UNIVERSIDADE D
COIMBRA

Aniana da Rosa de Brito da Cruz

MENTAL STATE MONITORING FOR IMPROVING THE BRAIN-COMPUTER INTERFACE ROBUSTNESS IN A HUMAN-CENTERED COLLABORATIVE CONTROL

Tese no âmbito do Doutoramento em Engenharia Biomédica orientada pelo Professor Doutor Urbano José Carreira Nunes e Professor Doutor Gabriel Pereira Pires e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Dezembro de 2020

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Mental state monitoring for improving the Brain-Computer Interface robustness in a human-centered collaborative control

Aniana da Rosa de Brito da Cruz

Thesis to obtain the degree of Doctor of Philosophy in

Biomedical Engineering

Supervisors:
Prof. Urbano José Carreira Nunes
Prof. Gabriel Pereira Pires

December 2020

Acknowledgments

To complete this thesis I had help from several persons at different levels and without whom this project would not be possible. So I would like to take this opportunity to express my gratitude to all these people.

First of all, I want to express my deepest gratitude to my supervisors, Prof. Gabriel Pires and Prof. Urbano Nunes. To Prof. Gabriel Pires, I would like to thank him for his dedicated guidance, motivation, support, and friendship. His extensive knowledge gave to me excellent insight to conduct this research. To Prof. Urbano Nunes I would like to thank him for his guidance, motivation, support, and friendship, he always finds time for discussions.

I would like to express my deepest gratitude to my entire family and in particular to my husband Aricson Cruz for always being there for me over all these years, with encouragement, patience, and love. To my lovely son Eric Cruz, who is our pride and joy. To my parents Olívio Brito and Cicília Brito who support me in every moment of my life. Special thanks to my aunt Augusta Barros who sadly died this year.

I would like to thank Prof. Ana Lopes for her support and friendship, in particular for the aid in the experiment with RobChair.

I would like to thank all my lab friends to be always ready to help me with any difficulty. Special thanks to João Perdiz who participated in almost all experimental tests that lasted for many hours, and Jorge Perdigão who helped me in the experiment with RobChair.

I would like to thank all the people who participated in the experiments, especially the participants from Associação de Paralisia Cerebral de Coimbra (APCC). I would like to thank all the staff of the APCC, specially Doutor Carlos Carona for the entire support during the experimental tests.

This study was conducted at the Institute of Systems and Robotics (ISR), which provided all the equipment resources and frameworks used during this research. I would like to thank all the staff of the ISR.

This thesis was financially supported by Portuguese foundation for science and technology (FCT) scholarship SFRH/BD/111473/2015 and under grants B-RELIABLE: PTDC/EEI-AUT/30935/2017, AMSHMI12: RECI/EEI-AUT/0181/2012, and by Institute for Systems and Robotics UID/EEA/00048/2019.

Abstract

A brain-computer interface (BCI) is a useful device for people with severe motor disabilities, and several BCI applications and devices have already been validated, such as communication spellers, wheelchairs, prosthetic devices, etc. However, BCI still has a very limited application in daily real-world tasks due to its low speed, low reliability, and other practical aspects, such as the high user's workload imposed by continuous focus, and the need for calibration in every session. BCI researchers have been striving to improve the reliability as well as other usability issues. The main goal of this PhD thesis was to research and develop new approaches and new methods to improve the reliability and usability of current BCI systems. This thesis contributes to the BCI field mainly on three topics.

The first main contribution is a novel approach based on double error-related potential (ErrP) detection for automatic error correction. We introduce a second ErrP to validate automatic correction. The approach demonstrates the possibility of using ErrPs in a closed-loop human-computer interaction, allowing the user to change or confirm system decisions. The proposed approach was assessed offline and online through a set of experimental tests. Results showed that the proposed approach is effective with a significant increase in classification accuracy and information transfer rate.

Secondly, a self-paced P300-based brain-controlled wheelchair (BCW) combined with a collaborative-controller and dynamic-time commands is proposed. The feasibility of this approach was analysed by measuring the impact of these three features on the system reliability, naturalness of interaction, and users' effort. The proposed method was validated through extensive experiments conducted with able-bodied individuals and individuals with severe motor disabilities. Results showed high feasibility for both able-bodied and motor-disabled participants.

The third main issue researched in this thesis is focused on model generalization across session and cross-subject variability. To achieve a good performance, the classification model is usually built from calibration data recorded at the beginning of each session. A novel statistical spatial filter is proposed that takes advantage of the Riemannian distance to extract features that are robust to the non-stationarity of Electroencephalography (EEG) signals. The results show that the proposed method improves gener-

alization across sessions and it is robust to the variation of the amount of training data.

Keywords

Brain-computer interface, electroencephalogram, P300, double error-related potentials, self-paced, dynamic time-window, collaborative control, spatial filter, Riemannian geometry, generalization, speller, brain-controlled wheelchair

Resumo

Uma interface cérebro-computador (ICC) é um dispositivo útil para pessoas com deficiências motoras graves e vários dispositivos e aplicações já foram validados, como por exemplo soletradores para comunicação, cadeiras de rodas, dispositivos protéticos, etc. No entanto, a aplicação das ICCs em tarefas do dia-a-dia ainda é limitada devido à sua baixa velocidade, baixa fiabilidade e outros aspetos práticos, como a alta carga de trabalho mental do utilizador imposta pelo foco contínuo e a necessidade de calibrar o sistema em cada sessão. Os investigadores das ICCs têm se esforçado para melhorar a fiabilidade e também outras questões relacionadas com usabilidade. Esta tese pretende investigar e desenvolver novas abordagens e novos métodos para melhorar a fiabilidade e usabilidade dos sistemas ICC. Esta tese contribui para a área das ICCs principalmente em três tópicos.

A primeira contribuição é uma nova abordagem baseada na deteção dupla do potencial de erro (ErrP) para correção automática de erros. Introduzimos um segundo ErrP para validar a correção automática. A abordagem demonstra a possibilidade de usar ErrPs numa interação homem-computador em malha fechada, permitindo ao utilizador alterar ou confirmar as decisões do sistema. A abordagem proposta foi avaliada offline e online através de um conjunto de testes experimentais. Os resultados demostraram que a abordagem proposta é eficaz com um aumento significativo na precisão da classificação e na taxa de transferência de informação.

Em segundo lugar, é proposta uma interface aplicada a uma cadeira de rodas atuada pelo cérebro baseada no potencial P300 ao ritmo do utilizador combinado com um controlador colaborativo e comandos com janela de tempo dinâmico. A viabilidade dessa abordagem foi testada através da medição do impacto dessas três características na fiabilidade do sistema, na naturalidade da interação e no esforço dos utilizadores. O desempenho dos métodos propostos foi avaliado através de um extenso conjunto de experiências realizadas com indivíduos saudáveis e indivíduos portadores de deficiência motora grave. Os resultados evidenciaram alta fiabilidade tanto para os participantes saudáveis como para os participantes com deficiência motora.

A terceira questão abordada nesta tese está focada na generalização do modelo de classificação ErrP entre sessões e entre utilizadores. Para se obter um bom desempenho, o modelo de classificação é normalmente construído com os dados de calibração obtidos no início de cada sessão. É proposto um novo método de filtro espacial que tira partido da distância Riemanniana para extrair características

que são robustas à não estacionariedade dos sinais eletroencefalográfico (EEG). Pode concluir-se dos resultados que o método proposto aumenta a generalização entre as sessões e é robusto à variação da quantidade de dados de treino.

Palavras Chave

Interface cérebro-computador, eletroencefalografia, P300, duplo potencial de erro, ao ritmo do utilizador, intervalo dinâmico, controlo colaborativo, filtro espacial, geometria Riemanniana, generalização, soletradores, cadeira de rodas atuada pelo cérebro

Contents

List of Figures

List of Tables

List of Acronyms

ACC	Anterior cingulate cortex
AIRM	Affine invariant Riemannian metric
ANS	Autonomic nervous system
APCC	Cerebral Palsy Association of Coimbra
ASD	Autistic spectrum disorder
BCI	Brain-computer interface
BCW	Brain-controlled wheelchair
CC	Control commands
CNS	Central nervous system
CP	Cerebral palsy
CRN	Correct response negativity
CSP	Common spatial pattern
DTT	Dynamic trial time
ECoG	Electrocorticography
ECS	Error correction system
EEG	Electroencephalography
ERD	Event-related desynchronization
ERN	Error-related negativity
ERP	Event-related potential
ErrP	Error-related potential
ERS	Event-related synchronization
eSPM	Effective symbols per minute

FBPD	Feature-based peak detection
FC	Fisher criterion
FCB	Fisher criterion beamformer
FLD	Fisher linear discriminant
fMRI	Functional magnetic resonance imaging
fNIRS	Functional near-infrared spectroscopy
FRN	Feedback-related negativity
GED	Generalized eigenvalue decomposition
GSR	Galvanic skin response
INR	Intracortical neuron recording
ITI	Inter-trial interval
ITR	Information transfer rate
LIS	Locked-in state
LS	Local scaling
LSC	Lateral single character
MEG	Magnetoencephalography
MI	Motor imagery
NB	Naïve Bayes
NCC	Non-control commands
Nrep	Number of repetitions of each trial
Pe	Error positivity
RFCB	Riemannian Fisher criterion beamformer
SCI	Spinal cord injury
SCL	Skin conductance level
SCR	Skin conductance response
SMR	Sensorimotor rhythm
SNR	Signal-to-noise ratio
SPD	Symmetric and positive definite
SSVEP	Steady-state visual evoked potential
STT	Static trial time

TL	Transfer learning
TSSF	Tangent space spatial filter
VEP	Visual evoked potential
WT	Wrong target

1

Introduction

Contents

Brain-computer interface (BCI) translates the subject's intention expressed through brain signals into commands to control external applications such as spelling and robotic devices [Wolpaw et al., 2002]. People suffering from conditions that affect neuromuscular structures and functions tend to gradually lose the muscle activity responsible for speaking, walking, and general movement of the body, becoming dependent on human assistance in daily activities. There are many assistive devices, such as head and eye trackers technologies, speech-based interfaces, smart walkers, and devices controlled by adapted joysticks that can be used by motor-impaired people to communicate and control devices [Gibbons and Beneteau, 2010]. However, many disabled people become unable to use these conventional interfaces, which require some form of muscular activity, as a result of the physical ability deterioration [Frank et al., 2000]. For those with severe motor impairments (including complete Locked-in state (LIS)), BCI emerges as a potential solution since it does not use conventional neuromuscular output pathways. BCI is independent of neuromuscular activity, thus it can restore communication and control skills of individuals with severe motor impairments arising from neuromuscular disorders such as amyotrophic lateral sclerosis, muscular dystrophy, and Spinal cord injury (SCI). BCI technology is not limited to communication and control applications for motor disabled people. In the last decade, BCI research has evolved considerably in different directions, reaching new applications and new target users. The BNCI (Brain/Neural Computer Interaction) Horizon 2020 project has identified several application scenarios in which BCI can be used [Brunner et al., 2015], namely: replace natural Central nervous system (CNS) output (e.g., Brain-controlled wheelchair (BCW)), restore lost natural CNS output (e.g., BCI-controlled neuroprosthesis), enhance natural CNS output (e.g., enhanced user experience in computer games), supplement natural CNS output (e.g., augmented reality glasses controlled with BCI), improve natural CNS output (e.g., upper limb rehabilitation after stroke), and as a research tool in cognitive neurosciences.

The main goal of this thesis was to research methods to improve the usability of BCI systems, aiming to decrease their calibration time, increase their reliability and improve the human-machine interaction naturalness.

This chapter contains the context, motivation, and goals of the study. Then, we provide the main contributions, the outline of the developed work, and publications.

1.1 Context and Motivation

Since the question: "Can these observable electrical brain signals be put to work as carriers of information in man-computer communication or for the purpose of controlling such external apparatus as prosthetic devices or spaceships?" [Vidal, 1973], BCI has received tremendous interest and efforts from the research community, with possibly hundreds of active groups around the world [Nam et al., 2018, page 1]. Efforts have been undertaken to increase the performance of BCI systems in research

labs around the world as well as in our Human-Centered Mobile Robotics research group at ISR-UC, by exploiting: (a) different neural signals such as Motor imagery (MI) [Pichiorri et al., 2015], P300 Event-related potential (ERP) [Pires et al., 2012], slow cortical potentials [Birbaumer et al., 2000], and Visual evoked potential (VEP) such as steady state evoked potentials (SSVEP) [Chen et al., 2015], time modulated VEP, frequency modulated VEP, and pseudorandom code modulated VEP; (b) different signal processing and classification methods [Pires et al., 2011a, Lotte et al., 2007, Lotte et al., 2018]; and (c) different designs of paradigms [Pires et al., 2012, Treder et al., 2011, Barbosa et al., 2016]. Despite the many efforts that have been made to create a reliable and usable BCI, current BCIs are still very limited to experimental tests, facing many challenges in order to be transferred into real-world scenarios. Some of these challenges will be discussed in the next section.

1.1.1 EEG-based BCIs challenges

Broader applicability of EEG-based BCI for real-world scenarios requires a plug-and-play and a reliable system, which has proved difficult to achieve. EEG records brain signals from the scalp, so the quality of the signal is affected by many artifacts and the low conductivity of the skull causes the spread of brain signals and attenuates the amplitude of neuronal activity. These effects produce an EEG signal with a very low Signal-to-noise ratio (SNR). Therefore, it is a challenging process to accurately decode users' intentions using these noisy EEG data. Moreover, the EEG signal has a great variability within and across subjects. This imposes a subject-specific calibration, which is performed whenever BCI is used. For real-world applications, it is crucial to avoid recalibration procedures, which is time-consuming and annoying. Today's BCI systems also require a high level of concentration and attention, which imposes a high mental workload. This workload can cause attention shifts and fatigue, resulting in even greater uncertainty of the decoded user's intentions. Therefore, a natural interface/interaction is needed that adapts according to the user's mental state. BCIs also face a technological challenge related to wearability, and system setup and preparation. Usually, wet electrodes are used because they produce better classification performance, but this requires skin preparation and conductive gel, which is cumbersome to the user and needs a long setup time. Although dry electrodes are available, they provide lower SNR compared to wet electrodes, worsening BCI's performance and making all the challenges described above even more difficult.

1.2 Objectives and Main Contributions

This thesis focuses on the development of approaches to make BCI systems more reliable and more user-friendly. These approaches were validated in BCIs based on P300 event-related potentials, namely, in a BCI communication-speller and a brain-controlled wheelchair (RobChair). This thesis contributes to

the BCI research field in several ways:

1. **BCI reliability.** A major drawback of current BCIs, which limits their use outside laboratories, is the low reliability in recognizing the subject's intent [Kübler et al., 2015]. One way to increase the BCI reliability is through robust signal processing methods or by decreasing the transfer rate. BCI research strives for ways to improve it. The Error-related potential (ErrP), an ERP that is generated naturally in the brain when the user perceives that an error has been made by himself/herself or by the system [Falkenstein et al., 2000, Ferrez and Millán, 2008a], has been proposed aiming to improve BCI reliability [Ferrez and Millán, 2008a, Schalk et al., 2000, Zeyl et al., 2016]. ErrP appears within a time window of 500 ms, thus it is possible to detect/correct the errors in real-time. Yet, automatic ErrP detection is challenging because it needs to be detected at a single trial level and ErrP has a very low SNR. We addressed this issue using a new approach with two contributions [Cruz et al., 2018a]:

    - **Double ErrP detection.** This approach proposes an interaction process with two ErrPs instead of one to identify and correct errors. Thus, two single-trial responses, instead of one, contribute to the final selection, resulting in increased reliability in detecting errors. This approach showed on one hand the possibility of increasing the performance of the system mitigating the single-trial issue, and on the other hand showed that ErrPs could be used in a closed-loop, increasing the level of human-machine interaction which can be used in diverse non-BCI applications;
    - **P300 and ErrP feature level concatenation.** In order to increase the reliability of error detection, the error detector combines the ErrP elicited by the feedback with the ERPs of target and non-targets, taking advantage of their correlation with error occurrence. The proposed feature fusion significantly improved the accuracy of the error detector.

2. **Natural and Adaptive interface.** Another BCI issue that was addressed was the variability of the user's performance while controlling the BCI. BCI operation requires continuous high levels of attention and focus, and therefore performance may oscillate due to different reasons, such as attention shifts, fatigue, and stress. A system able to automatically adapt to the user can provide a reliable performance independently of the user's mental state. One way to achieve a more reliable system with a more natural interaction is through a self-paced control operation, which provides the possibility for users to send BCI commands only when they wish, at their own pace.

    - **Self-paced paradigm.** We implemented a self-paced approach that frees the user from being continuously attentive and providing commands. This improves the naturalness of interaction, decreases the mental effort, and consequently improves the BCI reliability.

Another way to adapt the BCI is through the monitorization of mental state markers, such as stress and attention levels. Two approaches were explored:

- **A self-paced BCI with Dynamic trial time (DTT).** We proposed a new dynamic time-window approach in which the number of stimuli sequences is continuously adapted to the user's current state during online BCI use, that allow balancing the reliability and speed of the BCI [Cruz et al., 2021].
- **Stress detection.** We investigate the possibility of detecting stressful situations in which the user's state can change, and automatically adjust the system. We analyze whether Galvanic skin response (GSR) can be used to detect emotional arousals such as stress, and the possibility of automatically detect GSR peaks to adapt the BCI system. For the automatic detection of Skin conductance response (SCR), we proposed a new method called Feature-based peak detection (FBPD) which proved to be very effective compared to state-of-the-art methods [Cruz et al., 2019].

