_Y \mathbf{w}_Y}}, \end{aligned} \quad (3.1.15)$$

where Σ_{XY} is the between-set covariance matrix; Σ_X and Σ_Y are the within-set covariance matrices. CCA can be solved [as in Hardoon et al., 2004]:

$$\begin{aligned} &\underset{\mathbf{w}_X, \mathbf{w}_Y}{\text{maximise}} \quad \mathbf{w}_X^T \Sigma_{XY} \mathbf{w}_Y \\ &\text{subject to} \quad \mathbf{w}_X^T \Sigma_X \mathbf{w}_X = 1, \\ &\qquad \mathbf{w}_Y^T \Sigma_Y \mathbf{w}_Y = 1. \end{aligned}$$

A common way of generating the representation of the simulation signal at frequency f is:

$$Y_f = \begin{bmatrix} \sin(2\pi f) \\ \cos(2\pi fn) \\ \vdots \\ \sin(2\pi N_h fn) \\ \cos(2\pi N_h fn) \end{bmatrix}, n = \frac{1}{f_s}, \frac{2}{f_s}, \dots, \frac{N}{f_s} \quad (3.1.16)$$

Where f_s is the EEG sampling frequency, N_h is the number of harmonics, and N the number of sampling points. N_h is a parameter that can be defined by cross validation.

Support Vector Machine

Support Vector Machines (SVM) have been successfully used in classification of SSVEP signal, and in BCI in general. The binary SVM decision function is of the form:

$$y = f(x) = \text{sgn} \left(\sum_{i=1}^m y_i \alpha_i k(x, x_i) + b \right), y \in \{\pm 1\} \quad (3.1.17)$$

where x is the sample variable, x_i a sample in the training data with the label $y_i \in \{\pm 1\}$. m is the number of data samples in the training set, α_i the weight of sample x_i and b an offset. $k(\cdot, \cdot)$ is a kernel, *i.e.* a function that returns a real number characterising the similarity between its inputs. In a Euclidean space, the dot product would often be used as a linear kernel (3.1.21). Function (3.1.17) defines a hyperplane of decision boundary that separates samples in the negative class from samples in the positive class. SVM ignores the influence of training samples x_i that are very far away from the decision boundary by setting their corresponding weight α_i to zero. Thus, it only relies on a subset of data close to the decision boundary. They are called *Support Vectors*. This reduces model complexity and improves generalisation. Thought there could possibly exist many hyperplanes that accurately separate data into their specific classes, SVM finds the unique hyperplane that has maximum margin of

separation between the support vectors and itself. This fact gives SVM classifiers good generalisation performance and robustness to overfitting [Scholkopf and Smola, 2001; Ang et al., 2008, 2012]. To find such a hyperplane SVM solves the following problem:

$$\begin{aligned} \text{minimise}_{\mathbf{w} \in \mathcal{H}, b \in \mathbb{R}} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{subject to} \quad & y_i(\langle x_i, \mathbf{w} \rangle + b) \geq 1, i = 1, \dots, m. \end{aligned} \quad (3.1.18)$$

where \mathcal{H} is a dot product space, and m is the number of training samples.

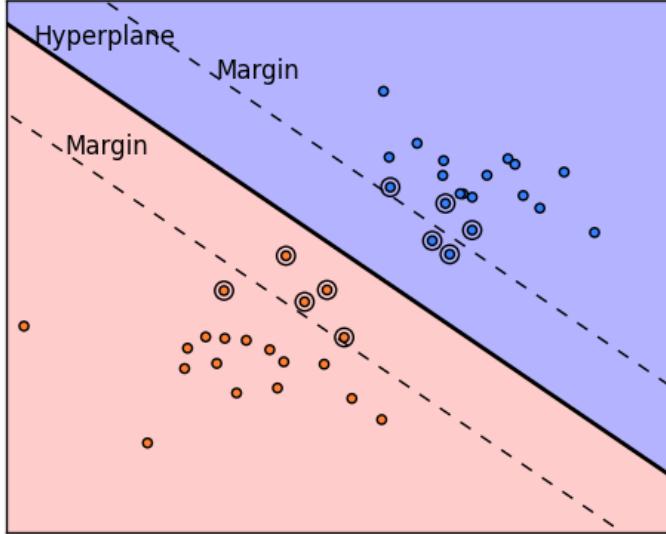


Figure 3.6: Examples of SVM classifiers. SVM is applied on 2D artificial data forming two classes represented in red and blue. The hyperplane $\langle x_i, \mathbf{w} \rangle + b = 0$ separating the two classes is shown, as well as the margins $\langle x_i, \mathbf{w} \rangle + b \geq \frac{1}{\|\mathbf{w}\|}$. Only few samples relatively close to the hyperplanes are used as support vectors; they are shown with big circles [Pedregosa et al., 2011].

To achieve nonlinear classification SVM uses kernels other than the dot product. Such kernels are needed in cases where (1) features are better separated in a higher-dimensional space, (2) features are defined in a space where the dot product is not defined, or (3) the separating line is not linear. Since SVM is

based on the dot product (Eq. (3.1.18)), the kernel used in the input space \mathcal{X} corresponds to dot product in \mathcal{H} via a mapping Φ ,

$$\begin{aligned} \Phi : \mathcal{X} &\rightarrow \mathcal{H} \\ x \rightarrow \mathbf{x} &:= \Phi(x), \end{aligned} \tag{3.1.19}$$

such that,

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle. \tag{3.1.20}$$

Examples of such kernels admitting the representation of the form (3.1.19) that are often used in SVM are the linear kernel (*i.e.* dot product), the polynomial kernel, and the radial basis function (RBF) kernel.

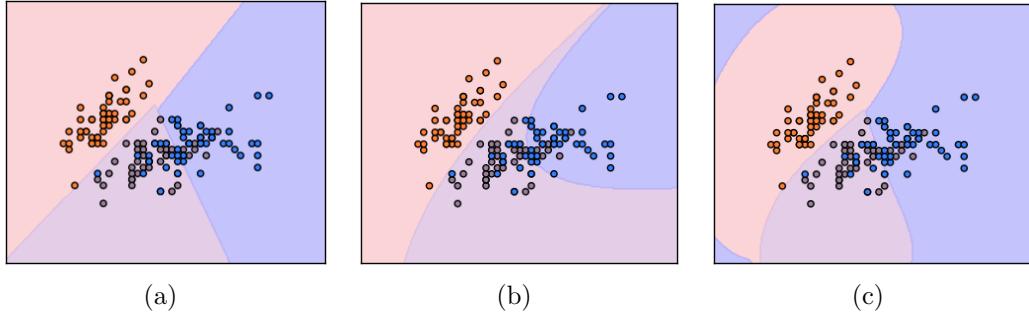


Figure 3.7: Multiclass SVM classification with different kernels on 2D projection of the iris dataset. The decision surface separating three classes are shown. The x-axis and y-axis represent sepal length and sepal width respectively. (a) Linear kernel, (b) polynomial kernel of order 3, (c) RBF kernel [Pedregosa et al., 2011].

Linear kernel:

$$k(x_1, x_2) = \langle x_1, x_2 \rangle = x_1^T x_2. \tag{3.1.21}$$

Polynomial kernels:

$$k(x, x') = \langle x, x' \rangle^d, \tag{3.1.22}$$

where d is the polynomial degree.

Radial basis function (RBF) kernels:

$$k(x, x') = \exp(-\|x - x'\|^2), \tag{3.1.23}$$

Details on various implementations of SVM classifiers can be found in [Chang and Lin, 2011]

3.1.3 P300 Processing

P300 as well as other ERP components are time-locked deflections in the EEG voltage in response to a sensory stimulus. The time-lock factor is the only known parameter in ERP signals and has been a key factor in their processing. Indeed other information about ERP components – such as phase, amplitude, and period, are either unknown or changing. They are influenced by concurrent or overlapping components.

Having a very small amplitude compared to the ongoing brain activity, ERP components are analysed using their occurrence time information. They are extracted through an averaging of many aligned signal segments of repeated trials. After averaging, phenomena that are time-locked to the stimulus will remain while unrelated EEG will cancel out. This requires that the experiment be repeated a couple of time [Rakotomamonjy et al., 2005].

A spatial filter that builds upon the time-locked characteristic of ERP was introduced by Rivet et al. [2009] and called xDAWN. It enhances a specific component in ERP by extracting spatial components that best describe the ERP features reconstructed through averaging of past trials. It is a major advance in P300-based BCI and ERP analysis in general [Rivet et al., 2011]. Several machine learning competition winners have relied on this approach [Barachant and Congedo, 2014; Barachant et al., 2015]. Using xDAWN, P300 can be processed online with a reduced number of trial repetitions, or even a single trial for ERP identification.

xDAWN spatial filters can be coupled with any binary classifier used in P300 identification. Classification algorithms based on SVM and LDA described in sections 3.1.1 and 3.1.2 have been particularly used in many successful P300 machine learning [Rakotomamonjy et al., 2005; Krusienski et al., 2008; Rivet et al., 2009; Jrad et al., 2011; Cecotti et al., 2011; Mak et al., 2011].

xDAWN

The first assumption made in xDAWN modelling is that the recorded EEG signal is composed of two typical patterns P_1 and P_2 , one evoked by the ERP stimuli (P_1), and another evoked by any stimulus including ERP stimuli (P_2). The second assumption is that the ERP patterns lie in an evoked subspace, hence could be enhanced by a spatial filter.

The first assumption yield the model

$$X = D_1 P_1 + D_2 P_2 + N, \quad (3.1.24)$$

where $D_1 \in \mathbb{R}^{N_t \times N_1}$ and $D_2 \in \mathbb{R}^{N_t \times N_2}$ are Toeplitz matrices with first column entries set to one at samples corresponding to ERP stimuli indexes and are zeros otherwise. They define a sort of ERP response temporal distribution in over all recorded EEG samples. N_1 and N_2 are the number of time samples considered for P_1 and P_2 respectively. N is the residual noise.

Based on the second assumption, xDAWN searches for a spatial filter $\mathbf{u}_1^* \in \mathbb{R}^{N_s}$ that maximises the signal-to-signal-plus-noise ratio (SSNR) $\rho(\mathbf{u})$:

$$\mathbf{u}_1^* = \arg \max_{\mathbf{u}} \rho(\mathbf{u}), \quad (3.1.25)$$

where the SSNR is estimated with

$$\hat{\rho}(\mathbf{u}) = \frac{\mathbf{u}^T \hat{\Sigma}_1 \mathbf{u}}{\mathbf{u}^T \hat{\Sigma}_X \mathbf{u}}, \quad (3.1.26)$$

where $\hat{\Sigma}_1$ is the estimation of the covariance matrix of the matrix $D_1 P_1$ and $\hat{\Sigma}_X$ is the estimation of the covariance matrix of the EEG signal X . The estimations of covariances are based on estimations of both P_1 and P_2 [Rivet et al., 2011, as described in].

In practice (3.1.26) can be solved for an estimate of the spatial filter $\hat{\mathbf{u}}_1$ with the generalised eigenvalue decomposition (GEVD) of $\hat{\Sigma}_1$ and $\hat{\Sigma}_X$ to obtain

$$\hat{\Sigma}_1 \hat{\mathbf{u}}_1 = \lambda_1 \hat{\Sigma}_X \hat{\mathbf{u}}_1, \quad (3.1.27)$$

where λ_1 is the largest eigenvalue returned by the GEVD, and $\hat{\mathbf{u}}_1$ the associated eigenvector.

P_1 can be factorised as $P_1 = \mathbf{a}_1 \mathbf{w}_1^T$ where $\mathbf{a}_1 \in \mathbb{R}^{N_1}$ is the temporal pattern and $\mathbf{w}_1 \in \mathbb{R}^{N_s}$ is its spatial distribution over channels. \mathbf{w}_1 is estimated as

$$\hat{\mathbf{w}}_1 = \hat{\Sigma}_X \hat{\mathbf{u}}_1. \quad (3.1.28)$$

As a side note, the appellation xDAWN came from the initial method modelling of the EEG signal X obtained in by [Rivet et al. \[2009\]](#):

$$X = DAW^T + N \quad (3.1.29)$$

where A is the pattern of the synchronised response to the ERP stimulus, D is the Toeplitz matrix defining the samples of the EEG epoch where the pattern of the synchronised response are active (like a temporal distribution), W is the spatial distribution of the ERP over channels, and N is the ongoing EEG.

3.1.4 Discussion

There is a variety of machine learning techniques that have been explored for classification in brain-computer interfaces [[Lotte et al., 2007](#); [Nicolas-Alonso et al., 2012](#); [Bashashati et al., 2007](#); [Khorshidtalab and Salami, 2011](#); [Krusienski et al., 2011](#)]. The ones described in this section are among the most successful and have thus been used recurrently in BCI research. While the design or the choice of spatial filters is guided by the neurological phenomenon used in a particular BCI type, the classification is achieved using standard classifiers that best separate the extracted features. Thus classifiers can be used interchangeably over various BCI types.

A remarkable fact about the discussed methods is that they all involve in a way or another, an estimate of covariance matrices or scatter matrices. Indeed, covariance matrices and scatter matrices capture a great deal of information about the signal of interest in the EEG. They contain information such as spatial patterns of neuronal activities involves in mental tasks, data distribution, and data variance, which are all crucial for the classification task. It is also noticed that nearly all algorithms – with the exception of kernel SVM – are developed from vector space or Euclidean space point of view. As has been said in chapter [1](#), and will be discussed in detail in chapter [5](#), covariance matrices lie on a curved space where Euclidean geometry is not suitable [[Congedo et al., 2013](#)].

BCI learning algorithms often face the *curse of dimensionality*, a phenomenon that describes the relationship between the sample size (*i.e.* number of observations in training set) and dimensionality (*i.e.* dimension of feature space): the amount of data needed to achieve sound statistical learning grows exponentially with the dimensionality [Fukunaga, 1990; Foley, 1972; Kanal and Chandrasekaran, 1971; Raudys and Jain, 1991]. If the sample size to dimensionality ratio is not large enough, the algorithms will be strongly biased. Although this is a general problem in machine learning, it is particularly present in EEG-based BCI, where a single observation is described by many features (e.g. time samples, frequency bands) from multiple sources. Such a big feature space would require very large training samples which are not usually attainable. The samples are recorded through thorough experiment protocols that can be conducted only for a relatively short period of time. A common way of alleviating the curse of dimensionality is through feature selection and dimensionality reduction techniques such as PCA and ICA.

A small training set may also lead to the problem of *overfitting*. When the training set is too small to represent the entire population, the model trained on such data will describe a separating line that is dependent on processes specific to the observed data rather than the global underlying discriminating factors [Hill et al., 2006]. This also happens when the model is overtrained or too complex for the task at hand. The fact that both spatial filter and classifier parameters are learned from the same training sample increases the risks of overfitting. In machine learning when the training set is deemed too small (or non-existent) to train a statistical model, notions of *domain adaptation* and *transfer learning* are used [Pan and Yang, 2010]. In domain adaptation, exiting data drawn from a different distribution are adapted and used to train a task on data from another distribution. In transfer learning, knowledge learned from previous data is used to lighten the learning process and alleviate the lack of training data. These two options are being investigated in machine learning for BCI [Kang et al., 2009; Wang et al., 2015].

3.2 Riemannian Approaches in Machine Learning

Information geometry provides useful tools for various machine learning and optimisation problems. In machine learning, Symmetric Definite-Positive (SPD) matrices have been used in various applications where features and data are only considered in the Euclidean space. A typical case where SPD could be found in machine learning is in covariance matrices which are of paramount importance in feature representation (Section 3.1). The covariance matrices are constrained to a special topology by their properties namely symmetry, positive definiteness, strict positivity of the diagonal elements, and the Cauchy-Schwarz inequalities (further discussed in Section 5.3.2). Indeed, covariance matrices lie in the space of SPD matrices which is a subset of the Euclidean space when considered with the scalar product. But the same space of SPD matrices, endowed with a differential structure, induces a Riemannian manifold.

Riemannian geometry can improve machine learning algorithms, taking into consideration the underlying structure of the considered space explicitly. Three kinds of approaches in the literature use the data geometry in machine learning. The first one relies on the mapping of the Riemannian manifold onto a Euclidean vector space. One such mapping, called logarithmic mapping, exists between the manifold and its tangent space, which is a Euclidean space, and has been used in classification tasks for BCI [Barachant et al., 2012b, 2013b]. Other kernels have been applied successfully to this end: Stein kernel, Log-Euclidean kernels as well as their normalised versions [Yger, 2013]. The main idea is to map the input data to a high-dimensional feature space, providing a rich and hopefully linearly separable representation. The so-called kernel trick is to provide a kernel function, which computes an inner product in the feature space directly from points lying in the input space, defining a Reproducing Kernel Hilbert Space (RKHS). The family of kernels defined on the Riemannian manifold allows implementing extension of all kernel-based methods, such as SVM, kernel-PCA or kernel k -means [Jayasumana et al., 2013]. Apart from the kernel approaches, once the data are mapped onto a vector space, any machine learning algorithm working in Euclidean space, such as LDA, could be applied [Barachant et al., 2012a].

A second kind of machine learning approach exploit the underlying geometry of the data. Instead of mapping the data to a Euclidean space, either a tangent space or an RKHS, the algorithms are adapted to Riemannian space. For instance, sparse coding algorithm has been adapted to Riemannian manifold, using the geodesic distance to estimate the data point and its sparse estimate [Xie et al., 2013]. Similarly nonlinear dimensionality reduction techniques have been adapted to Riemannian manifold, such as Laplacian Eigenmaps (LE), Locally Linear Embedding (LLE), and Hessian LLE. This adaptation was used to cluster data using their pdfs [Goh and Vidal, 2008b] or covariance matrices [Goh and Vidal, 2008a] as a feature. Another example is the adaptation of interpolation and filtering of data to Riemannian space performed in [Pennec et al., 2006], where an affine-invariant Riemannian metric is also proposed to offer a geodesically complete manifold *i.e.* a manifold with no edge and no singular point that can be reached in a finite time.

In the last kind of approach, instead of adapting existing algorithms from Euclidean to Riemannian geometry, new algorithms are developed directly for Riemannian manifolds. The *minimum distance to Riemannian mean* (MDRM) relies on a Riemannian metric to implement a multi-class classifier and have been applied on EEG. New EEG trials are assigned to the class whose average covariance matrix is the closest to the trial covariance matrix [Barachant et al., 2012a]. The MDRM classification can be preceded by a filtering of covariance matrices, like in [Barachant et al., 2010b] where covariance matrices are filtered with LDA component in the tangent space, then brought back to the Riemannian space for classification with MDRM. Another example is the *Riemannian Potato* [Barachant et al., 2013a], an unsupervised and adaptive artifact detection method, providing an online adaptive EEG filtering (*i.e* outlier removal). Incoming signals are rejected if their covariance matrix lies beyond a predefined distance z-score from the mean covariance matrix, computed from a sliding window. With the same objective of achieving robustness to noise that affect covariance matrices, Riemannian geometry is used to solve divergence functions of pdfs [Amari, 2010]. This allows to reformulate the CSP as the maximisation of the divergence between the distributions of data from two different classes corresponding to two cognitive states [Samek et al., 2013; Samek and Muller,

[\[2014\]](#). Using the *beta divergence* the obtained CSP is robust to outliers in sample covariance matrices and this algorithm is successfully applied to the EEG filtering for BCI. Riemannian metrics are also used for the EEG channel selection [\[Barachant and Bonnet, 2011\]](#) and the selection of the most discriminatory spatial filters in CSP [\[Barachant et al., 2010a\]](#).

Applications of Riemannian geometry to BCI mentioned thus far are focusing on motor imagery (MI) paradigm. In MI experiments, the subject is asked to imagine a movement (usually hand, feet or tongue), generating Event-Related Synchronisation and Desynchronisation (ERD/ERS) in pre-motor brain area. Riemannian BCI is well suited for MI experiments as the spatial information linked with synchronisation is directly embedded in covariance matrices obtained from multichannel recordings. However, for BCI that rely on Evoked Potential such as SSVEP or Event Related Potential (ERP), as P300, both frequential and temporal information are needed; the spatial covariance matrix does not contain this information. To apply Riemannian geometry to SSVEP and ERP, the sample covariance matrices can be defined from a rearrangement of the recorded data. The rearrangement is done such that the temporal or frequency information is captured [\[Congedo et al., 2013\]](#). With similar motivations, [\[Li et al. \[2009, 2012\]\]](#) defined a new Riemannian distance between SPD matrices that would take into account a weighting factor on matrices. They use this new distance as a dissimilarity between weighted matrices of power spectral density to classify EEG into different sleep state by k -nearest neighbours.

3.3 New Trend in BCI Systems

From the current state-of-the-art in BCI for control and communication ([Chapter 2](#)), it has become clear that the limitations in this field are such that BCI cannot replace traditional input modalities for human machine interface, nor match their performance. This restrains the use of BCI to a population with limited residual muscular ability to use traditional input devices.

Recently, research has been exploring ways of extending the use of BCI to a larger population – including healthy subject, in applications that will suffer less from BCI limitations such as the limited bandwidth (low information transfer rate), the BCI illiteracy, the training required to intentionally alter or generate

patterns of brain signals, as well as the cognitive workload involved in performing the BCI tasks. This effort has resulted in applications (or modalities) that tend to move away from fully relying on BCI as the sole input modality for human machine interfaces. The prominent examples in this trend are *hybrid brain-computer interfaces* (hBCI) [Millan et al., 2010] and *passive brain-computer interfaces* (pBCI) [Zander and Kothe, 2011]. In the following lines both of them are further discussed.

3.3.1 Hybrid BCI systems

One way of alleviating limitations in BCI is to combine multiple modalities or neurological phenomena. This has the potential of achieving higher information transfer rate and increasing degree of freedom. It is also a way to compensate a weakness in a particular type of brain-computer interface by relying on another one. Many combinations are possible: SSVEP and motor imagery, P300 and error related potential, etc. This has been suggested as a solution to BCI illiteracy as a subject who is illiterate toward a particular BCI type, e.g. SSVEP, might show efficiency in using another BCI type e.g. motor imagery.

The existing hBCI can be categorised according to (1) the type of signals combined and (2) how the signals are combined to achieve the desired task. According to the type of signal used, two types of hBCI are distinguished. In the first type, different brain signals (e.g. motor imagery, evoked potentials) are combined [Ferrez and Millan, 2008; Allison et al., 2010; Finke et al., 2011], while in the second a brain signal is combined with other biosignals e.g. ECG [Scherer et al., 2007] or EMG [Leeb et al., 2010]. The hBCI combining EMG and a brain signal is the only case where the residual muscular functionalities of the patients are used. Apart from this approach, residual muscular functionalities have been combined with BCI in a neuroprosthesis where the patient uses arm movement for reaching positions and BCI for grasping objects [Millan et al., 2009, 2004]. Though in this approach BCI is used as an additional channel to assistive technologies, using residual muscular functionality, BCI literature refers to it also as hBCI [Millan et al., 2010].

Depending on the combination of interfaces (or control channel), several control strategies are possible. The first one is to assign one specific task per

interface. Another possibility is to merge all interfaces in a weighted combination to achieve a unique task with higher accuracy. Finally, they can be used alternatively so as to allow users to smoothly switch from one interface to another depending on their performance or preference. The work of [Millan et al. \[2010\]](#) provides a comprehensive review of the existing hBCI approaches and their applications.

3.3.2 Passive BCI

While BCI relies on brain activities intentionally controlled by the user, there is a lot more information related to the user's states not intentionally expressed, that can be obtained from a real-time brain signal decoding (RBSD) and used in human computer/machine interaction (HCI or HMI). For decades, RBSD has been used for cognitive monitoring providing a way of looking into one's cognitive and affective states [\[Zander and Kothe, 2011\]](#). Passive BCI uses this implicit information as an additional input modality in HCI. Thus, the objective of BCI is moved from control and communication to improving HCI by using brain information not intentionally generated by the user. This opens BCI to all HCI applications and to a larger population.

In BCI for communication and control, the brain activities are consciously generated by the user either directly or indirectly resulting in a considerable cognitive workload which is not present in pBCI. [Zander and Kothe \[2011\]](#) categorises BCI in active, reactive, and passive. Active BCI and reactive BCI are used for control and communication. In the first, the brain patterns are directly and consciously controlled by the user (e.g. motor imagery). In the second, they are indirectly generated with the help of external stimulus (e.g. SSVEP).

pBCI does not require training users to generate specific patterns in brain signals, and thus not affected by problems of BCI illiteracy and heavy cognitive workload. Moreover pBCI does not suffer from the bandwidth limitation as it is used on top of other HCI input modalities. pBCI uses various brain signal features to infer information about the cognitive and affective state of the user [\[George and Lécuyer, 2010\]](#). Using this information, pBCI has been deployed in many HCI applications. In gaming for instance, pBCI has been used to

adapt the game interface and difficulty level to the affective and cognitive state of the player. This has crucial importance as the whole point of recreational activities such as games themselves is to relax, engage and entertain the user. With pBCI it is possible to measure this information and adapt while the user is still playing to ensure a balance between frustration, rewards, pleasure, etc. for overall satisfaction [Cutrell and Tan, 2008; Girouard et al., 2013].

Similarly, pBCI can be used for adaptive control in delicate applications such as aeroplane piloting, driving semi-autonomous cars, and automated industries. As control can be shared between man and machine, pBCI can detect when man control is likely to be defective—due to stress, fatigue, somnolence, etc., and allow the machine to take over [Cutrell and Tan, 2008; George and Lécuyer, 2010].

pBCI can also be used to detect and correct errors in HCI. When a user notices an erroneous response of a machine, an error related potential is evoked in their brain, and can be used in pBCI to correct the error. Errors are very common in active BCI and are more machine-made than human-made. With pBCI, BCI systems can fix their own mistakes [Ferrez and del R. Millan, 2008; Ferrez and Millan, 2008; George and Lécuyer, 2010].

pBCI is not only limited to human-machine interactions. Other applications include prevention of epileptic seizures [Liang et al., 2010], neurofeedback [Huster et al., 2014; Hao et al., 2014], and music creation [Yuksel et al., 2015], etc.

3.4 Proposed Approach

To address BCI problems raised in Section 1.2, namely user specificity, robustness of EEG representation and learning, and sufficiency of training data, two main avenues are explored: a hybrid BCI approach and a machine learning approach based on Riemannian geometry. The first aims at giving a solution to the problem of BCI users’ specificity while the second tackles problems of EEG representation and data sufficiency.

A Hybrid BCI Approach

The hybrid approach taken in this work combines brain signals (cerebral commands) with haptic command from user's residual motor skills. This approach offloads or takes a part of the control away from BCI by allowing the use of residual motor skills. The cerebral command limited by BCI state-of-the-art and the haptic command limited by user residual abilities complement each other. It has the advantage of reactive BCI systems; they only require the user to be attentive to the external stimuli; the pattern generated are merely a natural brain response. Thus, less training requires from the user, and lower cognitive workload involved. A touchless interface is used as the input modality for haptic commands. It is designed for a comfortable 3-D navigation, particularly for users with limited hand control.

The concept of hybrid BCI has been known mostly as a combination of various neurological phenomenon, combining various BCI types in one system. In the largest sense, hybrid BCI has combined brain signals and other biosignals [Millan et al., 2010]. In this work, the concept of hybrid BCI is stretched further to include muscular commands. A SSVEP based BCI is used for the cerebral command.

Machine Learning with Riemannian Framework

The proposed approach uses covariance matrices of EEG signals in a Riemannian framework. The covariance matrices are key in the representation of information contained in the EEG signal and constitute an important feature in their classification. They are handled with tools provided by the Riemannian geometry to alleviate difficulties in current BCI machine learning. Using covariance matrices as features, the machine learning pipeline depicted in Figure 3.8 is adopted. It consists of three main phases: offline model selection, training phase, and classification. Unlike the one in Figure 3.1, it does not include a spatial filtering phase.

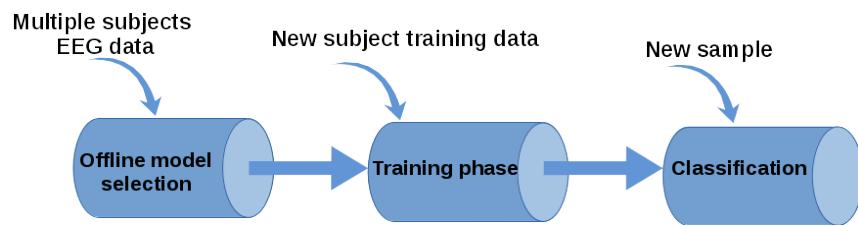


Figure 3.8: Adopted machine learning pipeline. It consists of 3 main phases: 1) Offline model selection: A number of preprocessing and classifiers are tested on a collection of data from multiple subjects, the best combination is selected. 2) Training phase: Classifier parameters are trained for each new subject before classification. 3) Classification: A new [unseen] sample is preprocessed similarly to training data, then classified.

Chapter 4

Hybrid Brain-Computer Interface

Anybody who has been seriously engaged in scientific work of any kind realises that over the entrance to the gates of the temple of science are written the words: Ye must have faith. It is a quality which the scientist cannot dispense with.

— Max Planck

4.1 Introduction

Rehabilitation and assistive technologies aim at developing solutions *adapted* to the subjects' disabilities. A crucial aspect is to take into account the specificities of each person and to propose technical solutions which make use of their residual motor capabilities. As a simple example, an electric wheelchair will not interest someone who has still some strength in his upper limbs but the same person could be interested by the assistance of an electric motor for driving a manual wheelchair.

BCI, in their essence, overlook the muscular system. They do not rely on subjects' residual motor capabilities. But because of the limited performances of current BCI systems, patients might desire to use their residual motor abilities. In this case a system that allows patients to use both their brain signals and

their residual motor abilities would be adequate and in-line with rehabilitation principles.

Hence, in this work, we propose a new methodology for disabled people, using a hybrid approach where a physical interface is complemented by a brain-computer interface.

The contribution is threefold : we define a new methodology for a shared control system, we introduce a new learning scheme for BCI and we propose an implementation of the whole system in two applications of rehabilitation robotics.

The proposed system makes use of user's residual motor abilities and offers BCI as an optional choice: the user chooses when to rely on BCI and could alternate between the muscular- and brain-mediated interface at the appropriate time. The hybrid system integrates a 3D touchless interface based on IR-sensors [Martin et al., 2012] that captures hand poses and an SSVEP-based BCI. Such an approach combines these two interfaces in a multimodal BCI-motor system that takes advantage of both the user's brain signals and her residual motor ability.

Regarding the touchless interface, the IR-based interface does not need to be held by the user, thus not requiring any grasping capability. It provides a three degrees of freedom controller. On the neural side, an SSVEP-based BCI is used and a novel algorithm based on Canonical Correlation Analysis (CCA) is used to classify SSVEP epochs.

4.2 Proposed Hybrid Interface

The proposed hybrid approach is illustrated in Figure 4.1. The user – a person with motor disabilities – is given two modalities to control the system. The first modality is an input device that takes a signal generated by users' motor action. This might be any type of device that is adapted to the subject's disability, allowing her to use her residual motor ability. This modality is used for the continuous control of the system.