3. **Single or zero calibration.** Due to the inherent non-stationarity of EEG signals, before each experimental session, BCI is calibrated to build the classification models, thus avoiding performance decay. This tedious re-calibration procedure is a limiting factor for real-world applications. Therefore, single-calibration or zero-calibration plays a crucial role in the use of BCIs in real contexts, outside the laboratory. We addressed the calibration issue in the context of ErrP classification, namely:

   - **ErrP Generalization analysis.** Automatic detection of ErrPs requires a long calibration. In order to reduce this calibration time, we analyzed whether a classification model obtained in a scenario with a high error-rate could generalize to a scenario with much less error-rate. We also analyzed the generalization of the classification model across sessions, that is, the classification model was built in one day and the validation session was held on a different day. The results demonstrated that the classification model generalizes well across error-rates [Cruz et al., 2018b].
   - **Subspace invariance based on Riemannian geometry.** We extended the statistical spatial filter based on the Fisher criterion (FC) proposed in [Pires et al., 2011a] to a Riemannian manifold [Jost and Jost, 2008, pages 1-45]. Using the invariance properties (e.g. invariance under congruent transformations) of Riemannian distance the proposed method allowed to re-use the data obtained from one session/subject to build the classification model to test in other sessions/subjects without great performance loss.

4. **Validation with clinical users.** Some of the proposed approaches were validated on individuals

with severe motor disabilities, particularly in the context of the BCI-wheelchair experiments, which is of most relevance to assess the effective use of BCI.

1.3 Thesis Outline

This thesis is divided into seven chapters including the present one, organized as follows:

- **Chapter 2.** This chapter is dedicated to the background information, essential to understand the content of this thesis. It starts with an introduction to BCI and a state-of-the-art. Then, it provides the brain signal acquisition techniques and neural signals for BCI control.
- **Chapter 3.** This chapter provides the BCI processing pipeline, namely, artifact rejection, feature extraction and selection, and classification methods. It also presents the most popular metrics used to measure BCI performance.
- **Chapter 4.** This chapter presents the proposed double ErrP detection and two case studies to assess the approach. The first case study was performed by able-bodied participants and one tetraplegic participant. The main goal of these experiments was to assess the feasibility of detecting ErrP to increase BCI performance. The second case-study analyses the generalization of the ErrP detector across error-rates.
- **Chapter 5.** This chapter contains approaches to increase the reliability and usability of the BCI system. Two case-studies were carried out by both healthy and severely motor-impaired individuals. In the first one, we validated the proposed P300-based BCW that combines in a single framework: self-paced control, dynamic adjustment of time window commands, and collaborative control to improve the overall reliability of the system with natural interaction and minimum-effort. In the second case-study, the GSR signal was used to detect stressful situations while driving a BCW.
- **Chapter 6.** This chapter presents the proposed method, called the Riemannian Fisher criterion beamformer (RFCB) for single/zero calibration. The method was assessed with the same datasets of chapter 4 and benchmark datasets.
- **Chapter 7.** This final chapter summarizes the main achievements of this thesis and suggests some future directions of investigation.

1.4 Publications

The results obtained in this thesis have been published in peer-reviewed journals and conferences.

Journal Articles Published

1. ”Double ErrP Detection for Automatic Error Correction in an ERP-Based BCI Speller” published in IEEE Transactions on Neural Systems and Rehabilitation Engineering, 99, January 2018 [Cruz et al., 2018a].

2. ”A self-paced BCI with a collaborative controller for highly reliable wheelchair driving: experimental tests with physically disabled individuals” published in IEEE Transactions on Human-Machine Systems [Cruz et al., 2021].

Journal Articles in preparation

1. Spatial filtering based on Riemannian distance to improve the generalization of ErrP classification (A. Cruz, G. Pires, and U. J. Nunes) in preparation for submission to a ISI journal.

Conference Articles published

1. ”Facial Expression Recognition based on EOG toward Emotion Detection for Human-Robot Interaction” published in 8th International Conference on Bio-inspired Systems and Signal Processing [Cruz et al., 2015],

2. ”Generalization of ErrP-calibration for different error-rates in P300-based BCIs” published in IEEE International Conference on Systems Man, and Cybernetics , SMC2018 [Cruz et al., 2018b]

3. ”Detection of Stressful Situations Using GSR While Driving a BCI-controlled Wheelchair” published in 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2019) [Cruz et al., 2019],

4. ”A Hybrid Brain-Computer Interface Fusing P300 ERP and Electrooculography” published in XV Mediterranean Conference on Medical and Biolgical Engineering and Computing (MEDICON 2019) [Perdiz et al., 2019].

Dataset Published

1. ”BCI double ErrP dataset: error-related potentials (primary and secondary ErrP) and P300 event related potentials” published online in IEEE DataPort [Cruz et al., 2020].

2

Neurophysiological background and state of the art

Contents

This chapter contains an overview of BCI systems, including neurophysiological and technical background for the rest of this thesis. We first introduce a BCI system and then we present some relevant historical events, current developments, future applications, and trends. Next, different methods used to measure brain signals are described with a special focus on EEG, which is the technique used in this study. This chapter also presents a short description of neural mechanisms used in EEG-based BCI with an emphasis on event-related potentials, in particular, P300 ERP and error-related potentials. Finally, the state-of-the-art of error-related potentials and the related work to the topics covered in this thesis are presented.

2.1 Brain-computer Interfaces

There is no single definition of a brain-computer interface and definitions have evolved to accommodate new advances and new applications. A commonly accepted definition is given by [Wolpaw and Wolpaw, 2012]: "A BCI is a system that measures CNS activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment." This means that BCI offers new output pathways that allow action without using any activity of peripheral nerves and muscles. As already mentioned in Chapter One, BCI can be used to control many applications (e.g robotic wheelchair, hand prosthesis) using only brain signals, which might assist individuals with diseases that affect neuromuscular channels.

A BCI system comprises mainly three steps: signal acquisition, signal processing, and feedback (see Figure 2.1). Brain activity can be recorded with different techniques as will be seen in the next section. These raw brain signals are first processed and then the most relevant features are extracted and are classified as a control option (e.g., direction, letter). Finally, the user is informed about the detected commands/actions through feedback.

BCI systems can be categorized as either invasive or non-invasive depending on the techniques that are used to record brain activity (for more details, see Section 2.2). BCIs can also be classified as dependent or independent. Dependent BCI needs both brain and muscular activity to control the BCI. Independent BCIs do not use any peripheral nerves or muscle activity, and for that reason, they are sometimes called "true" BCIs. The operating mode is another way to classify BCIs. They can be either non-self-paced or self-paced. A non-self-paced BCI provides a cue whenever the user should control the system. On the other hand, the self-paced BCI, are controlled whenever the user desires [Wolpaw et al., 2002], [Nam et al., 2018, pages 13-15].

BCI technology has grown rapidly and is becoming increasingly more robust with the computational advances and improvements in signal processing and machine learning techniques. Several events

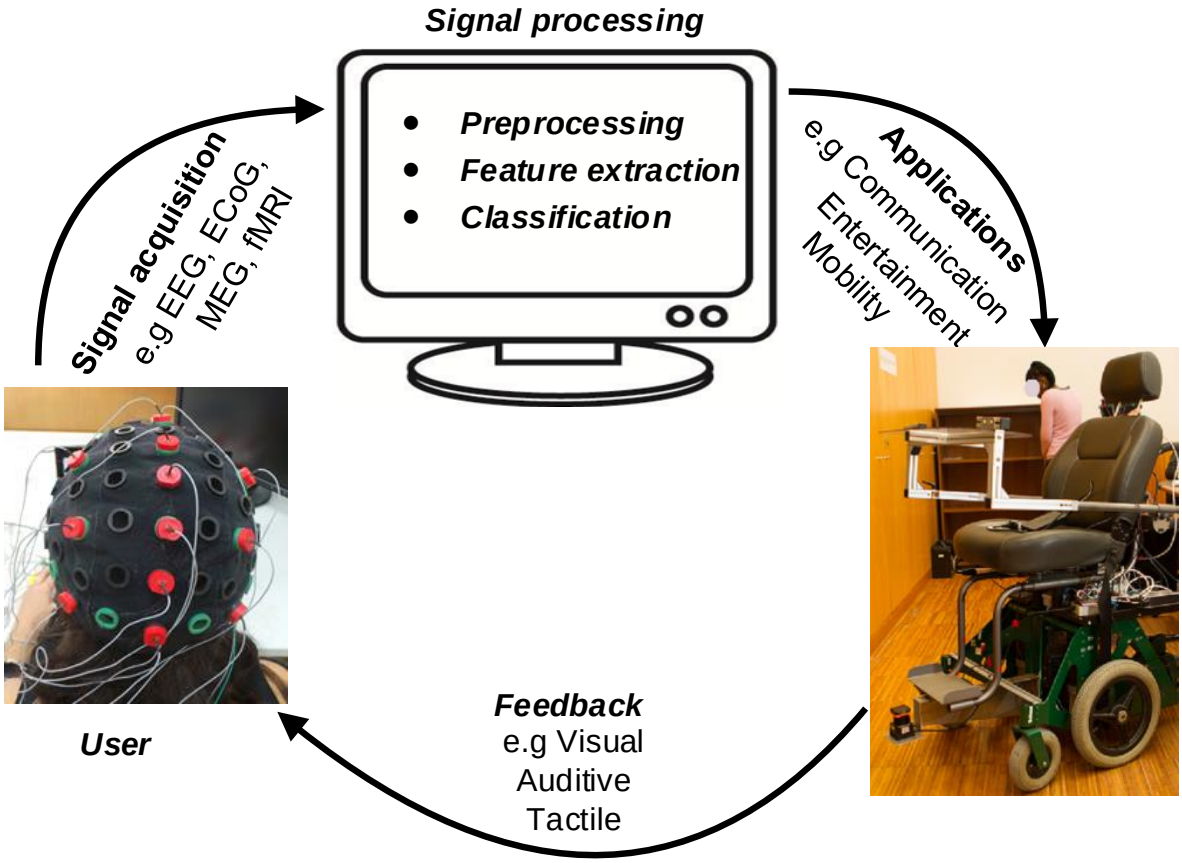


Figure 2.1: Generic representation of a BCI loop. Brain signal is measured through non-invasive (e.g., EEG, MEG, fMRI) or invasive (e.g., ECoG) recording methods. Then, these signals are processed, classified, and translated into some action. These actions can be the control of a communication system, a robotic device, or a game. The execution of these actions provides feedback (e.g., visual, auditory or tactile) to the user.

marked the history of the BCI research field, a brief history of BCIs is presented. It started in 1929 with the discovery of EEG as a tool to measure human brain activity [Jung and Berger, 1979]. Then, in 1973, Vidal introduced the term ”Brain-Computer Interface” describing the idea of utilizing the brain signals to control external devices [Vidal, 1973]. In 1988, Farwell and Donchin proposed the first speller system, referred to as Row-Column speller (RC), allowing the selection of letters by detecting P300 evoked potentials [Farwell and Donchin, 1988]. Currently, the RC speller is the most widely used BCI paradigm for spelling. This paradigm contains a 6-by-6 matrix that includes all letters of the alphabet and other symbols such as the ’spc’ and ’del’. The rows and columns of the matrix flash randomly according to an oddball paradigm. To select a target the user is instructed to focus on the target stimulus and mentally counting the times that the target flashes. Target symbols are expected to elicit a P300 ERP, which can be detected by a classifier. Combining the target row and target column it is possible to detect the

target symbol. In 1991, Sensorimotor rhythm (SMR) was used to control a 1D cursor control [Wolpaw et al., 1991]. In 1993 [Pfurtscheller et al., 1993] introduced a BCI based on the imagination of left and right-hand movements (MI-based BCIs). The use of Steady-state visual evoked potential (SSVEP) was proposed in [McMillan et al., 1995]. In [Chapin et al., 1999], an invasive BCI tested in rats showed that they could control a robot arm to obtain water by pressing a lever, through signals measured at motor cortex neurons. Later, an invasive BCI was tested on patients with epilepsy [Schalk et al., 2008].

Beyond advances in technology, efforts are also being made to create an official organization, in order to encourage discussion and collaboration. BCI research groups joined and started a BCI research community, organizing in 1999 the first International BCI meeting, and the most accepted definition of BCI was set: "A BCI is a communication system in which messages or commands that an individual sends to the external world do not pass through the brain's normal output pathways of peripheral nerves and muscles" [Wolpaw et al., 2002]. The Brain-Computer Interfaces journal's first issue was released in 2014. In 2015, with the motto "to foster research and development leading to technologies that enable people to interact with the world through brain signals" (bcisociety.org), the international BCI society was created.

2.1.1 BCI Applications

Nowadays, there is a wide range of BCI applications, aiming to broaden the target population that can benefit from BCIs. These applications can be divided into four main areas: 1) communication; 2) control of assistive devices; 3) rehabilitation; and 4) non-medical applications.

- **Communication -** Human is a social being, thus communication skill is crucial. Therefore, BCI applications for communication purposes is the main area in BCI research. It can help to restore the communication ability of people with severe motor impairments and has been considered a promise for people in the complete locked-in state. The most common applications of communication-related BCIs are spelling devices, which allow the disabled people to select letters of the alphabet displayed on a virtual keyboard on a computer screen. As mentioned above, the first spelling device was developed by Farwell and Donchin. Since then, many different paradigms have been proposed to create more reliable and faster speller systems [Treder et al., 2011, Pires et al., 2012]. Other applications for communication include web browsers and painting. The BCI web browsers allow disabled persons to access the Internet. Many browser prototypes have been developed, using different signal controls such as slow cortical potentials and P300 ERP [Karim et al., 2006, Muglerab et al., 2008]. An artistic BCI application known as "Brain Painting" enables disabled people to express their imagination through painting artworks [Holz et al., 2015].

- **Control of assistive devices -** BCI has been successfully used for controlling assistive de-

vices such as prostheses, wheelchairs, and mobile robots for navigation [Vilela and Hochberg, 2020, Lopes et al., 2016]. One of the most extensively researched applications is brain-controlled wheelchairs. BCWs can help individuals with severe motor disabilities, who are unable to use conventional interfaces, to increase their levels of mobility. Several studies have demonstrated that healthy and motor-impaired people can successfully control a brain-actuated robotic wheelchair, as long as it is supported by a navigation system with a human-machine collaborative controller [Lopes et al., 2013, Carlson and Millan, 2013]. In [Millán et al., 2004], the authors proposed the first brain-actuated mobile micro-robot. Since then, several researchers have developed brain-actuated robotic wheelchairs with different architectures. They mainly differ in the type of neural signal (e.g., motor imagery, P300 ERP), operation mode (e.g., self-paced or non-self-paced modes), and navigation commands (high-level commands or low-level commands) [Lopes et al., 2016, He et al., 2016, Tang et al., 2018]. Environmental control is also an application that may increase the autonomy of disabled people. BCI technologies can be useful in ambient assisted living to control domestic devices like TV, telephone, lights, motorized bed, among others [Cincotti et al., 2008, De Venuto and Mezzina, 2018].

- **Rehabilitation -** The use of BCI may have a strong impact on motor rehabilitation, for example, in stroke survivors. There has been increased interest in developing ways to apply BCI technologies to rehabilitation. BCI systems can facilitate stroke recovery by monitoring motor imagery and providing a way to determine whether the patients are performing the expected movement. BCI feedback may help users to perform movements more efficiently and increase their engagement [Reichert et al., 2016, Pichiorri et al., 2015]. Another way to enhance recovery is the use of BCI to ensure that subjects were optimally prepared to execute a particular movement [McFarland et al., 2015]. Although most of the works in rehabilitation focus on patients with stroke, there are many other applications. BCI can quantify attention level, which has been successfully used as a tool for the treatment of attention deficit hyperactivity disorder [Lim et al., 2019]. BCI can also be a promising tool for Autistic spectrum disorder (ASD) intervention, improving social cognition skills in ASD patients [Amaral et al., 2018].

- **Non-medical Applications -** Both, healthy and disabled people can benefit from these applications. BCI systems can assess and decode brain states in real-time, which allows the creation of diverse applications, such as user performance monitoring, attention assessment, brain fingerprinting and authentication, car accident prevention, etc. [Blankertz et al., 2010, Farwell et al., 2014, Sharma et al., 2018]. For example, brain fingerprinting is a new research area that can detect concealed information stored in the brain but can also be used as an identification and authentication mechanism. Therefore, it can be used in crime investigation for determining whether the suspects are familiar with some information, sound, or photo [Sharma et al., 2018]. BCI may

be effective in preventing car accidents by detecting drivers' lack of attention or drowsiness [Saichoo and Boobrahm, 2019]. The use of BCI for entertainment is also growing, including gaming and multimedia [Parafita et al., 2013, Congedo et al., 2011], art, and music [Holz et al., 2015, Vamvakousis and Ramirez, 2014]. BCI has been used to play several well-known games like tetris, pinball, puzzle, and spacecraft, as well as virtual reality games [Parafita et al., 2013, Pires et al., 2011b], see [Ahn et al., 2014] for review. Brain-computer music interfacing (BCMI, the use of BCI for music) has also been explored to perform in real-time music, musical composition, and control a musical score (choosing from short pre-composed musical phrases) [Eaton and Miranda, 2014].