The second modality, which is based on the EEG, is used to provide an additional command, giving alternative control options to the user, or a special command to activate a common and repetitive task. In this work, the continuous

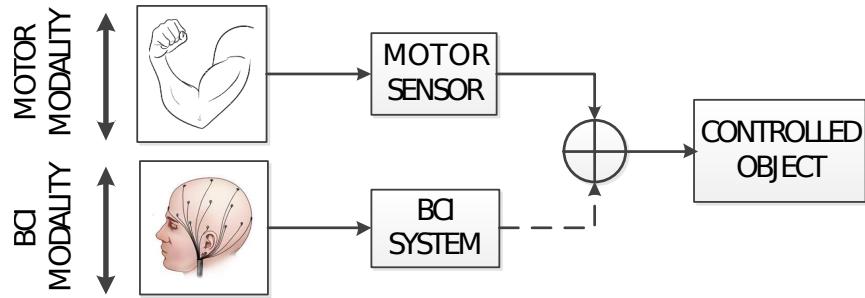


Figure 4.1: Hybrid BCI system integration. The motor abilities of the user are the primary controller of the system, using adapted interfaces (here: 3D touchless interface). The brain-computer interface (here: SSVEP) provides a complementary communication channel and is designed to trigger specific actions.

control is achieved by means of a touchless interface and the BCI modality relies on SSVEP.

4.3 Touchless Interface

There are a number of neurodegenerative diseases that affect large muscles. Some result in loss of strength in the hands and the fingers. For instance, people with amyotrophic lateral sclerosis may have their hands and fingers weakened and lose their grasping capacity. Such patients cannot use most of the existing computer input devices such joysticks and mouses that require a certain degree of grasping. The *touchless* interface used in this work is designed for such patients. It requires no physical contact and no grasping. The interface embeds five infrared (IR) sensors which could be setup in different spatial positions, according to the user requirements. Each IR-sensor consist of a pair of IR transmitter-receiver (Tx/Rx). The sensors are placed spatially around a user's arm such that a slight hand movement will alter the pattern of reflected IR beams. The interface is configured to detect six hand positions for six commands namely up, down, left, right, forward, and backward. To these is added a resting position. The position is chosen to be intuitive and coherent with the command. They are further tuned to fit user hand movement capacities through a training phase where the subject is asked to perform movements corresponding to the

six positions for a number of times.

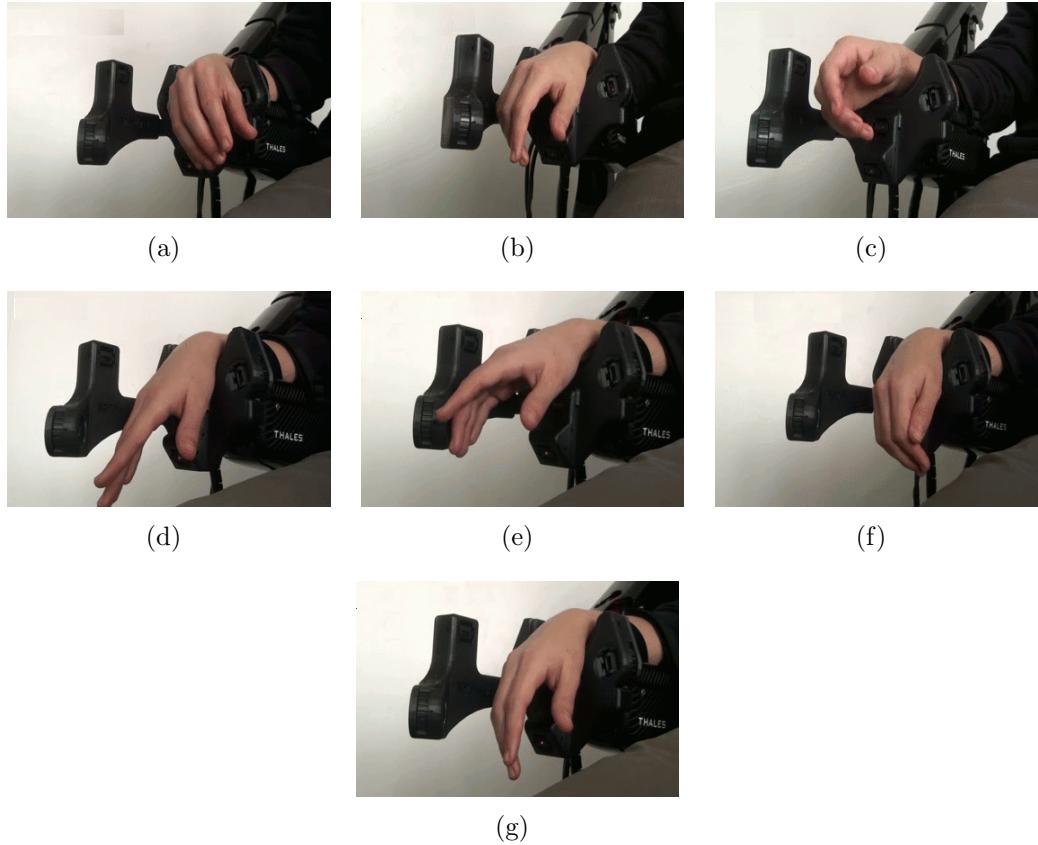


Figure 4.2: The 3D touchless interface and a user's hand positions for different commands: (a) left, (b) right, (c) up, (d) down, (e) forward, (f) backward, (g) rest. The IR-sensors are in the black plastic housing on the right side of the hand and around the wrist. Another symmetrical plastic housing has been realised for left-handed users.

The control system relies on an iterative k -Nearest Neighbours (kNN) scheme to learn hand poses of each user. Firstly, the iterative kNN scheme requires a fast calibration phase to learn the different hand poses, here seven (six for the directions and one for the resting position). The outliers and ambiguous examples are excluded from the training examples. Secondly, the algorithm continuously adapts to the received signal, labelling new examples change the set of neighbours. This algorithm is able to track the changes of the user's hand pose, providing an online adaptation to the behavioural modifications induced

by tiredness. For more details on the algorithm, see [Martin et al., 2012]. The six hand positions recognised by IR-interface is used to control a four degrees-of-freedom robotic arm exoskeleton developed in the *ESTA* project [Baklouti et al., 2008] to compensate for motor deficiency in the upper limb (Figure 4.4). More generally, it can be used by patients as well as healthy subjects in applications where navigation in a 3D Euclidean space is needed.

4.4 SSVEP-based BCI

4.4.1 Material and EEG Recording

The g.Mobilab+ device (shown in Figure 4.3) is used for recording EEG at 256 Hz on 8 channels.



Figure 4.3: Acquisition material, the EEG is recorded with electrodes, the signal is amplified and sent to a computer running OpenVIBE.

For SSVEP stimulation, flash stimulus technique has been chosen. To avoid limitations imposed by refresh rate of computer screens, a microcontroller is set up to flash stimuli with light-emitting diodes (LED) at frequencies $F = \{13, 17, 21\}$ Hz. The device has been controlled and the LED blinking is precise up to the millisecond. The eight electrodes are placed according to the 10/20 system on Oz, O1, O2, POz, PO3, PO4, PO7 and PO8. The ground was placed on Fz and the reference was located on the right (or left) ear lobe.

In this study, 12 male and female subjects aged between 20 and 32 years participated in the experiment. Informed consent was obtained from all subjects, each one has signed a form attesting their consent. The subject sits in an electric wheelchair, his right upper limb is resting on the exoskeleton (Figure 4.4). The exoskeleton is functional but is not used in the offline recordings.

A panel of size 20x30 cm is attached on the left side of the chair, with three groups of four LEDs blinking at different frequencies. Despite the panel being on

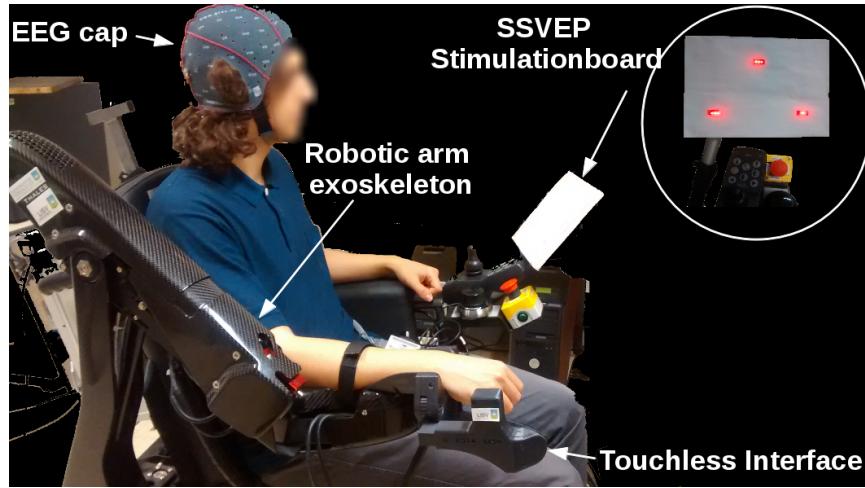


Figure 4.4: Experimental setting: subject sitting on an electric wheelchair equipped with a robotic arm exoskeleton. During offline recordings, the exoskeleton and the touchless interface are disabled; the subject performs the SSVEP task as prompted by audio cues.

the left side, users can see it without moving their head. The subjects were asked to sit comfortably in the wheelchair and to follow the auditory instructions, they could move and blink freely. A sequence of trials is proposed to the user. A trial begins by an audio cue indicating which LED to focus on, or to focus on a fixation point set at an equal distance from all LEDs for the reject class. A trial lasts five seconds and there is a three second pause between each trial. The evaluation is conducted during a session consisting of 32 trials, with 8 trials for each frequency and 8 trials for the reject class (or resting class), *i.e.* when the subject is not focusing on any specific blinking LED.

The experiments were conducted at the *Laboratoire d'Ingénierie des Systèmes de Versailles* (LISV) of the Université de Versailles Saint-Quentin-en-Yvelines, Paris-Saclay.

4.4.2 Signal Processing

The measured EEG signal is treated with a processing pipeline that offers state-of-the art BCI performance. EEG epochs of three seconds are gathered every 0.5 second. Each epoch is filtered between 12 Hz and 45 Hz to discard irrelevant

bands while allowing all stimulation frequencies and their first harmonics.

A spatial filter is then designed based on CCA method described in Section 3.1.2. Unlike the work of [Lin et al., 2006], which propose to rely on the correlation coefficient of CCA for processing SSVEP signals, in this work CCA is only applied to determine the spatial filter w_X .

With the signal being recorded at 256 Hz with eight electrodes, the size of a 3-second EEG trial is 8×768 ($X \in \mathbb{R}^{8 \times 768}$). And let $Y \in \mathbb{R}^{H \times N}$ be a multivariate signal representing the SSVEP stimulation signal. As described in Section 3.1.2, CCA finds the mappings $\mathbf{w}_X, \mathbf{w}_Y \in \mathbb{R}^8$ that maximises the correlation between the $\mathbf{w}_X^T X$, and $\mathbf{w}_Y^T Y$. The signal $x = \mathbf{w}_X^T X$ is a linear combination of all the electrodes and is expected to maximise the correlation with a hypothetically perfect neural response, that is the sinusoids of Y . A similar approach can be found in [Spüler et al., 2012] but in a different context and using x to generate exemplars for supervised learning. After filtering, a multiclass SVM classifier with RBF kernels is used for classification (refer to Section 3.1.2). It is given as input the power spectral densities extracted from the spatially filtered signal x , and output a class $k \in \{13\text{Hz}, 17\text{Hz}, 21\text{Hz}, \text{resting}\}$. The LIBSVM [Chang and Lin, 2011] package is used for SVM implementation.

4.5 Applications

The described approach is validated in two contexts: a Virtual Environment (VE) for the navigation of a helicopter shown in Figure 4.7, and an exoskeleton arm control task shown in Figure 4.8. In the VE, the user is asked to reach three waypoints. Three specific locations are identified in the VE to serve as *shortcuts*. In previous works, locations of this nature have been used as a predefined final destination [Lotte et al., 2010], while we only use them as shortcuts. After reaching these locations using BCI commands, the user could reach any position using the 3D-touchless interface.

The approach with the exoskeleton arm bears some similitude with the VE navigation task. The arm is controlled with the 3D-touchless interface. Common arm movements performed by the user are predefined (e.g. reaching the mouth or a resting position). The BCI shortcut trigger the automatic arm movement to these positions.

The hybrid scheme is especially well suited for exoskeleton arm control task: as the arm is continuously controlled by the 3D touchless interface, once the user has grabbed an object (e.g. a glass of water), she will no longer be able to move her hand freely to control the touchless interface. The BCI command allows to overcome this limitation by activating predefined movements.

4.6 Experimental Results

This section describes the results obtained with the proposed system. Out of twelve subjects who participated in the offline EEG recording, only five participated in the online experiment in the virtual environment. One of the subjects is hemiplegic and the four others are healthy. The first section is dedicated to the assessment of the proposed online detection of SSVEP algorithm. The second section provides the results obtained using the hybrid system for a navigation task in a virtual environment. The last section explained how the system has been implemented on an embedded system for an exoskeleton arm control task.

4.6.1 Validation of Proposed SSVEP Algorithm

Before using the BCI subsystem in online mode, a calibration phase is needed to compute the CCA spatial filter and training the SVM classifier. During the calibration phase, a sequence of trials is proposed to the user. The online classification is done every 0.5 second, using a $t_W = 3$ s window of EEG signals. An audio feedback indicates the predicted class to the user.

Figure 4.5 shows the online BCI classification performances for each prediction made every 0.5 second, starting at $t = t_0 + t_W$, that is three seconds after the beginning of the trial t_0 . The y-axis indicates the error rate for each of the five subjects. The results demonstrate that the proposed algorithm is very robust and provides a very reliable response after $t + 2$ s with a small mean error rate for all subjects.

To further evaluate the algorithm, it is important to consider that the loss function is not uniform. If the algorithm detects a reject class instead of a specific class, the consequences are not as bad as a wrong prediction: e.g. detecting

Table 4.1: Comparison with other algorithms

	Subject1	Subject2	Subject3	Subject4	Subject5
Baseline	81.3	88.3	80.0	75.0	79.2
ICA	100	88.3	91.7	93.3	95.0
CCA	100	100	97.5	93.3	96.7

13 Hz instead of 17 Hz, as the user needs only to concentrate half a second on the chosen LED before the system make another prediction. Thus we propose the following accuracy measure, similar to a precision score. For each trial, we consider the first class prediction at time t : if this is correct the accuracy is increased, if this is false the accuracy is decreased. If the prediction is the reject class, the accuracy measure is only postponed on the next time segment. Figure 4.6 displays the results of this measure for all subjects. The accuracy is above 70% for almost all subjects and it can be seen that the algorithm provides almost immediately the correct answer.

At last, we compare the proposed algorithm with classical SSVEP approaches in Table 4.1, using an offline evaluation for each subject. The baseline is a comparison with a SVM using the PSD of the EEG signal, that is without applying the CCA spatial filter. A classical methodology is to rely on ICA to extract the main components of the signal and to provide these components to the SVM classifier. Table 4.1 shows that the proposed algorithm yields the best results.

4.6.2 Experiments in Virtual Environment

For the navigation task in the virtual environment, the assessment is based on the time spent and the distance travelled during the experiment for four subjects. These results are shown in Table 4.2. The time is indicated in seconds and the distance in metric units. Each subject has performed three experiments: in the first experimental condition, the subject should rely only on the 3D touchless interface ('None' in Table 4.2). In the second one, shortcuts are available and are triggered by the BCI subsystem (BCI-S). In the last experimental condition, the subject could trigger a shortcut using a keyboard (KB-S). The fourth subject is hemiplegic and she could not use the keyboard with her

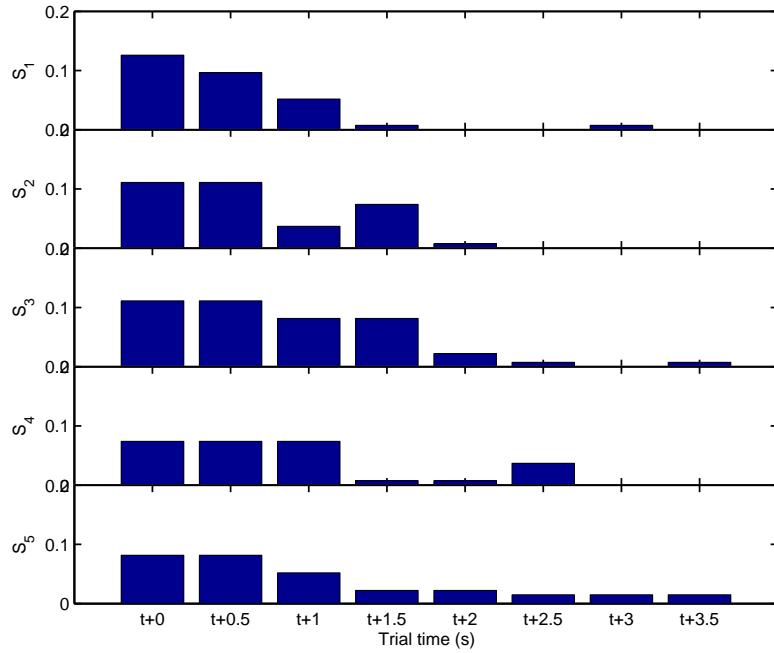


Figure 4.5: Evaluation of the online performances of the proposed BCI algorithm. The error rates for all five subjects are indicated as a function of time, with $t+0$ indicating the first prediction made (after $t_W = 3$ s). The error rates are averaged on all classes.

spare hand. Thus, her results do not include the last experimental condition.

In Table 4.2, next to the BCI and keyboard shortcut, a percentage indicates the relative improvement compared to the reference experiment (without shortcut). It could be seen that distance covered is almost equivalent with BCI shortcuts and keyboard shortcuts, which is the expected results as users have activated the shortcut each time it was possible. When the shortcuts are activated by the BCI, the task is slower than when using the keyboard. This effect is mainly caused because the subject need to focus at least three seconds on a blinking LED before triggering the shortcut.

4.6.3 Application to Exoskeleton Arm Control

The proposed system has been applied to the ESTA exoskeleton arm control. This assistive device is designed to compensate shoulder and elbow deficiencies occurring in degenerative diseases. The subject controls the exoskeleton arm

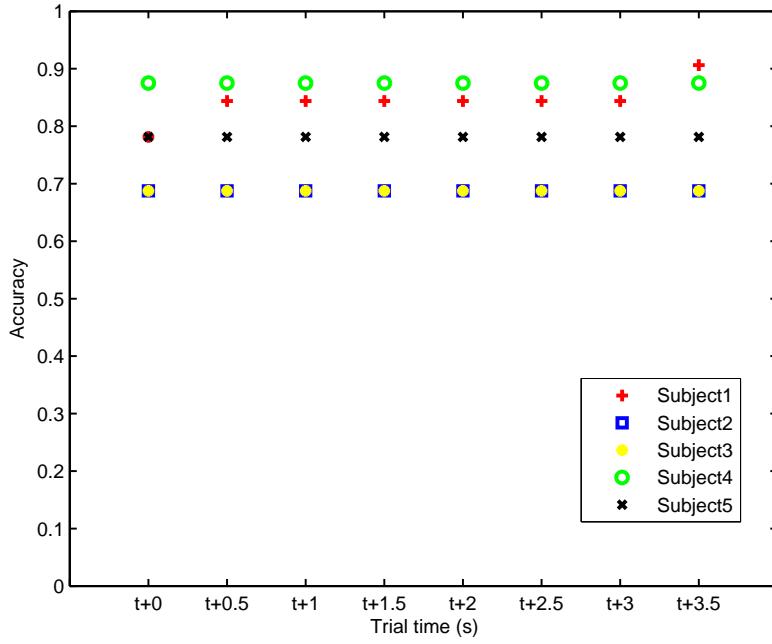


Figure 4.6: Assessment of the accuracy of classification depending on the time of the prediction. On x-axis, $t+0$ indicates the first prediction made three seconds after the start of a trial. The results are averaged on all trials for each subject. Subject 1 is the only one to present a slight increase of the classification accuracy. For all other subjects, the algorithm proposes a correct answer as the first prediction.

with the touchless interface and the BCI shortcuts allow to reach predefined positions, such as a resting or a close-to-mouth positions. In the case of the hemiplegic subject (who cannot use her left arm and hand), the BCI subsystem is the only possibility to control the exoskeleton with an object in hand. This example illustrates the complementary aspect of the two interfaces, the physical and the brain one.

Figure 4.8 illustrates an application of the proposed hybrid interface on the ESTA chair. The user is seated in an environment where he can perform daily life routine. In this case, next to a table where a phone and a glass of water are placed on (to represent objects that are commonly used). The user can reach to the table, and pick an object of his choice and manoeuvre it around. His arm is supported with the robotic exoskeleton and he is given the touchless interface and the BCI in the hybrid multimodal framework to control it. As with in the



Figure 4.7: Experiment in the virtual environment. Here the subject is using the 3D touchless interface with his right hand and the SSVEP LEDs are put in front of him. The screen displays a helicopter in the virtual environment. The subject should pass through all waypoints, materialised by red (or grey) disks on the screen. When the subject triggers a shortcut, the helicopter is automatically moved to a location materialised by the transparent ball.

VR environment, regions of interest can be defined in a daily life environment. These regions are the most visited and trajectories leading to them can be optimised and recorded. In the current experiment, the table and the user's face (mouth), and the resting positions are defined as regions of interest. Their trajectories are optimised, recorded and can be triggered automatically. In this way, the user can use BCI commands to trigger movement to regions of interest. The touchless interface can then be used for continuous control to reach local positions. The IR interface can be turned on and off using a BCI command. This will allow the user to move his hand even when he does not intend to send a command to the touchless interface.

Table 4.2: Distance covered and duration of experiments, without shortcuts (None), with BCI-activated shortcuts (BCI-S) and with keyboard-activated shortcuts (KB-S).

		None	BCI-S (inc. %)	KB-S (inc. %)
Subject 1	Time	108.9	68.3 (37.3%)	53.56 (50.8%)
	Distance	1367.7	538.2 (60.6%)	535.0 (60.9%)
Subject 2	Time	99.2	74.7 (24.7%)	50.5 (49.1%)
	Distance	1469.4	529.1 (64.0%)	549.0 (62.6%)
Subject 3	Time	105.5	63.4 (39.9%)	50.4 (52.2%)
	Distance	1447.3	627.6 (56.6%)	542.1 (62.5%)
Subject 4 (hemiplegic)	Time	125.6	70.4 (43.9%)	—
	Distance	1490.8	598.9 (59.8%)	—

4.7 Conclusion

In this chapter a new methodology for designing hybrid systems was proposed. It uses a brain interface and physical interface specifically design to fit the user's needs. The main goal of this hybrid system is to assist people with motor disabilities or muscular diseases, by proposing a system that adapts to their individual needs, and makes use of their residual abilities. The BCI is integrated in the system as a secondary modality, which is used to trigger specific behaviour or predefined actions.

A first contribution is to propose an implementation of such a system using a 3D touchless interface and a SSVEP-based BCI. This implementation gathers the two interfaces in a multimodal system which benefits from both the brain and motor signals. The second contribution is to describe a novel algorithm for processing SSVEP-based EEG signals, with stable results, even when computed in an online setup. This algorithm is compared to other existing solutions and an experimental assessment of its validity is conducted.

The full system is evaluated on a 3D navigation task in virtual environment. The results demonstrate that the system is functional and could be used to assist people in various contexts. The system is lastly used to control the ESTA arm exoskeleton: the system is functional and could be adapted for controlling

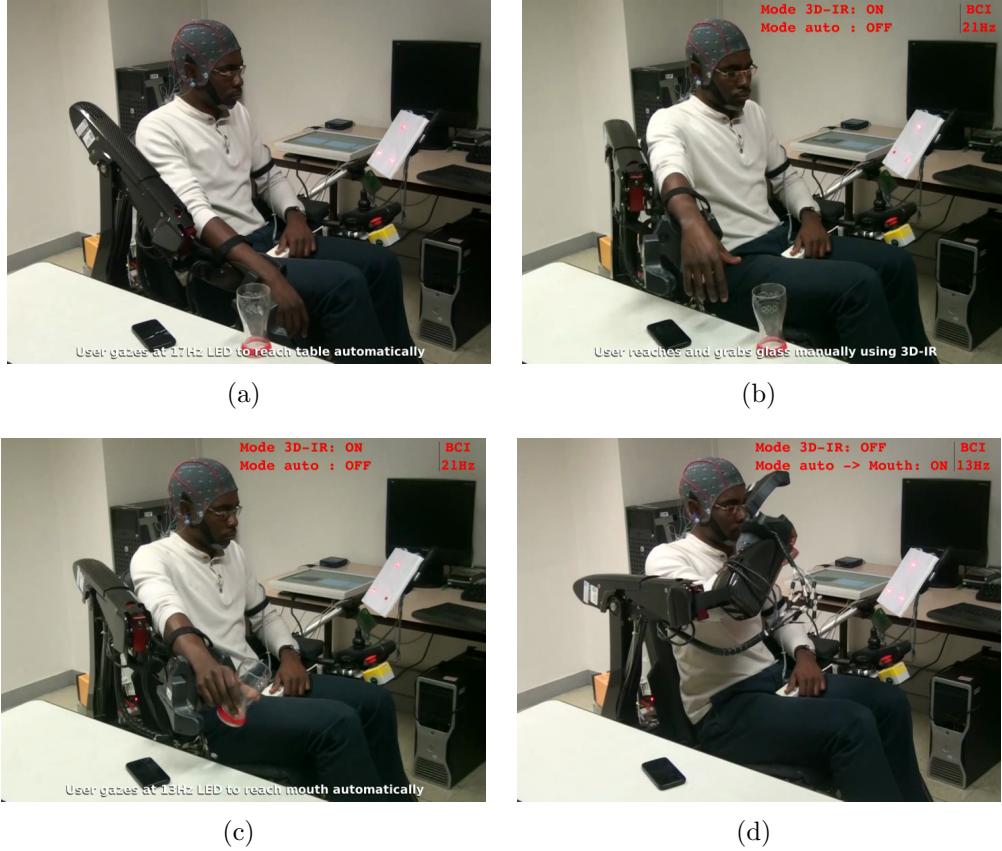


Figure 4.8: Subject sitting on the ESTA wheelchair. His arm is supported by the exoskeleton, and the left hand is lying on the touchless interface. On his right-hand side is the SSVEP stimulation board. He is fitted with an EEG cap for brain signals recording. Next to the exoskeleton, an object is put on a table. (a) The subject is in resting position. He is gazing at the 17 Hz LED to trigger an automatic trajectory to the table. (b) Subject has reached the table and is using the touchless interface to reach and grab the glass (c) Glass in hand, the user gazing at the 13 Hz LED to activate the automatic trajectory to mouth. (d) The arm reaches the mouth while the touchless interface is deactivated.

other assistive devices.

Although a good classification accuracy is achieved with the proposed method based on CCA and SVM, this work focuses more on BCI framework to improve the BCI usability and adaptability to the physical needs of subjects. The proposed framework answers the problem of variability in physical aptitude

amongst patients and users of BCI. Future work will be focused on the signal processing and machine learning methods that tackle variability in the brain response and in the experimental environment.

Chapter 5

Riemannian Geometry for Brain-Computer Interfaces

The history of science has proved that fundamental research is the lifeblood of individual progress and that the ideas that lead to spectacular advances spring from it.

— Sir Edward Appleton

5.1 Riemannian Manifold of Symmetric Positive-Definite Matrices

A m -dimensional *manifold* \mathcal{M} is a Hausdorff space for which every point has a neighbourhood that is homeomorphic to an open subset of \mathbb{R}^m [Jost, 2011]. When a tangent space is defined at each point, \mathcal{M} is called a differential manifold. A *geodesic* γ is the shortest smooth curve between two points, Σ_1 and Σ_2 . The tangent space $T_\Sigma\mathcal{M}$ at point Σ is the vector space spanned by the tangent vectors of all geodesics on \mathcal{M} passing through Σ . A *Riemannian* manifold is a manifold endowed with an inner product defined on the tangent space, which varies smoothly from point to point.

For the rest of this chapter, we will restrict to the analysis of the manifold \mathcal{M}_C of the $C \times C$ symmetric positive-definite (SPD) matrices, defined as:

$$\mathcal{M}_C = \left\{ \Sigma \in \mathbb{R}^{C \times C} : \Sigma = \Sigma^\top \text{ and } x^\top \Sigma x > 0, \forall x \in \mathbb{R}^C \setminus \{0\} \right\} .$$

The tangent space $T_\Sigma \mathcal{M}_C$ is identified to the Euclidean space of symmetric matrices:

$$\mathcal{S}_C = \{\Theta \in \mathbb{R}^{C \times C} : \Theta = \Theta^\top\} .$$

The dimension of the manifold \mathcal{M}_C , and its tangent space $T_\Sigma \mathcal{M}_C$, is $m = C(C + 1)/2$.

The mapping from a point Θ_i of the tangent space to the manifold is called the exponential mapping $\text{Exp}_\Sigma(\Theta_i) : T_\Sigma \mathcal{M}_C \rightarrow \mathcal{M}_C$ and is defined as:

$$\text{Exp}_\Sigma(\Theta_i) = \Sigma^{\frac{1}{2}} \text{Exp}(\Sigma^{-\frac{1}{2}} \Theta_i \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}} . \quad (5.1.1)$$

Its inverse mapping, from the manifold to the tangent space is the logarithmic mapping $\text{Log}_\Sigma(\Sigma_i) : \mathcal{M}_C \rightarrow T_\Sigma \mathcal{M}_C$ and is defined as:

$$\text{Log}_\Sigma(\Sigma_i) = \Sigma^{\frac{1}{2}} \text{Log}(\Sigma^{-\frac{1}{2}} \Sigma_i \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}} . \quad (5.1.2)$$

$\text{Exp}(\cdot)$ and $\text{Log}(\cdot)$ are the matrix exponential and matrix logarithm respectively. The computation of these operators is straightforward for SPD matrices of \mathcal{M}_C . They are obtained from their eigenvalue decomposition (EVD):

$$\begin{aligned} \Sigma &= U \text{diag}(\lambda_1, \dots, \lambda_C) U^\top , \\ \text{Exp}(\Sigma) &= U \text{diag}(\exp(\lambda_1), \dots, \exp(\lambda_C)) U^\top , \\ \text{Log}(\Sigma) &= U \text{diag}(\log(\lambda_1), \dots, \log(\lambda_C)) U^\top , \end{aligned}$$

where $\lambda_1, \dots, \lambda_C$ are the eigenvalues and U the matrix of eigenvectors of Σ . As any SPD matrix can be diagonalised with strictly positive eigenvalues, $\text{Log}(\cdot)$ is always defined. Similarly the square root $\Sigma^{\frac{1}{2}}$ is obtained as:

$$\Sigma^{\frac{1}{2}} = U \text{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_C^{\frac{1}{2}}) U^\top ,$$

and is unique. The same goes for $\Sigma^{-\frac{1}{2}}$.

The tangent vector of the geodesic $\gamma(t)$ between Σ_1 and Σ_2 , where $\gamma(0) = \Sigma_1$ and $\gamma(1) = \Sigma_2$ is defined as:

$$v = \overrightarrow{\Sigma_1 \Sigma_2} = \text{Log}_{\Sigma_1}(\Sigma_2) . \quad (5.1.3)$$

5.2 Covariance Matrix Estimation

When working with covariance matrices, a crucial point is to correctly estimate the covariance when the number of samples is small or heavily corrupted by noise. Several approaches have been proposed to build the covariance matrices, relying on normalisation or regularisation of the sample covariances. To assess the quality of the covariance matrices obtained from EEG samples, a comparative study of these estimators is conducted.