2.2 Signal Acquisition Methods

There are several techniques for brain activity measurement, which can be categorized according to their invasiveness. These recording techniques are now briefly presented, comparing their strengths and weaknesses.

2.2.1 Non-invasive Technologies

Non-invasive techniques measure brain activity without any device implantation in the user's brain, i.e., without requiring any surgery.

- **Electroencephalography** EEG is a non-invasive method that measures brain activity using electrodes placed on the scalp. Hans Berger was the first one who recorded the human brain activity [Berger, 1929]. A typical EEG signal has a frequency range between 0.5 and 100 Hz and amplitudes between 10 and 100 μV. EEG is the most popular brain recording techniques in BCI research, approximately 68% of all BCI studies [Hwang et al., 2013]. It is non-invasive, inexpensive, safe, and portable [Hwang et al., 2013, Corralejo et al., 2014]. Besides, EEG has a very high temporal resolution, typically in the order of milliseconds. However, EEG has some drawbacks, such as low spatial resolution and high susceptibility to artifact contamination. The EEG recording systems include an EEG cap with electrodes, an amplifier, and a computing device. Electrodes are typically small discs with low impedance made of silver, silver chloride, or gold. Dry electrodes are placed directly on the skin without gel, while wet electrodes require a skin preparation with an abrasive gel and a conductive gel between the scalp and the electrodes. Electrodes are usually positioned according to the International 10–20 system, which is independent of the size and shape of the head, proposed in 1958 by the American Electroencephalographic Society (Figure 2.2) [Jasper, 1958]. Four anatomical references need to be determined in order to identify the position of the electrode on the scalp: nasion, inion, and pre-auricular points (left and right). Nasion

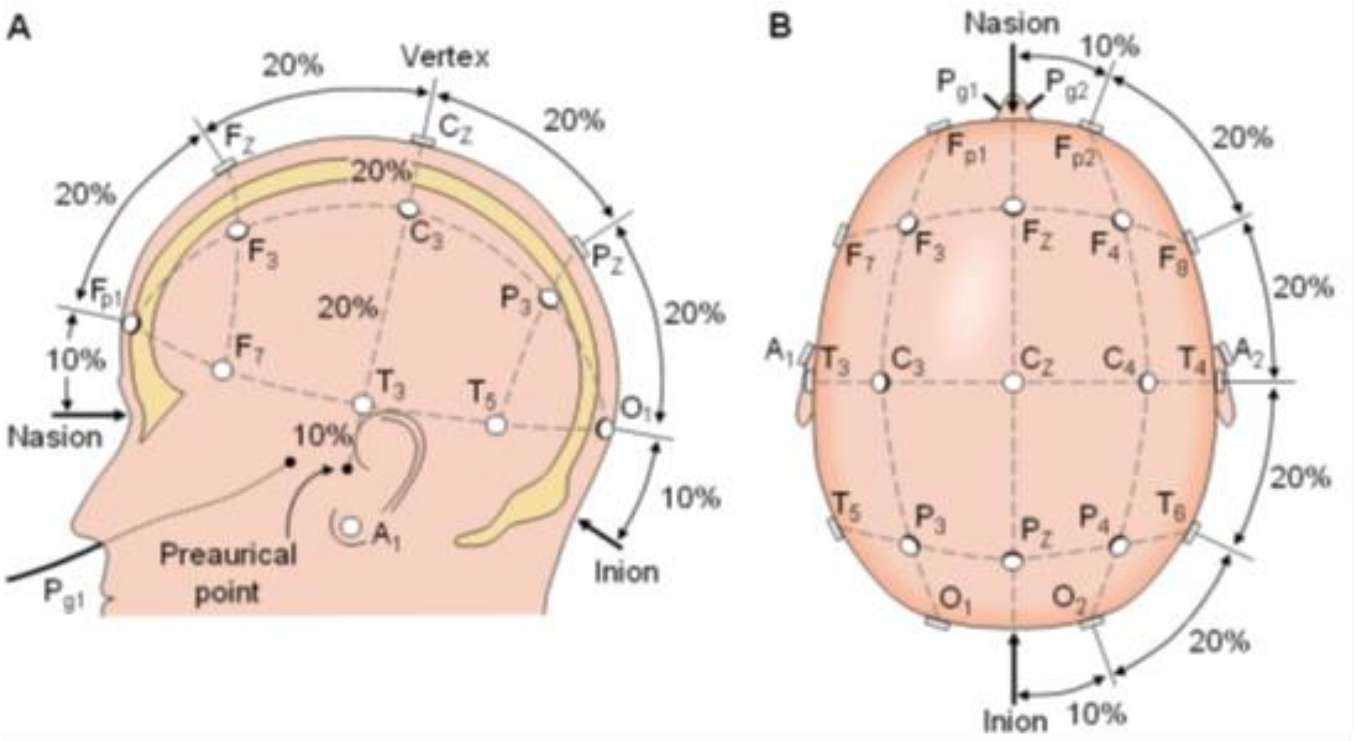


Figure 2.2: The 10-20 system for electrodes placement over the scalp [Malmivuo et al., 1995].

is the point located on the front of the face between the eyes, below the forehead, and above the nose. Inion is an external occipital protuberance located between the skull and neck. Pre-auricular point is the upper limit of the external auditory channel.

- **Magnetoencephalography (MEG)** is a non-invasive method that measures the magnetic fields induced by the electrical activity of the brain. This method has good characteristics, namely, very high temporal resolution similar to EEG and a better spatial resolution. However, it is very expensive, not portable, and requires highly sensitive devices to obtain quality recordings. These drawbacks make it impractical for applications in everyday life. It has been used in few BCI works, about 2% of the literature [Hwang et al., 2013].

- **Functional magnetic resonance imaging (fMRI)** is a non-invasive method that indirectly detects the neuronal activity based on blood oxygen level-dependent contrast. This technique presents very high spatial resolution, but its temporal resolution is quite low, typically in the order of seconds. Moreover, its size and high cost limits its use in BCI research and practical applications, accounted for only 2% of the literature [Hwang et al., 2013].

- **Functional near-infrared spectroscopy (fNIRS)** is a non-invasive method that indirectly measures neuronal activity using infrared light placed on the scalp. Oxygenated and deoxygenated

blood in the cerebral cortex has different optical properties. fNIRS uses these differences to detect brain activity. This method has poor temporal resolution compared to EEG, and it is susceptible to many types of artifacts, namely, heartbeats, motion, and pulse artifacts. However, it is portable and inexpensive, making it a possible alternative for BCI research. Yet, in 2013 only 3% of BCI studies used this technique [Hwang et al., 2013], but it has been increasingly used on many applications such as neurorehabilitation, psychiatric disorders (e.g. schizophrenia, anxiety, and depression), cognitive states [Ehlis et al., 2014, Pinti et al., 2020].

2.2.2 Invasive Technologies

Invasive methods record brain activity using electrodes or microelectrodes implanted inside the skull. These methods allow recording brain activity with high spatial and temporal resolution. However, they involve substantial clinical risk to the individual, and they do not provide a stable signal for a long period [Castermans et al., 2014]. The following is a brief description of two invasive methods, Electrocorticography (ECoG) and Intracortical neuron recording (INR).

- **Electrocorticography** is also called intracranial EEG, it measures neural activity with electrodes directly placed on the surface of the cortex. These electrodes are placed through intracranial surgery. Compared to EEG, this method has a higher temporal resolution, spatial resolution, and amplitude, wider frequency range, and much less susceptible to artifacts. Its main drawback is invasiveness that imposes clinical risk, such as postoperative infection. ECoG is used in patients with epilepsy to monitor and localize the seizure onset zone before surgery. Usually, the experiments with ECoG-based BCIs are conducted with these patients [Moses et al., 2018].
- **Intracortical Neuron Recording** records the neuronal activity in the gray matter of the brain through microelectrodes placed inside of the brain. It can detect single or multiple neural impulses, and local field potentials (LFPs). This method is highly invasive, thus it is mainly used in animals but it has also been used in humans with severe motor disorders. Invasive methods, ECoG and INR, accounted for 32% of BCI works in 2013 [Hwang et al., 2013].

Table 2.1 summarizes some of the advantages and disadvantages of the above recording techniques.

2.3 Neural Signals for EEG-based BCI Operation

This thesis will focus on an EEG-based BCI. As already shown in section 2.2.1, it is the most studied system due to its several advantages. Different neuromechanisms can be used for EEG-based BCI control. EEG signals could be modulated by either exogenous stimuli, visual and auditory, or by endogenous mental activities. The brain signals modulated by the exogenous stimulus can be generated

Table 2.1: Comparison between the different brain recording technologies, namely electroencephalography, magnetoencephalography, functional magnetic resonance imaging, functional near-infrared spectroscopy, electrocorticography, and intracortical neuron recording.

Recording technologies	Risk	Measurement	Temporal resolution	Spatial resolution	Portable
EEG	Non-invasive	Direct	High	Very low	Yes
MEG	Non-invasive	Direct	High	Low	No
fMRI	Non-invasive	Indirect	Very low	High	No
fNIRS	Non-invasive	Indirect	Very low	Low	Yes
ECoG	Invasive	Direct	Very high	High	Yes
INR	Invasive	Direct	Very high	Very high	Yes

Table 2.2: Summary of the main control signals used in EEG-based BCI: sensorimotor rhythms, visual evoked potentials, and P300 event-related potentials.

Signal control	Training	Task difficulty	External stimuli	Number of choices	Typical ITR
SMR	Yes	High concentration	No	Low	3–35 bits/min
VEPs	No	Low concentration	Yes	High	60–100 bits/min
P300	Yes	High attention	Yes	High	20–25 bits/min

mechanically without any process (e.g. VEPs) or generated through selective attention (e.g. P300, N200). P300 sometimes is referred to as endogenous component, because it requires selective attention to discriminate the stimuli [Gao et al., 2014]. The brain signals modulated endogenously can be generated as a response to mental tasks, such as motor imagination (e.g. sensorimotor rhythms and slow cortical potentials). Hybrid BCIs have also been used combining different neural mechanisms, for example, P300 with motor imagery. The main purpose of hybrid approaches is to increase the system reliability or to take full advantage of available users' functionalities [Zhang et al., 2015b, Wang et al., 2014].

Table 2.2 shows a summary of the three major neural signals (P300 ERP, VEPs, and SMR) comparing their main characteristics, regarding training, task difficulty, external stimuli, number of choices and typical Information transfer rate (ITR) [Nicolas-Alonso and Gomez-Gil, 2012]. These neural signals are described below with a special focus on event-related potentials, which were used in this thesis.

2.3.1 Sensorimotor Rhythms

Sensorimotor Rhythms arise from the sensorimotor cortex due to either real movement (motor execution) or imagined movement (motor imagery). These movements generate the Event-related desynchronization (ERD) and Event-related synchronization (ERS) of sensorimotor rhythms, namely Mu band (7–13 Hz) and the Beta band (14–30 Hz) [Fabiani et al., 2004]. There is a decrease in frequency band amplitude (ERD) when the subject is performing a motor execution or motor imagery, and an increase in frequency band amplitude (ERS) during motor relaxation [Graimann et al., 2009]. Sensorimotor rhythm

mu starts immediately before movement onset, reaching its maximum values shortly after movement onset, and a few seconds later it recovers the original level. On the other hand, the sensorimotor rhythm beta presents a low ERD when the movement starts, followed by ERS that reaches its maximum amplitudes when the subject is relaxed (after movement execution) [Nicolas-Alonso and Gomez-Gil, 2012]. Different tasks, such as left-hand or right-hand movements activate distinct sensorimotor areas. Therefore, the BCIs based on motor imagery use the imagined movement of left hand/foot, right hand/foot, or tongue for the control process [Graimann et al., 2009]. Other mental processes like calculus, writing and sound production have also been used [Song and Sepulveda, 2020, Qiu et al., 2017]. Extensive training is needed for the subject to learn how to control SMRs. This training process can last several days, weeks, or months, and some people are unable to control their rhythms.

2.4 Event-related Potentials

Event-related potentials are a scalp-recorded voltage change elicited by a population of neurons during a sensory, cognitive, affective, and motor processes [Luck and Kappenman, 2011, pages 4-8]. It is generated in response to an external stimulus, that can be visual, auditory, or even tactile.

2.4.1 Visual Evoked Potentials

Visual evoked potentials are EEG waveforms elicited in response to visual stimuli. They usually occur at approximately 100 ms from stimulus onset (e.g, N100). Currently, SSVEPs are the most popular VEPs for BCI research. SSVEPs are potentials elicited as a response to stimuli flickering at a frequency higher than 6 Hz [Faller et al., 2017]. Depending on the type of stimulus modulation there are different subtypes of VEPs: 1) time modulated VEP which are generated from stimulus presented to the user with different time slots; 2) frequency modulated VEP that are the response to stimulus derived from different frequencies; 3) pseudorandom code modulated VEP induced from a stimulus with different pseudorandom codes, and 4) space modulated VEP generated from a stimulus with different locations [Liu et al., 2018]. This stimulus is used as control option in BCI system. SSVEP is recorded mainly in electrodes over the occipital and parietal areas of the brain. BCIs based on VEPs are the most robust and fast BCIs [Volosyak, 2011]. However, they are not a good alternative for users with severe motor impairments, since it requires eye gaze and the stimuli can cause eye discomfort and fatigue.

2.4.2 P300 ERPs

P300 is an event-related potential that appears in response to a relevant and rare stimuli (target event) in an oddball paradigm, in which infrequent stimuli is interspersed with frequent stimuli (non-target

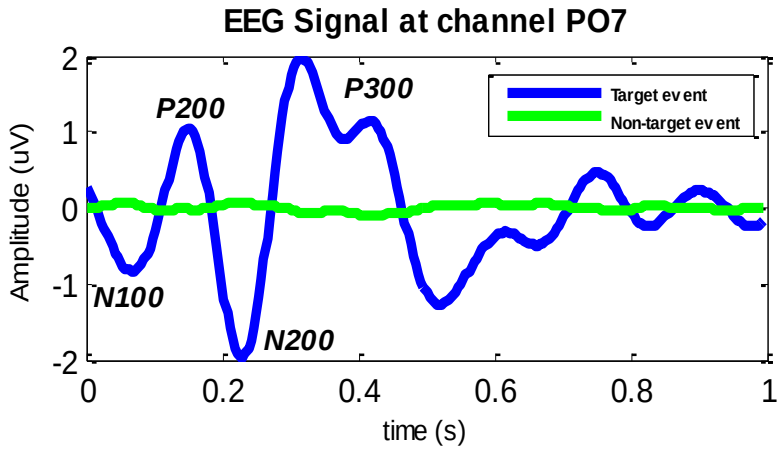


Figure 2.3: Event related potentials of target and non-target events, recorded in an oddball task.

event) [Huettel and McCarthy, 2004] (see Figure 2.3). These stimuli can be auditory, visual, tactile, or somatosensory. The P300 wave has a latency value of about 300 ms after stimulus onset, usually related to the difficulty in discriminating target and non-target events. Its amplitude value is around 10-20 microV, which varies inversely with the target probability [Picton, 1992, Comerchero and Polich, 1999, Friedman et al., 2001]. Moreover, P300 signal changes according to many experimental factors, such as stimulus modality, intensity, and duration, as well as the interstimulus interval and selective attention [Comerchero and Polich, 1999, Polich, 2007]. P300 wave has two sub-components: P3a related to novelty and P3b linked to attention and stimulus processing [Comerchero and Polich, 1999, Friedman et al., 2001]. P3a has a central maximum and P3b has a parietal maximum. The first ERP-based BCI [Farwell and Donchin, 1988] was based on the P300 component and it became one of the most well-known paradigms. Since then, P300 ERP is widely used in BCI research due to its several advantages, such as short training time and it offers a high number of target selections for the control process.

Besides the P300, an oddball paradigm also elicits other components, namely N100, P200, and N200. N100 and P200 are negative and positive peaks occurring around 100 and 150 ms post-stimulus onset and they reflect automatic stimulus processing caused by early attention and orientation processes. N200 is a negative peak at around 200 ms and it reflects the cognitive processes of stimulus identification [Mueller et al., 2008].

2.4.3 Error-related Potentials

Error-related potential is a potential generated when wrong actions are perceived [Gehring et al., 1993]. Recent finds demonstrated that ErrP reflects an unconscious process, i.e., it is elicited without consciousness of which action was performed (correct or wrong action). However, a consciousness of which action was required is needed, that is, users need to know what they should do [Dehaene, 2018].

ErrP was first observed in the 80's during speeded choice reaction tasks which is called error negativity or Error-related negativity (ERN) [Falkenstein et al., 1989, Gehring et al., 1993]. ERN shows a negative deflection approximately 50–100 ms after erroneous responses and it is typically distributed over the fronto-central region. ERN indexes the brain's self-monitoring system, and it seems that the Anterior cingulate cortex (ACC) plays an important role in this process, although there is no agreement on the exact role performed by the ACC [Luu et al., 2003, Holroyd and Coles, 2002]. The ACC has diverse functions, it has an important role in emotion regulation, and various aspects of cognition (e.g. decision-making). Several studies have reported the presence of other components, namely, Error positivity (Pe) and Correct response negativity (CRN), that will be described in the next sections. Four types of ErrPs are described in the literature, they differ in how the person perceives the error, and in the agent committing the error (Figure 2.4). These four ErrPs are: the response ErrP, the feedback ErrP, the recognition/observation ErrP, and the interaction ErrP [Teeuw, 2010].

The response ErrP is mostly studied in Go/Nogo tasks when the subject has to make decisions under pressure. Response ErrP occurs when an incorret selection is made by the subject and she/he recognizes it immediately. Response ErrP is characterized by a negative potential that appears between 50 and 100 ms after the incorrect selection followed by a larger positive peak that occurs between 200 and 500 ms after the incorrect response [Falkenstein et al., 1991, Falkenstein et al., 2000].

The feedback ErrP occurs in tasks where a subject has to make a choice, but she/he does not know the correct answer until she/he is informed through external feedback [Holroyd and Coles, 2002]. This ErrP is mostly studied in typical reinforcement learning tasks where the subject is asked to make a choice and adapt his/her strategy to minimize the occurrence of negative feedback (errors). The feedback ErrP appears after the feedback that indicates incorrect selection. The main component of this ErrP is a negative deflection that occurs between 200 and 300 ms after the feedback during a reinforcement learning task. This component is referred to as Feedback-related negativity (FRN) and is typically observed at central to frontal-central scalp regions. In this case, the brain reacts to the feedback indicating the error and not to the incorrect selection. The feedback can either be visual, auditory, or tactile.

The observation ErrP occurs when a subject is observing an operator performing a task [van Schie et al., 2004]. The observation ErrP emerges after the incorrect response of the operator. Like in the feedback ErrP, the main characteristic is a negative deflection that occurs between 200 and 300 ms after the errors were made by the operator.

More recently, in [Ferrez and Millán, 2005] the authors reported an ErrP generated when a subject interacts with a BCI system and it was called interaction ErrP. The Interaction ErrP occurs when the BCI makes an error, that is, the BCI feedback is different from what the subject intended to. This ErrP is more complex than the other ErrPs exhibiting a small positive peak around 200 ms, a negative peak around

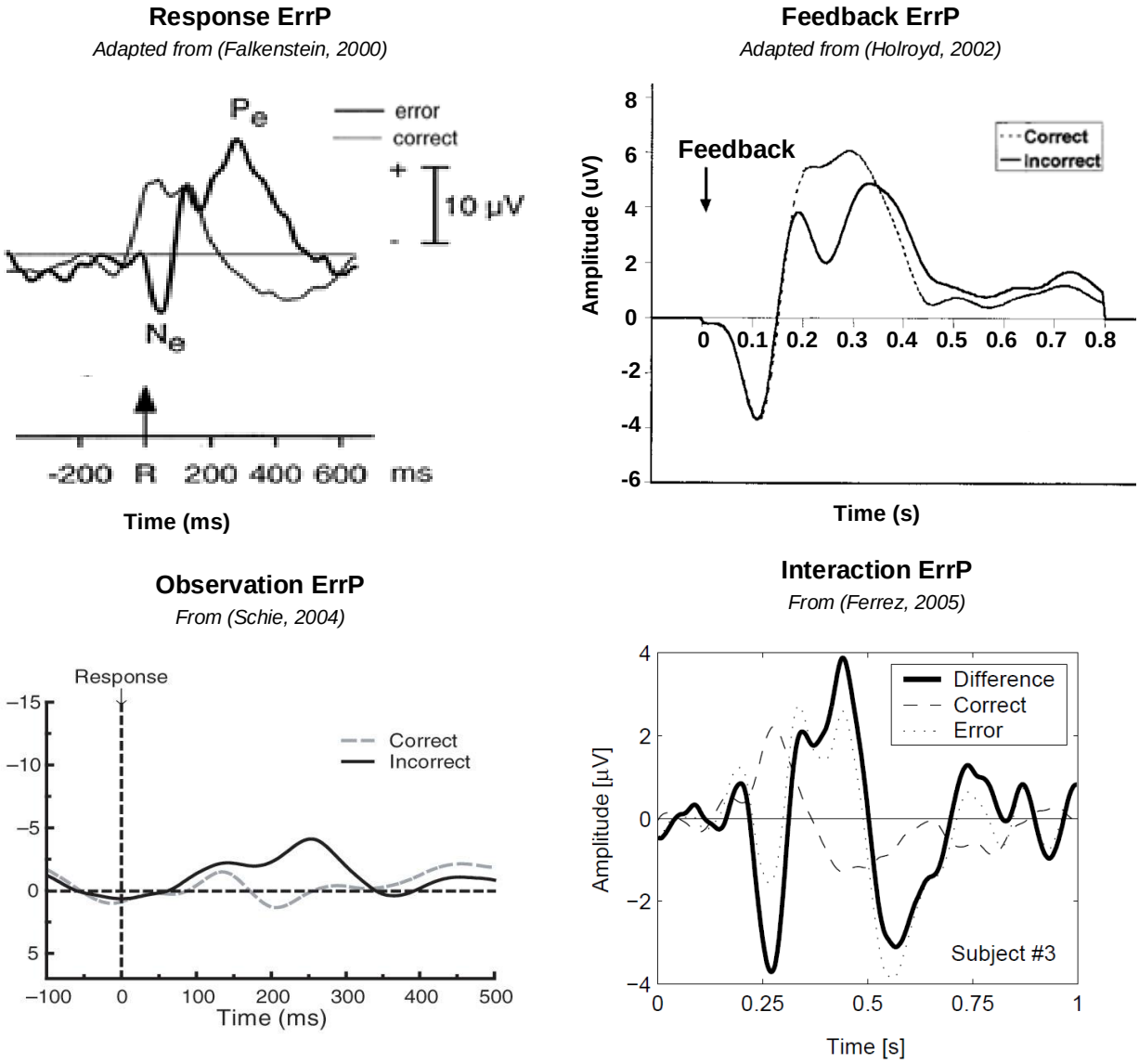


Figure 2.4: The four types of ErrPs reported in literature. Top row left: response ErrP occurs when an error is made by the subject and she/he recognizes it immediately [Falkenstein et al., 2000]. Top row right: the feedback ErrP occurs when an error is made by the subject and she/he is informed through external feedback [Holroyd and Coles, 2002]. Bottom row left: the observation ErrP occurs when an error is made by an operator (note that the authors presented the graph with an inverted scale) [van Schie et al., 2004]. Bottom row right: the Interaction ErrP occurs when an error is generated when a subject interacts with a BCI system [Ferrez and Millán, 2005].

250 ms, a positive peak between 350 and 450 ms, and a second negative peak around 550 ms after the BCI feedback. However, it has been shown that the characteristics of ErrPs, such as amplitude and latency can differ depending on the experimental paradigm [Iturrate et al., 2014]. Both feedback ErrP and interaction ErrP are elicited by the same process, that is, the ErrPs arise after feedback indicating

the error. However, in the case of feedback ErrP errors are made by the subject himself and in interaction ErrP errors are made by the interface. In the BCI system, wrong feedback can be due to the error of both the interface and the user himself. To make sure that the erroneous feedback is only caused by the interface, in [Ferrez and Millán, 2005] the subjects provide commands manually and not mentally (using BCI). The results showed that both experimental protocols with commands provided manually [Ferrez and Millán, 2005] or mentally [Ferrez and Millán, 2008a] elicited a very similar ErrP with the same shape but with different latency, and amplitude.

2.4.4 Functional Significance of the ERN

Since the first reports of an ERN, several theories have been developed regarding the functional significance of the ERN. There is still no consensus view of the ERN and with the emergence of CRN, the theories have evolved to include this component. The major theories are error detection/comparator theory, conflict-monitoring theory, reinforcement learning theory of the ERN, and motivational significance theory.

The first theory is the error-detection theory also called the mismatch theory, which is mainly tested on response ErrP. According to it, the ERN reflects a process that compares the actual response to the correct response [Falkenstein et al., 1991, Gehring et al., 1993], or the comparison process itself [Vidal et al., 2000]. Coles et al. [Coles et al., 2001] believe that, in speeded-response tasks, the errors usually occur due to fast guessing. Responses are executed before stimulus evaluation is complete and the stimulus processing continues and the correct response is derived further. A comparator computes the difference between the correct response and actual responses. The discrepancy between these two responses originates error signal. According to this, ERN amplitude should be modulated by the degree of the similarity between the correct and wrong response. Studies confirmed this hypothesis and found that ERN amplitude increased when the erroneous and the correct response are more dissimilar [Bernstein et al., 1995, Falkenstein et al., 1996]. On the other hand, [Gehring and Fencsik, 2001] achieved contradictory results, they obtained larger ERNs when the error response and the correct response are similar.

The conflict-monitoring theory was proposed as an alternative to the error-detection theory since it is unclear how the comparator knows the correct action and why the brain did not perform the correct action if it knows the correct response [Carter et al., 1998], [Yeung et al., 2004]. According to the conflict-monitoring theory, the ACC monitors the conflict during multiple competing responses. Therefore, the ACC detects situations when errors can occur rather than errors, per se. This theory suggests that the ACC is active in both error and correct response representations due to continued stimulus processing [Botvinick et al., 1999, Carter et al., 2000]. Researchers tried to manipulate the level of response conflict in the Flankers task and the results were not consistent. [Forster and Pavone, 2008] reported

that congruent stimuli elicited ERNs with greater amplitude than incongruent stimuli, and no difference in the ERNs amplitude is found in [Christ et al., 2000]. To take account of the CRN component, the proponents of the conflict monitoring theory suggested that might exist different levels of conflict, such as stimulus conflict and response conflict, and they occur in parallel. CRN might be associated with conflict at the response level [Van Veen and Carter, 2005].

The reinforcement learning theory proposes that a monitoring mechanism in the basal ganglia uses information from both the environment and internal actions to evaluate current action based on learned expectations, and ERN is elicited when the event is worse than expected [Holroyd and Coles, 2002]. The error signals are used to improve task performance by predicting future rewards and non-rewards [Schultz, 2002, Holroyd and Coles, 2002].

According to motivational significance theory, ERN reflects an affective response to errors and its amplitude may be influenced by the importance of an error [Hajcak et al., 2005]. Gehring et al. [Gehring et al., 1993] found that when participants were instructed to put more emphasis on performance, they elicited ERN with larger amplitude, and ERN with smaller amplitude was elicited when they put more emphasis on speed. Studies also suggested that the amplitude of ERN may be related to an individual's personality [Pailing and Segalowitz, 2004].

To sum up, these theories still have unsolved issues. Most of the studies focus on response ErrP, and the results are yet contradictory and inconclusive. The emergence of other components (e.g. CRN) and new kinds of ErrPs increase the degree of complication and challenge. Recent extensions to these theories have been proposed to include these components and answers some questions [Dehaene, 2018, Alexander and Brown, 2010]. New works are needed for solving some issues and inconsistencies among studies.

2.4.5 Error Positivity

is a positive deflection that occurs between 200 and 500 ms after the erroneous response and it usually follows the ERN [Falkenstein et al., 2000]. Contrarily to the functional significance of the ERN where there are several studies, just a few works have focused on the function of P_e. There is large variability in scalp locations of the P_e across studies, however, it generally appears over centroparietal regions [Overbeek et al., 2005]. There are different views regarding the functional significance of the P_e. Some researchers suggested that the P_e maybe a P300 associated with the importance of the error [Overbeek et al., 2005]. Supporting this, several studies showed a larger amplitude of P_e for errors that were detected by the subject than for errors that were not [Endrass et al., 2007, O'Connell et al., 2007, Vidal et al., 2000]. P_e can be divided into early frontocentral and posterior centroparietal components. The early P_e is suggested to be similar to a P3a and it can reflect error processing and the posterior P_e is suggested to related to P3b and it can reflect a conscious recognition of the error or a subjective

affective response [Veen and Carter, 2002], [Endrass et al., 2007]. Recently in [Di Gregorio et al., 2018], the authors demonstrated that ERN is not necessary for eliciting the P_e and they suggest that ERN and P_e may reflect two independent systems of error monitoring.

2.4.6 Correct Response Negativity

Studies found a negative potential after correct trials which is very similar to the ERN, thus it has been called the correct response negativity [Vidal et al., 2000]. The CRN and ERN have a similar latency, waveform, and scalp distribution. The observations of the CRN call into question theories regarding the functional significance of the ERN and this raises the question of how could ERN represents the detection of an error [Vidal et al., 2000]. Investigators have suggested that the CRN may reflect either a comparison process or uncertainty of a correct response or an emotional reaction [Coles et al., 2001, Vidal et al., 2000].

2.5 Related work

This section contains the state-of-the-art of the main issues addressed in this thesis. Therefore, it is composed of three parts, namely error-related potentials to improve BCI reliability, transfer-learning to increase BCI generalization, and self-paced BCI.

2.5.1 Error-related potentials

The idea of using automatic detection of ErrP to improve the performance of BCI has been well accepted by the BCI research community. ErrP appears within a time window of 500 ms, and thus its automatic detection could be used in myriad ways, in real-time, in human-machine interaction processes. In the context of brain-computer interfaces, i.e., when BCI is being used as the primary communication channel, ErrP detection was researched to increase the reliability and the information transfer rate in P300-based BCI spellers [Zeyl et al., 2015] and in BCIs based on motor-imagery [Ferrez and Millán, 2008b], by eliminating or even correcting errors. ErrPs were also researched in real-life tasks for monitoring decisions/actions of systems not controlled by the user [Zhang et al., 2015a,Salazar-Gomez et al., 2017]. In [Zhang et al., 2015a] ErrPs were used to validate the predictions of a driving assistance system (selection of a driving direction in intersections) in a simulated and a real car. In [Salazar-Gomez et al., 2017], the selections of a robot were corrected based on ErrPs elicited when a human operator observed the wrong actions of the robot. In [Chavarriaga and Millán, 2010], ErrPs were used in a reinforcement-learning loop to change the behavior of an agent in a simulated environment. Chavarriaga et al. review the use of ErrP in BCI over the last decade [Chavarriaga et al., 2014]. Here, we review only works using

ErrP to improve the performance of P300 BCIs, which are directly related to our approach proposed in this thesis. The first research reporting an online P300 speller system with an integrated ErrP-based correction was developed in [Dal Seno et al., 2010]. Although they demonstrated that it was possible to detect single-trial ErrP with an accuracy of around 60%, it did not lead to an improvement in the overall classification performance. This outcome is mostly because they did not tune the ErrP classifier to a low false-positive rate. False positives imply a decrease in transfer rate since correctly detected targets become incorrect. [Combaz et al., 2012] tested a standard speller matrix with nine participants to analyze the difference between correct and incorrect feedback and explored the possibility of classifying these feedback responses. They showed that the integration of ErrP classifiers into the P300 Speller system could produce a theoretical (assuming perfect ErrP detection) improvement around 15%. In [Schmidt et al., 2012] the online recognition of ErrP was used in the Center Speller [Treder et al., 2011] and was tested with twelve participants. They obtained a mean accuracy of single-trial ErrPs around 89% and an increase in the spelling speed of about 49%. [Spüler et al., 2012] observed that young, elderly, and motor-impaired individuals (participants with ALS and Duchenne muscular dystrophy) presented similar ErrPs and that the ErrP-based error correction system substantially increased the transfer rate (0.44, 0.73 and 0.35 bit/trial respectively).

While the approaches in previous works used the ErrP detection only to delete the wrong characters, [Margaux et al., 2012] and [Zeyl et al., 2015] extended the concept also allowing automatic error correction. While in [Margaux et al., 2012] the Error correction system (ECS) did not show improvement, [Zeyl et al., 2015] improved the selection by 13.67% for 2.54 Effective symbols per minute (eSPM). The contribution of automatic error detection to increase the BCI performance depends on the accuracy of error detection, which must be achieved from a single trial. Accuracy of the error detector lower than the P300 accuracy will lead instead to a degradation of the performance. Therefore, approaches to enhance error detection are essential to make error detection useful.