Let $x_n \in \mathbb{R}^C$, $n = 1, \dots, N$, denotes a sample of a multichannel EEG trial recorded on C electrodes. N is the trial length. Let $X \in \mathbb{R}^{C \times N}$ be the EEG trial such as $X = [x_1, \dots, x_N]$. Under the hypothesis that all N samples x_n are randomly drawn from a distribution, it follows that \mathbf{x} is a variable of random vectors and its expected vector is $\omega = E\{\mathbf{x}\}$ [Fukunaga, 1990]. The covariance matrix of the random variable \mathbf{x} is defined by $\Sigma = E\{(\mathbf{x} - \omega)(\mathbf{x} - \omega)^\top\}$ and is unknown, thus an estimate $\hat{\Sigma}$ should be computed. The choice of the appropriate estimator is crucial to verify that the obtained covariance matrices fulfil the following properties: they should be accurate, SPD, and well-conditioned. The last property requires that the ratio between the maximum and minimum singular value is not too large. Moreover, to ensure the computational stability of the algorithm, the estimator should provide full-rank matrices, and its inversion should not amplify estimation errors.

5.2.1 Sample Covariance Matrix Estimator

The most usual estimator is the empirical *sample covariance matrix* (SCM), defined as:

$$\begin{aligned}\hat{\Sigma}_{\text{scm}} &= \frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^\top \\ &= \frac{1}{N-1} X \left(\mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \right) X^\top ,\end{aligned}\tag{5.2.1}$$

where $\bar{x} \in \mathbb{R}^C$ is the sample mean vector $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$. In the matrix notation, \mathbf{I}_N is the $N \times N$ identity matrix and $\mathbf{1}_N$ is the vector $[1, \dots, 1]$. The

SCM is often normalised as [Fukunaga, 1990]:

$$\hat{\Sigma}_{\text{nsbcm}} = \frac{C}{N} \sum_{n=1}^N \frac{(x_n - \bar{x})(x_n - \bar{x})^\top}{\sigma_{x_n}^2}, \quad (5.2.2)$$

with the inter-channel variance at time n defined as $\sigma_{x_n}^2 = (x_n - \bar{x})^\top(x_n - \bar{x})$. Another normalisation techniques could be used.

This estimation is fast and computationally simple. However when $C \approx N$, the SCM is not a good estimator of the true covariance. In the case $C > N$, the SCM is not even full rank.

5.2.2 Shrinkage Covariance Matrix Estimators

To overcome the shortcomings of SCM, the shrinkage estimators have been developed as a weighted combination of the SCM and a target covariance matrix, which is often chosen to be close to the identity matrix, *i.e.* resulting from almost independent variables of unit variance.

$$\hat{\Sigma}_{\text{shrink}} = \kappa \Gamma + (1 - \kappa) \hat{\Sigma}_{\text{scm}}, \quad (5.2.3)$$

where $0 \leq \kappa < 1$. This estimator provides a regularised covariance that outperforms the empirical $\hat{\Sigma}_{\text{scm}}$ for small sample size, that is $C \approx N$. The shrinkage estimator has the same eigenvectors as the SCM, but the extreme eigenvalues are modified, *i.e.* the estimator is shrunk or elongated toward the average.

The different shrinkage estimators differ in their definition of the target covariance matrix Γ . Ledoit and Wolf [Ledoit and Wolf, 2004] ($\hat{\Sigma}_{\text{shrink_ledoit}}$ on Figure 5.2) have proposed $\Gamma = v \mathbf{I}_C$, with $v = \text{Tr}(\hat{\Sigma}_{\text{scm}})$. Blankertz [Blankertz et al., 2011] ($\hat{\Sigma}_{\text{shrink_blank}}$) defines Γ also as $v \mathbf{I}_C$ but with $v = \frac{\text{Tr}(\hat{\Sigma}_{\text{scm}})}{C}$. Schäfer ($\hat{\Sigma}_{\text{shrink_schaf}}$) proposes several ways of defining Γ depending on the observed $\hat{\Sigma}_{\text{scm}}$ [Schäfer and Strimmer, 2005].

5.2.3 Fixed-Point Covariance Matrix Estimator

The fixed-point covariance matrix [Pascal et al., 2005] is based on the maximum likelihood estimator $\hat{\ell}$ which is a solution to the following equation:

$$\hat{\Sigma}_{\text{fp}} = \hat{\ell} = \frac{C}{N} \sum_{n=1}^N \left(\frac{(x_n - \bar{x})(x_n - \bar{x})^\top}{(x_n - \bar{x})^\top \hat{\ell}^{-1} (x_n - \bar{x})} \right). \quad (5.2.4)$$

As there is no closed form expression to Eq. (5.2.4), it can be written as a function of $\hat{\ell}$: $g(\hat{\ell}) = \hat{\Sigma}_{\text{fp}}$. g admits a single *fixed point* $\hat{\ell}^*$, where $g(\hat{\ell}^*) = \hat{\ell}^*$, which is a solution to Eq. (5.2.4). Using $\hat{\ell}_0 := \hat{\Sigma}_{\text{nscm}}$ as the initial value of $\hat{\ell}$, it is solved recursively as $\hat{\ell}_t \xrightarrow[t \rightarrow \infty]{} \hat{\ell}^*$.

5.3 Classification of SSVEP Covariance Matrices

5.3.1 Machine Learning Approach for Classification

From multiple labelled observations, belonging to two or more classes, and a new unlabelled observation, the classification task objective is to assign to the class whose elements share similar properties with the considered observation. In this article, we make two hypotheses that are commonly acknowledged in EEG: the data distribution is Gaussian and classes have identical dispersions. Given labelled observations x_i drawn from two classes ($y_i = 1$ or $y_i = -1$), a simple classification algorithm consists in assigning a previously unseen observation to the class with the closest mean. This simple algorithm requires only to define a computable distance and mean. Assuming that the observations are embedded into a dot product space, *e.g.* Euclidean space, the mean can be expressed as:

$$c_+ = \frac{1}{m_+} \sum_{\{i|y_i=+1\}} x_i , \quad (5.3.1)$$

$$c_- = \frac{1}{m_-} \sum_{\{i|y_i=-1\}} x_i , \quad (5.3.2)$$

where $y_i \in \{+1, -1\}$ is the label of the training observation x_i , m_+ and m_- the number of positive and negative observations respectively. An unseen observation x is assigned to the class whose mean is the closest. This simple geometric classification framework is the founding principle of more complex algorithms such as support vector machines. It can be formulated in terms of the dot product $\langle \cdot, \cdot \rangle$. If $c = (c_+ + c_-)/2$ is the point lying halfway between c_+ and c_- , and $w = c_+ - c_-$ the vector connecting c_+ to c_- , the class label y of the unseen observation x is determined by checking whether the vector $x - c$ connecting c

to x make an angle $\alpha < \pi/2$ with w . This is expressed as:

$$\begin{aligned} y &= \text{sgn} \langle (x - c), w \rangle \\ &= \text{sgn} \langle (x - (c_+ + c_-)/2), (c_+ - c_-) \rangle \\ &= \text{sgn}(\langle x, c_+ \rangle - \langle x, c_- \rangle + b) \end{aligned} \quad (5.3.3)$$

where sgn is the sign function. The offset b vanishes if class means are equidistant to the origin [Scholkopf and Smola \[2001\]](#).

This thesis focuses on a simple classification scheme, which assigns a previously unseen observation to the class with closest mean. The observations are considered in a different feature space, through their covariance matrices, which is not usual in signal processing for BCI. Most approaches favour more or less complex scheme relying on the estimation of covariance matrices, but consider only Euclidean metrics for the practical computations. The proposed approach is built upon Riemannian metrics and divergences, and their associated mean.

5.3.2 Means of Covariance Matrices

The covariance matrix of X which can be estimated with the sample covariance estimator as

$$\hat{\Sigma} = \frac{1}{N} X X^\top \quad (5.3.4)$$

is symmetric positive-definite (SPD). Other estimators seen in Section 5.2 are also producing SPD matrices. The properties of SPD matrices constrain them to a convex cone:

- (i) Symmetry: $\Sigma = \Sigma^\top$
- (ii) Positive definiteness: $x^\top \Sigma x > 0, \forall x \in \mathbb{R}^C \setminus 0$
- (iii) Strict positivity of diagonal elements: $\Sigma_{ij} > 0 | i = j, \forall i, j \in \{1, \dots, C\}$ i.e. positive variance.
- (iv) Cauchy-Schwarz inequalities: $|\Sigma_{ij}| \leq (\Sigma_{ii}\Sigma_{jj})^{1/2}, \forall i, j \in \{1, \dots, C\}$

The mean of SPD matrices can be computed as a centre of mass: given a set of covariance matrices $\{\Sigma_i\}_{i=1,\dots,I}$, the centre of mass $\bar{\Sigma}$ is the covariance matrix

that minimises the dispersion of matrices Σ_i :

$$\bar{\Sigma} = \mu(\Sigma_1, \dots, \Sigma_I) = \arg \min_{\Sigma \in \mathcal{M}_C} \sum_{i=1}^I d^p(\Sigma_i, \Sigma) , \quad (5.3.5)$$

where $d^p(\cdot, \cdot)$ can either be a distance ($p = 2$) between two matrices, or a divergence ($p = 1$).

In the literature, this mean is shown to have a unique solution and is at times designated as the geometric mean, *Cartan mean*, *Frechet mean*, or *Karcher mean*¹ [Ando et al. \[2004\]](#); [Lim and Pálfa \[2012\]](#). Depending on the divergence or distance used, different means can be defined from (5.3.5). Those considered in this study are presented below and summarised in Table 5.1.

Distances and Divergences

Divergences and distances are measures of dissimilarity between two points in a space. Here the Riemannian space will be considered. A distance function $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^+$ has the following properties for all $\Sigma_1, \Sigma_2, \Sigma_3 \in \mathcal{M}$:

- (i) $d(\Sigma_1, \Sigma_2) \geq 0$ non-negativity
- (ii) $d(\Sigma_1, \Sigma_2) = 0$ iff $\Sigma_1 = \Sigma_2$ identity
- (iii) $d(\Sigma_1, \Sigma_2) = d(\Sigma_2, \Sigma_1)$ symmetry
- (iv) $d(\Sigma_1, \Sigma_3) \leq d(\Sigma_1, \Sigma_2) + d(\Sigma_2, \Sigma_3)$ triangular inequality

Divergences are very similar to distances with the difference that properties (iii) and (iv) do not have to be satisfied. In the context of Covariance matrices, divergences and distances should both induce a Riemannian metric on the manifold of SPD matrices.

In this work, we consider several existing distances and their associated mean, namely the Euclidean distance, the Harmonic distance [Lim and Pálfa \[2012\]](#), the Affine-invariant Riemannian distance [Pennec et al. \[2006\]](#), the Log-Euclidean distance [Arsigny et al. \[2007\]](#), the Wasserstein distance [Villani \[2008\]](#), and divergences, such as the Kullback-Leibler divergence [Nielsen and Nock \[2009\]](#), the S-divergence [Sra \[2016\]](#), the α -divergence [Nielsen et al. \[2014\]](#), the Bhattacharyya divergence [Chebbi and Moakher \[2012\]](#).

¹This appellation has been recently criticised by Karcher himself [Karcher \[2014\]](#).

Euclidean Distance

The Euclidean distance between two matrices is represented by the Frobenius norm of their difference:

$$d_E(\Sigma_1, \Sigma_2) = \|\Sigma_1 - \Sigma_2\|_F = \sqrt{\text{tr}((\Sigma_1 - \Sigma_2)^T(\Sigma_1 - \Sigma_2))} . \quad (5.3.6)$$

In (5.3.5), this yields the arithmetic mean:

$$\bar{\Sigma}_E = \frac{1}{I} \sum_{i=1}^I \Sigma_i . \quad (5.3.7)$$

The arithmetic mean is drawn from a family of power means [Lim and Pálfa \[2012\]](#), defined as:

$$\Sigma_t = \left(\frac{1}{I} \sum_{i=1}^I \Sigma_i^t \right)^{\frac{1}{t}}, \quad t \in [-1, +1] . \quad (5.3.8)$$

and could be expressed as $\bar{\Sigma}_E = \Sigma_{t|t=1}$. From the same family, one can also define the *harmonic* mean as $\bar{\Sigma}_H = \Sigma_{t|t=-1}$.

We consider the arithmetic mean $\bar{\Sigma}_E$, as a baseline. This arithmetic mean is not adequate in the space of SPD matrices for two main reasons. First, the Euclidean distance and the arithmetic mean does not guarantee invariance under inversion known as *duality*, and thus could not guarantee that a matrix and its inverse are at the same distance from the identity matrix. Second, the arithmetic mean of covariance matrices leads to a *swelling effect*: the determinant of the arithmetic mean of SPD matrices can be larger than the determinant of its individual components. And since the determinant of a covariance matrix is a direct measure of the dispersion of the multivariate variable, the swelling effect introduces a large distortion of the data dispersion [Arsigny et al. \[2007\]](#). For these reasons, it is more appropriate to rely on a mean that adapt to the geometry of the SPD matrices.

Affine-Invariant Riemannian Distance

The convex cone of SPD matrices is a manifold that can be endowed with a Riemannian metric; such manifolds are called Riemannian manifold. The

affine-invariant Riemannian (AIR) distance between two points is defined by the length of the curve connecting these points on the Riemannian manifold known as the *geodesic* [Pennec et al. \[2006\]](#).

Let \mathcal{M} be a Riemannian manifold, and $T_\Sigma\mathcal{M}$ its tangent space defined on the point Σ . A Riemannian metric d is a family of inner product defined on the tangent space defined on each point Σ of the manifold. This inner product varies smoothly from point to point on the manifold,

$$d_\Sigma : T_\Sigma\mathcal{M} \times T_\Sigma\mathcal{M} \rightarrow \mathbb{R}$$

d is a function that assigns, for each point $\Sigma \in \mathcal{M}$, an inner product in the tangent space $T_\Sigma\mathcal{M}$. The Riemannian metric allows to compute the length of vectors or distance between two points on the tangent space.

The affine-invariant Riemannian distance is the distance between two points of a Riemannian manifold and is defined as:

$$d_{\text{AI}}(\Sigma_1, \Sigma_2) = \|\text{Log}(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2})\|_F = \left(\sum_{c=1}^C \log^2 \lambda_c \right)^{1/2}, \quad (5.3.9)$$

where Log is the matrix logarithm and λ_c , $c = 1, \dots, C$, are the eigenvalues of $\Sigma_1^{-1}\Sigma_2$. This expression is obtained by solving the geodesic equations on the space of SPD matrices.

Inserting (5.3.9) in (5.3.5) yields the mean $\bar{\Sigma}_{\text{AI}}$ associated to the affine-invariant Riemannian metric. It is the solution to:

$$\sum_{i=1}^I \text{Log}(\bar{\Sigma}_{\text{AI}}^{-1/2}\Sigma_i\bar{\Sigma}_{\text{AI}}^{-1/2}) = 0. \quad (5.3.10)$$

It has no closed form solution and can be solved iteratively through a gradient descent algorithm [Fletcher et al. \[2004\]](#).

This distance and mean are invariant to affine transformations. Some of these invariances are particularly interesting for the SPD matrices Let f be an affine-invariant Riemannian function defined on $\mathcal{M} \times \mathcal{M}$ (e.g. distance or mean), it displays the following properties:

- (i) *Invariance under congruent transformation, for any invertible matrix W*

$$f(\Sigma_1, \Sigma_2) = f(W^\top \Sigma_1 W, W^\top \Sigma_2 W) \quad (5.3.11)$$

(ii) *Invariance under inversion*

$$f(\Sigma, \mathbf{I}) = f(\Sigma^{-1}, \mathbf{I}) \quad (5.3.12)$$

implying self-duality

$$f(\Sigma_1, \Sigma_2) = f(\Sigma_1^{-1}, \Sigma_2^{-1}) \quad (5.3.13)$$

Another interesting property of the affine-invariant Riemannian metric is its invariance to left- and right-multiplication by a positive matrix [Arsigny et al. \[2007\]](#):

$$f(\Sigma_1, \Sigma_2) = f(\Sigma\Sigma_1, \Sigma\Sigma_2) = f(\Sigma_1\Sigma, \Sigma_2\Sigma) . \quad (5.3.14)$$

Log-Euclidean Distance

The Log-Euclidean is another distance that takes into consideration the topology of Riemannian manifold. It was introduced to alleviate the complexity involved in the computation of the affine-invariant Riemannian distance and its related mean [Arsigny et al. \[2007\]](#). The mean associated to the Log-Euclidean distance corresponds to an arithmetic mean in the domain of matrix logarithms. The distance between two SPD matrices is expressed as:

$$d_{\text{LE}}(\Sigma_1, \Sigma_2) = \|\text{Log}(\Sigma_1) - \text{Log}(\Sigma_2)\|_F , \quad (5.3.15)$$

and its associated mean is defined explicitly:

$$\bar{\Sigma}_{\text{LE}} = \text{Exp} \left(\frac{1}{I} \sum_{i=1}^I \text{Log}(\Sigma_i) \right) . \quad (5.3.16)$$

Unlike the affine-invariant Riemannian mean, the Log-Euclidean mean has a closed form expression, providing a large computational advantage. Moreover, the obtained mean is, to a large extent, similar to the affine-invariant Riemannian mean:

- (i) they have the same determinants which correspond to the geometric mean of the determinants of their matrices [\[Arsigny et al., 2007\]](#):

$$\det \bar{\Sigma}_{\text{LE}} = \det \bar{\Sigma}_{\text{AI}} = \prod_{i=1}^I (\det \Sigma_i)^{1/I} = \exp \left(\frac{1}{I} \sum_{i=1}^I \log(\det \Sigma_i) \right) ;$$

- (ii) they are similar, in some cases identical, and if not, $\text{tr } \bar{\Sigma}_{\text{LE}} > \text{tr } \bar{\Sigma}_{\text{AI}}$;
- (iii) Log-Euclidean mean has properties close to affine-invariance, *i.e.* similarity-invariance instead of congruent-invariance.

Wasserstein Distance

The Wasserstein distance, also known as the earth mover's distance, is a measure of distance between two probability distributions. It is the optimal cost of moving one probability distribution into another. If the two probability distributions are pictured as two different ways of piling up a mass of sand, then the Wasserstein distance can be seen as the optimal cost involved in transporting sand from one piling to another [Villani \[2008\]](#).

Let $P(\mathcal{X})$ and $P(\mathcal{Y})$ two spaces of probability measures, the optimal transport between two masses (or probability distributions) $\eta \in P(\mathcal{X})$ and $\nu \in P(\mathcal{Y})$ is defined as [Villani \[2008\]](#):

$$C(\eta, \nu) = \inf_{\pi \in \Pi(\eta, \nu)} \int c(x, y) d\pi(x, y), \quad (5.3.17)$$

where $\Pi(\eta, \nu)$ is the set of all joint probabilities on $\mathcal{X} \times \mathcal{Y}$; and $c(x, y)$ is the cost for transporting one unit of mass from x to y . In the Wasserstein distance, the cost $c(x, y)$ is defined as a distance. The Wasserstein distance of order p is defined as:

$$W_p(\eta, \nu) = \left(\inf_{\pi \in \Pi(\eta, \nu)} \int d(x, y)^p d\pi(x, y) \right)^{1/p}. \quad (5.3.18)$$

Following the development in [Barbaresco \[2011\]](#), the Wasserstein distance between multivariate Gaussian measures, with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 , which are noted $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$, is expressed as:

$$d_W(\mathcal{N}(\mu_1, \Sigma_1), \mathcal{N}(\mu_2, \Sigma_2)) = |\mu_1 - \mu_2|^2 + \text{tr } \Sigma_1 + \text{tr } \Sigma_2 - 2 \text{tr} \left(\left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \right). \quad (5.3.19)$$

Considering that $\mu_1 = \mu_2 = 0$, the Wasserstein distance between two covariance matrices is:

$$d_W(\Sigma_1, \Sigma_2) = \text{tr } \Sigma_1 + \text{tr } \Sigma_2 - 2 \text{tr} \left(\left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \right). \quad (5.3.20)$$

Bregman Divergences

Divergences have been considered for the computation of mean in applications of clustering and classification of SPD matrices due to the fact that they induce a Riemannian metric [Amari \[2010\]](#). Consider a strictly convex and differentiable function $f : \mathbb{R} \rightarrow \mathbb{R}$; then $f(x) > f(y) + f'(y)(x - y)$ and $f(x) = f(y) + f'(y)(x - y) \Leftrightarrow x = y$ for all $x, y \in \mathbb{R}$. The Bregman divergence [Bregman \[1967\]](#) is the difference between the left and right sides of the inequality:

$$D_f(x, y) = f(x) - f(y) - f'(y)(x - y) , \quad (5.3.21)$$

where f is called a *seed function*. It is shown that D_f verifies the non-negativity and the identity properties. When the seed function is quadratic, it can also be symmetric. The other properties of D_f are reported in [Bregman \[1967\]](#). Geometrically, the Bregman divergence can be seen as the measure of the difference between $f(x)$ and its representation on the plane tangent to f at y as illustrated on Fig. 5.1.

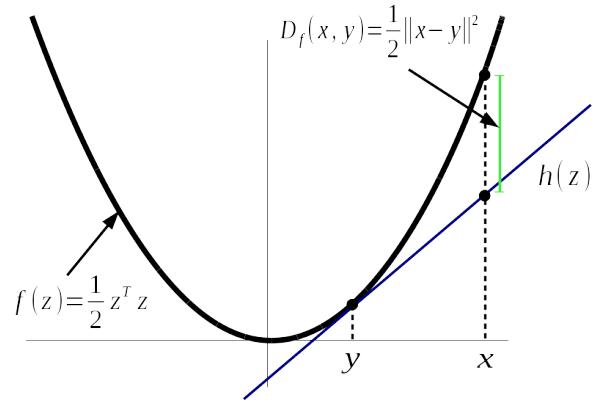


Figure 5.1: Geometry of the Bregman divergence with the seed function $f(z) = \frac{1}{2}z^T z$. $h(z)$ is a hyperplane tangent to $f(z)$ at y . While it accurately represents $f(y)$, it underestimates $f(x)$. The Bregman divergence measures how much the representation of $f(x)$ on $h(z)$ *diverges* from $f(x)$ (in green).

The scalar divergence can be directly adapted to SPD matrices as:

$$D_f(\Sigma_1, \Sigma_2) = \varphi(\Sigma_1) - \varphi(\Sigma_2) - \varphi'(\Sigma_2)(\Sigma_1 - \Sigma_2) , \quad (5.3.22)$$

where the seed function f is combined with a function $g : \mathcal{M} \rightarrow \mathbb{R}^C$ that maps an SPD matrix to a vector containing its eigenvalues: $\varphi = f \circ g$. g can also be the trace function, $g : \mathcal{M} \rightarrow \mathbb{R}$ that maps an SPD matrix to its trace. For convenience, $f \circ g$ will be referred to as $f(X)$ or $f(\Sigma)$ for matrices.

Depending on the seed function used, various divergences can be defined from the Bregman divergence. However, the mean induced by a Bregman divergence is independent of the seed function. It always correspond to the center of mass, i.e. the arithmetic mean [Nielsen and Nock \[2009\]](#).

Euclidean divergence

A first Bregman divergence could be defined from the Frobenius norm [Dhillon and Tropp \[2007\]](#), with $f(x) = \frac{1}{2}\|x\|_F^2$:

$$D_E(\Sigma_1, \Sigma_2) = \frac{1}{2}\|\Sigma_1 - \Sigma_2\|_F^2. \quad (5.3.23)$$

In the Euclidean case, this divergence is equivalent to the square distance and consequently the mean of SPD matrices based on the Euclidean divergence corresponds to their arithmetic mean, see Eq. (5.3.7).

Kullback-Leibler divergence

Using the *negative Shannon entropy* $f(x) = \sum_i x_i \log x_i$ yields the *Kullback-Leibler* divergence [Nielsen and Nock \[2009\]](#). It is also known as the *relative entropy* or *discrimination information*. The Kullback-Leibler divergence was introduced in information theory to measure the difference between two probability distributions over the same alphabet. Given a set $\mathfrak{X} = \{x, X, P\}$, where:

- $x \in \mathbb{R}^C$ is a variable,
- $X \in \mathbb{R}^{C \times N}$ is the set of all possible values of x , i.e. the *alphabet*, and
- P is the probability distribution of x over X

The Kullback-Leibler divergence measure the different between $P_1(x)$ and $P_2(x)$, both defined over X :

$$D_{KL}(P_1(x), P_2(x)) = \sum_i^N P_1(x_i) \log \frac{P_1(x_i)}{P_2(x_i)} \quad (5.3.24)$$

if both distribution are normal,

$$P(x) = \mathcal{N}(\mu, \Sigma) = \frac{1}{(2\pi)^{C/2} \det(\Sigma)^{1/2}} \left\{ \exp \left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right) \right\} \quad (5.3.25)$$

(5.3.25) in (5.3.24),

$$D_{\text{KL}}(P_1(x), P_2(x)) = \frac{1}{2} \left((\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) - \log \det(\Sigma_2^{-1} \Sigma_1) + \text{tr}(\Sigma_2^{-1} \Sigma_1) - C \right). \quad (5.3.26)$$

When $P_1(x)$ and $P_2(x)$ are zero-centered, (5.3.26) becomes:

$$D_{\text{KL}}(P_1(x), P_2(x)) = \frac{1}{2} \left(\log \det(\Sigma_1^{-1} \Sigma_2) + \text{tr}(\Sigma_2^{-1} \Sigma_1) - C \right) \quad (5.3.27)$$

The Kullback-Leibler divergence correspond to the Bregman divergence of covariance matrices with the seed function $f(\Sigma) = -\log \det(\Sigma)$:

$$D_{\text{KL}}(P_1(x), P_2(x)) = D_{\text{KL}}(\Sigma_1, \Sigma_2) = \frac{1}{2} \left(\log \det(\Sigma_1^{-1} \Sigma_2) + \text{tr}(\Sigma_2^{-1} \Sigma_1) - C \right). \quad (5.3.28)$$

S-divergence

An example of a symmetric divergence is the S-divergence. It is obtained from the *Jensen-Shannon* divergence which is a symmetrised Bregman divergence:

$$\begin{aligned} D_{\text{J-S}}(\Sigma_1, \Sigma_2) &= \frac{1}{2} \left(D_f(\Sigma_1, \frac{\Sigma_1 + \Sigma_2}{2}) + D_f(\frac{\Sigma_1 + \Sigma_2}{2}, \Sigma_2) \right) \\ &= \frac{1}{2} (\text{tr } f(\Sigma_1) + \text{tr } f(\Sigma_2)) - \text{tr } f(\frac{\Sigma_1 + \Sigma_2}{2}). \end{aligned} \quad (5.3.29)$$

The S-divergence is obtained by using the logarithmic barrier function for the positive-definite cone $f(\Sigma) = -\log \det(\Sigma)$ as seen in $D_{\text{J-S}}$, and the S-divergence between two SPD matrices corresponds to the Bhattacharyya divergence between them [Sra \[2016\]](#):

$$D_S(\Sigma_1, \Sigma_2) = \log \det(\frac{\Sigma_1 + \Sigma_2}{2}) - \frac{1}{2} \log \det(\Sigma_1 \Sigma_2). \quad (5.3.30)$$

Despite its symmetry, S-divergence is not a metric: it does not satisfy the triangular inequality criterion. However, its squared root has been shown to be a distance [Sra \[2016\]](#).

Other symmetric divergences can be obtained in the same fashion; for instance, the *Jeffreys divergence* which is a symmetrised Kullback-Leibler divergence: $D_J(\Sigma_1, \Sigma_2) = D_{KL}(\Sigma_1, \Sigma_2) + D_{KL}(\Sigma_2, \Sigma_1)$ [Sra \[2016\]](#).

Log-det α -divergence

Another family of divergence is defined when the right- and left-sided divergences are mixed in a weighted manner. One such family is the α -divergence [Nielsen et al. \[2014\]](#), and it is defined in this work by [Chebbi and Moakher \[2012\]](#):

$$D_f^\alpha(\Sigma_1, \Sigma_2) = \frac{4}{1-\alpha^2} \left[\frac{1-\alpha}{2} f(\Sigma_1) + \frac{1+\alpha}{2} f(\Sigma_2) - f\left(\frac{1-\alpha}{2}\Sigma_1 + \frac{1+\alpha}{2}\Sigma_2\right) \right], \alpha^2 \neq 1 \quad (5.3.31)$$

D_f^α can be expressed in terms of Bregman divergence as:

$$D_f^\alpha = \frac{4}{1-\alpha^2} \left[\frac{1-\alpha}{2} D_f\left(\Sigma_1, \frac{1-\alpha}{2}\Sigma_1 + \frac{1+\alpha}{2}\Sigma_2\right) + \frac{1+\alpha}{2} D_f\left(\Sigma_2, \frac{1-\alpha}{2}\Sigma_1 + \frac{1+\alpha}{2}\Sigma_2\right) \right], \alpha^2 \neq 1 \quad (5.3.32)$$

α -divergences at $\alpha = \pm 1$ are obtained through the limit values $\lim_{\alpha \rightarrow \pm 1} D_f^\alpha$.

Using the logarithmic-barrier function yields:

$$\begin{aligned} D_{LD}^\alpha(\Sigma_1, \Sigma_2) &= \frac{4}{1-\alpha^2} \log \det \left(\frac{1-\alpha}{2} (\Sigma_1 \Sigma_2^{-1})^{\frac{1+\alpha}{2}} + \frac{1+\alpha}{2} (\Sigma_2 \Sigma_1^{-1})^{\frac{1-\alpha}{2}} \right), \quad -1 < \alpha < 1 \\ D_{LD}^1(\Sigma_1, \Sigma_2) &= \text{tr}(\Sigma_2^{-1} \Sigma_1 - \mathbf{I}) - \log \det(\Sigma_2^{-1} \Sigma_1) \\ D_{LD}^{-1}(\Sigma_1, \Sigma_2) &= \text{tr}(\Sigma_1^{-1} \Sigma_2 - \mathbf{I}) - \log \det(\Sigma_1^{-1} \Sigma_2). \end{aligned} \quad (5.3.33)$$

D_{LD}^1 and D_{LD}^{-1} are right- and left-sided Bregman divergences respectively. At $\alpha = 0$, the log-det α divergence yields a symmetric divergence corresponding to the Bhattacharyya divergence [Chebbi and Moakher \[2012\]](#); [Sra \[2016\]](#).

All these distances and divergences are summed up in Table 5.1.

5.3.3 Minimum Distance to Mean Classifier for SSVEP

The considered classifier is described in section 5.3.1. It is given the name *Minimum Distance to Mean* or MDM, and was inspired by [\[Barachant et al., 2012a\]](#) where it is limited to Riemannian mean. The covariance matrices of

	Distance/Divergence	Mean	References
Euclidean	$d_E(\Sigma_1, \Sigma_2) = \ \Sigma_1 - \Sigma_2\ _F$	$\bar{\Sigma}_E = \frac{1}{I} \sum_{i=1}^I \Sigma_i$	
Harmonic	$d_H(\Sigma_1, \Sigma_2) = \ \Sigma_1^{-1} - \Sigma_2^{-1}\ _F$	$\bar{\Sigma}_H = \left(\frac{1}{I} \sum_{i=1}^I \Sigma_i^{-1} \right)^{-1}$	Lim and Pálfa [2012]
Log-Euclidean	$d_{LE}(\Sigma_1, \Sigma_2) = \ \log(\Sigma_1) - \log(\Sigma_2)\ _F$	$\bar{\Sigma}_{LE} = \exp \left(\frac{1}{I} \sum_{i=1}^I \log(\Sigma_i) \right)$	Arsigny et al. [2007]
Affine-invariant	$d_{AI}(\Sigma_1, \Sigma_2) = \ \log(\Sigma_1^{-1} \Sigma_2)\ _F$	Algorithm 3 in Fletcher et al. [2004]	Fletcher et al. [2004]
Kullback-Leibler	$D_{KL}(\Sigma_1, \Sigma_2) = \frac{1}{2} (\log \det(\Sigma_1^{-1} \Sigma_2) + \text{tr}(\Sigma_2^{-1} \Sigma_1) - C)$	$\bar{\Sigma}_{KL} = \frac{1}{I} \sum_{i=1}^I \Sigma_i$	Moakher [2005]; Fletcher et al. [2004]
S-divergence	$D_S(\Sigma_1, \Sigma_2) = \log \det \left(\frac{\Sigma_1 + \Sigma_2}{2} \right) - \frac{1}{2} \log \det(\Sigma_1 \Sigma_2)$	Eq. (17-20) in Cherian et al. [2011]	Chebbi and Moakher [2012]; Kang et al. [2009]
α -divergence	$D_{LD}^\alpha(\Sigma_1, \Sigma_2)$ from Eq. (5.3.33)	Algorithm 1 in Chebbi and Moakher [2012]	Sra [2016]; Cherian et al. [2011]
Bhattacharyya	$D_B(\Sigma_1, \Sigma_2) = \left(\log \frac{\det \frac{1}{2}(\Sigma_1 + \Sigma_2)}{(\det(\Sigma_1) \det(\Sigma_2))^{1/2}} \right)^{1/2}$	Algorithm 1 in Chebbi and Moakher [2012]	Chebbi and Moakher [2012]
Wasserstein	$d_W = \text{tr} \left(\Sigma_1 + \Sigma_2 - 2 \left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \right)$	Eq. 6.2 in Agueh and Carlier [2011]	Nielsen and Bhatia [2012]; Chebbi and Moakher [2012]
			Agueh and Carlier [2011]; Barbaresco [2011]

Table 5.1: Distances, divergences and means considered in the experimental study.

EEG trials are classified based on their distance to the centres of the classes (i.e. means or centroids). To embed frequency information in the covariance matrices, we use a construction of matrices proposed in [Congedo et al., 2013]. Let $X \in \mathbb{R}^{C \times N}$ be an EEG trial measured on C channels and N samples in an SSVEP experiment with F stimulus blinking at different frequencies. The covariance matrices are estimated from a modified version of the input signal X :

$$X \in \mathbb{R}^{C \times N} \rightarrow \begin{bmatrix} X_{\text{freq}_1} \\ \vdots \\ X_{\text{freq}_F} \end{bmatrix} \in \mathbb{R}^{FC \times N}, \quad (5.3.34)$$

where X_{freq_f} is the input signal X band-pass filtered around frequency freq_f , $f = 1, \dots, F$. Thus the resulting covariance matrix Σ belongs to \mathcal{M}_{FC} . Henceforth, all SSVEP EEG signals will be considered as filtered and modified by Eq. (5.3.34).

For ERP paradigm with a number E of different ERPs, the modified signal is the concatenation of the original signal and the grand averages of trials containing the target ERPs \bar{X}_e , $e = 1, \dots, E$:

$$X \in \mathbb{R}^{C \times N} \rightarrow \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_E \\ X \end{bmatrix} \in \mathbb{R}^{(E+1)C \times N}, \quad (5.3.35)$$

The resulting covariance matrix will be of size $((E + 1) \times C)^2$.

For SSVEP classification, $K = F+1$ classes are considered: one class for each target frequency, and one for the resting state. As described in Algorithm 1, from I labelled training trials $\{X_i\}_{i=1}^I$ recorded per subject, K centres of classes $\bar{\Sigma}^{(k)}$ are estimated (step 3). In this step, outliers matrices are removed to have a reliable mean estimation [Barachant et al., 2013a]. A new unlabelled test trial Y is predicted to belong to the class whose mean $\bar{\Sigma}^{(k)}$ is the closest to the trial covariance matrix, w.r.t. one of the distances from Table 5.1 (step 5).

Algorithm 1 Minimum Distance to Mean Classifier

Inputs: $X_i \in \mathbb{R}^{FC \times N}$, for $i = 1, \dots, I$, a set of labelled EEG trials.

Inputs: $\mathcal{I}(k)$, a set of indices of trials belonging to class k .

Input: $Y \in \mathbb{R}^{FC \times N}$, an unlabelled test EEG trial.

Output: k^* , the predicted label of Y .

- 1: Compute covariance matrices Σ_i of X_i
 - 2: **for** $k = 1$ **to** K **do**
 - 3: Compute centre of class : $\bar{\Sigma}^{(k)} = \mu(\Sigma_i : i \in \mathcal{I}(k))$
 - 4: **end**
 - 5: Compute covariance matrix Σ of Y , and classify it : $k^* = \arg \min_k d(\Sigma, \bar{\Sigma}^{(k)})$
 - 6: **return** k^*
-

where X_{freq_f} is the input signal X band-pass filtered around frequency freq_f , $f = 1, \dots, F$.

From I labelled training trials $\{X_i\}_{i=1}^I$ recorded per subject, K centres of class $\Sigma_\mu^{(k)}$ are estimated using Algorithm 2. When an unlabelled test trial Y is given, it is classified as belonging to the class whose centre $\Sigma_\mu^{(k)}$ is the closest to the trial's covariance matrix (Algorithm 1, step 5).

5.4 Online Classification

5.4.1 Curve-Based Online Classification

In offline synchronous BCI paradigm, cue onset is used as reference for the localisation of a brain response, e.g. an evoked potential. Nonetheless most of the BCI applications are online and asynchronous; cue onsets are not known,

Algorithm 2 Offline Estimation of Riemannian Centres of Classes

Inputs: $X_i \in \mathbb{R}^{FC \times N}$, for $i = 1, \dots, I$, a set of labelled trials.

Inputs: $\mathcal{I}(k)$, a set of indices of trials belonging to class k .

Output: $\Sigma_{\mu}^{(k)}$, $k = 1, \dots, K$, centres of classes.

```
1: Compute covariance matrices  $\hat{\Sigma}_i$  of  $X_i$ 
2: for  $k = 1$  to  $K$  do
3:    $\Sigma_{\mu}^{(k)} = \mu(\hat{\Sigma}_i : i \in \mathcal{I}(k))$  , Eq. (5.3.5)
4: end
5: return  $\Sigma_{\mu}^{(k)}$ 
```

Algorithm 3 Curve-based Online Classification

Inputs: hyper-parameters w , Δn , D , and ϑ .

Inputs: $\Sigma_{\mu}^{(k)}$, $k = 1, \dots, K$, centres of classes from Algorithm 2 (offline training).

Inputs: Online EEG recording $\chi(n)$.

Output: $\tilde{k}(n)$, online predicted class.

```
1:  $d = 1$ 
2: for  $n = w$  to  $\mathcal{N}$  step  $\Delta n$ 
3:   Epoch  $X_d$ , Eq. (5.4.1), and classify it with Algorithm 1
4:   if  $d \geq D$ 
5:     Find the most recurrent class in  $\mathcal{K} = k_{j \in \mathcal{J}(d)}^*$ :  $\bar{k} = \arg \max_k \rho(k)$ ,
      Eq. (5.4.2)
6:     if  $\rho(\bar{k}) > \vartheta$ 
7:       Compute  $\tilde{\delta}_{\bar{k}}$ , Eq. (5.4.3)
8:       if  $\tilde{\delta}_{\bar{k}} < 0$ 
9:         return  $\tilde{k} = \bar{k}$ 
10:    end
11:   end
12:   end
13:    $d = d + 1$ 
14: end
```

thus designing online version of BCI algorithms is not a trivial task. The approach introduced here identifies a period (*i.e.* time interval) in the online EEG $\chi \in \mathbb{R}^{FC \times N}$, where N is the number of recorded samples, associated with a high probability (above the threshold) of observing an SSVEP at a specific frequency, as illustrated in Algorithm 3.

To locate this interval, we focus on the last D recorded EEG overlapping epochs $\{X_j \in \mathbb{R}^{FC \times w}\}_{j \in \mathcal{J}(d)}$, with the set of indices $\mathcal{J}(d) = d-D+1, \dots, d-1, d$; where d is the index of the current epoch X_d in the online recording $\chi(n)$. Epochs have size w , and the interval between two consecutive epochs is Δn , with $w > \Delta n$:

$$X_d = \chi(n - w, \dots, n). \quad (5.4.1)$$

To obtain the first D epochs $X_{j \in \mathcal{J}(d)}$, at least $w + (D - 1)\Delta n$ samples of χ should be recorded (step 4).

The classification outputs $k_{j \in \mathcal{J}(d)}^*$ obtained in step 3 by applying Algorithm 1 on $X_{j \in \mathcal{J}(d)}$ are stored in a vector \mathcal{K} , which always contains the latest D classification outputs. The class that occurs the most in \mathcal{K} (step 5), with an occurrence probability $\rho(k)$ above a defined threshold ϑ , is considered to be the class, denoted \bar{k} , of the ongoing EEG recording $\chi(n)$. The vector ρ is defined as:

$$\rho(k) = \frac{\#\{k_{j \in \mathcal{J}(d)}^* = k\}}{D}, \text{ for } k = 1, \dots, K, \quad (5.4.2)$$

with $\bar{k} = \arg \max_k \rho(k)$; then $\rho(\bar{k})$ is compared to the threshold ϑ . If ϑ is not reached within the last D epochs, the classification output is held back, and the sliding process continues until ϑ it is reached. In the last D epochs, once a class \bar{k} has been identified, a curve direction criterion is introduced to enforce the robustness of the result. For class \bar{k} to be validated, this criterion requires that the direction taken by the displacement of covariance matrices $\hat{\Sigma}_{j \in \mathcal{J}(d)}$ be toward the centre of class $\Sigma_{\mu}^{(\bar{k})}$. Hence $\tilde{\delta}_{\bar{k}}$, the sum of gradients (*i.e.* differentials) of the curve made by distances from $\hat{\Sigma}_{j \in \mathcal{J}(d)}$ to $\Sigma_{\mu}^{(\bar{k})}$ should be negative (step 8):

$$\begin{aligned} \tilde{\delta}_{\bar{k}} &= \sum_{j \in \mathcal{J}(d)} \frac{\Delta \delta_{\bar{k}}(j)}{\Delta j} = \sum_{j=d-D+2}^d \delta_{\bar{k}}(j) - \delta_{\bar{k}}(j-1) < 0 \\ \text{with } \delta_{\bar{k}}(j) &= \frac{\delta(\hat{\Sigma}_j, \Sigma_{\mu}^{(\bar{k})})}{\sum_{k=1}^K \delta(\hat{\Sigma}_j, \Sigma_{\mu}^{(k)})}. \end{aligned} \quad (5.4.3)$$

The occurrence criterion is inspired by the dynamic stopping of [Verschore et al., 2012]; there is no fixed trial length for classification. The occurrence criterion ensures that the detected user intention is unaffected by any short time disturbances due to noise or subject inattention, as presented in Algorithm 3. This approach offers a good compromise to obtain robust results within a short and flexible time.

The curve direction criterion solves both the problems of latency in the EEG synchronisation and of the delays inserted by the EEG epochs processing. Indeed, some EEG epochs gather signals from different classes and might be wrongfully classified if the decision is solely based on the distance with the centre of the class. This situation and the effect of the curve direction criterion are well shown in Section 5.5.4. Ensuring that the covariance matrices are displaced toward the centre of the detected class provides a guaranty that it matches with the current EEG state. Inversely, if the direction of the curve is moving away from the centre of the detected class, it might indicate that there have been a change in the EEG state that has not been detected.

The Algorithm 3 has four hyperparameters: w , Δn , D , and ϑ . The values of w , D , and ϑ are set through cross validation and are given in Section 5.5.4. Although a large window size w is expected to increase the classification accuracy, it increases the response time, thus reducing the time resolution, and extends the overlap between different EEG states. The step size Δn should be set to a minimum value to allow a maximum number of overlapping epochs (D) within a short time. However, it should be large enough to avoid too many calculations within a time interval with small or nonexistent changes in EEG states. If the number of the epoch D is too small, the classification will be sensitive to non-intentional and abrupt changes in the EEG. A too large D will increase the momentum and reinforce the influence of the past EEG signals. It should also be mentioned that both the occurrence and the curve direction criteria cannot have a significant impact if the value of D is too small. The probability threshold parameter ϑ acts like a rejection parameter: high ϑ values correspond to high rejection rate.

5.4.2 Outliers Removal with Riemannian Potato

Outliers in the training data might affect the Riemannian mean of classes in the MDRM classification scheme. To alleviate this effect, an approach called the Riemannian potato, introduced in [Barachant et al., 2013a], is exploited. In this approach, all trials are represented by their covariance matrices Σ_i . A reference covariance matrix is estimated, e.g. Riemannian mean of all trials Σ_μ . The Riemannian distances δ_i between each Σ_i and Σ_μ are computed. Any trial that lies too far, *i.e.* beyond a certain threshold, from the reference matrix Σ_μ in terms of Riemannian distance is rejected. In [Barachant et al., 2013a], the distance z-score thresholding is defined as:

$$z(\delta_i) = \frac{\delta_i - \mu}{\sigma} > z_{th} \quad (5.4.4)$$

where μ and σ are respectively the mean and standard deviation of distances $\{\delta_i\}_{i=1}^I$. In other words, any trial Σ_i whose z-score $z(\delta_i)$ is larger than the threshold $z_{th} = 2.5$ is rejected.

In this work, we propose a slightly different application of the Riemannian potato where the outliers are removed per class. Hence for K class, K Riemannian potatoes are defined $\{\Sigma_\mu^k, \mu^k, \sigma^k\}_{k=1}^K$. Since Riemannian distances to geometric mean do not have a Gaussian distribution, we make use of the geometric mean for μ , the geometric standard deviation for σ and the geometric z-score. They are defined as follows [Congedo, 2013]:

$$\begin{aligned} \mu^k &= \exp \left(\frac{1}{I} \sum_i \ln(\delta_i^k) \right) \\ \sigma^k &= \exp \left(\sqrt{\frac{1}{I} \sum_i (\ln(\delta_i^k / \mu^k))^2} \right) \\ z(\delta_i^k) &= \frac{\ln(\delta_i^k / \mu^k)}{\ln(\sigma^k)}. \end{aligned} \quad (5.4.5)$$

Through cross-validation, the z-score threshold is set to $z_{th} = 2.2$. Moreover, outliers are removed iteratively. Each time outliers are rejected, a new centre of class is computed and used as reference for the next iteration. The iterations continue until convergence, *i.e.* no more outlier found.

5.5 Experimental Validation

5.5.1 Covariance Estimators Comparison

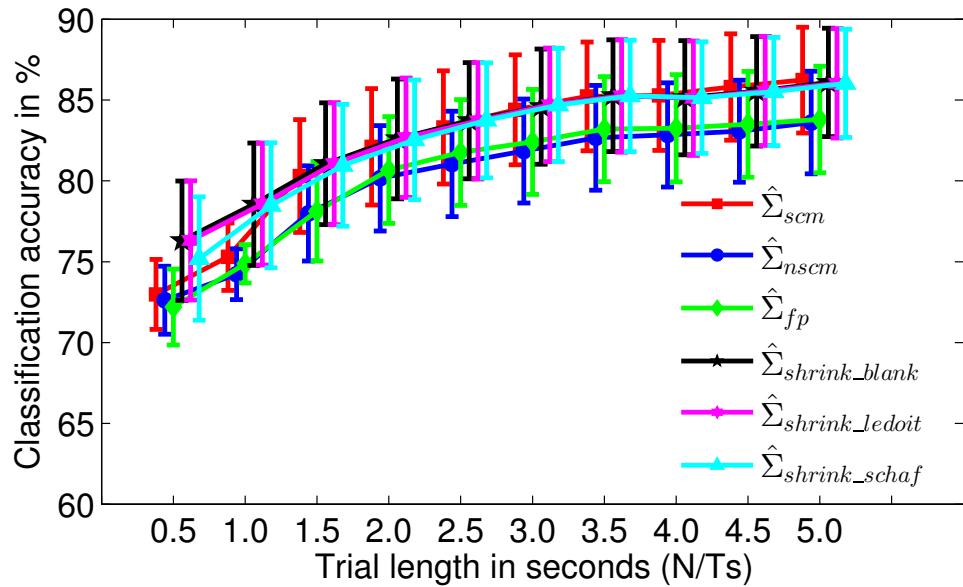


Figure 5.2: Comparison of covariance estimators in terms of classification accuracy obtained with MDRM with increasing EEG trial length. For each trial length, the average accuracy across all subjects and across all replication is shown. Bars indicate the error of the mean, i.e. standard deviation divided by the square root of $n - 1$, n = number of samples.

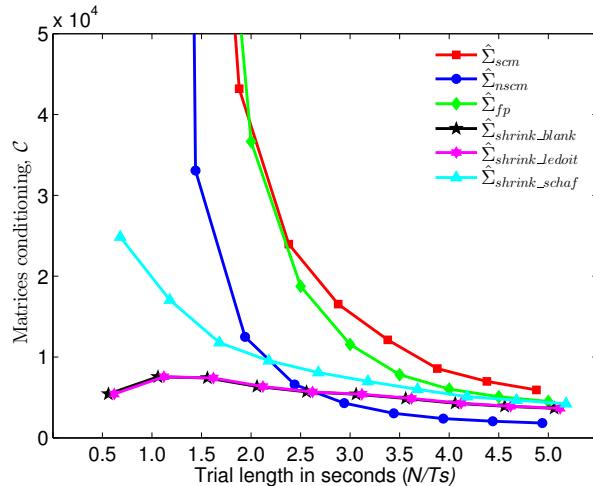
In this section, the effectiveness of covariance matrix estimators is evaluated for SSVEP signals. The evaluation is done in terms of classification accuracy and integrated discrimination improvement (IDI), obtained by each estimator (see Section 5.2) with respect to the SCM estimator while using the offline MDRM classifier. The different conditioning of covariance matrices are also investigated.

A bootstrapping with 1000 replications is performed to assess the performances of each estimator. Estimators are compared on 10 trial lengths $t \in \{0.5, 1.0, \dots, 5.0\}$ seconds, as these are known to affect the estimators performance. Here $N \in \{128, 256, \dots, 1280\}$ is computed as $N = t \times T_s$.

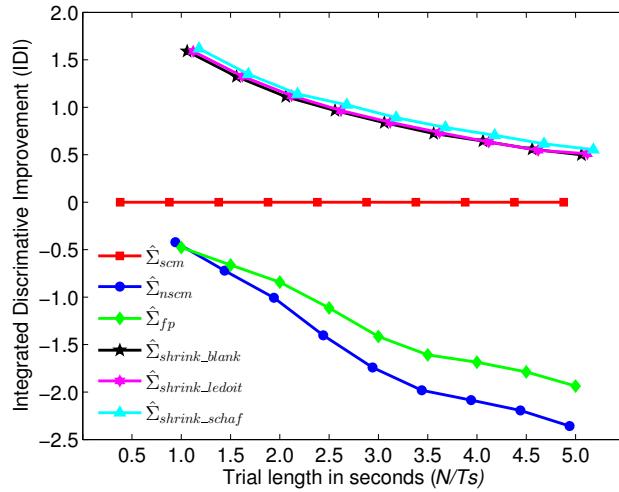
Figure 5.2 shows the classification accuracy of each estimator. The increase in the accuracy can be attributed to the fact that the relevant patterns in EEG accumulate with the trial length, producing better estimation of the covariance matrices. This is known to be particularly true for the SCM estimator and it could be seen in Figure 5.2. It appears that shrinkage estimators (especially Ledoit and Schäfer) are less affected by the reduction of epoch sizes than the other estimators. This is a direct consequence of the regularisation between the sample covariance matrices and the targeted (expected) covariance matrix of independent variables.

For computational purposes, it is important to look at the matrix conditioning. Figure 5.3(a) shows the ratio \mathcal{C} between the largest and smallest eigenvalues: in well-conditioned matrices, \mathcal{C} is small. Shrinkage estimators offer better conditioned matrices whereas the SCM, NSCM, and Fixed Point matrices are ill-conditioned below two seconds of trial length, and may result in singular matrices.

On Figure 5.3(b), the Integrated Discrimination Improvement (IDI), as defined in Pencina et al. [2008], is computed for the different estimators and trial lengths. The SCM is used as a reference for improvement, as this is the most popular estimator in the literature. Negative IDI means a deterioration in the method discrimination ability. It is clear that shrinkage estimators increase the discrimination power of the classifier. However, despite being more complex than the SCM, the NSCM and the Fixed Point estimators decrease the discrimination ability of classifiers. From Figures 5.2 and 5.3(b), it is apparent that the difference in performance between the SCM and shrinkage estimators reduces as the trial length increases. The simplicity of the SCM plays a favourable role: it is an attractive method for longer trials. The p -values under the hypothesis that there is no improvement (*i.e.* $IDI = 0$) from one estimator to another are all inferior to 10^{-47} , ($p < 10^{-3}$ indicating a statistically significant discriminatory improvement); hence the improvement is significant. It should be noted that



(a)



(b)

Figure 5.3: (a) Covariance matrices condition expressed as the ratio \mathcal{C} between largest and smallest eigenvalues for the different covariance estimators. The comparison is made for increasing EEG trial length. (b) Integrated discrimination improvement brought to the classification task by various estimators along varying trail length. The indicated IDI values are multiplied by 10^2 . $\hat{\Sigma}_{scm}$ is used as a baseline.

the estimation of covariance matrices is a trade-off between the quality of the estimate and the computation time required; this should be considered for real time processing.

5.5.2 Effect of Outliers on Centre Estimation

Outliers can affect the offline training of the K centres of class $\Sigma_{\mu}^{(k)}$ by Algorithm 2, which is crucial for the evaluation phase and online application. Figure 5.4 shows representations of training covariance matrices Σ_i in the tangent space (Θ_i), projected at the mean of all training trials, for the subjects with the lowest (5.4(a) and 5.4(b)) and the highest (5.4(c) and 5.4(d)) BCI performance. To obtain this visualisation, the first two principal components of a PCA applied on $\{\Theta_i\}_{i=1}^I$ are selected. In Figures 5.4(b) and 5.4(d), the Riemannian potato presented in Section 5.4.2 is applied; outliers in each class are removed. The interest of using a Riemannian potato is well seen in Figure 5.4(a) and 5.4(b). In 5.4(a), the outliers are so distant from the rest of the class matrices that the centre of class is stretched away. Applying a Riemannian potato removes the outliers, and the centre of class is better estimated (5.4(b)).

When training trials are not noisy, their covariance matrices are compact around their Riemannian mean. In this case the removal of outliers by the Riemannian potato does not influence, at least not significantly, the Riemannian mean. This is the case in Figure 5.4(c) and 5.4(d). Thus, applying the Riemannian potato is crucial for noisy data and will have a limited effect on clean data. The impact of the Riemannian potato on the classification accuracy is discussed in Section 5.5.4.

5.5.3 From Euclidean to Riemannian Centres of Class

The covariance matrices obtained from SSVEP data extended with Eq. (5.3.34) have interesting features, allowing the discrimination between signals of identical sources but with different frequencies. Figure 5.5 shows the K classes mean covariance matrices $\bar{\Sigma}^{(k)}$ from subjects with the highest (a) and lowest (b) classification accuracy. The three 8×8 diagonal blocks hold the covariance matrices of the $F = 3$ target frequencies. Inter-frequencies covariance blocks are almost null. In each mean covariance matrix, the block holding the covariance of the target frequency has the largest values. For the resting class, all F blocks tend to have similar and small values. These features are more visible in the subject with the highest classification accuracy, and less visible in the one with lowest classification accuracy. It is interesting to see that features used for classification

have a physiological meaning allowing an intuitive understanding, as opposed to *black-boxes* approaches such as LDA or SVM. EEG processing complexity is encoded by the distance and not by machine learning.

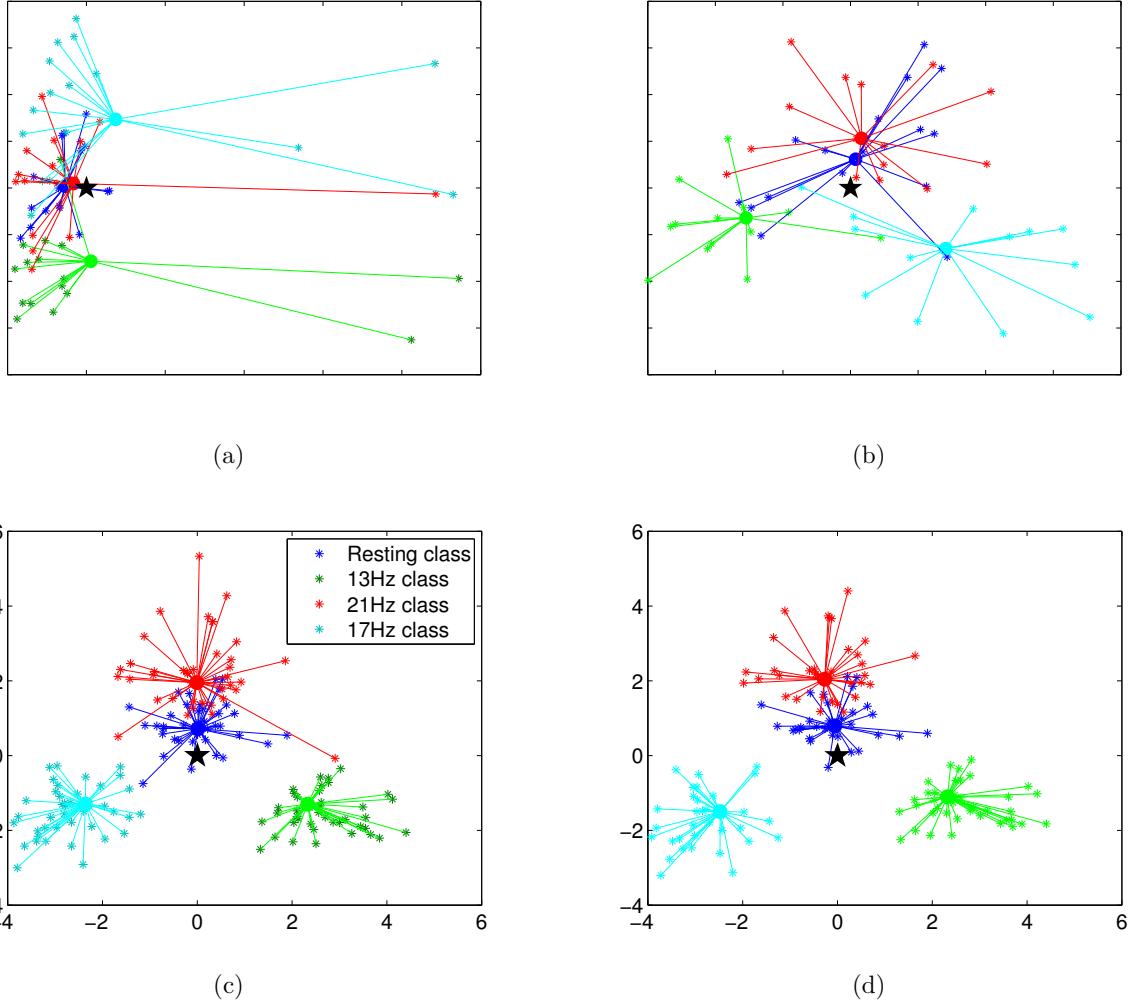


Figure 5.4: Scatter plot of covariance matrices for all trials mapped on the tangent space. The distance between each trial covariance matrix Σ_i and its Riemannian mean class $\Sigma_\mu^{(k)}$ is shown as connection line. The black star represents the Riemannian mean of all trials. Subject with lowest BCI performance, (5.4(a)) before and (5.4(b)) after Riemannian potato filtering. Subject with highest BCI performance, (5.4(c)) before and (5.4(d)) after Riemannian potato filtering.



Figure 5.5: Representation of covariance matrices: each image is the covariance matrix mean $\bar{\Sigma}^{(k)}$ of the class k , for one session of the recording. The diagonal blocks show the covariance in different frequency bands, i.e. 13 Hz in the upper-left block, 21 Hz in the middle, and 17 Hz in the bottom-right. Subjects with highest (a) and lowest (b) BCI performance.

Based on those covariance matrices, the different distances and means of Table 5.1 are compared in terms of classification accuracy and average CPU time elapsed on a trial classification, which involves the computation of four means of class and a distance to each mean. Table 5.2 summarises results obtained for each subject and each distance/divergence. Euclidean distance yields drastically low accuracy. This support the fact that using Euclidean distance and Arithmetic mean on SPD matrices is not appropriate. This is generally attributed to the invariance under inversion and the fact that the determinant of the Arithmetic mean of SPD matrices can be larger than the determinant of its parts; it is referred to as the *swelling effect*. Since the value of the determinant is a direct measure of dispersion of the multivariate variables (i.e. EEG channels and frequency bands), it leads to poor discrimination in the classification task. The swelling effect of Arithmetic mean is shown in Figure 5.6(a): the determinant of the Arithmetic mean is strictly larger than other means, the Log-Euclidean, Affine-Invariant and Bhattacharyya ones yielding similar determinants, close to trial values. Another observation is that the Bhattacharyya distance and the S-divergence yield similar results. In the S-divergence section, it was stipulated the square root of the S-Divergence was a distance, and it

is seen here that it correspond to the Bhattacharyya distance.

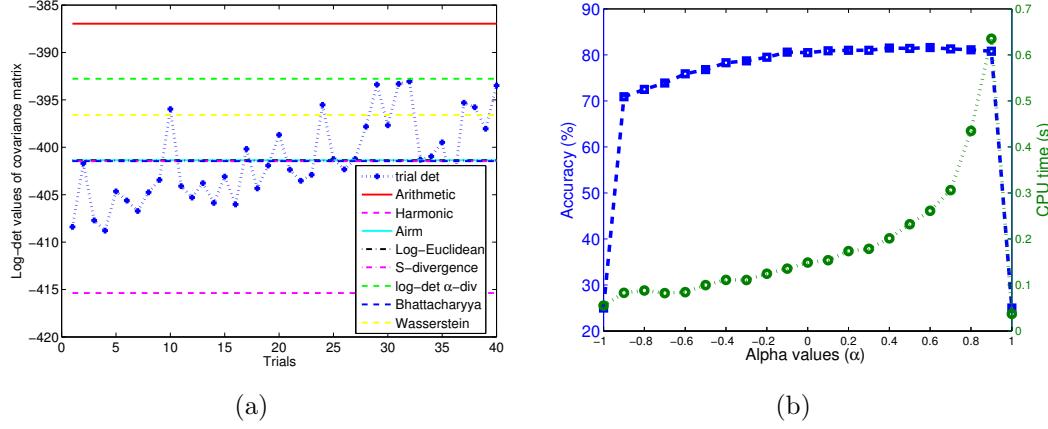


Figure 5.6: (a): Swelling effect of Arithmetic mean shown through log-determinant values. Training trials are taken from the 13 Hz class of the subject with the highest BCI performance. Log-determinant values are given for each trial covariance (points), and for means of Table 5.1 (horizontal lines). (b): Classification accuracy and CPU time, obtained with α -divergence for $-1 \leq \alpha \leq 1$.

Riemannian distances significantly improve classification performances, with α -divergence yielding the best results (81.56%). The value of α was set to 0.6 through cross-validation. This procedure lasted 225.42 seconds and makes α -divergence the most costly method, due to the optimisation of its parameter α . Log-Euclidean yields lower classification accuracy (average 78.98%) but could be computed faster than α -divergence or Affine-Invariant distance. However, the Bhattacharyya distance has the lowest computational cost of the considered Riemannian distances (average CPU time 0.140s), with a higher average accuracy of 80.51%. So, it is a good trade-off between efficiency and speed. The accuracy and CPU time of the α -divergence at different values of α are shown in Figure 5.6(b). It is seen that for $\alpha = \pm 1$, where α -divergence represents a Bregman divergence associated with the log-determinant function, the classification accuracy are at the lowest accuracy (25%). For all other values of alpha, the expected accuracy is $78.85 \pm 3.3\%$ and one can choose $-1 < \alpha < 1$ without any major impact on classification results. This experiment on real EEG data shows that it is crucial to process covariance matrices with dedicated

Riemannian tools, impacting the efficiency of the classification.