2.5.2 Transfer-learning to improve BCI generalization

The poor classification generalization between sessions and across subjects is one of the major BCI challenges. Transfer learning (TL) techniques can be used to address this issue. Transfer learning methods have been successfully used in BCIs, which can be divided into feature space learning and model space learning [Jayaram et al., 2016]. TL approaches based on feature space learning try to find a transformation to a data space in which the features are invariant between either sessions or subjects and a single classification rule can classify all the data [Lee et al., 2006, Arvaneh et al., 2014, Morioka et al., 2015, Raza et al., 2016]. On the other hand, model space approaches attempt to learn how the decision rules differ across sessions or subjects [Alamgir et al., 2010, Kang and Choi, 2014]. Feature space learning are by far the most used approaches in the literature [Jayaram et al., 2016], [Nam et al.,

2018, pages 426-440]. The feature space based on trial covariance matrices plays a strong role in signal processing, since it provides compact and relevant information [Cherian et al., 2011]. Covariance matrices are widely used through spatial filters, such as Common spatial pattern (CSP), which is one of the most popular techniques for feature extraction in BCI. Recently, several approaches based on Riemannian geometry that works on the Riemannian manifold of the Symmetric and positive definite (SPD) matrix (e.g. covariance matrix) have been proposed for transfer learning [Zanini et al., 2017], spatail filtering [Xu et al., 2020], dimentionality reduction [Harandi et al., 2017], outliers detection [Yamamoto et al., 2020], and adaptive classification [Kumar et al., 2019] see [Yger et al., 2016, Lotte et al., 2018] for reviews. Riemannian geometry has been shown to be effective for cross-session and cross-subject BCI learning due to its useful invariance properties, such as, invariance under congruent transformations [Congedo et al., 2017, Zanini et al., 2017]. Currently, zero calibration has attracted a lot of attention and a variety of algorithms have been proposed to deal with EEG non-stationarity [Krauledat et al., 2008, Kindermans et al., 2014, Schönleitner et al., 2019]. In order to find robust filters several modifications of CSP have been proposed, which are done on the level of the covariance estimation or on optimisation function [Kang et al., 2009, Lotte and Guan, 2010, Samek et al., 2012, Devlaminck et al., 2011]. The variability of EEG signals across sessions or even subjects can be understood as domain shift problem. Thus, some authors minimize the distance between the target and source domains using data alignment approaches [Sun et al., 2016, Zanini et al., 2017, He and Wu, 2019]. More precisely, in [Zanini et al., 2017] they consider that the inter-session/subject variability induces shifts of covariance matrices with respect to a reference state. The authors proposed a TL approach to align the covariance matrices from target and source domains by moving them to a common reference point using the Riemannian mean of the reference state (resting period) covariance matrices. In [He and Wu, 2019], the authors extended this concept to the Euclidean space. So, instead of aligning the covariance matrices, they aligns the EEG trials and they suggested that the reference matrix should be the average of covariance matrices of all trials.

2.5.3 Self-paced BCI: Brain-controlled wheelchairs

Brain-controlled wheelchairs are more complex than other BCIs such as spellers and games because they combine BCI and assistive navigation systems, and require higher levels of reliability and usability to ensure safety and natural human-machine interaction. Brain-controlled wheelchairs have been researched for more than one decade. They are mainly based on three neural mechanisms: motor imagery [Carlson and Millan, 2013, Zhang et al., 2015b, Yu et al., 2018], steady-state visual evoked potential [Diez et al., 2013, Duan et al., 2014, Ng et al., 2014, Li et al., 2016], and P300 event-related potential [Iturrate et al., 2009, Kaufmann et al., 2014, Lopes et al., 2016, He et al., 2016]. Hybrid brain-actuated wheelchairs have also been proposed combining different neural mechanisms or combining

brain signals with other physiological signals [Rebsamen et al., 2010, Wang et al., 2014, Choi, 2012, Tang et al., 2018, Yu et al., 2017]. They may be used to increase the system reliability or to adapt to users functionality. For example, in [Wang et al., 2014] it is proposed the use of MI for selecting left, right, forward and backward commands, P300 and MI for acceleration or deceleration, and eye-blinking to issue stop commands. Most proposed BCI systems consider a fixed time interval to select the desired command, meaning that a fixed number of stimuli sequences (P300-based BCIs) or a fixed time-window (MI and SSVEP-based BCIs) are required for decoding the user's intention [Carlson and Millan, 2013, Choi, 2012, Diez et al., 2013, Duan et al., 2014, Ng et al., 2014, Lopes et al., 2016, He et al., 2016]. The use of a dynamic time-window to issue BCI commands is also a desirable feature, as the speed of the system can be adjusted online to the user's performance, thereby increasing BCI accuracy, however very few have used this approach [Kaufmann et al., 2014]. BCWs can use either high-level commands [Choi, 2012, Ng et al., 2014, Lopes et al., 2016, Rebsamen et al., 2010, Zhang et al., 2015b, He et al., 2016, Tang et al., 2018] or low-level commands [Carlson and Millan, 2013, Diez et al., 2013, Kaufmann et al., 2014, Duan et al., 2014, Wang et al., 2014]. The use of high-level commands requires the robot to be able to perform autonomously safe and effective navigation without user's aid (commands can be either global, such as 'kitchen', 'wc', or local, such as 'door', 'go-left'). With low-level commands, the user can steer the wheelchair with raw commands (e.g., 'forward', 'left', 'increase speed'). Although this approach is flexible, as the user can control any specific motion, it is highly demanding and almost impossible to use in real-world environments, even with a collaborative controller. So far, the reported experimental tests combining user's intent and context awareness in a collaborative controller have been conducted in very structured environments [Rebsamen et al., 2010, Diez et al., 2013, Lopes et al., 2013, Duan et al., 2014, Ng et al., 2014, Tang et al., 2018, Li et al., 2016] or open spaces [Choi, 2012], and in semi-structured environments [Iturrate et al., 2009, Carlson and Millan, 2013, Wang et al., 2014, Lopes et al., 2016, Zhang et al., 2015b]. Moreover, just a few works report experiments conducted with motor impaired participants [He et al., 2016, Tang et al., 2018]. For more extensive surveys comparing different BCWs approaches, please refer to [Fernandez-Rodriguez et al., 2016] and [Bi et al., 2013].

3

ERP-based BCI: Methods

Contents

This chapter provides the background methods for implementing the ErrP and P300 detection approaches proposed in the remaining chapters. All classification framework are described, namely, the pre-processing, feature extraction/selection, and classification methods. Additionally, it presents the performance metrics used to measure BCI efficiency.

3.1 P300-based BCI

Figure 3.1 shows a generic view of a P300-based BCI system. Two BCI applications are used in this thesis. The first is a communication speller called Lateral single character (LSC), which was introduced in [Pires et al., 2012]. The other application is an interface to drive a robotic wheelchair (RobChair prototype) [Pires, 2011, Chapter 8].

3.1.1 LSC speller Paradigm

LSC minimizes some effects usually occurring in the classical row-column speller paradigm (e.g., distractors, high target probability, low target-to-target interval, eyestrain) (see details in [Pires et al., 2012]). On average, participants usually reach higher accuracies with LSC than with the classical row-column speller, for the same number of repetitions. For that reason, it is usually our preferred BCI-speller. LSC contains the letters of the alphabet, and the 'spc' and 'del' symbol, which are spatially arranged as shown in 3.2. These symbols flash randomly according to an oddball paradigm. A target symbol is expected to elicit a P300 ERP. Each symbol flashes individually alternating between left and right sides, and with no inter-stimulus interval (ISI). The Number of repetitions of each trial (Nrep) can be adjusted individually to each user according to the desired BCI performance for online operation. Data segments (epochs) of one second are recorded for each event from stimulus onset and then classified as target or non-target (standard). The symbol detected as the target is fed back to the user. An interval between trials (Inter-trial interval (ITI)) allows the user to shift his/her attention to the next desired symbol. The overall time

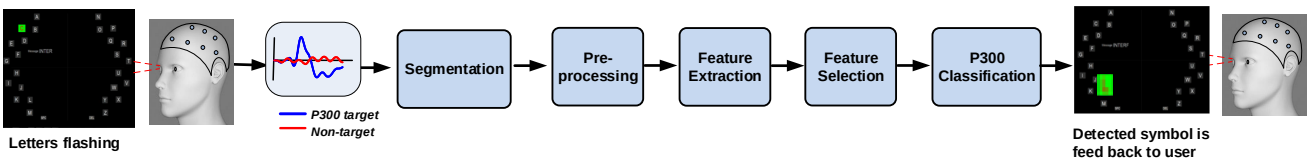


Figure 3.1: Squematic representation of the P300-based BCI system. Letters flash according to the oddball paradigm and the user focuses on the target letter. The EEG signal is acquired and pre-processed. Then features are extracted, selected, and classified with the P300 detector. The detected letter is shown to the user.

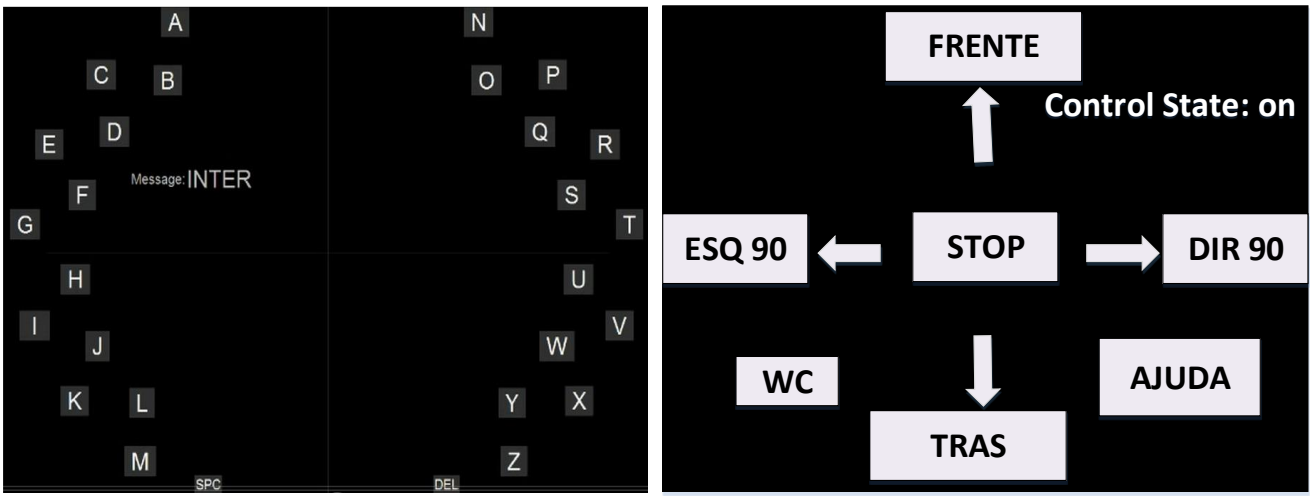


Figure 3.2: Left: Screenshot of the lateral single character speller. Right: Screenshot of the Robchair paradigm.

for one trial is:

$$TT = N_{rep} \times N_s \times SOA + CT + ITI \tag{3.1}$$

where N_s is the number of symbols, SOA is the stimulus onset asynchrony and CT is the time associated with the last flash of the trial.

3.1.2 Robchair Paradigm

RobChair paradigm interface was designed in our research lab specifically to provide commands to drive a robotic wheelchair (RobChair). The initial visual paradigm had 11 symbols which included assistance words and information to steer the robotic wheelchair [Pires, 2011, Chapter 8]. Then, the paradigm was slightly re-arranged in [Lopes et al., 2013] which is the version currently used. The paradigm comprises 7 steering commands: 'FORWARD', 'BACK', 'LEFT90', 'RIGHT90', 'STOP', 'WC', and 'HELP', in Portuguese ('FRENTE', 'TRAS', 'ESQ90', 'DIR90', 'STOP', 'WC', 'AJUDA', respectively), as depicted in Fig. 3.2. Similarly to the LSC, these symbols flash individually and randomly to obtain an oddball paradigm. The only difference is that RP requires an inter-stimulus interval.

3.2 Pre-processing

The raw EEG signal is affected by undesired components (artifacts). These artifacts can be categorized as either technical or physiological [Wolpaw and Wolpaw, 2012]. Technical artifacts can be caused by electrical and mechanical interference, such as power lines, incorrect electrode contact, and impedance

fluctuations [Gutberlet et al., 2009]. Artifacts due to power line interference have frequency values of 60 Hz (North America) or 50 HZ (Europe). A notch filter is used to remove power line data and a high-pass filter is used to eliminate slow oscillations of signal for example resulting from small electrodes movements. Physiological artifacts can be caused by body activities. The most common are ocular movements (e.g, eye blinking, which affects mostly signals recorded on frontal and front polar regions), muscular artifacts (due to body movements or muscular tension), and heartbeat [Gutberlet et al., 2009]. These artifacts can be minimized using for example, the independent component analysis [Hyvärinen and Oja, 2000]. To enhance signal quality (e.g, increase the SNR of the signal) some spatial filtering techniques like common average reference, surface Laplacian, and bipolar derivation are also used [McFarland et al., 1997]. The above pre-processing procedures are the first step to remove noisy or redundant information.

3.3 Feature Extraction

To accurately decode EEG patterns, discriminative features need to be extracted from the pre-processed EEG signals. Feature extraction methods should be adjusted to the neural mechanisms being used. Accordingly, they use information in time, frequency, time-frequency, and spatial domains. Temporal methods extract the temporal evolution of the signals. They are useful for signals with timing-specific properties, usually evoked potentials, such as ERPs. There are several feature extraction techniques applied in time-domain, such as the amplitude of raw EEG signals [Hoffmann et al., 2005], autoregressive parameters [Schlögl et al., 1997], and matched filters [Serby et al., 2005]. Frequency domain methods use the information on specific frequency bands. These methods are suitable for SMR and SSVEP BCIs since their patterns are based on brain rhythms activities. Examples of frequency domain and time-frequency methods include Fourier transform and periodogram-based techniques [Brunner et al., 2006], short-time Fourier, and wavelets, which provide information both in temporal and frequency variations of the signal [Kevric and Subasi, 2017]. Spatial domain algorithms extract information based on the combination of EEG channels. Several approaches have been proposed in the literature such as common spatial patterns, xDAWN algorithm [Rivet et al., 2009], and Fisher criterion beamformer (FCB) [Pires et al., 2011a]. FCB has been used in different applications in our research group with successful results compared to the state-of-the-art methods. For this reason, we extended the FCB to a Riemannian manifold in order to increase its robustness to deal with EEG variability between sessions/subjects.

3.3.1 Statistical spatial filter: Fischer Criterion Beamformer

Consider an EEG epoch $X = [x(t_1) \quad x(t_2) \quad ... \quad x(t_{Ns})]$ of N_c channels with L time samples, which is represented by a spatio-temporal matrix denoted by $X \in \Re^{N_c \times L}$. A spatial filter is defined as a weighting

vector, that combines the data of N_c channels at each time instant

$$Y = v^T X \tag{3.2}$$

where Y denotes the output projection of vector X.

Spatial filters differ by the criterion or criteria they use to discriminate classes. We briefly review the statistical spatial filter called Fisher criterion beamformer, one of the spatial filters proposed in our research lab [Pires et al., 2011a], which extends the concept of Fisher linear discriminant to the spatial domain. FLD aims to optimize class separability based on the within and between class spatial scatter matrices.

Let S_w and S_b denote the spatial within-class scatter matrix and the between-class scatter matrix defined by:

$$\mathbf{S}_w = \sum_i \sum_{k \in C_i} (\mathbf{X}_{i,k} - \mathbf{m}_i)(\mathbf{X}_{i,k} - \mathbf{m}_i)^T \tag{3.3}$$

$$\mathbf{S}_b = \frac{1}{K} \sum_i kc(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \tag{3.4}$$

where m_i and m are the mean of the epochs in class C_i and the mean of all epochs respectively defined as

$$m_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \mathbf{X}_{i,k} \quad \text{and} \quad m = \frac{1}{N} \sum_{k=1}^{N} \mathbf{X}_k \tag{3.5}$$

where N_i denotes the number of epochs in class C_i and N denotes the number of all epochs. Using S_w and S_b, the Fisher's criterion (FC) is defned as the Rayleigh quotient:

$$J(V) = \frac{V^T S_b V}{V^T S_w V} \tag{3.6}$$

The optimal filter V should simultaneously maximize between-class scatter matrix, and minimize within-class scatter matrix. The selected filter is the eigenvector associated with the largest eigenvalue and it is obtained by solving the generalized eigenvalue problem:

$$\mathbf{S}_b V = [(I - \theta)\mathbf{S}_w + \theta I] V \Lambda \tag{3.7}$$

where I is the identity matrix, Λ is the eigenvalue matrix, and θ denotes the parameter obtained from training data, which increases class discrimination.

3.3.2 Riemannian Manifold of Symmetric Positive Definite Matrices

Riemannian manifolds are smooth manifolds in which the tangent space is endowed with a smoothly varying inner product called the Riemannian metric [Absil et al., 2009, pages 45–50]. Recently, Riemannian geometry has attracted increasing attention from the BCI research community, due to its robustness and important invariance properties. It has been used to solve transfer learning issues [Zanini et al., 2017], to design classifiers [Barachant et al., 2010], and decrease calibration time [Lotte, 2015], see a review in [Yger et al., 2016]. Covariance matrix lies in a Riemannian manifold of the symmetric positive definite (SPD) matrices. Riemannian metrics can be used directly in the native space of covariance matrices, thus some approaches have been developed to increase spatial filters robustness which are based on covariance matrices [Xu et al., 2020]. Here, we will explore the congruence invariance of the Riemannian distance to extract "invariant" features that might improve the generalization across sessions/subjects.

In this section, we introduce some basic definitions and properties of SPD manifolds, presenting tools to manipulate EEG data in this manifold through their spatial covariance matrices used in spatial filters. It provides the basis for the method called Riemannian Fisher criterion beamformer presented in chapter 6.