5.5.4 Classification Results and Analysis

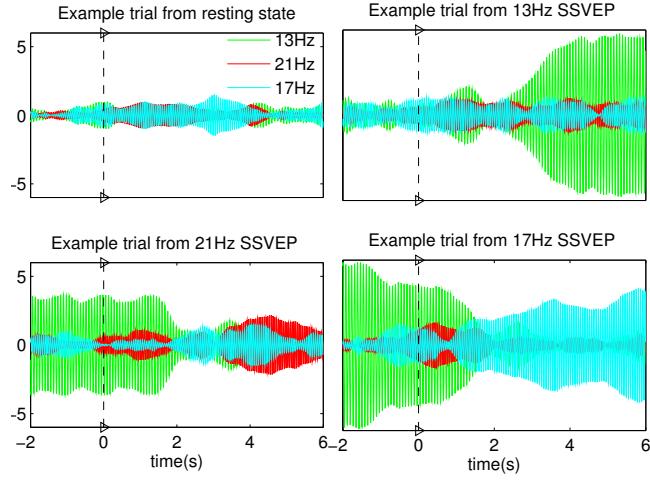
In this section, the performance of the proposed method is presented. First, the performance of the MDRM approach in an offline setup is analysed, then the results of the online algorithm are presented. In the offline analysis, the relevance of identifying the latency between cue onset and SSVEP response is shown. The results of the MDRM approach are compared to two state-of-the-art methods, [Lin et al., 2006] and [Nakanishi et al., 2014]. The online evaluation is divided in two parts: in the first one the algorithm discriminates between $K = F = 3$ SSVEP classes (*i.e.* 13, 17 and 21 Hz) and in the second one is applied on $K = 4$ classes, *i.e.* the $F = 3$ SSVEP class and the resting class.

Offline Analysis

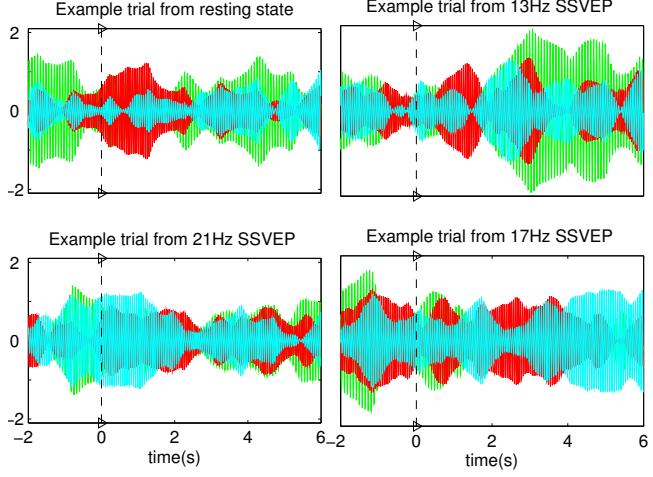
A close inspection of the filtered signals shows that almost all signals are synchronised with the trial frequency 2 seconds after cue onset $\tau_0 = 0$, as shown in Figure 5.7. This delay is mainly due to protocol design and user-specific cognitive processes. The protocol is aimed to provide an asynchronous setup close to real applications. The user is not required to look at a fixation point or to directly gaze toward the target, as in [Kimura et al., 2013; Nakanishi et al., 2014], during inter-trial periods. This is a tentative explanation for the higher delay observed in our study and it is consistent with literature observations [Vialatte et al., 2010; Bakardjian et al., 2010]

In fact, before $\tau_0 + 2\text{s}$, for some users the signal could still be synchronised with the previous trial frequencies. An important increase in average classification accuracy (almost 10%) could be obtained by taking the trial from 2 seconds after cue onset. It is therefore crucial to consider the latency between the cue onset of trial and the actual synchronisation of SSVEP at stimulus frequency. Thus in the offline synchronous processing, the confident window for classification is set 2 seconds after the cue onset ($\tau_0 + 2$).

Table 5.3 shows the offline classification accuracy for each subject obtained by the application of the *MDRM* as described in Algorithm 1, with the epochs taken at $\tau_0 + 2$. Column $MDRM(\tau_0)$ shows the results obtained when the epochs



(a)



(b)

Figure 5.7: Signal amplitude at each stimulus frequency, showing synchronisation of EEG with respect to time (seconds). The raw signal of the trial measured on Oz is band filtered using a Butterworth of order 8 at each stimulus frequency and the resulting signals are shown in blue (dark grey), green (grey), and red (light grey) for the same signal filtered respectively at 13, 17, and 21 Hz. The cue onset τ_0 at time 0 on the x-axis is shown with a vertical discontinued line. 4 trials are shown, one for each class. Signals are from the subjects with the highest (5.7(a)) and with the lowest BCI performance (5.7(b)).

are taken from cue onset. The Riemannian potato technique presented in Section 5.4.2 was applied for outlier removal (*MDRM-Potato*). The performance of

the MDRM approach is compared to two CCA-based state-of-the-art methods proposed by Lin et al. [2006] and Nakanishi et al. [2014] respectively. In the implementation of these methods, the epochs are also taken from $\tau_0 + 2$.

The MDRM approach outperforms both CCA-based method with an average classification accuracy of $90.4 \pm 7.8\%$ and ITR of 16.3 ± 5.3 bits/min. Lin et al. [2006] ranks second with $87.5 \pm 15.1\%$ and 15.5 ± 6.8 bits/min. The method proposed by Nakanishi et al. [2014], which could be expected to achieve better results as reported in their work, only ranks third. This is mainly due act that this method requires information on the phase of the stimuli. In fact, Nakanishi et al. [2014] use the average of all training trials belonging to a unique class as a reference signal in the CCA. When SSVEP trials belonging to a unique trial are not in-phase, which is the case in the current work, averaging them will cancel the signal.

Within the MDRM approach, it is shown that taking into account the latency between the cue onset and the SSVEP response significantly increases the classification performances: accuracy and ITR rise from $75.9 \pm 11.4\%$ and 6.0 ± 3.1 bits/min to $90.4 \pm 7.8\%$ and 16.3 ± 5.3 bits/min. In turn removing outliers with the Riemannian potato does not bring significant change. This could be attributed to the fact that the recording have been conducted in controlled environment, with small or little external noise.

Online Analysis without Resting Class

In an online asynchronous experiment, there is no cue onset, and the delay before SSVEP synchronisation might differ from one trial to another and from one subject to another. To locate the trust EEG region for the classification, D and ϑ are set respectively to 5 and 0.7 through cross-validation. The performance of this online setup is analysed and Figure 5.8 shows the results. From the analysis shown in Figure 5.8(d), the epoch size is set to $w = 2.6$ seconds. The step size is set to $\Delta n = 0.2$ s, that is a new epoch is classified every 0.2 second.

On Figure 5.8(a), the classification error is plotted against the epoch index. It shows that the error decreases as epochs move from the beginning of the trial. The error increases in the last epochs of the trial, corresponding to the end of the SSVEP task. Figure 5.8(b) details the evolution of the probability for each

class as epochs index increases. It appears clearly that the class of the EEG trial (thick-and-star line) has the largest probability only a few epochs after the beginning of the trial. Moreover, one can see that this is an increasing trend over the whole trial. Thus by setting an appropriate probability threshold ϑ , the actual class can be identified with enough confidence. Figure 5.8(c) shows the influence of the probability threshold ϑ on the classification error. The error is reduced when the probability threshold ϑ is increased. Figure 5.8(d) shows how the average online performance varies with respect to the epoch size (w). Both the classification accuracy and the ITR are shown. With short w values, the epoch size does not capture enough feature for a correct classification, and with long w , the epoch losses temporal resolution. The ITR increases with the classification rate but drops sensibly after a peak value.

The observation of Figure 5.9 provides a visualisation of the principle guiding the online implementation of Equation (5.4.3). This figure shows the trajectory on the tangent space taken by covariance matrices during a 4-class SSVEP experiment, and how they are classified epoch by epoch. It can be seen (encircled in Figure 5.9(a)) that a change in the SSVEP stimulus might not be detected instantaneously by the classifier. The trials are erroneously attributed with confidence to the previous class.

The proposed online algorithm, described in Algorithm 3, mitigates this issue and increases the classification accuracy as shown in Table 5.4. The “Online ($\rho(\bar{k}) > \vartheta$)” column shows the results of the online algorithm without the curve direction criterion (*i.e.*, without steps 6 to 11), and “Online (full algo. 3)” shows the improvement brought by this criterion. The performances are in terms of average classification accuracy (acc (%)), average time taken into the trial before classification (delay (s)), and the ITR (itr (bits/min)).

The curve direction criterion increases the rejection of epochs that could be wrongly classified, it thus significantly increases the classification accuracy of the online algorithm ($70.8\pm13\%$ to $87.3\pm9.8\%$), while increasing the delay (0.7s to 1.1s) before classification. When compared to the state-of-the-art off-line MDRM, the online curve-based classification yields better results in terms of ITR as the delay before classification is much shorter in the latter than the trial length used in the former; classification outputs are reached faster with

the online algorithm. Moreover, the online algorithm can be applied in both synchronous and asynchronous paradigms, whereas the offline algorithms are limited to synchronous paradigms which provide strongly limited user interaction.

Last, the impact of the Riemannian potato is analysed. A bootstrapping with 50 replications was performed on the offline data to assess the effect of applying the Riemannian potato. The results show that for most subjects the results are unchanged when the Riemannian potato is applied: due to the fact that data are recorded in a controlled environment, most of them are thus clean. It does, however, improve the results of few subjects. It was then applied in the training phase of the online application, and a similar observation is made. We can conclude that the Riemannian potato can be used as a safety guard to ensure that the Riemannian mean used in the MDRM classification scheme is not affected by outliers, especially for BCI used in less controlled environment.

Online Analysis with Resting Class

Using the MDRM approach it is possible to identify the resting class. In fact, covariance matrices of signals recorded during resting periods can be characterised with their own Riemannian mean. As such, they can be identified as any other class using the MDRM approach. The state-of-the-art methods, [Lin et al., 2006] and [Nakanishi et al., 2014], are both based on CCA where a reference signal is needed. These methods do not handle resting class, since there is no reference signal for them. In this section, the performance of the proposed approach including the identification of the resting class is presented. Table 5.5 summarises the classifier performance in the same format as Table 5.4, in terms of classification accuracy, delay before valid classification and ITR. Like in Table 5.4, the best performance is achieved by the complete online algorithm preceded with outlier removal with the Riemannian potatoes (*i.e.* *Online-Potato*). The identification of the resting class induces a drop of the overall classification accuracy by 8.2%, and a drop of ITR from 52.5 ± 25.5 to 49.2 ± 18.2 .

The effect of the resting class is seen with more details in Figure 5.10. Figure 5.10(a) shows the classification confusion matrix. There are few misclassifications between SSVEP classes compared to the misclassifications between the

resting class and any SSVEP class: the largest percentages are located in the first row and the first column, apart from the diagonal block. Figure 5.10(b) displays a ROC curve showing how the classifier performs in discriminating each class versus the others depending on the value of the ϑ parameter. On this ROC curve, the performance of the Online-Potato algorithm is indicated in terms of False Positive Rate (FPR) and True Positive Rate (TPR).

Confirming the observation from the confusion matrix, the ROC curve indicates that the resting is the most prone to false positive. Despite the drop in performance, the identification of resting class is crucial for online BCI setup, allowing the subject to use the system at his own pace.

5.6 Conclusion

This chapter investigated the efficiency of Riemannian geometry when dealing with covariance matrices as classification features. Existing covariance matrix estimators were investigated and their robustness was assessed on multichannel SSVEP signals to ensure that the obtained matrices are accurate estimates of data covariance, are well conditioned, and verify the positive-definiteness property. The Schäfer shrinkage estimator was found to be the best as it yielded the highest classification accuracy with the MDRM algorithm. The chapter demonstrated the interest in moving from Euclidean to Riemannian geometry in the design of machine learning algorithms applied to EEG signal and SSVEP in particular. Various distances and divergences as well as their corresponding means were presented and evaluated. Riemannian metric/divergences and their means are shown to be more appropriate on the structure of SPD matrices, and yield better results than their Euclidean counterparts in machine learning algorithms for classification. A novel algorithm based on MDRM, enhanced by class probability and the curve direction in the space of covariance EEG signals, was introduced and applied on a SSVEP classification task for a 4-class brain-computer interface.

The MDRM approach is first analysed in an offline classification setup. To prevent the effect of noisy signals on the MDRM approach, outliers in the training set are removed using a modified version of the *Riemannian potato*. This approach is compared to two CCA-based state-of-the-art methods. The results

show that offline MDRM achieves better classification performances than any of the CCA-based methods.

In the online setup, the proposed online algorithm enhances the stability of the BCI system, balancing between classification speed and prediction accuracy. The evaluation of the classification confidence over several epochs mitigates the short term perturbations in the experimental conditions and the attentional variations of the subject. The curve direction overcomes the misclassification of EEG trials that are still synchronised with past stimuli frequencies at classification time.

Unlike the CCA-based state-of-the-art methods considered in this work, the proposed online algorithm is capable of identifying the resting periods during an online EEG recording. These resting periods are considered as an additional class in the classification task.

All these contributions help to pave the way towards BCI used in non-controlled, assistive environment.

Euclidean										Riemannian									
Arithmetic					Harmonic					Log-Euclidean					Affine-invariant				
Sub.	acc (%)	time(s)	acc (%)	time(s)	acc (%)	time(s)	acc (%)	time(s)	acc (%)	acc (%)	time(s)	acc (%)	time(s)	acc (%)	acc (%)	time(s)	acc (%)	time(s)	acc (%)
1	53.12	0.025	40.62	0.030	71.88	0.150	73.44	0.194	59.37	0.155	68.75	0.225	60.94	0.025	68.75	0.220	54.69	0.630	
2	43.75	0.020	57.81	0.055	78.13	0.160	79.69	0.190	81.25	0.065	73.44	0.200	81.25	0.020	81.25	0.225	54.69	0.285	
3	67.19	0.020	73.44	0.040	85.94	0.120	85.93	0.205	95.31	0.155	85.94	0.100	95.31	0.040	85.94	0.200	76.56	0.280	
4	54.68	0.030	50.312	0.030	84.38	0.225	87.50	0.315	89.07	0.250	85.94	0.100	90.62	0.035	85.94	0.120	65.62	0.310	
5	37.50	0.020	35.94	0.040	62.50	0.115	68.75	0.290	73.44	0.140	65.62	0.125	70.31	0.035	65.63	0.110	45.31	0.660	
6	34.37	0.015	62.50	0.035	84.38	0.120	85.94	0.210	87.50	0.145	82.81	0.100	85.94	0.025	82.81	0.130	53.13	0.300	
7	60.42	0.027	67.71	0.037	87.50	0.267	88.54	0.410	91.66	0.417	86.46	0.137	94.79	0.020	86.46	0.243	69.79	0.777	
8	67.19	0.035	78.12	0.035	90.63	0.215	92.19	0.290	92.19	0.290	92.19	0.125	95.31	0.030	92.19	0.165	85.94	0.335	
9	57.81	0.035	43.75	0.035	70.31	0.275	70.31	0.380	75.00	0.300	76.56	0.134	76.56	0.035	67.19	0.160	62.50	0.310	
10	38.28	0.035	42.19	0.035	75.00	0.254	80.47	0.514	82.03	0.510	78.13	0.160	82.81	0.045	78.13	0.263	51.56	0.650	
11	48.44	0.025	48.44	0.030	60.94	0.144	65.63	0.235	57.81	0.150	75.00	0.105	48.44	0.030	75.00	0.195	56.25	0.575	
12	71.25	0.032	63.12	0.040	96.25	0.292	96.69	0.534	95.62	0.634	96.88	0.300	94.37	0.040	96.88	0.466	82.50	1.042	
Avg.	52.83	0.027	55.56	0.037	78.98	0.194	81.27	0.314	81.56	0.279	80.51	0.140	80.74	0.040	80.51	0.210	63.21	0.513	

Table 5.2: Subject classification accuracy (acc(%)) and average CPU time (time(s)) elapsed for the classification of a single trial. Classification is performed with MDM using either Euclidean or Riemannian means (see Table 5.1).

	Offline algorithms									
	Lin <i>et al.</i>		Nakanishi <i>et al.</i>		MDRM(τ_0)		MDRM		MDRM-Potato	
	acc(%)	itr(bpm)	acc(%)	itr(bpm)	acc(%)	itr(bpm)	acc(%)	itr(bpm)	acc(%)	itr(bpm)
S1	91.7	16.3	84.7	12.2	67.6	3.5	84.7	12.2	84.5	12.1
S2	45.8	0.7	47.9	1.0	66.0	3.2	79.4	9.7	79.3	9.6
S3	100.0	23.8	93.0	17.2	90.2	10.3	99.3	22.7	99.3	22.7
S4	97.9	21.3	96.6	20.0	78.3	6.1	89.7	15.0	89.7	15.0
S5	83.3	11.5	82.2	11.0	76.0	5.5	89.5	14.9	89.4	14.9
S6	77.1	8.7	76.2	8.3	72.2	4.5	87.2	13.6	87.2	13.6
S7	98.6	22.0	96.7	20.1	90.0	10.2	99.8	23.5	99.8	23.4
S8	97.9	21.3	65.5	4.7	90.4	10.3	99.7	23.2	99.7	23.2
S9	91.7	16.3	77.9	9.0	64.0	2.8	85.8	12.8	85.7	12.7
S10	80.2	10.0	76.9	8.6	79.2	6.4	93.1	17.3	93.0	17.2
S11	89.6	15.0	82.7	11.2	54.8	1.4	78.2	9.2	78.2	9.1
S12	95.8	19.4	93.8	17.8	82.3	7.4	98.6	22.0	98.6	22.0
Mean	87.5±15.1	15.5±6.8	81.2±14.1	11.8±6.0	75.9±11.4	6.0±3.1	90.4±7.8	16.3±5.3	90.4±7.8	16.3±5.3

Table 5.3: Offline performance in terms of accuracy and ITR. Five methods are compared: (1) CCA approach introduced by [Lin *et al.*, 2006], (2) CCA approach introduced by [Nakanishi *et al.*, 2014], (3) MDRM described in Section 5.3.3 (Algorithm 1), (4) MDRM where processed epochs are taken 2 seconds from the beginning of the trial, and (5) MDRM-Potato, where outliers are removed using the Riemannian potato approach described in Section 5.4.2.

	Online ($\rho(k) > \vartheta$)			Online (full algo. 3)			Online-Potato		
	acc(%)	delay(s)	itr(bpm)	acc(%)	delay(s)	itr(bpm)	acc(%)	delay(s)	itr(bpm)
S1	68.8	0.8	26.3	77.1	1.1	27.9	77.1	1.1	27.9
S2	64.6	0.7	21.6	77.1	1.2	26.8	77.1	1.2	26.8
S3	81.2	0.7	54.3	95.8	1.0	73.0	95.8	1.0	73.0
S4	83.3	0.8	53.2	91.7	1.0	58.6	95.8	1.0	69.2
S5	72.9	0.7	37.1	83.3	1.0	42.5	83.3	1.0	42.5
S6	66.7	0.7	24.5	72.9	1.1	24.3	72.9	1.1	24.3
S7	93.1	0.7	89.6	98.6	0.9	87.0	98.6	0.9	86.8
S8	87.5	0.6	76.2	100.0	0.9	95.9	100.0	0.9	95.9
S9	60.4	0.7	15.7	77.1	1.2	27.6	77.1	1.2	27.6
S10	64.6	0.7	21.5	87.5	1.1	45.3	87.5	1.1	45.3
S11	54.2	0.7	9.9	87.5	1.3	38.9	87.5	1.3	38.9
S12	52.5	0.7	8.0	99.2	1.2	71.7	99.2	1.2	71.8
Mean	70.8±13	0.7±0.0	36.5±26.3	87.3±9.8	1.1±0.1	51.6±25.1	87.7±10	1.1±0.1	52.5±25.5

Table 5.4: Classification performances (accuracy in %, delay before valid and confident classification in seconds, and ITR in bits/min) achieved using the online algorithm. The first column indicates the subjects. The following three columns show the results obtained without the curve direction criterion (Algorithm 3 up to 6): by stopping at step 6, \bar{k} is taken to be the valid class. The next three columns contain the results of the complete online algorithm. The last three columns report the results obtained when outliers are removed in the training phase using the Riemannian potato technique described in Section 5.4.2.

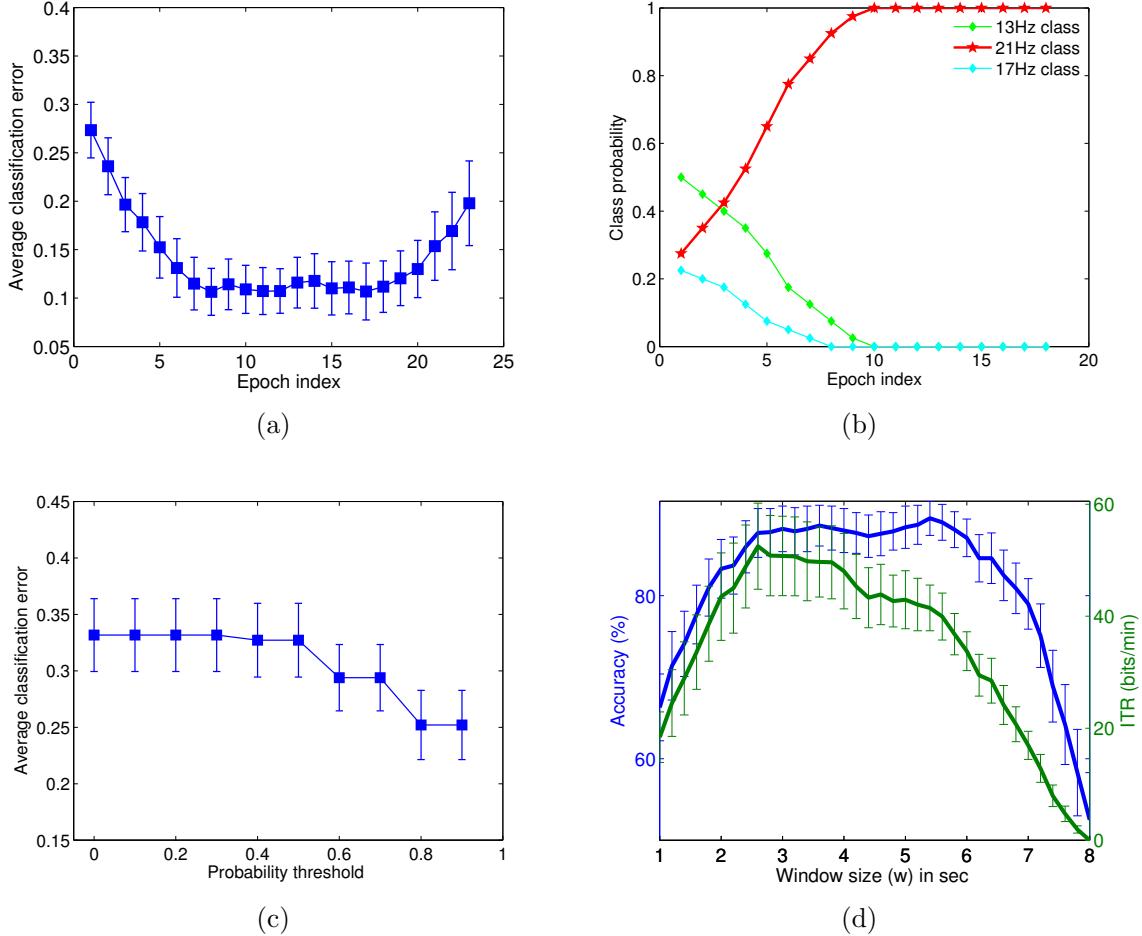


Figure 5.8: Evaluation of the online algorithm parameters. 5.8(a) shows the decrease of the average classification error over all subjects during the successive epochs after the beginning of the trial. 5.8(b) is an example taken from the subject with the best performance showing how the probability of the actual class varies with epoch position from beginning of trial. The groundtruth class probability is represented with a thick-and-star line, while other classes probability lines are thin-and-diamond. 5.8(c) shows the variation of the average classification error for different probability threshold ($0 \leq \vartheta < 1$) and its influence on the classifier output (Algorithm 3 step 6). 5.8(d) shows how the average online performance varies with respect to the epoch size (w). It shows both the classification accuracy (left y-axis) and the ITR (right y-axis). In 5.8(a), 5.8(c), and 5.8(d), the bars represent the error of the mean i.e. standard deviation divided by the square root of $n - 1$, $n = \text{number of samples}$.

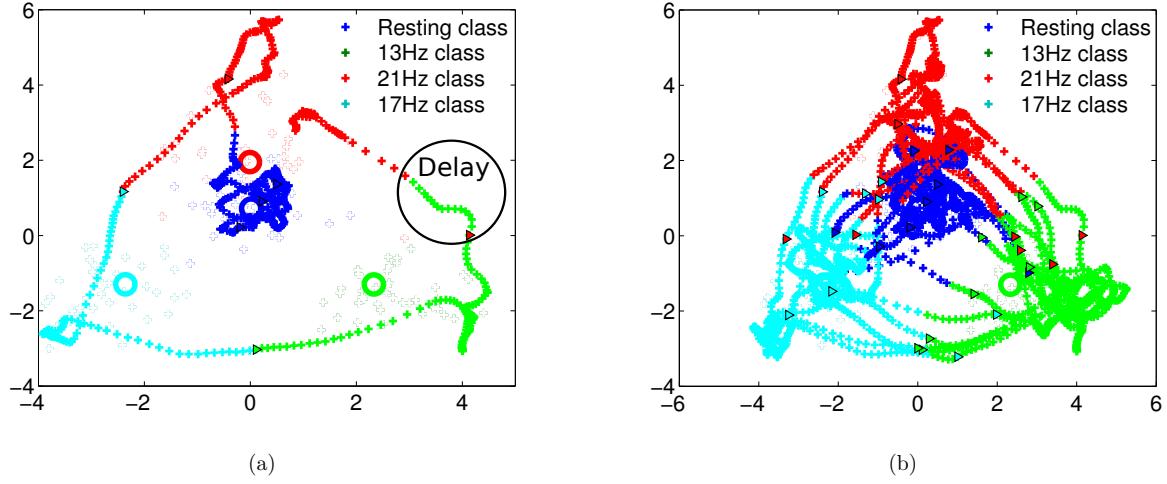


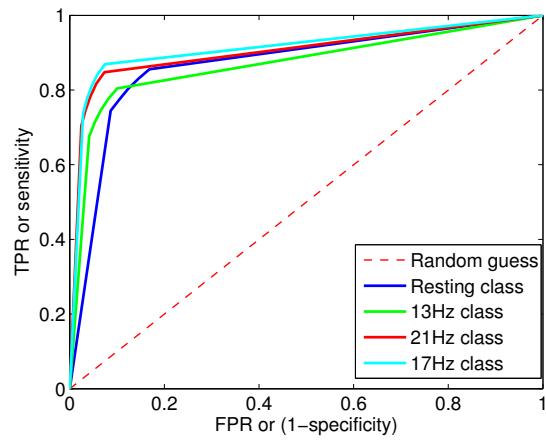
Figure 5.9: The covariance matrices trajectory during a 4-class SSVEP online recording. The circles represent class centres. The triangles mark the beginning of the experiment of a new trial whose class is indicated by the triangle's colour. 5.9(a) shows the first 7 trials. The first 3 trials are from the resting class, the remaining are respectively class 13 Hz, 17 Hz, and 21 Hz. 5.9(b) shows the entire recording. Data are taken from the subject with the highest BCI performance.

	Online ($\rho(k) > \vartheta$)			Online (full algo. 3)			Online-Potato		
	acc(%)	delay(s)	itr(bpm)	acc(%)	delay(s)	itr(bpm)	acc(%)	delay(s)	itr(bpm)
S1	67.2	0.7	37.6	71.4	1.1	32.4	71.4	1.1	32.4
S2	78.1	0.7	59.0	75.0	1.0	39.2	75.0	1.0	39.2
S3	89.1	0.8	85.2	89.1	1.0	67.6	89.1	1.0	67.6
S4	75.0	0.7	52.2	75.0	0.9	42.9	75.0	0.9	43.4
S5	71.9	0.7	46.7	70.3	1.1	31.0	70.3	1.1	31.0
S6	87.5	0.8	80.2	87.3	1.1	58.7	87.3	1.1	58.7
S7	84.4	0.7	76.3	85.4	1.0	62.5	88.5	1.0	69.1
S8	85.9	0.8	76.4	89.1	1.0	68.1	89.1	1.0	68.1
S9	67.2	0.7	37.2	75.0	1.0	39.6	76.6	1.1	40.3
S10	62.5	0.7	30.3	69.5	1.0	32.0	69.5	1.0	32.0
S11	59.4	0.8	23.5	68.8	1.1	29.1	68.8	1.1	29.1
S12	69.4	0.7	44.8	93.8	1.0	79.4	93.8	1.0	79.9
Mean	74.8±10.2	0.7±0.0	54.1±21.0	79.1±9.1	1.0±0.1	48.6±17.6	79.5±9.3	1.0±0.1	49.2±18.2

Table 5.5: This table summarises the performance achieved with the online algorithm with resting class identification, as in Table 5.4.

	Resting	13Hz	21Hz	17Hz	
Resting	202 21.1%	45 4.7%	31 3.2%	26 2.7%	66.4% 33.6%
13Hz	18 1.9%	177 18.5%	8 0.8%	6 0.6%	84.7% 15.3%
21Hz	10 1.0%	8 0.8%	191 19.9%	6 0.6%	88.8% 11.2%
17Hz	10 1.0%	9 0.9%	9 0.9%	202 21.1%	87.8% 12.2%
	84.2% 15.8%	74.1% 25.9%	79.9% 20.1%	84.2% 15.8%	80.6% 19.4%

(a)



(b)

Figure 5.10: (a) Confusion matrix for $K = 4$ classes with Online-Potato. (b): ROC curve indicating the influence of the ϑ parameter.

Chapter 6

Perspectives for Riemannian Approaches

Tell me and I forget, teach me and I may remember, involve me and I learn.

— Benjamin Franklin

6.1 Introduction

For efficient learning in EEG based BCI, as in most machine learning applications, an important amount of training data is needed. However the amount of data available within the BCI community is little [Delorme, 2015]. Another particularity with BCI is that the inter-subject variability requires that data used for training come from the same subject that the testing ones. Because of the difficulties in acquiring long signals from users and the need to keep the calibration time short, such training data are usually not available. Moreover, in some BCI applications the number of trials per class cannot be determined by the experimental paradigm, resulting in a class imbalance that disturbs the learning process.

A possible way of solving these problems related to data scarcity is *data augmentation*. In this approach, artificial data are generated by applying a transformation to the recorded data [Van Dyk and Meng, 2001; Grandvalet, 2000]. This technique has been successfully applied on image classification, when the number of samples in each class is small. The common practice is to identify

a set of possible transformations that could affect input images, e.g. rotation, translation, scaling, flipping, brightness adjustment, and to randomly applied those transformations to each training example [Dieleman et al., 2015]. In the context of handwritten character recognition, an elastic distortion emulating uncontrolled oscillation of hand muscles is applied [Simard et al., 2003]. Figure 6.1 shows an example of images where translation, scaling, rotation and elastic distortion have been applied. Data augmentation works well when combined



Figure 6.1: Hand-written digits from MNIST dataset. The original data are on the first row, the other rows are artificially created images from distorted version of the original digits. [Image taken from Ciresan et al., 2012]

with artificial neural network [Duda et al., 2001; Ciresan et al., 2012; Krizhevsky et al., 2012]. In BCI applications, a similar approach has been used to reduce calibration time in a motor imagery based BCI system [Lotte, 2011]. Each recorded trial is segmented and segments from the original set are randomly selected and concatenated to form new artificial trials.

In this chapter a novel data augmentation method based on non-Euclidean geometry is proposed. Unlike those mentioned above, data are not generated in the input space. Each training trial is represented in the space of SPD matrices by its covariance matrix. The augmented data lives on the manifold and within the convex hull defined by their class set. As a result, the convex hull of the class is densified with transformed versions of the original data. The augmented data are fed to a classifier, here we consider a multi-layer perceptron. This method is evaluated on two experimental datasets. The first one is an SSVEP-based BCI where only a limited number of training examples are available. The second one is an error detection application of ERP-based BCI to generate artificial

trials to balance the number of positive and negative trials. In the error related potential (ErrP) application paradigm, the number of trials with and without ErrP variable and not controlled.

Other than data augmentation, another way to make up for missing training data is to use data or parameters learnt from other subjects with sufficient training data. This is referred to as *transfer learning*. Machine learning algorithms aim at learning a task from training data. Once a task has been learnt, it can then be applied to future data (also referred to as test data). These algorithms work on the assumption that the training data and the future data are drawn from the same feature space and the same distribution. However, in real life applications, it is not always possible to have training data available, which are drawn from the same feature space and same distribution as the test data. Moreover, the task to be performed on new data can differ from the task learned from the training (or previous) data. Transfer learning thus aims at transferring the knowledge learnt from the previous task and data to a new task and data.

In BCI the need of transfer learning is important due to *inter-subject* variability and *intersession* variability. *Inter-subject* variability is expressed by the difference of brain signals recorded from different subjects despite them being involved in the similar mental activities. This difference is mostly attributed to anatomical differences among users. BCI algorithms are thus trained on brain signals recorded from a user to be letter used for the same task and on the same user. *Inter-session* variability is manifest between distinct recording sessions of a unique subject. This variability is attributed to changes in the mental states of the user, fatigue, and changes in experimental settings e.g. electrodes placement, environment, stimulations, etc. To make up for these setbacks, training data should be recorded for every BCI user in a controlled environment and the experimental settings meticulously noted. To have enough training data, lengthy recordings are needed, which is not always achievable due to BCI illiteracy, fatigue and discomfort. Being able to use data recorded from previous BCI users via transfer learning will

- 1) eliminate or shorten the recording of training data, and
- 2) improve BCI performance for users with limited training data.

6.2 Data Augmentation

This section presents the proposed approach of augmenting training data examples from their covariance matrices using Riemannian geometry.

6.2.1 Generating Artificial Points on Riemannian Manifold

Each trial's covariance matrix being represented as a point on the manifold, artificial trials are generated by interpolating new points between original trials' covariance matrices belonging to one class. This interpolation is done on the geodesic connecting each pair of original trials such that the generated point remains on the manifold and within the convex hull of the set of the class original data. This approach is similar to tensor linear interpolation introduced in [Pennec et al., 2006]. Given the definition of the tangent vector $\overrightarrow{\Sigma_1 \Sigma_2}$ between Σ_1 and Σ_2 in (5.1.3), the geodesic γ on the manifold can be obtained by the exponential mapping of $\overrightarrow{\Sigma_1 \Sigma_2}$ defined in (5.1.1) as: $\gamma = \text{Exp}_{\Sigma_1}(\text{Log}_{\Sigma_1}(\Sigma_2))$. Defining $t \in [0; 1]$, points lying on the geodesic are defined by:

$$\begin{aligned}\Sigma(t) &= \text{Exp}_{\Sigma_1}(t \text{ Log}_{\Sigma_1}(\Sigma_2)) \\ &= \Sigma_1^{\frac{1}{2}} (\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}})^t \Sigma_1^{\frac{1}{2}}\end{aligned}\tag{6.2.1}$$

with $\Sigma_1 = \Sigma(0)$ and $\Sigma_2 = \Sigma(1)$. Remark that the interpolation (6.2.1) is equivalent to $(1 - t)\Sigma_1 + t\Sigma_2$ in Euclidean space. Artificial points for data augmentation are obtained between original points by setting t in (6.2.1) to any value other than 0 and 1. In our experiments, interpolated matrices between each pair Σ_1, Σ_2 are linearly spaced on the geodesic between 0 and 1, and all possible pairs are considered.

Outliers in the pool of original data covariance matrices can distort the convex hull of classes, resulting in misclassification of new trials. To alleviate these effects, outliers are rejected from the original data before the generation of artificial data using an offline Riemannian potato described in section 5.4.2. The Riemannian mean of matrices belonging to one class is used as the centre of the Riemannian potato for that class. For each class, all matrices beyond the

z-score of 1 from the class centre are rejected. This value has been chosen after careful cross-validation.

6.2.2 Classification

To evaluate the benefit of applying the proposed data augmentation method, three classifiers are considered: a multi-layer perceptron (MLP) neural network [Duda et al., 2001] which is used on original data and then on augmented data, a tangent space linear discriminant analysis (TSLDA) [Barachant et al., 2012a] and a Riemannian-kernel support vector machine (RK-SVM) [Yger, 2013]. The choice for a MLP is motivated by the fact that neural networks are known to be sensitive to the amount and diversity of examples of data they are presented with [Ciresan et al., 2012; Krizhevsky et al., 2012]. On the other hand, RK-SVM and TSLDA are versions of SVM and LDA adapted to data lying on a Riemannian space. They are arguably the state-of-the-art concerning EEG covariance classification in tangent space [Barachant et al., 2012a, 2013b]. The classification features $w \in \mathbb{R}^{C(C+1)/2}$ are obtained projecting matrices on the tangent space at their mean $\bar{\Sigma}$:

$$\Theta_i = \bar{\Sigma}^{-\frac{1}{2}} \text{Log}_{\bar{\Sigma}}(\Sigma_i) \bar{\Sigma}^{-\frac{1}{2}} = \text{Log}(\bar{\Sigma}^{-\frac{1}{2}} \Sigma_i \bar{\Sigma}^{-\frac{1}{2}}), \quad (6.2.2)$$

and then extracting the upper triangular part of a symmetric matrix Θ_i and vectorising it (applying $\sqrt{2}$ weight for out-of-diagonal elements). These 3 classification methods are offline since the feature extraction (6.2.2) requires the projection on the global mean. However, online extensions are possible [Barachant et al., 2013b; Kalunga et al., 2015b].

6.2.3 Experimental Data Description

The assessment of the proposed data augmentation method is conducted on two datasets. The first one is from the SSVEP-based experiment described in section 4.4.1. The second dataset is an error-related potential detection, where the number of positive examples (the error potential) is smaller than the number of negative examples, that is a problem with unbalanced classes. Here only the ERP dataset is described.

ERP Dataset

The dataset, available for the NER Kaggle competition, was recorded during an online P300 speller experiment for error detection in the speller Perrin et al. [2012]. 16 healthy subjects participated in the experiment, the brain activity was recorded on $C = 56$ channels. Subjects have to spell a series a letter in under two spelling conditions: a fast, more error-prone condition (each item is flashed 4 times), and a slower, less error-prone (each item is flashed 8 times). The subjects had to go through five spelling sessions. Each session consisted of twelve 5-letter words, except the fifth which consisted of twenty 5-letter words making up for a total of 340 letters. For each spelled letter, the feedback of the result of the speller is displayed on a screen. The time of feedback is recorded and the label of feedback (correct or incorrect) is also recorded. In case of a spelling error, an error evoked potential occurs in the EEG. In the current work we focus on the detection of the error in spelling based on this *a priori*. The task of learning algorithms is to detect errors, *i.e.* to classify trials as incorrect or correct ($K = 2$, positive or negative). In such experiments, the number of positive and negative trials is not balanced. In case of a good speller, the number of positive trials are very limited. In this dataset the number of positive trials is largely inferior to the number of negative trials creating a problem of class unbalance in training set. To balance training set from this experiment, artificial data can be generated in the class with less number of trials.

6.2.4 Results and Discussion

SSVEP dataset

SSVEP training set is augmented with different numbers of artificial samples for each class. One to five samples are interpolated between each pair of original samples belonging to a single class. Figure 6.2 shows the densification effect resulting from the augmentation process. Original covariance matrices of each class are projected on the tangent space computed at the mean of all the matrices, and the two principal components (obtained by applying PCA) are shown on Fig. 6.2(a). Similarly, Fig. 6.2(b) shows the augmented covariance matrices after interpolation of 5 points between each pair of covariance matrices within

each class. The augmented data are within the convex hull of the original data.

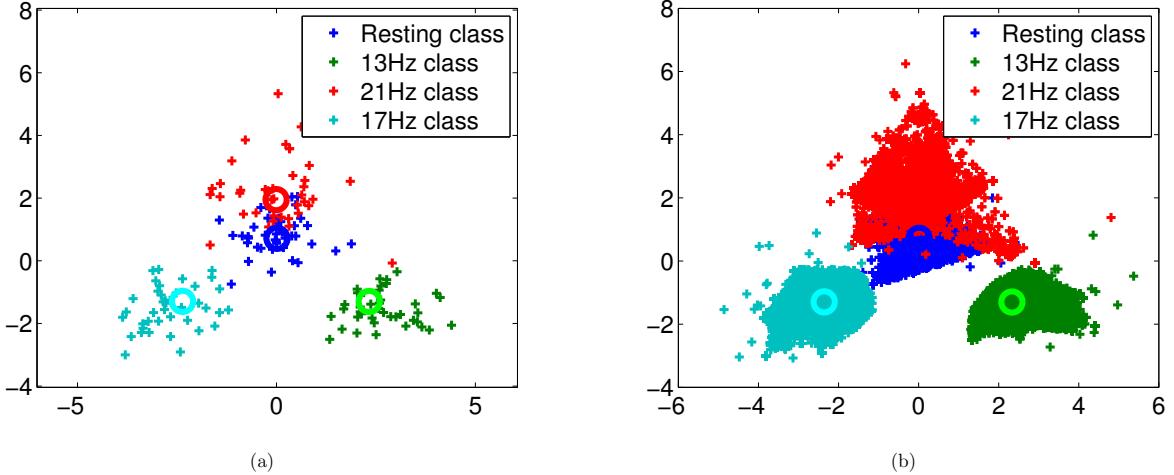


Figure 6.2: Mapping of covariance matrices of trials from each class on the tangent space (5.1.2). Matrices on the tangent space are vectorised and the 2 most significant components from PCA are used to obtain the 2-D representation. The covariance matrices of original data (a) and augmented data (b).

The performance of the augmentation approach is evaluated in terms of classification accuracy obtained with an MLP classifier and the results are compared with those obtained with TSLDA and RK-SVM classifiers. The MLP inputs are trial covariance matrices mapped on the tangent space. The MLP has 108 input units, one hidden layer with 50 neurons, and 4 output units. The classification obtained with each number of interpolated points are compared to the performance without training set augmentation. Figure 6.3 shows the classification performances from zero interpolated point (no training set augmentation) to 5 points interpolated. Due to the non-convexity of MLP optimisation, results averaged over subjects, are then averaged over 10 repetitions. Significant p-values show that average classification across all subjects is improved by the data augmentation. The effect of data augmentation varies depending on the quality of training examples from individual subjects. In Figure 6.4, the effect of augmenting training data in the subject with the lowest BCI performance and the subject with highest BCI performances are put side by side. In Table 6.1, the classification accuracy (in %) of the MLP preceded with data augmentation

are compared with RK-SVM and TSLDA.

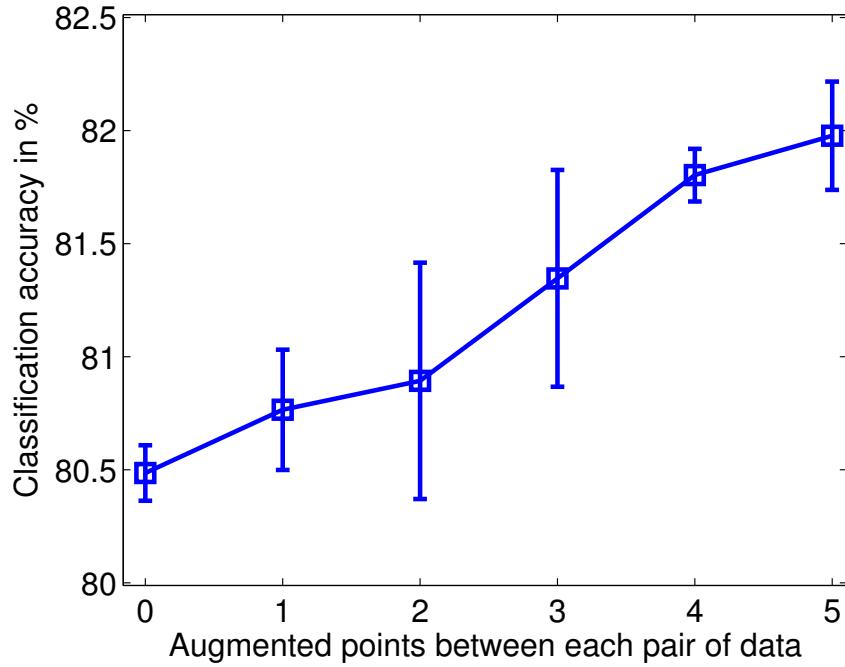


Figure 6.3: Mean classification accuracy in % across all subjects for different levels of data augmentation. At 0, there is no augmented data. At 1, one artificial data is interpolated between each pair of original data within each class, and so forth

ERP Dataset

On the ERP dataset the data augmentation is done to balance the number of positive trials (incorrect P300 feedback where ErrP is present) and negative trials (feedback with no error) in the training set. Each subject has 240 or 280 trials in the training set. The number of positive trials can be as low as 2% of the training set. The number of generated artificial trials g is determined by the gap between the number of positive trials and negative trials in the training set. To generate g trials, a covariance matrix is interpolated between g pairs

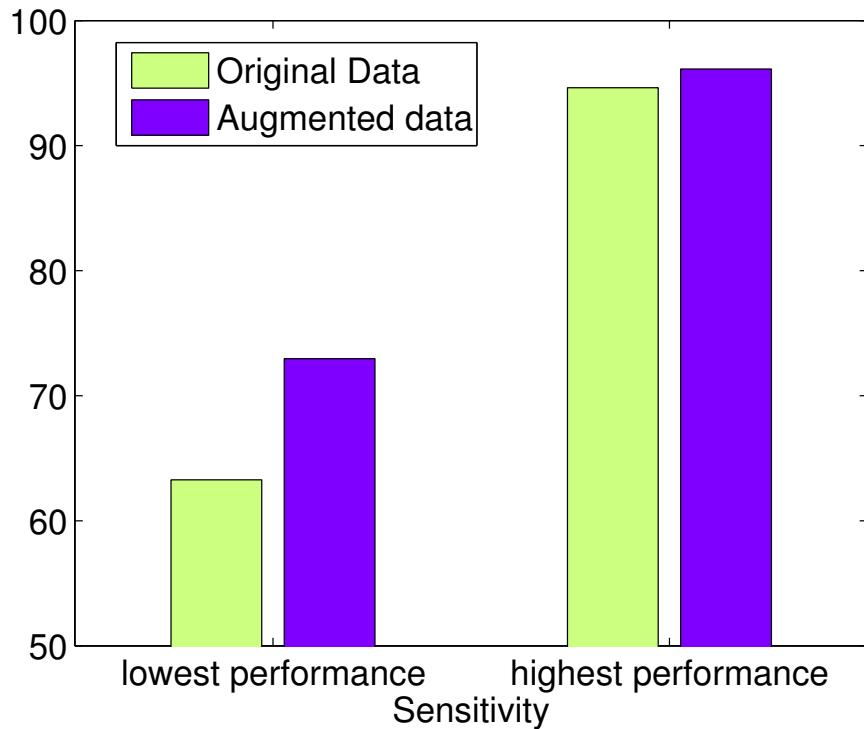


Figure 6.4: Classification accuracy of subject with lowest BCI performance versus subject with highest BCI performance, using original training set and using augmented training set with 5 interpolated points between each pair of original data within each class.

of randomly selected original matrices. The effect of balancing classes with artificial trials is evaluated with the three classifiers (*i.e.* MLP, TSLDA and RK-SVM). The MLP has 10 input units, one hidden layer with 50 neurons and two output units. The number of MLP units is chosen after a cross-validation phase.

Since the class unbalance is still present in the evaluation set, the classification performances are evaluated in terms of sensitivity. Figure 6.5 shows the performance achieved when classes are balanced by augmenting data in the positive class. They are compared to the results achieved when using unbalanced

	MLP	aug+MLP	RK-SVM	TSLDA
Sub 1	70.63	70.63	68.75	73.44
Sub 2	71.25	78.28	82.81	76.56
Sub 3	94.22	95.00	93.75	93.75
Sub 4	84.06	86.72	92.19	93.75
Sub 5	73.75	67.50	73.44	71.88
Sub 6	84.84	87.66	82.81	84.38
Sub 7	90.73	91.67	89.58	90.63
Sub 8	89.22	92.19	89.06	90.63
Sub 9	70.78	68.28	62.50	67.19
Sub 10	78.44	76.72	78.91	78.13
Sub 11	63.28	72.97	71.88	70.31
Sub 12	94.62	96.13	95.63	93.13
Average	80.49	81.98	81.78	81.98

Table 6.1: Comparison of classification accuracy (in %) using the MLP on original dataset, MLP with data augmentation (aug+MLP), RK-SVM and TSLDA.

training set. A t-test was performed and the p-values reveal significant improvement after data augmentation. Table 6.2 shows details of classifiers performance per subject in terms of sensitivity with and without data augmentation.

6.3 Transfer Learning

6.3.1 User Specificity as Domain in Transfer Learning

Exposed to the same stimuli, BCI users do not produce similar EEG response. To users specificities should be added changes induced by different environmental conditions during recording. In this work, we consider this specificity of the recorded EEG as being different domains in transfer learning.

Definition 6.3.1. (*Transfer Learning*) Given a source domain \mathcal{D}_S and learning task \mathcal{T}_S , a target domain \mathcal{D}_T and learning task \mathcal{T}_T , transfer learning aims at improving the learning of the target predictive function $f_T(\cdot)$ in \mathcal{D}_T using the knowledge in \mathcal{D}_S and T_{trS} , where $\mathcal{D}_S \neq \mathcal{D}_T$, or $\mathcal{T}_S \neq \mathcal{T}_T$ [Pan and Yang, 2010].

Considering the above definition of transfer learning, a domain is a pair $\mathcal{D} = \{\mathcal{X}, P(X)\}$ consisting of a feature space X and a marginal probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. A task is defined as a pair

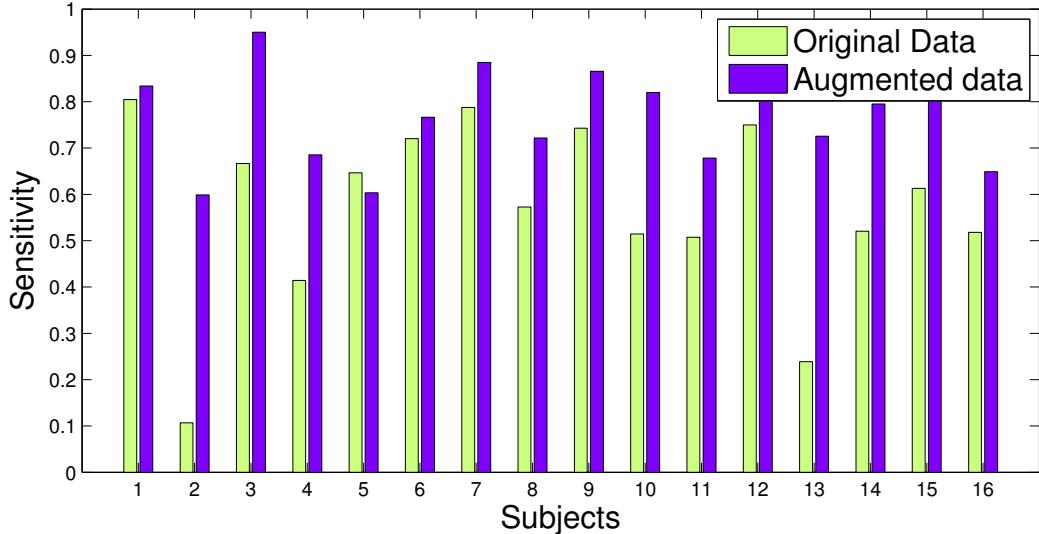


Figure 6.5: Classification performance in terms of sensitivity. For each of the 16 subjects these measures are given for classification based on training on original unbalance training set and training on augmented and balanced training set.

$\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$ consisting of a label space \mathcal{Y} and an objective predictive function $f(\cdot)$ that can be learned from the training data, which consist of pairs $\{x_i, y_i\}$, where $x_i \in X$ and $y_i \in \mathcal{Y}$. From a probabilistic viewpoint, $f(x)$ can be written as $P(y|x)$. Thus a task can be defined as $\mathcal{T} = \{\mathcal{Y}, P(Y|X)\}$.

6.3.2 Category of Proposed Transfer Learning

In definition 6.3.1, $\mathcal{D}_S \neq \mathcal{D}_T$, implies that either $\mathcal{X}_S \neq \mathcal{X}_T$ or $P(X_S) \neq P(X_T)$; and $\mathcal{T}_S \neq \mathcal{T}_T$ implies that either $\mathcal{Y}_S \neq \mathcal{Y}_T$ or $P(Y_S|X_S) \neq P(Y_T|X_T)$. Depending on each case, the following categories of transfer learning are defined [Pan and Yang, 2010]:

1. $\mathcal{T}_S = \mathcal{T}_T$ and $\mathcal{D}_S = \mathcal{D}_T$: Traditional Machine Learning (no transfer)
2. $\mathcal{T}_S \neq \mathcal{T}_T$: *Inductive transfer learning* and *Unsupervised transfer learning*
3. $\mathcal{D}_S \neq \mathcal{D}_T$: *Transductive transfer learning*

Sub.	Imbalanced classes			Balanced classes		
	MLP	RK-SVM	TSLDA	MLP	RK-SVM	TSLDA
1	0.85	0.76	0.79	0.83	0.77	0.85
2	0.11	0	0.32	0.60	0.07	0.57
3	0.67	0.60	0.72	0.95	0.63	0.95
4	0.41	0.42	0.63	0.69	0.32	0.70
5	0.65	0.51	0.61	0.60	0.49	0.68
6	0.72	0.71	0.74	0.77	0.70	0.76
7	0.79	0.70	0.78	0.88	0.70	0.89
8	0.57	0.33	0.63	0.72	0.25	0.70
9	0.74	0.59	0.77	0.87	0.58	0.89
10	0.51	0.34	0.59	0.82	0.34	0.90
11	0.51	0.27	0.57	0.68	0.27	0.61
12	0.75	0.65	0.82	0.97	0.65	0.99
13	0.24	0	0.57	0.73	0.08	0.75
14	0.52	0.47	0.62	0.80	0.43	0.75
15	0.61	0.51	0.65	0.81	0.60	0.83
16	0.52	0.46	0.54	0.65	0.42	0.53
Average	0.570	0.459	0.648	0.773	0.46	0.772

Table 6.2: Sensitivity analysis of performances obtained with 3 classifiers trained with imbalanced training set versus trained with balanced training set. The class imbalance of the ERP dataset is solved with data augmentation.

In BCI classification task, both inductive transfer learning and transductive transfer learning can be applied. In transductive learning, no labelled data are needed from the target domain, while few or all unlabelled data are needed at training time in order to obtain the marginal probability for the target data. This situation is suitable in offline BCI applications, and completely eliminate the need for training data, i.e. no recording phase.

In inductive learning few labelled data from the target data are needed to induce the predictive function. In this case just a small training set is needed shortening the recording of training data. This type of transfer can be used in both offline and online applications.

In this work, we are interested in a transductive transfer learning.

6.3.3 Composite Riemanian Mean

Composite Riemanian Mean is an *instance transfer* technique, *i.e.* re-weighting of labelled data in source domain for use in the target domain [Pan and Yang, 2010], inspired from the composite common spatial patterns method proposed by [Kang et al., 2009] as a feature representation transfer technique. Data from subjects in source domain are weighted based on the subject's similarity to the subject in the target domain. The measure of subjects' similarity is based on the KullbackLeibler divergence (KL-divergence) as proposed by [Kang et al., 2009]. Additionally the Affine Invariant Riemannian metric (AIRM) is also used for analysis. Other distances and divergences introduced in Section 5.3.2 might be used. These weights are obtained in an unsupervised way; no labels are required nor in the target domain, nor in the source domain. We rewrite the definition of the KL-divergence of multivariate Gaussian distribution X_1 and X_2 , with covariance matrices Σ_1 and Σ_2 respectively from (5.3.28) as:

$$D_{\text{KL}}(X_1, X_2) = \frac{1}{2} \left(\log \frac{\det \Sigma_2}{\det \Sigma_1} - C + \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1) \right)$$

In the preprocessing, the DC component of EEG signals is removed: $X \in N(0, \Sigma)$; and the covariance matrices are det-normalized. Therefore the KL-divergence can be expressed by:

$$D_{\text{KL}}(X_1, X_2) = \frac{1}{2} (\text{tr}(\Sigma_2^{-1} \Sigma_1 - C))$$

Where C is the dimension of Σ , and $\text{tr}(\cdot)$ denotes the trace of a matrix.

In Section 5.3.2, a symmetrised version of the KL-divergence was presented as the Jeffreys divergence. It can be expressed as:

$$D_{\text{KL}}^s(X_1, X_2) = \sqrt{\frac{1}{2} (\text{tr}(\Sigma_2^{-1} \Sigma_1 + \Sigma_2 \Sigma_1^{-1} - 2\mathbf{I}_C))} \quad (6.3.1)$$

Where \mathbf{I}_C is the identity matrix of size C .

The AIRM ($d_{\text{AI}}(\Sigma_1, \Sigma_2)$) is defined by (5.3.9)

The similarity between two subjects is defined as the inverse of the KL-divergence of their recorded EEG signals (or AIRM of their covariance matrix):

$$\mathbf{s}_{j,k} = \frac{1}{Z^k} \cdot \frac{1}{D_{\text{KL}}(X_j, X_k)} \quad (6.3.2)$$

Where Z^k is a normalisation factor for all distances to subject k :

$$Z^k = \sum_{l \neq k} \frac{1}{D_{\text{KL}}(X_l, X_k)} \quad (6.3.3)$$

For the AIRM, $D_{\text{KL}}(\cdot, \cdot)$ is replaced by $d_{\text{AI}}(\cdot, \cdot)$ in Equations (6.3.2) and (6.3.3).

To classify samples from the target subject k using MDRM, *composite Riemannian means* of classes are obtained as:

$$\mathbf{C}_c^k = (1 - \lambda) \mathbf{C}_c^k + \lambda \sum_{j \neq k} \mathbf{s}_{j,k} \mathbf{C}_c^j \quad (6.3.4)$$

where \mathbf{C}_c is the individual subject's mean of class c , and $\lambda \in [0, 1]$.

Defining parameter λ

In (6.3.4), the parameter λ defines how much the classification relies on data from other subjects. When a subject has enough and clean training data, it might not be necessary to use data from other subject; when the subject has noisy or little data, or his or her training data cannot be trusted, it might be safer to rely more on data from other subjects. The balance between this cases is determined by λ . We identified that λ will depend mainly on the number of data available for the test subject, and the proximity (or similarity) of this subject to other subjects. In other words, the proximity is also a measure of data transferability. We define λ as a flipped logistic function of the similarity of the test subject to other subjects and the number of training samples available per class in the training data of the test subject:

$$\lambda = \frac{1}{1 + e^{az(n-n_0)}} \quad (6.3.5)$$

a : $a \geq 1$, parameters controlling the decay. It can be learnt through cross-validation process.

z : Normalised z-score used as proximity measure.

n : Number of labelled samples per class.

n_0 : $n_0 > 1$ Shift in logistic function. It can be learnt through cross-validation process.

6.3.4 Experimental Results

Subjects similarity and weighting

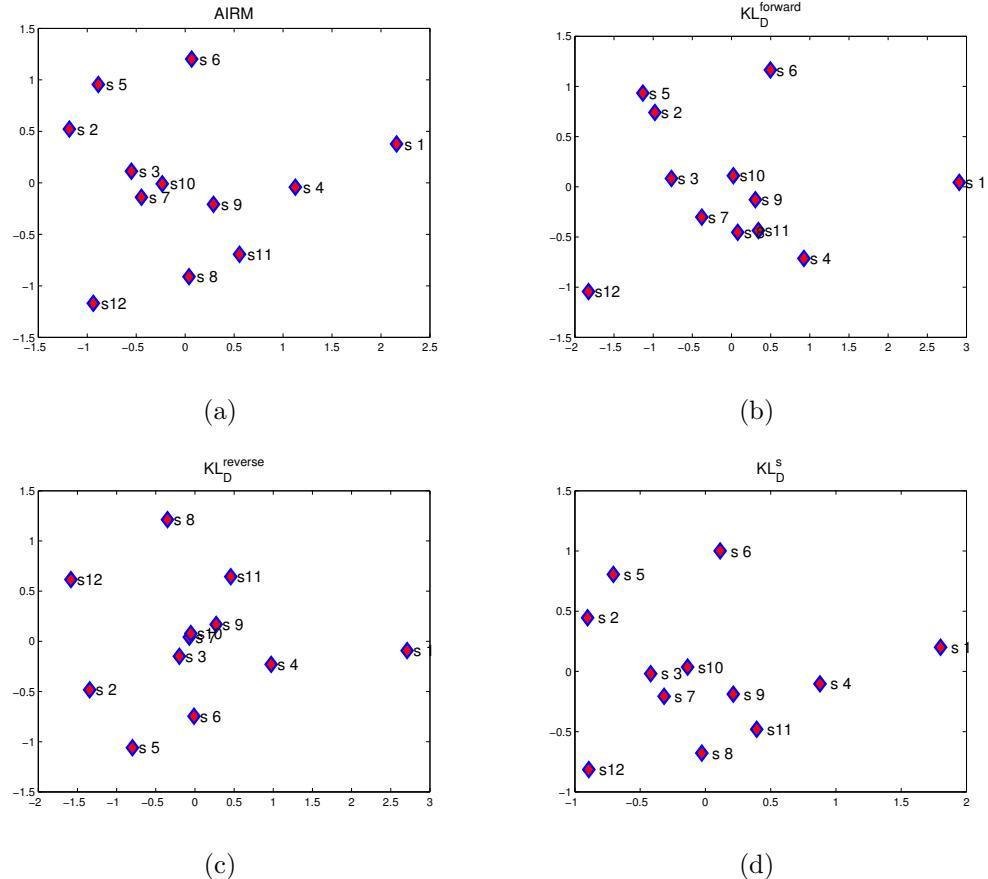


Figure 6.6: 2-D representation of affinity (or similarity) between subjects based on the 4 metrics: 6.6(a): Affine Invariant Riemannian distance, 6.6(b): Kullback-Leibler using forward divergences, 6.6(c): Kullback-Leibler using reverse divergences, 6.6(d): the symmetric version of KL divergence.

Grid search

A grid search is performed for $\lambda \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The number of labelled training examples from the test subject – used in the computation of \mathbf{C}_c^k in Eq. (6.3.4), is also varied: $n \in \{4, 8, 12, 16, 20, 24, 28, 32\}$.