3.3.2.A Notation

Consider the spatial covariance matrix $C_i \in \Re^{N_c \times N_c}$ of i-th trial computed from

$$C_i = X_i X_i^T \tag{3.8}$$

where X is the epoch defined in section 3.3.1, and N_c is the number of channels. Let:

- $M(n) = \{M \in \Re^{n \times n}\}$ be the space of $n \times n$ square matrices;
- $S(n) = \{S \in M(n), S = S^T\}$ the set of all $n \times n$ symmetric matrices in the space of $M(n)$;
- $P(n) = \{P \in S(n), P > 0\}$ the space of all symmetric positive-definite (SPD) matrices;
- $\mathcal{S}_{++}^n$ the SPD manifold;
- $GL(n)$ the set of real invertible $n \times n$ matrices;
- $I_n \in \Re^{n \times n}$ the identity, and
- ^{T} the transpose operator.

3.3.2.B Riemannian Distance

Riemannian manifolds are smooth differentiable manifolds equipped with metrics for estimation of similarity or dissimilarity between points. The Affine invariant Riemannian metric (AIRM) is the Riemannian distance between two SPD matrices $P1, P2 \in P(n)$ which is defined as the minimum length of the curves between them, called geodesic [Moakher, 2005]

$$\delta_R(P_1, P_2) = \|log(P_1^{-1}P_2)\|_F = \left[\sum_i log^2\lambda_i\right]^{\frac{1}{2}} \tag{3.9}$$

where $log(.)$ denotes the matrix logarithm, $\|.\|_F$ is the Frobenius norm of a matrix, and λ_i are the real eigenvalues of $P_1^{-1}P_2$.

The AIRM $\delta_R(.,.)$ has many attractive invariance properties and two of them can be particularly relevant in BCI [Zanini et al., 2017]:

1) Invariance under congruent transformations: $\delta(CP_1C^T, CP_2C^T) = \delta(P_1, P_2) \quad \forall C \in GL(n)$.

2) Invariance under inversion: $\delta(P_1^{-1}, P_2^{-1}) = \delta(P_1, P_2)$.

Congruence invariance property is very important for across sessions/subjects BCI learning because it implies that the distance between two SPD matrices remains unchanged after a linear invertible transformation in the data. For example, the change in set up configurations between sessions (e.g., electrode positions and electrode impedance) produce a different EEG signal that consequently produce a different covariance matrices, however the AIRM is very similar between these sessions.

3.3.2.C Riemannian Mean

The Euclidean (arithmetic) mean minimizes the sum of the squared Euclidian distances of a set points. Similarly, the Riemannian mean (also known as Fréchet/Karcher/geometric mean) of a set of SPD matrices P_1 ... P_N is defined as the SPD matrix that minimizes the sum of the squared Riemannian distances to the given SPD matrices:

$$\mathcal{G}(P_1 \ ... \ P_N) = \arg min_{P \in P(n)} \sum_i^N \delta_R^2(P, P_i) \tag{3.10}$$

There is no closed-form expression for Riemannian mean ($N > 2$). It is usually computed iteratively, and several algorithms have been proposed in the literature [Barachant et al., 2010, Jeuris et al., 2012]. The Riemannian mean also has invariance under congruent transformations: $\mathcal{G}(CP_1C^T, CP_2C^T) = C\mathcal{G}(P_1, P_2)C^T \quad \forall C \in GL(n)$.

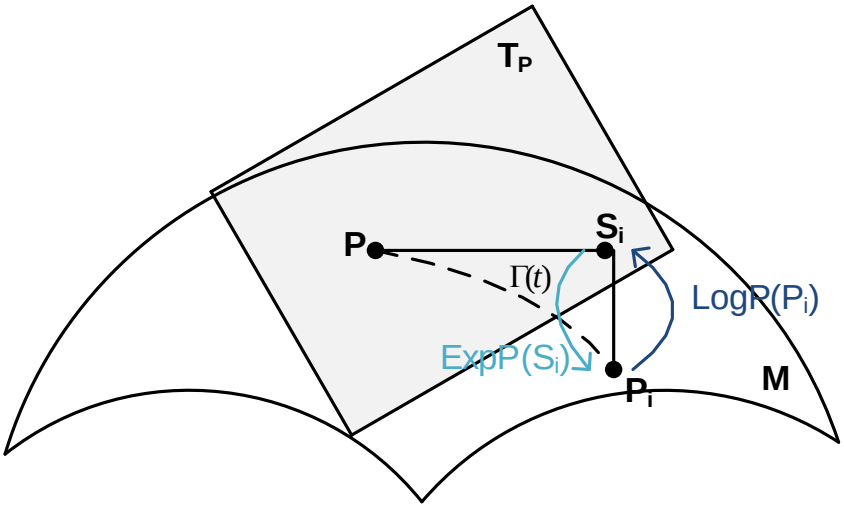


Figure 3.3: Tangent space at point P, and the relations between tangent space T_P, tangent vectors S_i, and geodesics $\Gamma(t)$ between P and P_i.

3.3.2.D Tangent Space

For each point P (covariance matrix in this case) of the manifold M, a tangent space T_P is a vector space that contains all tangent vectors at the point P (Fig. 3.3). The tangent space is Euclidean, it provides a bridge to a local vector space representation. Considering S_1 and S_2 two tangent vectors in the tangent space, the inner product at P is defined as:

$$\langle S_1, S_2 \rangle_P = Tr(S_1 P^{-1} S_2 P^{-1}) \tag{3.11}$$

where, $Tr(.)$ represents the trace operator.

The tangent space is endowed with an exponential mapping $ExpP : T_P \rightarrow M$ that projects a tangent vector S_i into a point (covariance matrix, P_i) of the manifold, defined as:

$$ExpP(S_i) = P_i = P^{\frac{-1}{2}} exp(P^{\frac{1}{2}} S_i P^{\frac{-1}{2}}) P^{\frac{1}{2}} \tag{3.12}$$

The inverse mapping is called logarithmic mapping $LogP : M \rightarrow T_P$ which enables going back to the tangent space, that is, projects a point on the manifold back to the tangent space, given by:

$$LogP(P_i) = S_i = P^{\frac{-1}{2}} log(P^{\frac{1}{2}} P_i P^{\frac{-1}{2}}) P^{\frac{1}{2}} \tag{3.13}$$

These two mappings allow switching between the manifold and tangent space.

3.4 Feature Selection

The redundancy of the features hinders the classification, making it more complex and less reliable. Therefore, the goal of feature selection is to reduce the number of features by selecting just the most relevant. There are two approaches, namely, wrappers and filter methods. The wrapper algorithms select the features based on the accuracy of the specific classifier. On the other hand, the filter methods select the features before the training of the classifier and its metric is independent of the classifier error rate. Filter methods have usually better generalization and lower computational costs. In this thesis, we used the r^2 correlation (filter method).

3.4.1 r^2 correlation

The r^2 coefficient is a bi-serial correlation coefficient with a value between 0-1 that quantifies the relevance of each feature for class discrimination. Let us consider the feature vector $X = [x_1, ..., x_K]$ and a corresponding class label $C_i, i \in [+, -]$. Then, X_+ and X_- are the samples of classes C_+ and C_-, respectively. The coefficient r is computed as

$$r(X_+, X_-) = \frac{\sum_{k=1}^{K}(X_+^k - \overline{X_+}) \times (X_-^k - \overline{X_-})}{\sqrt{\sum_{k=1}^{K}(X_+^k - \overline{X_+})^2} \times \sqrt{\sum_{k=1}^{K}(X_-^k - \overline{X_-})^2}} \tag{3.14}$$

where $\overline{X_i}$ denotes the average of the samples in class C_i. The selected features are the ones with the highest r^2 score.

3.5 Classification Methods

After feature extraction and selection, a classification algorithm is used to automatically determine the user's intention. Based on training data, a classifier aims to predict the classes for the new instances. There are a wide variety of classification methods employed in BCI systems. Among these classification algorithms, most notably are the linear discriminant analysis and extensions (e.g, Fisher LDA) [Hoffmann et al., 2008, Blankertz et al., 2011], the support vector machines [Schröder et al., 2005, Rakotomamonjy and Guigue, 2008], the artificial neural networks [Coyle et al., 2010, Hazrati and Erfanian, 2010], the naïve Bayes [Kohlmorgen and Blankertz, 2004, Machado and Balbinot, 2014], and more recently the Riemannian classifier [Barachant et al., 2011], and deep learning methods [Li et al., 2020]. Detailed reviews can be found in [Lemm et al., 2011, Nicolas-Alonso and Gomez-Gil, 2012, Lotte et al., 2018]. Here, in the experiments, a Naïve Bayes (NB) classifier and a Fisher linear discriminant (FLD) classifier are used. A comparative analysis made in [Pires, 2011, section 5.5.1] using several classical classifiers showed that when the classifiers are applied to statistical spatial filtering projections, all classifiers per-

formed similarly. Considering the ease of implementation of NB and LDA and the good interpretability of their probability/likelihood scores, we opted for these classifiers against other more complex methods such as the support vector machine.

3.5.1 Naïve Bayes Classifier

NB classifier is a supervised learning method based on Bayes' theorem, and it assumes that the features are strongly independent [Friedman et al., 1997]. Considering a class $C_i, i \in [+, -]$ and feature vector $x = [x_1, x_2, ..., x_n]$, the Bayes theorem is calculated as:

$$P\left(C_i|x_1, \ldots, x_n\right) = \frac{P(C_i)P\left(x_1, \ldots, x_n|C_i\right)}{P(x_1, \ldots, x_n)} \tag{3.15}$$

where $P\left(C_i \mid x_1, ..., x_n\right)$ denotes the probability of a given feature $[x_1, x_2, ..., x_n]$ being in class C_i (posterior conditional probability), $P\left(x_1, ..., x_n| \, C_i\right)$ indicates the probability of feature given class C_i (conditional probability), $P(C_i)$ is the prior probability of the class (the frequency of class C_i), and $P\left(x_1, ..., x_n\right)$ is the probability of features $[x_1, x_2, ..., x_n]$ occurring (prior probability of the feature), and it is defined as

$$P\left(x_1, ..., x_n\right) = \sum_i P\left(x_1, \ldots, x_n|y\right) P(C_i) \tag{3.16}$$

It is computationally expensive to compute $P(X|Ci)$ with large amount of features. Using the naive assumption which assumes the independence of attributes, the $P(X|Ci)$ is given approximately by

$$P(X|Ci) = \prod_{i=1}^{n} P(x_i|C_i) \tag{3.17}$$

Thus, one can label new instances x_i with a class C_i that achieves the highest posterior probability:

$$\hat{y} = arg \ \max P(C_i) \prod_{i=1}^{n} P(x_i|C_i) \tag{3.18}$$

NB classifier has many advantages, namely, simplicity, low computational cost and does not need regularization. Despite its simplicity, NB classifier performs better then other sophisticated classifiers in many situations [Ding and Peng, 2005], [Pires, 2011, section 5.5.1].

3.5.2 Fisher's linear discriminant classifier

The Fisher's linear discriminant classifier tries to optimize class separability based on the within and between class scatter matrices [Bishop, 2006, pages 186-189]. Considering a d-dimensional input vector $x = [x_1, x_2, ..., x_n]$, we can projecting the vector x into a line as:

$$y = w^T x \tag{3.19}$$

where w denotes the weight vector. The two scatter matrices are defined by:

$$\mathbf{S}_i = \sum_i (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \tag{3.20}$$

$$\mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2 \tag{3.21}$$

$$\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m_2})(\mathbf{m}_1 - \mathbf{m_2})^T \tag{3.22}$$

where S_w and S_B denotes the within and between class matrix respectively, and m_i indicates the mean of vector x of class i. The FLD selects the optimal projection w in order to maximize the range of the scalars and to obtain a better discrimination. The vector w that yielding the maximum ratio between-class scatter to within -class scatter is computed as:

$$\mathbf{w} = \mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m_2}) \tag{3.23}$$

3.6 Performance Metrics

In binary classification, the classifier needs to predict two classes, for example, target and non-target. Based on the classifier's outputs a true positive (TP), a false positive (FP), a true negative (TN), and a false-negative (FN) can occur. Using this information the BCI system can be evaluated through the following metrics:

$$Sensitivity\ (Sens) = \frac{TP}{TP + FN} \tag{3.24}$$

$$Specificity\ (Spec) = \frac{TN}{TN + FP} \tag{3.25}$$

$$Accuracy\ (Acc) = \frac{TP + TN}{FN + TN + FP + TP} \tag{3.26}$$

Classification accuracy is the simplest metric for evaluating the BCI performance used in the BCI literature. However, in P300-based BCI this measurement requires careful treatment since the probability of non-target class is much higher than the target class. This produces highly unbalanced classes, leading to biased evaluations. To overcome this limitation, in the offline analysis we used a balanced accuracy

$$Balanced\ Acc = \frac{Sens + Spec}{2} = 0.5 \times \frac{TP}{TP + FN} + 0.5 \times \frac{TN}{TN + FP} \tag{3.27}$$

The information transfer rate, or bit rate, is another popular metric for evaluating the BCI performance [Wolpaw et al., 2000]. ITR is derived from Shannon's theory [Shannon, 1948], the BCI system is seen as a noisy communication channel. It represents the maximum capacity of the channel in bits per minute (bpm). Wolpaw, et al., 2000 proposed its theoretical formula:

$$ITR = rSPM \times \left[log_2\left((N_s\right) + p \times log_2\left(p\right) + \left(1-p\right) \times log_2 \frac{1-p}{N_s - 1} \right] \tag{3.28}$$

where p is the online accuracy of the overall system, N_s is the number of symbols, and $rSPM$ is the raw rate of symbols per minute defined as

$$rSPM = \frac{60}{N_{rep} \times N_s \times SOA + ITI} \tag{3.29}$$

where ITI is the inter-trial interval and SOA is the stimulus onset asynchrony. The ITR is the most common metric used in the literature, it includes both the number of encoded symbols and accuracy, and it does not depend on any specific paradigm. However, it has some limitations. The ITR may lead to inexact bit rates in particular cases since it does not considers the correction of errors. Some researches considered that ITR may be an unrealistic metric for evaluating the performance of practical BCI systems [Seno et al., 2010, Spüler et al., 2012].

For BCI systems contemplating error correction, other metrics have been proposed to measure the information transfer rate and considering the impact of errors to convey error-free messages. From a practical point-of-view, when we use a spelling BCI system with error correction strategies, a bit rate represents the number of correct letters spelled per minute. Therefore, more recently, Spüler and colleagues suggested a particular formula to assess the efficacy of a P300 spelling interface [Spüler et al., 2012]:

$$B = log_2\left(\left(N_s - 1\right) \times \left(2p - 1\right)\right. \tag{3.30}$$

For a P300 speller with the error correction system the information transfer rate is defined as:

$$B_{ECS} = \left(1 - Pos\right) \times log_2\left(\left(N_s - 1\right) \times \left(2P_{ECS} - 1\right)\right. \tag{3.31}$$

where Pos and P_{ECS} represent the percentage of error produced by the interface and the error accuracy respectively:

$$Pos = \frac{\left(FP + TP\right)}{\left(TN + FP + TP + FN\right)} \qquad P_{ECS} = \frac{TN}{\left(TN + FN\right)} \tag{3.32}$$

Therefore, we can rewrite the information transfer rate for ECS as:

$$B_{ECS} = log_2\left(N_s - 1\right) \times \frac{TN - FN}{TN + FP + TP + FN} \tag{3.33}$$

B_{ECS} is found to be equivalent to the Utility metric proposed in [Seno et al., 2010].

The effective symbols per minute, i.e., the number of symbols per minute conveyed without errors, is also a straightforward and meaningful metric that encompasses speed and error impact [Seno et al., 2010]

$$eSPM = rSPM \times (2p - 1) \tag{3.34}$$

4

Error-Related Potential: Double Detection Approach

Contents

In this chapter, we propose and describe a P300-based BCI speller combined with a double ErrP detection to automatically correct erroneous decisions [Cruz et al., 2018a]. This novel approach introduces a second error detection to infer whether wrong automatic correction also elicits a second ErrP. Thus, two single-trial responses, instead of one, contribute to the final selection, improving the reliability of error detection. Moreover, to increase error detection, the evoked potential detected as a target by the P300 classifier is combined with the evoked error potential at a feature-level. Discriminable error and positive potentials (response to correct feedback) were clearly identified. The proposed methods were validated through a set of experimental tests (referred to as Case-study I). Results achieved in this case-study showed that the integration of ErrP detection into the P300 Speller system improves the classification accuracy, however it requires a long calibration time in order to gather enough error-samples to train the classifier. The long training process is a very time consuming, tedious and it limits the applicability of ErrP. Therefore we replicated the protocol using little ErrP-calibration time by calibrating the BCI with a high rate of errors (referred to as Case-study II). The goal of these experiments is to analyze the generalization of ErrP classifier across error-rates [Cruz et al., 2018b].

4.1 Proposed Methodologies

The proposed approach integrated the error detection/correction classifiers into the P300 Speller system in order to increase the reliability of the BCI system. The BCI application is the LSC speller described in section 3.1.1. The feedback procedure of the original LSC speller was slightly adapted to accommodate error detection during online experiments, that is, the ITI is set to 4 seconds, to have time for double-error detection and correction and the time for the user to shift his/her attention to the next desired symbol.