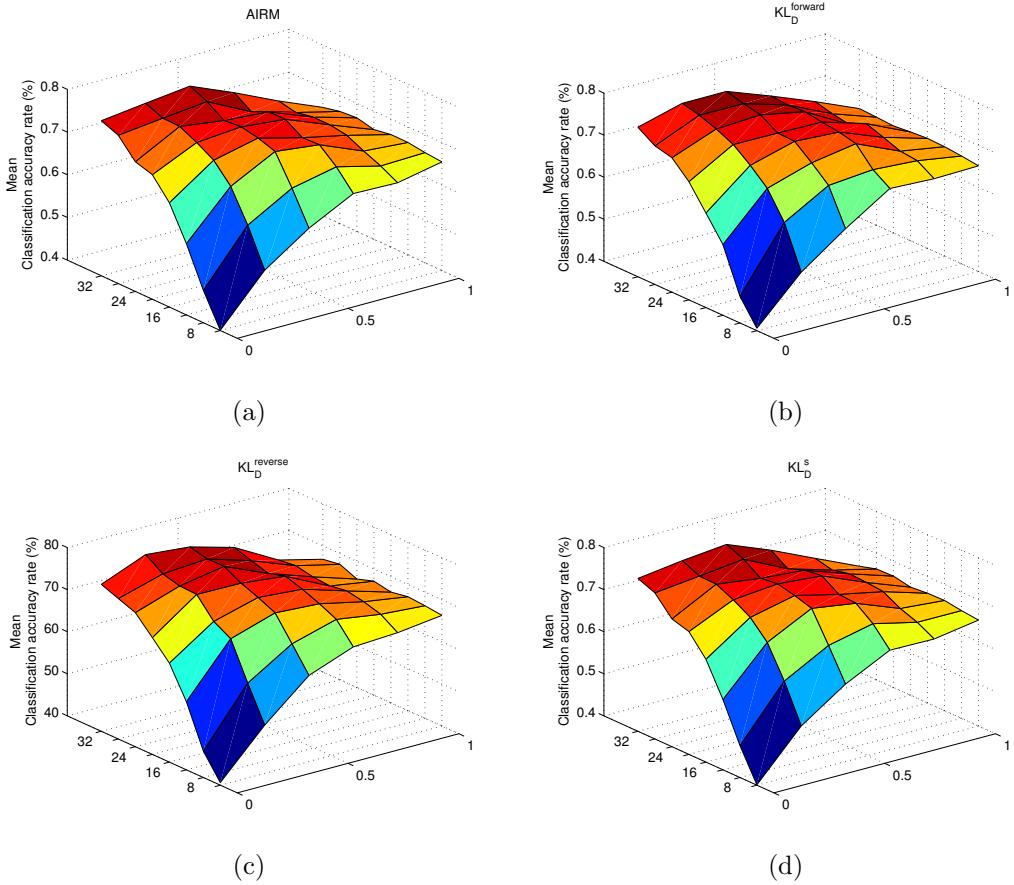


Figure 6.7: Mean classification accuracy for 12 subjects. Grid search with different values of $n \in \{4, 8, 12, 16, 20, 24, 28, 32\}$ on the x-axis and different values of $\lambda \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ on the y-axis. The results are obtained using 4 metrics to measure similarity between subjects: 6.7(a) AIRM, 6.7(b) $D_{KL}^{forward}$, 6.7(c) $D_{KL}^{reverse}$, 6.7(d) $D_{KL}^{symmetric}$

Evaluation λ

Through the grid search, different values of λ were tested with different numbers of samples. The optimal performances, i.e. the classification performance obtained with optimal λ , with each number available sample n are compared with the performance obtained with λ as defined in (6.3.5). To this end a Pareto analysis is performed, with the Pareto front being the optimal classification accuracy obtained through a grid search. Figure 6.9 shows the Pareto front against the performance obtained with $\lambda = 1$, meaning only data from other

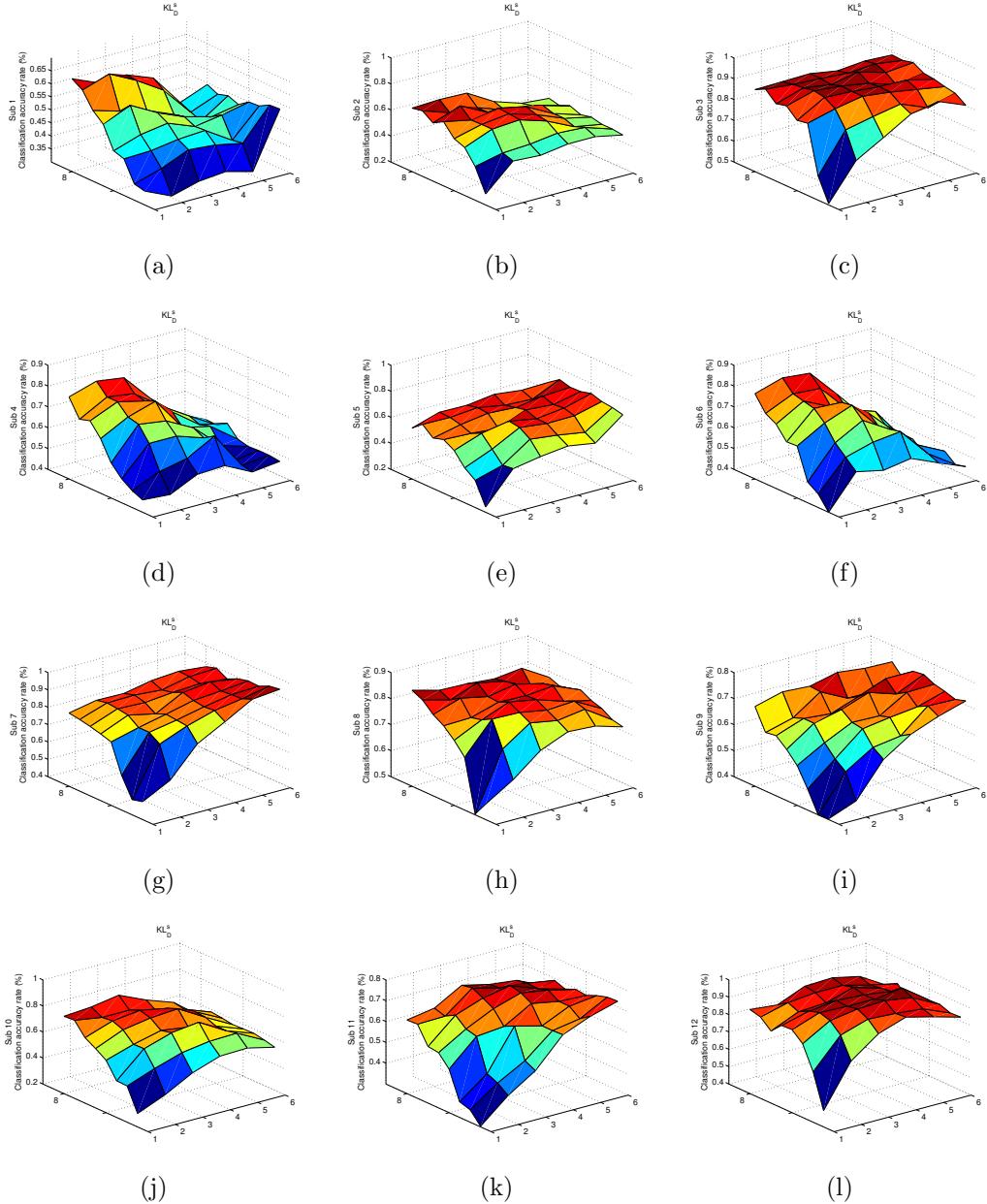


Figure 6.8: Individual subject classification accuracy. Grid search with different values of $n \in \{4, 8, 12, 16, 20, 24, 28, 32\}$ on the x-axis and different values of $\lambda \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ on the y-axis. (a) to (l) correspond to subjects 1 to 12 respectively. The results are obtained using $D_{KL}^{symmetric}$

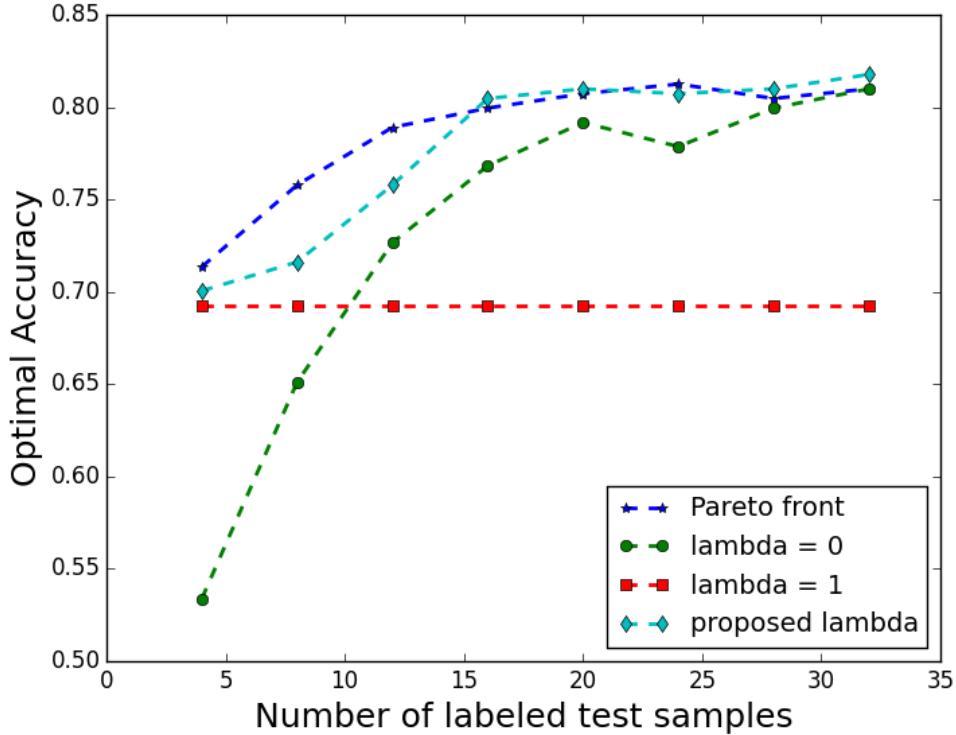


Figure 6.9: Optimal performance (Pareto front) obtained through a grid search is compared with the performance with $\lambda = 0$, $\lambda = 1$, and λ as proposed in (6.3.5).

subjects are considered for training, $\lambda = 0$, only using test subjects' data, and λ of (6.3.5).

It is seen that the defined λ performs well compared to the Pareto front. For $n > 4$, it even outperforms it. Indeed it allows for values between the intervals of those picked in the grid search.

6.4 Conclusion

In BCI, cases of insufficient training data and/or unbalanced classes are frequent. The section demonstrated that Riemannian geometry can be applied to address these issues through data augmentation and transfer learning. A data

augmentation scheme based on the geometry of covariance matrices was introduced. From the geodesics passing through pairs of samples, new samples are drawn and fed to a neural classifier. The data augmentation allows to enhance the classification accuracy when there is only a few number of samples per class. Data augmentation can compensate for dataset with unbalanced classes as it is often the case in event-related potential paradigm. The choice of the classifier is important when dealing with this augmented data; neural networks yield the best results. Future works will focus on the optimisation of the neural networks: determining the best architecture (in terms of layers and neurons) for processing covariance matrices and the investigation of common deep learning methods to improve results (dropouts, ReLU units, etc). The perspective of transfer learning yield promising results. Further work should be done on the optimisation of parameters in the logistic function defining lambda through a cross validation process. Other functions that could better describe the relationship between lambda, the proximity, and the number of training samples of the test subject should be explored.

Chapter 7

Conclusion

Better is the end of a thing than the beginning thereof

— Ecclesiastes 7:8

This work presented the current state-of-the art in brain-computer interfaces, and identified the challenges thereof. It focused on improving BCI performances and adaptivity and address problems related to the adaptability of BCIs to users' muscular abilities, to the robustness of EEG representation and learning, and to the insufficiency of samples in the training data. To address the adaptability to physical needs and muscular abilities of the user, a new methodology for designing hybrid systems was proposed. It uses a brain interface and motor interface specifically design to fit the user's needs and abilities. The main goal of these hybrid system is to assist people with motor disabilities or muscular diseases, by proposing a system that adapts to their individual needs, and makes use of their residual skills. The BCI is integrated in the system as a secondary modality, which is used to trigger specific behaviour or predefined actions. The proposed approach is implemented using a 3D touchless interface and a SSVEP-based BCI. This implementation gathers the two interfaces in a multimodal system which benefits from both the brain and motor signals. It is validated on a 3D navigation task in virtual environment and on the ESTA chair for the control of a robotic arm exoskeleton. To ensure robust EEG representation and learning, this work explores the Riemannian geometry of covariance matrices. It studies the necessary tools required for analysis of covariance matrices as elements of a Riemannian space. Methods of covariance estimation are studied to ensure the

quality and positive definiteness of the obtained covariance matrices. The notion of distance and mean being central to classification algorithms, metrics used for measure of distance (or divergence) and mean of covariance matrices are studied. The work shows that Riemannian metrics and their mean significantly improves the classification performances. Using the studied tools, an online algorithm for SSVEP classification was proposed, and was evaluated successfully. It provides, for the first time, an online approach to classification of covariance matrices of EEG in particular, and SPD matrices in general, using Riemannian geometry. Tools of Riemannian geometry offer many perspectives in BCI machine learning. This work proposes two other areas where they can be successfully applied, namely data augmentation and transfer learning. These two techniques address the problem of insufficient samples in the training data. By generating artificial training samples that are constrained to the manifold of SDP matrices defined by the original data, the proposed data augmentation technique can provide larger and more representative training data, and solve the problem of class imbalance in EEG classification particularly in ERP BCI. The proposed transfer learning approached enlarge the training set of a test subject by appropriately using data from other subjects. It increases the performance of classifiers, particularly when the test subject has a very small training set.

Seeing the benefit and perspective brought by Riemannian geometry from simple classification algorithms such as MDM, it is encouraging to apply them to other methods that are currently designed with linear Euclidean algebra. They can foreseeably be applied to adapt dictionary learning to Riemannian geometry, and with further investigation to methods such as artificial neural networks.

Bibliography

- Abdulkader, S. N., Atia, A., and Mostafa, M.-S. M. (2015). Brain computer interfacing: Applications and challenges. *Egyptian Informatics Journal*, 16(2):213–230.
- Acqualagna, L. and Blankertz, B. (2013). Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). *Clinical Neurophysiology*, 124(5):901–908.
- Acqualagna, L., Treder, M. S., Schreuder, M., and Blankertz, B. (2010). A novel brain-computer interface based on the rapid serial visual presentation paradigm. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 2686–2689.
- Agueh, M. and Carlier, G. (2011). Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924.
- Allison, B. Z., Brunner, C., Kaiser, V., Muller-Putz, G. R., C Neuper, and Pfurtscheller, G. (2010). Toward a hybrid brain computer interface based on imagined movement and visual attention. *Journal of Neural Engineering*, 7(2):026007.
- Allison, B. Z. and Neuper, C. (2010a). Could Anyone Use a BCI? In Tan, D. S. and Nijholt, A., editors, *Brain-Computer Interfaces*, Human-Computer Interaction Series, pages 35–54. Springer London.
- Allison, B. Z. and Neuper, C. (2010b). Could Anyone Use a BCI? In Tan, D. S. and Nijholt, A., editors, *Brain-Computer Interfaces*, Human-Computer Interaction Series, pages 35–54. Springer London. DOI: 10.1007/978-1-84996-272-8_3.
- Aloise, F., Aric, P., Schettini, F., Riccio, A., Salinari, S., Mattia, D., Babiloni, F., and Cincotti, F. (2012). A covert attention P300-based braincomputer interface: Geospell. *Ergonomics*, 55(5):538–551.

- Amari, S.-I. (2010). Information geometry in optimization, machine learning and statistical inference. *Frontiers of Electrical and Electronic Engineering in China*, 5(3):241–260.
- Ando, T., Li, C.-K., and Mathias, R. (2004). Geometric means. *Linear algebra and its applications*, 385:305–334.
- Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 2390–2397.
- Ang, K. K. K., Chin, Z. Y. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter Bank Common Spatial Pattern Algorithm on BCI Competition IV Datasets 2a and 2b. *Frontiers in neuroscience*, 6.
- Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347.
- Badcock, N. A., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., and McArthur, G. (2013). Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs. *PeerJ*, 1:e38.
- Bakardjian, H., Tanaka, T., and Cichocki, A. (2010). Optimization of SSVEP brain responses with application to eight-command braincomputer interface. *Neuroscience letters*, 469(1):34–38.
- Baklouti, M., Guyot, P. A., Monacelli, E., and Couvet, S. (2008). Force controlled upper-limb powered exoskeleton for rehabilitation. In *Intelligent Robots and Systems (IROS)*, page 4202.
- Barachant, A., Andreev, A., Congedo, M., and others (2013a). The Riemannian Potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry. In *Proceedings of TOBI Workshop IV*, pages 19–20.
- Barachant, A. and Bonnet, S. (2011). Channel selection procedure using Riemannian distance for BCI applications. In *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*, pages 348–351. IEEE.
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2010a). Common spatial pattern revisited by Riemannian geometry. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pages 472–476. IEEE.

- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2010b). Riemannian geometry applied to BCI classification. In *Latent Variable Analysis and Signal Separation*, pages 629–636. Springer.
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2012a). Multiclass braincomputer interface classification by Riemannian geometry. *Biomedical Engineering, IEEE Transactions on*, 59(4):920–928.
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013b). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, 112:172–178.
- Barachant, A., Bonnet, S., Congedo, M., Jutten, C., and others (2012b). BCI Signal Classification using a Riemannian-based kernel. In *Proceeding of the 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 97–102.
- Barachant, A. and Congedo, M. (2014). A Plug&Play P300 BCI Using Information Geometry. *arXiv preprint arXiv:1409.0107*.
- Barachant, A., Cycon, R., and Gouy-Pailler, C. (2015). P300-speller: Gomtrie Riemannienne pour la dtection multi-sujets de potentiels d’erreur. In *GRETISI 2015*.
- Barbaresco, F. (2011). Geometric radar processing based on Frchet distance: information geometry versus optimal transport theory. In *2011 12th International Radar Symposium (IRS)*.
- Bashashati, A., Fatourechi, M., Ward, R., and Birch, G. (2007). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *journal of neural engineering*, 4(2).
- Bayliss, J. D. and Ballard, D. H. (1998). Single Trial P300 Recognition in a Virtual Environment. *University of Rochester*, pages 22–25.
- Bédard, C., Kröger, H., and Destexhe, A. (2006). Model of low-pass filtering of local field potentials in brain tissue. *Nonlinear Soft Matter Physics*, 73:051911.
- Berger, P. D. H. (1929). ber das Elektrenkephalogramm des Menschen. *Archiv fr Psychiatrie und Nervenkrankheiten*, 87(1):527–570.
- Bhattacharyya, S., Khasnobish, A., Chatterjee, S., Konar, A., and Tibarewala, D. (2010). Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data. *Systems in Medicine and Biology (ICSMB), 2010 International Conference on*, pages 126–131.

- Bin, G., Gao, X., Wang, Y., Li, Y., Hong, B., and Gao, S. (2011). A high-speed BCI based on code modulation VEP. *Journal of Neural Engineering*, 8(2):025015.
- Bin, G., Gao, X., Yan, Z., Hong, B., and Gao, S. (2009). An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. *Journal of Neural Engineering*, 6(4).
- Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Mller, K.-R. (2011). Single-trial analysis and classification of ERP components: a tutorial. *NeuroImage*, 56(2):814–825.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K. R. (2008). Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *Signal Processing Magazine, IEEE*, 25(1):41–56.
- Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217.
- Brunner, C., Naeem, M., Leeb, R., Graimann, B., and Pfurtscheller, G. (2007). Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. *Pattern Recognition Letters*, 28(8):957–964.
- Bundy, D. T., Pahwa, M., Szrama, N., and Leuthardt, E. C. (2016). Decoding three-dimensional reaching movements using electrocorticographic signals in humans. *Journal of Neural Engineering*, 13(2):026021.
- Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., and Gross, J. (2011). Steady-State Visual Evoked Potentials Can Be Explained by Temporal Superposition of Transient Event-Related Responses. *PLoS ONE*, 6(1):e14543.
- Caton, R. (1875). Electrical Currents of the Brain. *The Journal of Nervous and Mental Disease*, 2(4):610.
- Cecotti, H., Rivet, B., Congedo, M., Jutten, C., Bertrand, O., Maby, E., and Mattout, J. (2011). A robust sensor-selection method for P300 braincomputer interfaces. *Journal of neural engineering*, 8(1):016001.
- Chang, C. C. and Lin, C. J. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.*, 2(3).

- Chebbi, Z. and Moakher, M. (2012). Means of Hermitian positive-definite matrices based on the log-determinant -divergence function. *Linear Algebra and its Applications*, 436(7):1872–1889.
- Chen, L., Jin, J., Zhang, Y., Wang, X., and Cichocki, A. (2015). A survey of the dummy face and human face stimuli used in BCI paradigm. *Journal of Neuroscience Methods*, 239:18–27.
- Cheng, M., Gao, X., Gao, S., and Xu, D. (2002). Design and implementation of a brain-computer interface with high transfer rates. *Biomedical Engineering, IEEE Transactions on*, 49(10):1181–1186.
- Cherian, A., Sra, S., Banerjee, A., and Papanikolopoulos, N. (2011). Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence. In *2011 International Conference on Computer Vision*, pages 2399–2406. IEEE.
- Cincotti, F., Scipione, A., Timperi, A., Mattia, D., Marciani, A., Millan, J., Salinari, S., Bianchi, L., and Bablioni, F. (2003). Comparison of different feature classifiers for brain computer interfaces. *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*, pages 645–647.
- Ciresan, D., Meier, U., and Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649. IEEE.
- Congedo, M. (2013). *EEG source analysis*. HDR, Université de Grenoble.
- Congedo, M., Barachant, A., and Andreev, A. (2013). A New Generation of Brain-Computer Interface Based on Riemannian Geometry. *arXiv preprint arXiv:1310.8115*.
- Courchesne, E., Hillyard, S. A., and Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, 39(2):131–143.
- Cutrell, E. and Tan, D. (2008). BCI for passive input in HCI. In *Proceedings of CHI*, volume 8, pages 1–3. Citeseer.
- Daly, J. J. and Wolpaw, J. R. (2008). Brain-Computer interfaces in neurological rehabilitation. *The Lancet Neurology*, 7(11):1032–1043.
- Delorme, A. (2015). EEG data available for public download.

- Dhillon, I. S. and Tropp, J. A. (2007). Matrix nearness problems with Bregman divergences. *SIAM Journal on Matrix Analysis and Applications*, 29(4):1120–1146.
- Dieleman, S., Willett, K. W., and Dambre, J. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly Notices of the Royal Astronomical Society*, 450(2):1441–1459.
- Dingyin, H., Wei, L., and Xi, C. (2011). Feature extraction of motor imagery EEG signals based on wavelet packet decomposition. *Complex Medical Engineering (CME), 2011 IEEE/ICME International Conference on*, pages 694–697.
- Donchin, E. (1969). Discriminant analysis in average evoked response studies: the study of single trial data. *Electroencephalography and clinical neurophysiology*, 27(3):311–314.
- Donchin, E. (1981). Surprise! Surprise? *Psychophysiology*, 18(5):493–513.
- Donchin, E., Spencer, K. M., and Wijesinghe, R. (2000). The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Neural Systems and Rehabilitation]*, 8(2).
- Dornhege, G., Blankertz, B., Curio, G., and Klaus-Robert Muller (2004a). Increase information transfer rates in BCI by CSP extension to multi-class. *Advances in Neural Information Processing Systems*, pages 733–740.
- Dornhege, G., Blankertz, B., Curio, G., and Muller, K.-R. (2004b). Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *Biomedical Engineering, IEEE Transactions on*, 51(6):993–1002.
- Duda, R., Hart, P., and Stork, D. (2001). *Pattern classification*. Wiley, 2 edition.
- Elshout, J. A. (2009). Review of brain-computer interfaces based on the P300 evoked potential. *Tech. Rep. PR-TN 2009/00066, Koninklijke Philips Electronics (2009)*.
- Faradji, F., Ward, R., and Birch, G. (2009). A brain-computer interface based on mental tasks with a zero false activation rate. *Neural Engineering, 2009. NER '09. 4th International IEEE/EMBS Conference on*, pages 355–358.

- Farwell, L. A. and Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70:510–523.
- Ferrez, P. and del R. Millan, J. (2008). Error-Related EEG Potentials Generated During Simulated Brain-Computer Interaction. *Biomedical Engineering, IEEE Transactions on*, 55(3):923–929.
- Ferrez, P. W. and Millan, J. d. R. (2008). Simultaneous Real-Time Detection of Motor Imagery and Error-Related Potentials for Improved BCI Accuracy. *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course*, pages 197–202.
- Finke, A., Knoblauch, A., Koesling, H., and Ritter, H. (2011). A hybrid brain interface for a humanoid robot assistant. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 7421–7424.
- Fisher, R. S., Harding, G., Erba, G., Barkley, G. L., Wilkins, A., and Epilepsy Foundation of America Working Group (2005). Photic- and pattern-induced seizures: a review for the Epilepsy Foundation of America Working Group. *Epilepsia*, 46(9):1426–1441.
- Fitzgerald, P. G. and Picton, T. W. (1967). Temporal and sequential probability in evoked potential studies. *Canadian Journal of Psychology*, 35(2):188–200.
- Fletcher, P. T., Lu, C., Pizer, S. M., and Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *Medical Imaging, IEEE Transactions on*, 23(8):995–1005.
- Foldes, S. T., Weber, D. J., and Collinger, J. L. (2015). MEG-based neurofeedback for hand rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 12:85.
- Foley, D. H. (1972). Considerations of sample and feature size. *Information Theory, IEEE Transactions on*, 18(5):618–626.
- Friedman, J. H. (1989). Regularized Discriminant Analysis. *Journal of the American Statistical Association*, 84(405):165–175.
- Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. Academic press.

- Gaur, P., Pachori, R. B., Wang, H., and Prasad, G. (2015). An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface. In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–7. IEEE.
- Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2013). *Cognitive Neuroscience: The Biology of the Mind, 4th Edition*. W. W. Norton & Company, 4th edition edition.
- Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., and Donchin, E. (1993). A neural system for error detection and compensation. *Psychological science*, 4(6):385–390.
- George, L. and Lécuyer, A. (2010). An overview of research on “passive” brain-computer interfaces for implicit human-computer interaction. In *International Conference on Applied Bionics and Biomechanics ICABB 2010-Workshop W1 “Brain-Computer Interfacing and Virtual Reality”*.
- Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., Shukla, M., and Aslin, R. N. (2011). Near-infrared spectroscopy: A report from the McDonnell infant methodology consortium. *Developmental Cognitive Neuroscience*, 1(1):22–46.
- Girouard, A., Solovey, E. T., and Jacob, R. J. (2013). Designing a passive brain computer interface using real time classification of functional near-infrared spectroscopy. *International Journal of Autonomous and Adaptive Communications Systems*, 6(1):26–44.
- Goh, A. and Vidal, R. (2008a). Clustering and dimensionality reduction on Riemannian manifolds. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–7. IEEE.
- Goh, A. and Vidal, R. (2008b). Unsupervised Riemannian clustering of probability density functions. In *Machine Learning and Knowledge Discovery in Databases*, pages 377–392. Springer.
- Golub, M. D., Yu, B. M., Schwartz, A. B., and Chase, S. M. (2014). Motor cortical control of movement speed with implications for brain-machine interface control. *Journal of Neurophysiology*, 112(2):411–429.
- Gosseries, O., Demertzi, A., Noirhomme, Q., Tshibanda, J., Boly, M., Op de Beeck, M., Hustinx, R., Maquet, P., Salmon, E., Moonen, G., Luxen, A., Laureys, S., and De Tige, X. (2008). [Functional neuroimaging (fMRI, PET and MEG): what do we measure?]. *Revue Medicale De Lige*, 63(5-6):231–237.

- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., and Hmlinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, 86:446–460.
- Grandvalet, Y. (2000). Anisotropic Noise Injection for Input Variables Relevance Determination. In *IEEE TRANSACTIONS ON NEURAL NETWORKS*, pages 463–468. Springer.
- Grosse-Wentrup, M. and Buss, M. (2008). Multiclass common spatial patterns and information theoretic feature extraction. *Biomedical Engineering, IEEE Transactions on*, 55(8):1991–2000.
- Grbler, G., Al-Khadairy, A., Leeb, R., Pisotta, I., Riccio, A., Rohm, M., and Hildt, E. (2014). Psychosocial and Ethical Aspects in Non-Invasive EEG-Based BCI ResearchA Survey Among BCI Users and BCI Professionals. *Neuroethics*, 7(1):29–41.
- g.Tec (2012). intendix.com.
- Güçlütürk, Y., Güçlü, U., and Samraj, A. (2010). An online single trial analysis of the P300 event related potential for the disabled. *Electrical and Electronics Engineers in Israel (IEEEEI), 2010 IEEE 26th Convention of*, pages 338–341.
- Hammer, E. M., Halder, S., Blankertz, B., Sannelli, C., Dickhaus, T., Kleih, S., Mller, K.-R., and Kbler, A. (2012). Psychological predictors of SMR-BCI performance. *Biological Psychology*, 89(1):80–86.
- Hao, Y., Budd, J., Jackson, M. M., Sati, M., and Soni, S. (2014). A visual feedback design based on a brain-computer interface to assist users regulate their emotional state. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2491–2496. ACM.
- Hardoon, D. R., Szedmak, S. R., and Shawe-Taylor, J. R. (2004). Canonical Correlation Analysis: An Overview with Application to Learning Methods. *Neural Comput.*, 16(12):2639–2664.
- Haselsteiner, E. and Pfurtscheller, G. (2000). Using time-dependent neural networks for EEG classification. *Rehabilitation Engineering, IEEE Transactions on*, 8(4):457–463.
- Henson, R. N., Wakeman, D. G., Litvak, V., and Friston, K. J. (2011). A parametric empirical Bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration. *Frontiers in Human Neuroscience*, 5:76.

- Herculano-Houzel, S. (2009). The Human Brain in Numbers: A Linearly Scaled-up Primate Brain. *Frontiers in Human Neuroscience*, 3.
- Herculano-Houzel, S. (2012). The Remarkable, Yet Not Extraordinary, Human Brain as a Scaled-Up Primate Brain and Its Associated Cost. *Proceedings of the National Academy of Sciences*, 109(Supplement 1):10661–10668.
- Herrmann, C. S. (2001). Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, 137(3-4):346–353.
- Hill, N., Lal, T., Schrder, M., Hinterberger, T., Wilhelm, B., Nijboer, F., Mochty, U., Widman, G., Elger, C., Schlkopf, B., Kbler, A., and Birbaumer, N. (2006). Classifying EEG and ECoG Signals without Subject Training for Fast BCI Implementation: Comparison of Non-Paralysed and Completely Paralysed Subjects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):183–186.
- Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., Haddadin, S., Liu, J., Cash, S. S., van der Smagt, P., and Donoghue, J. P. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375.
- Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., and Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164–171.
- Hoffmann, U., Garcia, G., Vesin, J.-M., Diserens, K., and Ebrahimi, T. (2005). A Boosting Approach to P300 Detection with Application to Brain-Computer Interfaces. *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, pages 97–100.
- Hoffmann, U., Vesin, J.-M., Ebrahimi, T., and Diserens, K. (2008). An efficient P300-based braincomputer interface for disabled subjects. *Journal of Neuroscience Methods*, 167(1):115–125.
- Homer, M. L., Nurmikko, A. V., Donoghue, J. P., and Hochberg, L. R. (2013). Implants and Decoding for Intracortical Brain Computer Interfaces. *Annual review of biomedical engineering*, 15:383–405.
- Huettel, S. A., Song, A. W., and McCarthy, G. (2004). *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland.