An example of the online operation of the P300-ErrP BCI is illustrated in Fig. 4.1 (a video [1] is also provided for a better understanding). After the P300 classification, the detected letter is visually shown to the user (feedback is provided at the position of the letter). If the system detects an ErrP (feedback of incorrect letter) the system corrects the letter replacing it by the letter classified with the second highest classification score. The new letter is fed back to the user at the respective position. If the system detects wrong feedback (e.g., a correct letter is changed to a wrong letter), the last letter is changed again to the first detected letter, otherwise, the corrected letter is confirmed, the letter is written at the center of the screen, and the spelling system continues to the next trial. The feedbacks occur within the ITI, as shown in the temporal diagram of Fig. 4.2. Users were instructed to keep always focused on the desired letter without moving the eyes regardless of the position of the two feedbacks until the letter was written at the center. After the letter was presented at the center, participants still had 2 seconds to

[1] https://home.isr.uc.pt/~gpires/ErrP/video_ErrP.mp4

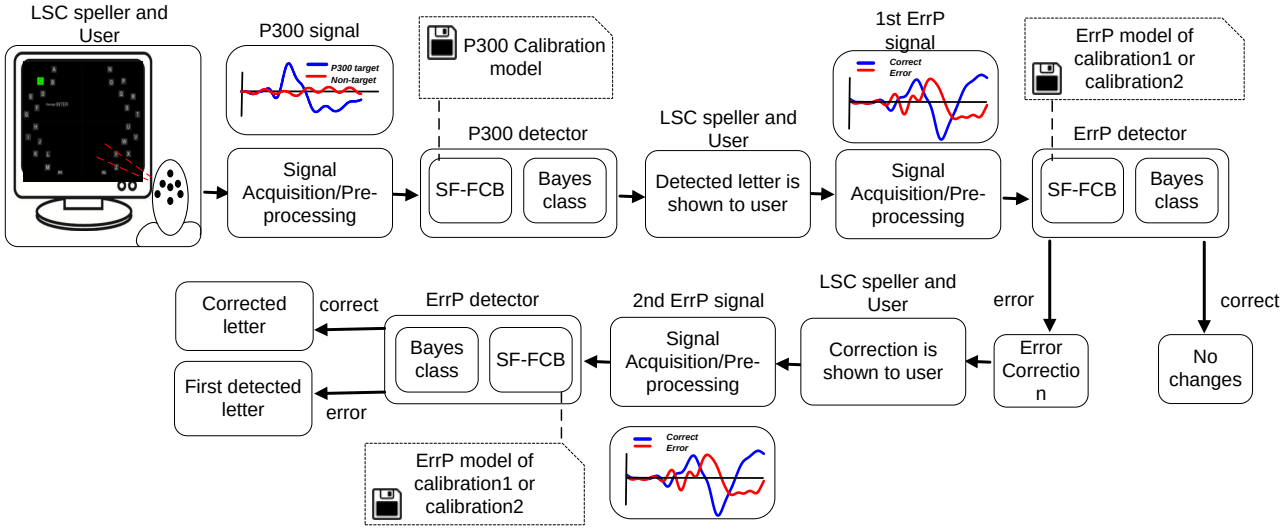


Figure 4.1: Schematic representation of the P300-ErrP BCI system. The user focuses on the target letter of the P300-LSC speller. The EEG signal is acquired and pre-processed, and the target event is detected with the P300 detector. The detected letter is shown to the user and the ErrP-detector evaluates whether an error has occurred. If an ErrP is detected, the letter with the second highest classification score is shown to the user, and the ErrP-detector checks for a possible error again. If the system detects an ErrP, the final classification is the initial detected letter, otherwise the corrected letter is selected (see a demonstrative video in [Pires, 2018].

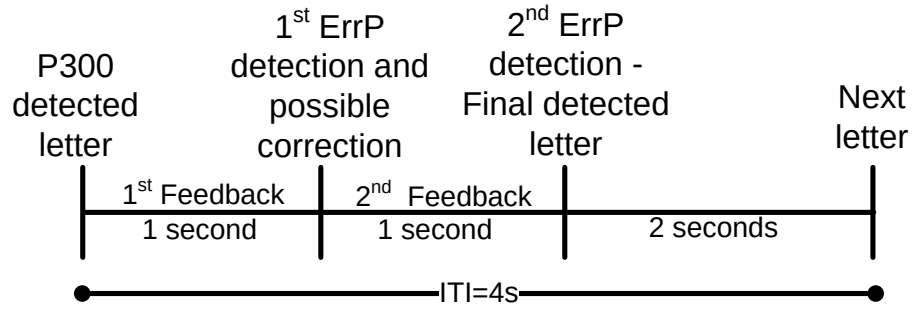


Figure 4.2: Temporal diagram of the double feedback ERPs. First, the P300 system detects the target letter, 1 second after occurs the 1^{st} ErrP detection. If an error potential is detected then an auto-correction is made. The 2^{nd} ErrP detection occurs 1 second after. Then, users have 2 seconds to focus on the new letter.

focus on the next letter.

4.2 ErrP Detection in the Context of the LSC P300-based BCI Speller (case-study I)

Experiments were conducted using the LSC speller in order to assess the feasibility of detecting ErrP in a single trial level and using this information to correct wrong decisions and thus increase the BCI perfor-

mances. The double ErrP detection approach was experimentally validated on nine healthy participants and one tetraplegic participant.

4.2.1 Participants

The experiments were carried out by nine healthy participants (S1-S9) and one tetraplegic participant (P1) with medullar injury (C4/C5 level) with ages between 24 and 43 years old. Five participants (S1, S2, S5, S8 and S9) had previous experience with P300 BCI, and the others had never used a BCI before. The tetraplegic participant has slight hand movements allowing him to interact with the computer and control a powered wheelchair. Each participant signed an informed consent which included the description and purpose of the research, the experimental procedure, the potential risks and the permission to publish the results.

4.2.2 Data Acquisition

The electric brain potentials were recorded with a g.USBamp bioamplifier, from electrodes Fz, Cz, C3, C4, CPz, Pz, P3, P4, PO7, PO8, POz, and Oz according to the international extended 10-20 standard system. The electrodes were referenced to the right ear lobe and the ground was located at AFz. The EEG signals were acquired with active Ag/AgCl electrodes, except for participant S2 who used passive electrodes. The EEG signals were pre-processed using a notch filter at 50 Hz and a band-pass filter with lower cutoff frequency of 1 Hz and a higher cutoff frequency of 10 Hz and sampled at 256 Hz.

4.2.3 Calibrations and Online Session

The experiment comprised two sessions which included three phases as detailed next: **calibration 1**, **calibration 2** and final **online operation**. In session 1, participants made two calibrations to acquire labeled data associated with target and non-target events, and with correct and error feedback responses (see Fig. 4.3). In **calibration 1**, participants were asked to attend the 10 letters of the word “INTERFACES” which were successively provided at the center of the screen. For each letter, all symbols flashed 9 times (9 complete rounds), lasting about 5 minutes. The acquired dataset was composed of 90 target epochs and 2430 non-target epochs, which were used to train the P300 classifier to be used in online operation of calibration 2.

Calibration 2 served to acquire labeled data when the user received positive (expected symbol) and negative (wrong symbol) feedback returned by the P300 classifier. Participants had to write online several times the Portuguese sentence ”ESTOU-A-ESCREVER-COM-UMA-INTERFACE-BCI” (38 characters) without either interruption or correction. This sentence is here referred to as a block. The number of repetitions per trial was adjusted individually so that all participants controlled the LSC speller with

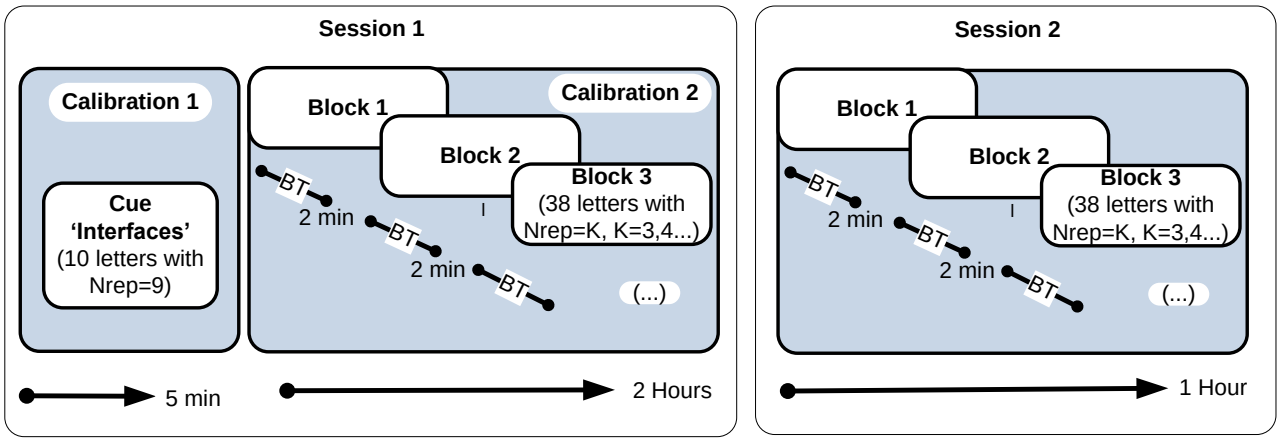


Figure 4.3: Schematic representation of data acquisition sessions. Session 1 included calibration 1 to gather data associated with target and non-target events (approx. 5 min), and calibration 2 to gather ErrP and correct ERP data from online feedback (approx. 2 hours). Each block corresponded to a sentence of 38 characters, with an interval of 2 min between blocks. In session 2, users wrote the sentences during one hour using the online error detection and correction BCI.

similar accuracies. Thus, N_{rep} was selected according to users' performance in **calibration 1** (considering a P300 offline accuracy around 90%). The sentence was written several times until a limit of two hours occurred. The duration of each block (BT-block time) varied according to N_{rep}, and is given by

$$BT = N_{sc} \times TT \tag{4.1}$$

where $N_{sc} = 38$ is the number of spelled characters and TT is the overall time for one trial (eq. 3.1). Between each block, the participants rested two minutes. For example, if N_{rep}= 5 each block lasted about 10 minutes, and participants performed 10 blocks. It was expected that a long experiment would lead to a natural increase of errors. The number of errors varied across participants (the minimum number of errors was 31 and the maximum was 86). Labeled data allowed to infer the existence of ErrPs and to train the Error detection classifier, as described in section 4.2.4.B.

In session 2, the final P300-ErrP system combined the two classifiers to detect targets and wrong selections. This session was held on a different day of the first session for all participants except S9, who made the two sessions on the same day. Two participants (S7 and S8) left the laboratory and took part only in the first phase (session 1) of the experiments. Participants were asked to spell the same sentence as in **calibration 2**, repeating the sentence during approximately one hour.

4.2.4 Feature Extraction and Classification

The 12 EEG signals are segmented into 1-second epochs, with onset on each stimulus. At the end of a trial, the P300 classifier predicts which event elicited a P300 ERP. The selected letter is fed back to the

user and an epoch of 1 s is recorded immediately after. This single trial epoch is classified by the error detector, which can be applied once or twice as exemplified in Fig. 4.1.

4.2.4.A P300 Classification

The epochs of the N_{rep} rounds are averaged for each channel. Then, a statistical spatial filter called Fisher criterion beamformer, proposed in [Pires et al., 2011a] obtains a single discriminative projection from the 12 EEG channels. Spatial filtering is a common feature extraction technique in EEG-based BCIs that simultaneously allows to increase the signal-to-noise ratio and reduce the dimension of the feature data. Considering a spatio-temporal matrix $E_{N_c \times L}$ representing the epochs of N_c channels with L time samples ($L = 256$), the spatial filter projection is obtained from $y = v_1^T E$, where v_1 is the optimal spatial filter, obtained from calibration 1, and T denotes the transpose operator. The 100 most relevant features are selected using the r-square correlation method. Features are then classified by a Bayes classifier, which returns a target probability (score) for each symbol $j, P_j, \in \{1, \cdots, N_s\}$. Consider $S1$ and $S2$ the symbols with the highest and second highest probabilities, respectively

$$\mathrm{S}_1 \equiv \arg\max P_j \quad , j \in \{1, \cdots, N_s\} \tag{4.2}$$

$$\mathrm{S}_2 \equiv \arg\max P_j \quad , j \in \{1, \cdots, N_s\} \backslash S_1 \tag{4.3}$$

The symbol $S1$ is chosen as the primary target event fed back to the user, and S2 is used as a secondary target in case an error is detected.

4.2.4.B Error Detection

The error detector uses the same classification framework of the P300 detector. Two approaches were tested, the first one, referred to as ErrP classifier, classifies the EEG data segment $R_{N_c \times L}$ of the single trial response elicited by the feedback. The second approach, called ErrP-P300 classifier, illustrated in Fig. 4.4, classifies the EEG data segment $R_{N_c \times L}$ of the single trial response elicited by the feedback, concatenated with the EEG data segment of the supposed target event ($E_{N_c \times L}$), resulting in a spatio-temporal matrix $F = [E : R]_{N_c \times 2L}$. Each vector of F corresponds to one channel with 512 time samples $[e_1 e_2 \ ... \ e_L r_1 r_2 \ ... \ r_L]$. It is hypothesized that the concatenation of these two data segments may improve error detection, since they correlate to each other, thereby helping in the challenging task of single trial classification. It is expected that a correct symbol detected as a true P300 originates positive feedback, whereas a false target with similarities to the P300 waveform, but still distinct from this one, leads to an ErrP feedback response, as exemplified in Fig. 4.5. The diagram of the ErrP-P300 classifier is illustrated in Fig. 4.4. After concatenation, the FCB spatial filter is applied to F, obtaining the projection $z = v_2^T F$,

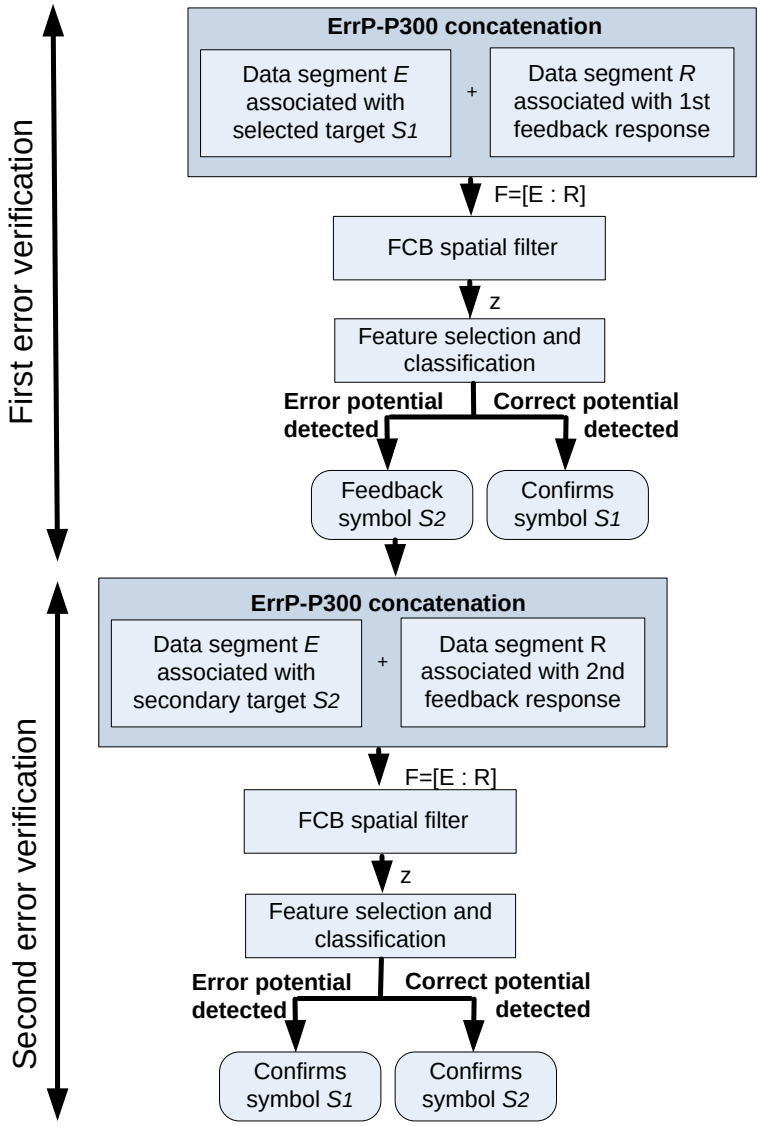


Figure 4.4: Online ErrP-P300 classification algorithm with double error verification. First and second error verifications use the single trial response elicited by the feedback concatenated with the EEG data segment of the supposed target event.

where v_2 is the optimal spatial filter, obtained from **calibration 2** . The 200 most relevant features selected from z with r-square correlation are classified by a Bayes classifier whose model was obtained in **calibration 2**. This process is the same in the first and second steps of error verification, as shown in Fig. 4.4.