- Huggins, J. E., Levine, S. P., BeMent, S. L., Kushwaha, R. K., Schuh, L. A., Passaro, E. A., Rohde, M. M., Ross, D. A., Elisevich, K. V., and Smith, B. J. (1999). Detection of event-related potentials for development of a direct brain interface. *Journal of clinical neurophysiology*, 16(5):448.
- Huster, R. J., Mokom, Z. N., Enriquez-Geppert, S., and Herrmann, C. S. (2014). Braincomputer interfaces for EEG neurofeedback: Peculiarities and solutions. *International journal of psychophysiology*, 91(1):36–45.
- Ishita, H., Sakai, M., Watanabe, J., Chen, W., and Darning Wei (2007). Development of P300 Detection Algorithm for Brain Computer Interface in Single Trial. *Computer and Information Technology, 2007. CIT 2007. 7th IEEE International Conference on*, pages 1100–1105.
- Jayasumana, S., Hartley, R., Salzmann, M., Li, H., and Harandi, M. (2013). Kernel methods on the Riemannian manifold of symmetric positive definite matrices. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 73–80. IEEE.
- Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, 33(11):1419–1432.
- Jeunet, C., Jahanpour, E., and Lotte, F. (2016). Why Standard Brain-Computer Interface (BCI) Training Protocols Should be Changed: An Experimental Study. *Journal of Neural Engineering*.
- Jia, C., Gao, X., Hong, B., and Gao, S. (2011). Frequency and Phase Mixed Coding in SSVEP-Based BrainComputer Interface. *Biomedical Engineering, IEEE Transactions on*, 58(1):200–206.
- Jost, J. (2011). *Riemannian geometry and geometric analysis*, volume 62011. Springer.
- Jrad, N., Congedo, M., Phlypo, R., Rousseau, S., Flamary, R., Yger, F., and Rakotomamonjy, A. (2011). sw-SVM: sensor weighting support vector machines for EEG-based braincomputer interfaces. *Journal of Neural Engineering*, 8:056004+.
- Kalunga, E., Chevallier, S., and Barthlemy, Q. (2015a). Data augmentation in Riemannian space for Brain-Computer Interfaces. In *ICML Workshop on Statistics, Machine Learning and Neuroscience (StamLins 2015)*.

- Kalunga, E., Djouani, K., Hamam, Y., Chevallier, S., and Monacelli, E. (2013). SSVEP enhancement based on Canonical Correlation Analysis to improve BCI performances. In *AFRICON, 2013*, pages 1–5.
- Kalunga, E. K., Chevallier, S., and Barthlemy, Q. (2015b). Using Riemannian geometry for SSVEP-based Brain Computer Interface. Technical report, Tshwane University of Technology & Universit de Versailles Saint-Quentin.
- Kalunga, E. K., Chevallier, S., Barthlemy, Q., Djouani, K., Hamam, Y., and Monacelli, E. (2015c). From Euclidean to Riemannian Means: Information Geometry for SSVEP Classification. In Nielsen, F. and Barbaresco, F., editors, *Geometric Science of Information*, number 9389 in Lecture Notes in Computer Science, pages 595–604. Springer International Publishing. DOI: 10.1007/978-3-319-25040-3_64.
- Kalunga, E. K., Chevallier, S., Barthlemy, Q., Djouani, K., Monacelli, E., and Hamam, Y. (2016). Online SSVEP-based BCI using Riemannian geometry. *Neurocomputing*.
- Kalunga, E. K., Chevallier, S., Rabreau, O., and Monacelli, E. (2014). Hybrid interface: Integrating BCI in multimodal human-machine interfaces. In *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 530–535.
- Kanal, L. and Chandrasekaran, B. (1971). On dimensionality and sample size in statistical pattern classification. *Pattern recognition*, 3(3):225–234.
- Kang, H., Nam, Y., and Choi, S. (2009). Composite common spatial pattern for subject-to-subject transfer. *Signal Processing Letters, IEEE*, 16(8):683–686.
- Karcher, H. (2014). Riemannian center of mass and so called karcher mean. *arXiv preprint arXiv:1407.2087*.
- Käthner, I., Ruf, C. A., Pasqualotto, E., Braun, C., Birbaumer, N., and Halder, S. (2013). A portable auditory P300 braincomputer interface with directional cues. *Clinical Neurophysiology*, 124(2):327–338.
- Kaufmann, T., Holz, E. M., and Kbler, A. (2013a). Comparison of tactile, auditory, and visual modality for brain-computer interface use: a case study with a patient in the locked-in state. *Frontiers in Neuroscience*, 7:129.
- Kaufmann, T., Schulz, S. M., Grnzinger, C., and Kbler, A. (2011). Flashing characters with famous faces improves ERP-based braincomputer interface performance. *Journal of neural engineering*, 8(5):056016.

- Kaufmann, T., Schulz, S. M., Kblitz, A., Renner, G., Wessig, C., and Kbler, A. (2013b). Face stimuli effectively prevent braincomputer interface inefficiency in patients with neurodegenerative disease. *Clinical Neurophysiology*, 124(5):893–900.
- Kennedy, P. R. and Bakay, R. A. (1998). Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport*, 9(8):1707–1711.
- Kennedy, P. R., Bakay, R. A., Moore, M. M., Adams, K., and Goldwaithe, J. (2000). Direct control of a computer from the human central nervous system. *Rehabilitation Engineering, IEEE Transactions on*, 8(2):198–202.
- Kennedy, P. R., Kirby, M. T., Moore, M. M., King, B., and Mallory, A. (2004). Computer control using human intracortical local field potentials. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 12(3):339–344.
- Khorshidtalab, A. and Salami, M. J. E. (2011). EEG signal classification for real-time brain-computer interface applications: A review. *Mechatronics (ICOM), 2011 4th International Conference On*, pages 1–7.
- Kimura, Y., Tanaka, T., Higashi, H., and Morikawa, N. (2013). SSVEP-based braincomputer interfaces using FSK-modulated visual stimuli. *IEEE Transactions on Biomedical Engineering*, 60(10):2831–2838.
- Koles, Z. J., Lazar, M. S., and Zhou, S. Z. (1990). Spatial patterns underlying population differences in the background EEG. *Brain topography*, 2(4):275–284.
- Kottaimalai, R., Rajasekaran, M. P., Selvam, V., and Kannapiran, B. (2013). EEG signal classification using principal component analysis with neural network in brain computer interface applications. In *Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 2013 International Conference on*, pages 227–231. IEEE.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105.
- Krusienski, D. J., Grosse-Wentrup, M., Galn, F., Coyle, D., Miller, K. J., Elliott Forney, and Anderson, C. W. (2011). Critical issues in state-of-the-art braincomputer interface signal processing. *Journal of Neural Engineering*, 8(2):025002.

- Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2008). Toward Enhanced P300 Speller Performance. *Journal of Neuroscience Methods*, 167(1):15–21.
- Kübler, A., Winter, S., Ludolph, A. C., Hautzinger, M., and Birbaumer, N. (2005). Severity of depressive symptoms and quality of life in patients with amyotrophic lateral sclerosis. *Neurorehabilitation and neural repair*, 19(3):182–193.
- Kumar, A., Mohanty, M., and Routray, A. (2010). Design of Support Vector Machines with Time Frequency Kernels for classification of EEG signals. *Students' Technology Symposium (TechSym), 2010 IEEE*, pages 330–333.
- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411.
- Lee, H. and Choi, S. (2003). PCA+HMM+SVM for EEG pattern classification. *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, 1:541–544.
- Leeb, R., Sagha, H., Chavarriaga, R., and del R Millan, J. (2010). Multi-modal Fusion of Muscle and Brain Signals for a Hybrid-BCI. *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4343–4346.
- Lenhardt, A., Kaper, M., and Ritter, H. (2008). An Adaptive P300-Based Online Brain Computer Interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16(2):121–130.
- Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., and Moran, D. W. (2004). A brain-computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering*, 1(2):63–71.
- Li, M. and Lu, B.-L. (2009). Emotion classification based on gamma-band eeg. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 1223–1226. IEEE.
- Li, Y., Wong, K., and De Bruin, H. (2012). Electroencephalogram signals classification for sleepstate decision: A Riemannian geometry approach. *Signal Processing, IET*, 6(4):288–299.

- Li, Y., Wong, K. M., and De Bruin, H. (2009). EEG signal classification based on a Riemannian distance measure. In *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, pages 268–273. IEEE.
- Liang, S.-F., Shaw, F.-Z., Young, C.-P., Chang, D.-W., and Liao, Y.-C. (2010). A closed-loop brain computer interface for real-time seizure detection and control. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 4950–4953. IEEE.
- Lim, Y. and Pálfa, M. (2012). Matrix power means and the Karcher mean. *Journal of Functional Analysis*, 262(4):1498–1514.
- Lin, Z., Zhang, C., Wu, W., and Gao, X. (2006). Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs. *Biomedical Engineering, IEEE Transactions on*, 53(12):2610–2614.
- Liu, Y. H., Cheng, C. A., and Huang, H.-P. (2011). Novel feature of the EEG based motor imagery BCI system: Degree of imagery. *System Science and Engineering (ICSSE), 2011 International Conference on*, pages 515–520.
- Looney, D., Kidmose, P., Park, C., Ungstrup, M., Rank, M. L., Rosenkranz, K., and Mandic, D. P. (2012). The in-the-ear recording concept: User-centered and wearable brain monitoring. *Pulse, IEEE*, 3(6):32–42.
- Lopez-Gordo, M., Pelayo, F., and Prieto, A. (2010). A high performance SSVEP-BCI without gazing. *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–5.
- Lotte, F. (2011). Generating artificial EEG signals to reduce BCI calibration time. In *5th International Brain-Computer Interface Workshop*, pages 176–179.
- Lotte, F., Bougrain, L., and Clerc, M. (2015). Electroencephalography (EEG)-Based BrainComputer Interfaces. *Wiley Encyclopedia of Electrical and Electronics Engineering*.
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2):R1.
- Lotte, F., Van Langhenhove, A., Lamarche, F., Ernest, T., Renard, Y., Arnaldi, B., and Lcuyer, A. (2010). Exploring large virtual environments by thoughts

- using a brain-computer interface based on motor imagery and high-level commands. *Presence: teleoperators and virtual environments*, 19(1):54–70.
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Mak, J. N., Arbel, Y., Minett, J. W., McCane, L. M., Yuksel, B., Ryan, D., Thompson, D., Bianchi, L., and Erdogmus, D. (2011). Optimizing the P300-based braincomputer interface: current status, limitations and future directions. *Journal of neural engineering*, 8(2):025003.
- Martin, H., Chevallier, S., and Monacelli, E. (2012). Fast calibration of hand movement-based interface for arm exoskeleton control. In *European Symposium on Artificial Neural Networks (ESANN)*, pages 573–578.
- Matthews, P. M. and Jezzard, P. (2004). Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(1):6–12.
- Mattout, J., Perrin, M., Bertrand, O., and Maby, E. (2013). Improving non-invasive BCI for possible clinical application: Example of the P300-speller. *Annals of Physical and Rehabilitation Medicine*, 56:e374.
- Mellinger, J., Schalk, G., Braun, C., Preissl, H., Rosenstiel, W., Birbaumer, N., and Kbler, A. (2007). An MEG-based Brain-Computer Interface (BCI). *NeuroImage*, 36(3):581–593.
- melomind (2016). Melomind |.
- Mensia (2016). Mensia Technologies.
- Millan, J., Galan, F., Vanhooydonck, D., Lew, E., Philips, J., and Nuttin, M. (2009). Asynchronous non-invasive brain-actuated control of an intelligent wheelchair. *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 3361–3364.
- Millan, J. D. R., Renkens, F., Mourino, J., and Wulfram Gerstner (2004). Non-invasive brain-actuated control of a mobile robot by human EEG. *IEEE Transactions on Biomedical Engineering*, 51:1026–1033.
- Millan, J. d. R., Rupp, R., Mueller-Putz, G., Murray-Smith, R., Giugliemma, C., Tangermann, M., Vidaurre, C., Cincotti, F., Kubler, A., Leeb, R., Neuper, C., Mueller, K. R., and Mattia, D. (2010). Combining Brain-Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges. *Frontiers in Neuroscience*, 4(161).

- Milln, J. d. R. and Carmena, J. (2010). Invasive or noninvasive: understanding brain-machine interface technology. *IEEE Engineering in Medicine and Biology Magazine*, 29(EPFL-ARTICLE-150426):16–22.
- Miltner, W. H. R., Braun, C. H., and Coles, M. G. H. (1997). Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task: Evidence for a Generic Neural System for Error Detection. *Journal of Cognitive Neuroscience*, 9(6):788–798.
- Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747.
- Morgan, S. T., Hansen, J. C., and Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences*, 93(10):4770–4774.
- Müller, M. M., Keil, A., Gruber, T., and Elbert, T. (1999). Processing of affective pictures modulates right-hemispheric gamma band eeg activity. *Clinical Neurophysiology*, 110(11):1913–1920.
- Muller-Putz, G. R. and Pfurtscheller, G. (2008). Control of an Electrical Prosthesis With an SSVEP-Based BCI. *IEEE Transactions on Biomedical Engineering*, 55(1):361–364.
- Musallam, S., Corneil, B. D., Greger, B., Scherberger, H., and Andersen, R. A. (2004). Cognitive Control Signals for Neural Prosthetics. *Science*, 305(5681):258–262.
- Mller, M. M., Andersen, S., Trujillo, N. J., Valds-Sosa, P., Malinowski, P., and Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 103(38):14250–14254.
- Nakanishi, M., Wang, Y., Wang, Y.-T., Mitsukura, Y., and Jung, T.-P. (2014). A high-speed brain speller using steady-state visual evoked potentials. *International journal of neural systems*, 24(06):1450019.
- Naseer, N. and Hong, K.-S. (2013). Classification of functional near-infrared spectroscopy signals corresponding to the right- and left-wrist motor imagery for development of a braincomputer interface. *Neuroscience Letters*, 553:84–89.

- Nicolas-Alonso, Fernando, L., and Gomez-Gil, J. (2012). Brain Computer Interfaces, a Review. *Sensors*, 12(2):1211–1279.
- Nielsen, F. and Bhatia, R. (2012). *Matrix Information Geometry*. Springer Publishing Company, Incorporated.
- Nielsen, F. and Nock, R. (2009). Sided and symmetrized Bregman centroids. *IEEE transactions on Information Theory*, 55(6):2882–2904.
- Nielsen, F., Nock, R., and Amari, S.-i. (2014). On clustering histograms with k-means by using mixed -divergences. *Entropy*, 16(6):3273–3301.
- Nunez, P. L. and Srinivasan, R. (2005). *Electric Fields of the Brain: The Neurophysics of EEG, 2nd Edition*. Oxford University Press, USA, 2 edition.
- Obermaier, B., Guger, C., Neuper, C., and Pfurtscheller, G. (2001). Hidden Markov models for online classification of single trial EEG data. *Pattern Recognition Letters*, 22(12):1299–1309.
- Oralhan, Z. and Tokmaki, M. (2016). The Effect of Duty Cycle and Brightness Variation of Visual Stimuli on SSVEP in Brain Computer Interface Systems. *IETE Journal of Research*, pages 1–9.
- Ozmen, N. and Ktu, L. (2011). Discrimination between mental and motor tasks of EEG signals using different classification methods. *Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on*, pages 143–147.
- Paetau, R. (2002). Magnetoencephalography in pediatric neuroimaging. *Developmental Science*, 5(3):361–370.
- Pan, J., Gao, X., Duan, F., Yan, Z., and Gao, S. (2011). Enhancing the classification accuracy of steady-state visual evoked potential-based brain-computer interfaces using phase constrained canonical correlation analysis. *Journal of neural engineering*, 8(3):036027.
- Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.
- Panicker, R. C., Puthusserypady, S., and Sun, Y. (2010). Adaptation in P300 braincomputer interfaces: A two-classifier cotraining approach. *Biomedical Engineering, IEEE Transactions on*, 57(12):2927–2935.

- Pascal, F., Forster, P., Ovarlez, J. P., and Arzabal, P. (2005). Theoretical analysis of an improved covariance matrix estimator in non-Gaussian noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4.
- Pastor, M. A. (2003). Human cerebral activation during steady-state visual-evoked responses. *The Journal of Neuroscience*, 23(37):11621–11627.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-
sos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pencina, M. J., D’Agostino, R. B., and Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in medicine*, 27(2):157–172.
- Pennec, X., Fillard, P., and Ayache, N. (2006). A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66.
- Penny, W., Roberts, S., Curran, E., and Stokes, M. (2000). EEG-based communication: a pattern recognition approach. *Rehabilitation Engineering, IEEE Transactions on*, 8(2):214–215.
- Perrin, M., Maby, E., Daligault, S., Bertrand, O., and Mattout, J. (2012). Objective and subjective evaluation of online error correction during P300-based spelling. *Advances in Human-Computer Interaction*, 2012:4.
- Pfurtscheller, G. (1977). Graphical display and statistical evaluation of event-related desynchronization (ERD). *Electroencephalography and Clinical Neurophysiology*, 43(5):757–760.
- Pfurtscheller, G. and Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and Clinical Neurophysiology*, 42(6):817–826.
- Pfurtscheller, G., Graimann, B., Huggins, J. E., Levine, S. P., and Schuh, L. A. (2003). Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement. *Clinical neurophysiology*, 114(7):1226–1236.

- Pfurtscheller, G. and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110(11):1842–1857.
- Pfurtscheller, G. and Neuper, C. (1994). Event-related synchronization of mu rhythm in the EEG over the cortical hand area in man. *Neuroscience Letters*, 174(1):93–96.
- Pfurtscheller, G. and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123–1134.
- Pfurtscheller, G., Solis-Escalante, T., Ortner, R., Linortner, P., and Muller-Putz, G. (2010). Self-Paced Operation of an SSVEP-Based Orthosis With and Without an Imagery-Based Brain Switch: A Feasibility Study Towards a Hybrid BCI. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(4):409–414.
- Pfurtscheller, G., Stanck, A., and Edlinger, G. (1997). On the existence of different types of central beta rhythms below 30 Hz. *Electroencephalography and Clinical Neurophysiology*, 102(4):316–325.
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128–2148.
- Popescu, F., Fazli, S., Badower, Y., Blankertz, B., and Mller, K.-R. (2007). Single trial classification of motor imagination using 6 dry EEG electrodes. *PloS one*, 2(7):e637.
- Power, S. D., Kushki, A., and Chau, T. (2012). Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state NIRS-BCI. *BMC Research Notes*, 5:141.
- Proverbio, A. M. and Zani, A. (2003). Electromagnetic manifestations of mind and brain. *The cognitive electrophysiology of mind and brain*, 2:13–37.
- Purves, D. (2008). *Neuroscience, Fourth Edition*. Sinauer Associates, Inc., 4th edition edition.
- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., and Williams, S. M., editors (2001). *Neuroscience*, chapter Excitatory and Inhibitory Postsynaptic Potentials. Sinauer Associates.

- Rakotomamonjy, A., Guigue, V., Mallet, G., and Alvarado, V. (2005). Ensemble of SVMs for improving brain-computer interface P300 speller performances. *15th International Conference on Artificial Neural Networks*, pages 45–50.
- Raudys, S. J. and Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(3):252–264.
- Ritaccio, A., Beauchamp, M., Bosman, C., Brunner, P., Chang, E., Crone, N., Gunduz, A., Gupta, D., Knight, R., Leuthardt, E., and others (2012). Proceedings of the Third International Workshop on Advances in Electrocorticography. *Epilepsy & Behavior*, 25(4):605–613.
- Ritter, W. and Vaughan, H. G. (1969). Averaged evoked responses in vigilance and discrimination: a reassessment. *Science*, 164(3877):326–8.
- Rivet, B., Cecotti, H., Souloumiac, A., Maby, E., and Mattout, J. (2011). Theoretical analysis of xDAWN algorithm: application to an efficient sensor selection in a P300 BCI. In *Signal Processing Conference, 2011 19th European*, pages 1382–1386. IEEE.
- Rivet, B., Souloumiac, A., Attina, V., and Gibert, G. (2009). xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface. *Biomedical Engineering, IEEE Transactions on*, 56(8):2035–2043.
- Roland, P., Larsen, B., Lassen, N., and Skinhøj, E. (1980). Supplementary motor area and other cortical areas in organization of voluntary movements in man. *Neurophysiology*, 43(1):118–136.
- Samek, W., Blythe, D., Müller, K.-R., and Kawanabe, M. (2013). Robust spatial filtering with beta divergence. In *Advances in Neural Information Processing Systems*, pages 1007–1015.
- Samek, W. and Müller, K.-R. (2014). Information geometry meets BCI spatial filtering using divergences. In *Brain-Computer Interface (BCI), 2014 International Winter Workshop on*, pages 1–4. IEEE.
- Santhanam, G., Ryu, S. I., Yu, B. M., Afshar, A., and Shenoy, K. V. (2006). A high-performance braincomputer interface. *Nature*, 442(7099):195–198.
- Schalk, G. and Leuthardt, E. C. (2011). Brain-computer interfaces using electrocorticographic signals. *Biomedical Engineering, IEEE Reviews in*, 4:140–154.

- Scherer, R., Mller-Putz, G. R., and Pfurtscheller, G. (2007). Self-initiation of EEG-based brain-computer communication using the heart rate response. *Journal of Neural Engineering*, 4(4):L23.
- Schettini, F., Aloise, F., Aric, P., Salinari, S., Mattia, D., and Cincotti, F. (2014). Self-calibration algorithm in an asynchronous P300-based braincomputer interface. *Journal of neural engineering*, 11(3):035004.
- Schlgl, A., Lee, F., Bischof, H., and Pfurtscheller, G. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *Journal of Neural Engineering*, 2(4):14–22.
- Scholkopf, B. and Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Schoot, G. (1993). Penfield’s homunculus: a note on cerebral cartography. *Journal of Neurology, Neurosurgery, and Psychiatry*, 56(4):329–333.
- Schfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1).
- Sellers, E. W., Krusienski, D. J., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2006). A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. *Biological Psychology*, 73(3):242–252.
- Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., and Donoghue, J. P. (2002). Brain-machine interface: Instant neural control of a movement signal. *Nature*, 416(6877):141–142.
- Silvoni, S., Ramos-Murgialday, A., Cavinato, M., Volpato, C., Cisotto, G., Turolla, A., Piccione, F., and Birbaumer, N. (2011). Brain-computer interface in stroke: a review of progress. *Clinical EEG and Neuroscience*, 42(4):245–252.
- Simard, P. Y., Steinkraus, D., and Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *2013 12th International Conference on Document Analysis and Recognition*, volume 2, pages 958–958. IEEE Computer Society.
- Singer, W. (1993). Synchronization of Cortical Activity and its Putative Role in Information Processing and Learning. *Annual Review of Physiology*, 55(1):349–374.

- Sitaram, R., Caria, A., Veit, R., Gaber, T., Rota, G., Kuebler, A., and Birbaumer, N. (2007). fMRI Brain-Computer Interface: A Tool for Neuroscientific Research and Treatment. *Computational Intelligence and Neuroscience*, 2007.
- Sitaram, R., Weiskopf, N., Caria, A., Veit, R., Erb, M., and Birbaumer, N. (2008). fMRI Brain-Computer Interfaces. *IEEE Signal Processing Magazine*, 25(1):95–106.
- Spüler, M., Rosenstiel, W., and Bogdan, M. (2012). One class SVM and Canonical Correlation Analysis increase performance in a c-VEP based Brain-Computer Interface (BCI). In *Proceedings of 20th European Symposium on Artificial Neural Networks (ESANN 2012), Bruges, Belgium*, volume 4, pages 103–108.
- Spüler, M., Walter, A., Ramos-Murguialday, A., Naros, G., Birbaumer, N., Gharabaghi, A., Rosenstiel, W., and Bogdan, M. (2014). Decoding of motor intentions from epidural ecog recordings in severely paralyzed chronic stroke patients. *Journal of neural engineering*, 11(6):066008.
- Sra, S. (2016). Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society*, 144(7):2787–2797.
- Sturm, I., Bach, S., Samek, W., and Müller, K.-R. (2016). Interpretable Deep Neural Networks for Single-Trial EEG Classification. *arXiv preprint arXiv:1604.08201*.
- Sutter, E. E. (1992). The brain response interface: communication through visually-induced electrical brain responses. *Journal of Microcomputer Applications*, 15(1):31–45.
- Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential Correlates of Stimulus Uncertainty. *Science*, 150:1187–1188.
- Sutton, S., Tueting, P., Zubin, J., and John, E. R. (1967). Information Delivery and the Sensory Evoked Potential. *Science*, 155(3768):1436–1439.
- Szuromi, B., Czobor, P., Komlosi, S., and Bitter, I. (2010). P300 deficits in adults with attention deficit hyperactivity disorder: a meta-analysis. *Psychol Med.*, 41(7):1529–1538.
- Taylor, D. M., Tillery, S. I. H., and Schwartz, A. B. (2002). Direct Cortical Control of 3d Neuroprosthetic Devices. *Science*, 296(5574):1829–1832.

- Team, E. W. (2016). Elekta Neuromag TRIUX Functional Mapping.
- Toro, C. and Deuschl, G and Thatcher, R and Sato, S. and Kufta, C and Hallett, M. (1994). Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 93(5):380–389.
- Treder, M. S., Schmidt, N. M., and Blankertz, B. (2011). Gaze-independent braincomputer interfaces based on covert attention and feature attention. *Journal of neural engineering*, 8(6):066003.
- Van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1).
- Van Erp, J. B., Lotte, F., and Tangermann, M. (2012). Brain-computer interfaces: beyond medical applications. *Computer*, 45(4):26–34.
- Verschore, H., Kindermans, P.-J., Verstraeten, D., and Schrauwen, B. (2012). Dynamic stopping improves the speed and accuracy of a P300 speller. In *Artificial Neural Networks and Machine Learning-ICANN 2012*, pages 661–668. Springer.
- Vialatte, F.-B., Maurice, M., Dauwels, J., and Cichocki, A. (2010). Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Progress in neurobiology*, 90(4):418–438.
- Vidal, J. J. (1973). Toward direct brain-computer communication. *Annual review of biophysics and bioengineering*, 2(1):157–180.
- Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer Science & Business Media.
- Wang, L., Xu, G., Wang, J., Yang, S., and Yan, W. (2011). Motor Imagery BCI Research Based on Hilbert-Huang Transform and Genetic Algorithm. *Bioinformatics and Biomedical Engineering, (iCBBE) 2011 5th International Conference on*, pages 1–4.
- Wang, P., Lu, J., Zhang, B., and Tang, Z. (2015). A review on transfer learning for brain-computer interface classification. In *Information Science and Technology (ICIST), 2015 5th International Conference on*, pages 315–322. IEEE.
- Wang, S. and James, C. J. (2006). Enhancing Evoked Responses for BCI Through Advanced ICA Techniques. In *Advances in Medical, Signal and Information Processing (MEDSIP)*, pages 1–4.

- Wang, W., Collinger, J. L., Degenhart, A. D., Tyler-Kabara, E. C., Schwartz, A. B., Moran, D. W., Weber, D. J., Wodlinger, B., Vinjamuri, R. K., Ashmore, R. C., Kelly, J. W., and Boninger, M. L. (2013). An Electrocorticographic Brain Interface in an Individual with Tetraplegia. *PLoS ONE*, 8(2).
- Wang, Y. and Jung, T.-P. (2013). Improving Brain-computer Interfaces Using Independent Component Analysis. *Towards Practical Brain-Computer Interfaces*, pages 67–83.
- Wilson, J., Felton, E., Garell, P., Schalk, G., and Williams, J. (2006). ECoG factors underlying multimodal control of a brain-computer interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 14(2):246–250.
- Wolpaw, J. R., Birbaumer, N., Heetderks, W., McFarland, D. J., Peckham, P., Schalk, G., Donchin, E., Quatrano, L., Robinson, C., and Vaughan, T. (2000). Brain-computer interface technology: a review of the first international meeting. *Rehabilitation Engineering, IEEE Transactions on*, 8(2):164–173.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791.
- Wolpaw, J. R. and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17849–17854.
- Xie, Y., Ho, J., and Vemuri, B. (2013). On A Nonlinear Generalization of Sparse Coding and Dictionary Learning. In *Proceedings of the 30th International Conference on Machine Learning*, page 1480. NIH Public Access.
- Yan, T., Jingtian, T., Andong, G., and Wei, W. (2008). Classifying EEG Signals Based HMM-AR. *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, pages 2111–2114.
- Yang, Y., Chevallier, S., Wiart, J., and Bloch, I. (2014). Time-frequency optimization for discrimination between imagination of right and left hand movements based on two bipolar electroencephalography channels. *EURASIP Journal on Advances in Signal Processing*, 2014(1):1–18.
- Yger, F. (2013). A review of kernels on covariance matrices for BCI applications. In *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, pages 1–6. IEEE.

- Yin, E., Zhou, Z., Jiang, J., Chen, F., Liu, Y., and Hu, D. (2013). A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm. *Journal of Neural Engineering*, 10(2).
- Yoo, S.-S., Fairneny, T., Chen, N.-K., Choo, S.-E., Panych, L. P., Park, H., Lee, S.-Y., and Jolesz, F. A. (2004). Braincomputer interface using fMRI: spatial navigation by thoughts. *Neuroreport*, 15(10):1591–1595.
- Yu, X., Chum, P., and Sim, K.-B. (2014). Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system. *Optik-International Journal for Light and Electron Optics*, 125(3):1498–1502.
- Yuksel, B. F., Afergan, D., Peck, E. M., Griffin, G., Harrison, L., Chen, N. W., Chang, R., and Jacob, R. J. (2015). Implicit Brain-Computer Interaction Applied to a Novel Adaptive Musical Interface. Technical report, Technical Report TR-2015-01, Department of Computer Science, Tufts University, 2015. www.cs.tufts.edu/Technical-Reports.html.
- Zander, T. O. and Kothe, C. (2011). Towards passive braincomputer interfaces: applying braincomputer interface technology to humanmachine systems in general. *Journal of Neural Engineering*, 8(2):025005+.
- Zhang, H., Guan, C., Ang, K. K., and Wang, C. (2012). BCI Competition IV Data Set I: Learning Discriminative Patterns for Self-Paced EEG-Based Motor Imagery Detection. *Frontiers in Neuroscience*, 6.
- Zhang, H., Guan, C., and Wang, C. (2008). Asynchronous P300-Based Brain-Computer Interfaces: A Computational Approach With Statistical Models. *Biomedical Engineering, IEEE Transactions on*, 55(6):1754–1763.
- Zhang, J.-C., Xu, Y.-Q., and Yao, L. (2007). P300 Detection Using Boosting Neural Networks with Application to BCI. *Complex Medical Engineering, 2007. CME 2007. IEEE/ICME International Conference on*, pages 1526–1530.
- Zhang, Y., Ji, X., and Zhang, Y. (2015a). Classification of EEG signals based on AR model and approximate entropy. In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–6. IEEE.
- Zhang, Y., Zhang, Y., Wang, J., and Zheng, X. (2015b). Comparison of classification methods on EEG signals based on wavelet packet decomposition. *Neural Computing and Applications*, 26(5):1217–1225.

Zhu, D., Bieger, J., Molina, G. G., and Aarts, R. M. (2010). A survey of stimulation methods used in SSVEP-based BCIs. *Intell. Neuroscience*, 2010:1–12.