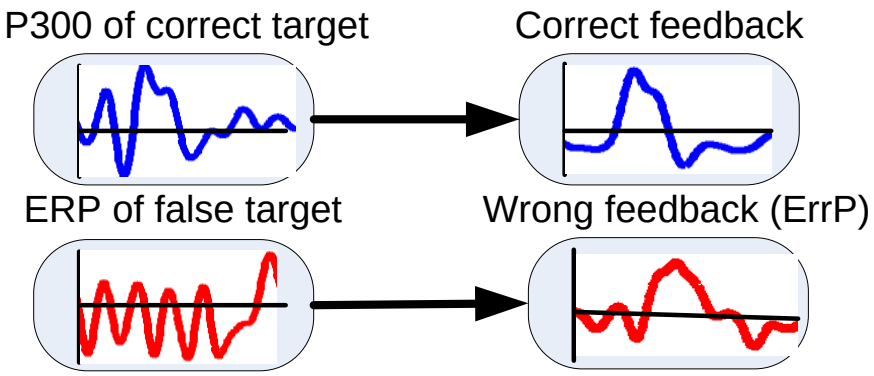


Figure 4.5: Expected ERP combinations: correct feedback is related with a typical P300 ERP (true target), and a wrong feedback is related to an ERP with waveform similarities to P300 ERP but still different from this (false target).

4.2.5 Double-check Error Correction

To be effective, the accuracy of the error detection must be higher than the P300 detection, otherwise the false positives will introduce a degradation of the overall performance. Moreover, the error correction rests on the premise that in case of error, the second event with the highest score is the target. This assumption may of course not be true, and therefore a second error feedback verification can avoid a performance degradation due to both wrong correction and wrong detection. The theoretical overall accuracy of the system with error detection and correction, but without using the second feedback is given by

$$P_f = P_{p3} \times \frac{TN}{TN+FP} + (1-P_{p3}) \times \frac{TP}{TP+FN} \times\ P_{cor} = P_p 3 \times\ Spec_1 + (1-P_p 3) \times\ Sens_1 \times\ P_{cor} \quad (4.4)$$

where $Spec_1$ and $Sens_1$ are the specificity and sensitivity of the first error detector, Pp3 is the accuracy of the P300 speller (without correction), and P_{cor} is the correction rate of the errors well identified, i.e., those errors that were identified and corrected by selecting target $S2$ (4.3). By introducing the second error checking in the system, it is given the opportunity to eliminate some False Positives, attaining the new overall accuracy

$$P_f^{'} = \ P_{p3} \times \frac{TN + FP \times Sens_2}{TN+FP} \ + (1-P_{p3}) \times Sens_1 \times P_{cor} \times Spec_2 \quad (4.5)$$

where $Spec_2$ and $Sens_2$ are the specificity and sensitivity of the second error detector. If an ErrP is evoked when the secondary target is shown, a percentage of the False Positives may be transformed into True Negatives ($FP \times\ Sens_2$), although some well corrected symbols may be also altered (second term of (4.5)). Therefore, the second feedback yields an improvement

$$\triangle P = \frac{P_{p3} \times FP \times Sens_2}{TN + FP} - (1 - P_{p3}) \times Sens_1 \times P_{cor} \times (1 - Spec_2) \tag{4.6}$$

The classification procedure is schematically illustrated in Fig. 4.4. If an error is detected when the spelled symbol $S1$ is fed back to the user, the symbol is deleted and replaced by $S2$ which has the second highest target probability. The user reaction to this second symbol will provide a second verification of the existence of an error. If the new target is accompanied by a second error, it is discarded and the final classification corresponds to the first target, $S1$. On the other hand, if the new target is accompanied by a non-ErrP, target $S2$ is confirmed as the final classification. Four scenarios can occur, as exemplified in Fig. 4.6.

4.2.6 Performance metrics

The following metrics were used to evaluate the binary offline data obtained from calibration of the ErrP-P300 system: the sensitivity, specificity, and the balanced accuracy. The online performance was evaluated using the Sens, Spec, and Acc, the information transfer rate, the bandwidth of error correction systems, and the effective symbols per minute, thereby facilitating a direct comparison with state-of-the-art studies.

4.2.7 Results

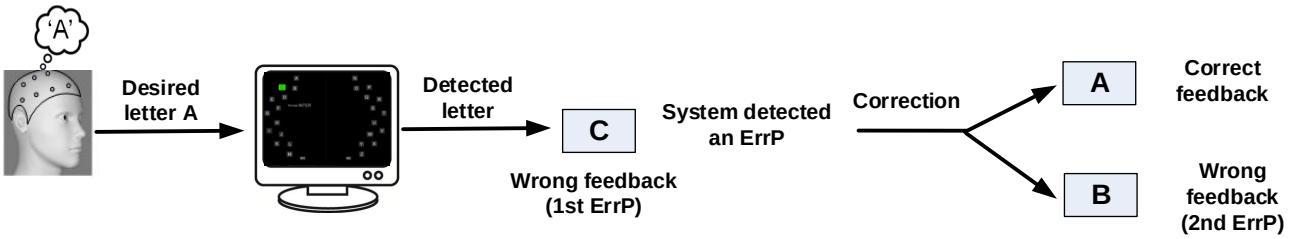
4.2.7.A Offline Performance of P300 and ErrP Detection

An offline analysis was performed to infer the efficiency of the extracted features for the binary P300, ErrP and ErrP-P300 classifiers. This analysis was based on the datasets collected during **calibration 2** (see Table 4.1). The P300 average accuracy was 87.8%, and the ErrP average accuracy was 84.5%. Concatenating the P300 and ErrP feature vectors in the ErrP-P300 classifier, the error detection accuracy increased about 7% (paired t-test, p = 0.005) over the ErrP classifier. These offline results showed that the feature concatenation improves error detection as initially hypothesized. These results agree with those reported in [Zeyl et al., 2015] and [Margaux et al., 2012], where a different fusion approach, was applied that combines the classification scores of P300 and ErrP classification, instead of the feature combination used in our approach.

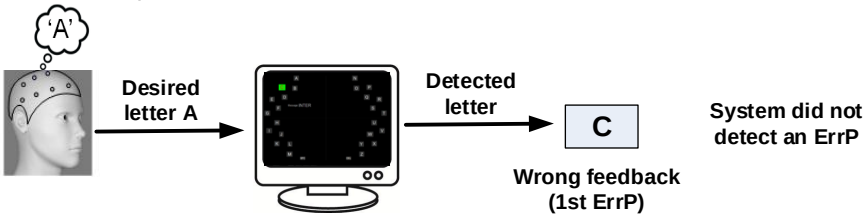
4.2.7.B Online Performance

Session 2 used the P300 and ErrP-P300 classifier models obtained in calibrations of session 1. The online experiments were carried out by 8 participants since participants $S7$ and $S8$ had to leave the laboratory. The number of repetitions per trial was set individually to values between 3 and 7 (average

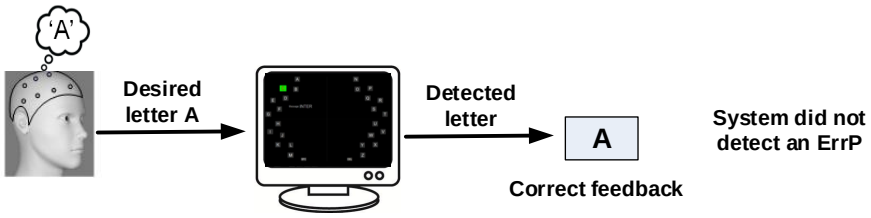
Scenario1: User wants to spell letter 'A' but the system detects letter 'C'. An ErrP is detected and an auto-correction is performed.



Scenario2: User wants to spell letter 'A' and system detects letter 'C', but system does not detect an ErrP, so no correction is performed.



Scenario3: User wants to spell letter 'A' and the system correctly detects letter A, so no correction is performed.



Scenario4: User wants to speller letter 'A' and the system correctly detects letter 'A', but system detects an ErrP, and an auto-correction is performed.

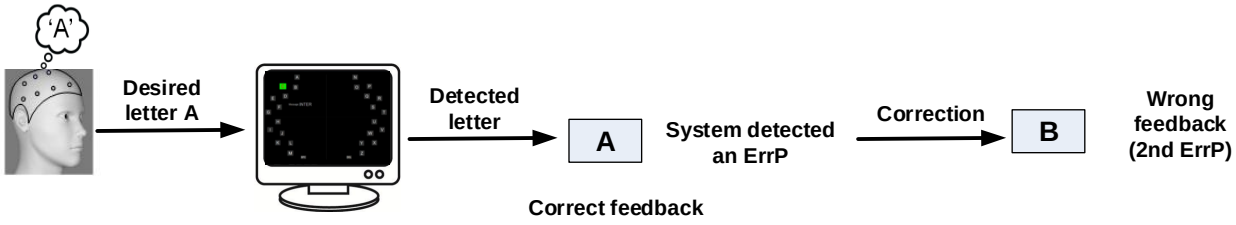


Figure 4.6: Schematic representation of of four scenarios that can happen using the double error detection approach.

of 5.8 repetitions) according to users' offline performances attained in Session 1 (yielding offline performance around 90%). The online results, obtained after one hour spelling out letters, are presented in Table 4.2. We opted to set the same ITI of 4 seconds for both conditions (with and without error correction) to keep the same raw rate of symbols, although a shorter ITI could have been considered

Table 4.1: Offline binary classification results for p300 and errp features used separately and concatened. Data obtained from calibration 2.

	Target detection			**Error detection**					
	P300			**ErrP**			**ErrP-P300**		
	Sens	Spec	Pp3 [a]	Sens	Spec	Pep [b]	Sen	Spec	Pp3e [c]
S1	82.7	83.7	83.2	90.7	86.3	88.5	92	86.9	89.5
S2	70.9	96.6	83.8	70.9	83.4	77.1	90.9	96.9	93.9
S3	87.9	96.2	92	93.9	92.1	93	98.5	94.6	96.6
S4	81.4	79.6	80.5	93	76.8	84.9	93	79.6	86.3
S5	92.4	86	89.2	75.8	88.5	82.1	83.3	91.5	87.4
S6	93.5	90.2	91.9	71	86.8	78.9	83.9	87.2	85.6
S7	85.9	93.8	89.9	87.3	80.5	83.9	94.4	92.3	93.3
S8	97.1	100	98.5	91.2	94.2	92.7	94.1	100	97.1
S9	86.9	92.5	89.7	98.4	95.9	97.1	100	97.2	98.6
P1	76.1	83.4	79.7	59.2	75.2	67.2	91.5	85.4	88.4
Mean	**85.5**	**90.2**	**87.8**	**83.1**	**86**	**84.5**	**92.2**	**91.2**	**91.7**
STD	**8**	**6.8**	**5.9**	**13**	**7.1**	**8.9**	**5.4**	**6.3**	**4.8**

[a] Pp3 is the balanced accuracy of P300 detector.
[b] Pep is the balanced accuracy of the error detector using only ErrP features.
[c] Pp3e is the balanced accuracy of the error-detector using the concatenation of P300 and ErrP features.

in the P300-speller without error correction. The online accuracy without error correction (Pre-Acc) was 84.8%. With error detection and correction, the online accuracy (Post-Acc), which included the detection of the first and second ErrP, as well as the error correction, increased the accuracy of 5.1% to 89.9%. The online accuracy was computed as the ratio between the number of correct symbols and the total number of spelled symbols. All participants, except S6, had an improvement, however this participant had the best initial performance (95.8%). Subject S1 had the highest improvement (11%) followed by the subject S9 (10%). The paired t-test shows that the improvement between the two conditions is statistically significant (paired t-test, p = 0.003). The ITR (3th and 4th columns) and BECS (5th and 6th columns) were for the Pre-correction condition 12.79 bpm and 3.31 bpt, respectively, and for the Post-correction the enhancement was of 1.40 bpm and 0.29 bpt respectively (paired t-test, p = 0.006 and p = 0.002). The average detection of the 1^{st} and 2^{nd} error were 88.4% and 84.8% respectively. The discrepancy of accuracy between the two detections can be explained by the lack of generalization of the classifier, once it was trained with the responses of the 1^{st} feedback. As will be shown in Section 4.2.7.C, the correct ERP and ErrP waveforms of the 1^{st} feedback responses differ from those of the 2^{nd} feedback, and therefore the classifier model may be overfitted to the 1^{st} feedback responses. In order to evaluate if there was a correlation between the speed of the speller paradigm (measured by the $rSPM$) and the detection accuracy of the ErrP, we calculated the Pearson correlation between them. Its value

Table 4.2: Online classification performance using the double error detection approach.

	Pre-Acc (%) [a]	Post-Acc (%)	Pre-ITR (bpm)[b]	Post-ITR (bpm)	Pre-BECS (bpt)[c]	Post-BECS (bpt)	*Pre-eSPM* [d]	*Post-eSPM*	Nrep [e]
S1	81.6	92.8	9.87	12.45	3	3.57	1.92	2.6	7
S2	93.2	94.7	15.96	16.49	4.1	4.2	3.34	3.46	5
S3	93.2	96.3	15.96	17.05	4.1	4.25	3.34	3.59	5
S4	84.7	89.5	11.81	13.03	3.3	3.6	2.37	2.69	6
S5	73.7	79.6	8.3	9.47	2.25	2.66	1.44	1.8	7
S6	95.8	95.8	14.85	14.85	4.35	4.35	3.12	3.12	6
S9	81.1	90.5	17.02	20.73	2.95	3.5	3.3	4.3	3
P1	75	79.6	8.55	9.47	2.38	2.63	1.52	1.8	7
Mean	**84.8**	**89.9**	**12.79**	**14.19**	**3.31**	**3.6**	**2.54**	**2.92**	**5.8**
STD	**8.5**	**6.8**	**3.58**	**3.9**	**0.81**	**0.7**	**0.83**	**0.9**	**1.4**

[a]Pre-Acc=pre-correction accuracy, Post-Acc=post-correction accuracy.
[b]Pre-ITR=pre-correction ITR, Post-ITR=post-correction ITR.
[c]pre-BECS=Pre-correction ITR with ECS, Post-BECS=post-correction ITR with ECS.
[d]Pre-eSPM=pre-number of symbols per minute, Post-eSPM=post-number of symbols per minute.
[e]Nrep is the number of repetitions used for P300 detection.

($r = 0.04$) shows that the performance of ErrP detection does not correlate with the rSPM value, which may suggest that the ErrP detection was not influenced by the BCI speed, as also suggested in [Iwane et al., 2016].

Table 4.3 shows the confusion matrix for the 1^{st} and 2^{nd} error detection and its sensitivity and specificity. The mean sensitivity and specificity for the 1^{st} error detector are respectively 91.9% and 88.0%, however, there is some variability across subjects. Subjects S4 and S6 had the highest values of false positives and consequently the lowest specificity. Analyzing the errors for these two subjects in the 2^{nd} error-detector, we verify that they present a high rate of true positives showing that most of the correct targets detected as errors in the 1^{st} error-detector are re-corrected in the 2^{nd} error-detector (confirmed in Table 4.2), emphasizing the importance of the double ErrP detection. The mean sensitivity and specificity for the 2^{nd} error-detector were respectively 85.8% and 73.4%. Fig. 4.7 compares the (3.34) with and without error correction. For the Pre-correction condition, the average eSPM was 2.54 symbols per minute and for the Post-correction condition it was 2.92, with a significant increase of 0.38 (paired t-test, p = 0.006). The highest eSPM was 4.3 attained by subject S9 who had the greatest increase between the two conditions. It is important to note that the eSPM was computed including the 4-seconds ITI. The Pcor (the correction rate of the errors well identified) was 44%. Replacing Pcor in (4.4) we obtain an improvement of 6.4% (second term). However, since the true negative rate (Spec) is not 100%, there are several false positives degrading the overall accuracy. Eq. (4.5) defines the performance $P_f^{'}$, obtained

Table 4.3: Confusion matrix of the online classification for the first and second error detectors using errp-p300 classifier.

	TP	FP	FN	TN	Sens	Spec	Acc-ErrP
First Error detection							
S1	25	11	3	113	89.3	91.1	90.8
S2	11	5	2	172	84.6	97.2	96.3
S3	13	20	0	157	100	88.7	89.5
S4	29	46	0	115	100	71.4	75.8
S5	35	7	5	105	87.5	93.8	92.1
S6	7	44	1	138	87.5	75.8	76.3
S9	35	5	1	149	97.2	96.8	96.8
P1	34	12	4	102	89.5	89.5	89.5
Mean	**23.6**	**18.8**	**2**	**131.4**	**91.9**	**88**	**88.4**
STD	**11.6**	**16.9**	**1.9**	**26.3**	**6.1**	**9.5**	**8.1**
Second Error detection							
S1	13	3	2	18	86.7	85.7	86.1
S2	10	2	1	3	90.9	60	81.3
S3	22	1	2	8	91.7	88.9	90.9
S4	40	1	14	20	74.1	95.2	80
S5	22	3	6	11	78.6	78.6	78.6
S6	46	3	1	1	97.9	25	92.2
S9	14	0	6	20	70	100	85
P1	32	6	1	7	97	53.8	84.8
Mean	**24.9**	**2.4**	**4.1**	**11**	**85.8**	**73.4**	**84.8**
STD	**13.2**	**1.8**	**4.5**	**7.6**	**10.5**	**25.4**	**4.9**

after applying the second error detector, which allows to eliminate some of the false positives. The improvement is defined by (4.6), yielding a value of 8.1% over the value obtained from the first correction in (4.4).

4.2.7.C Evoked Potentials After Correct and Wrong Feedbacks

Previous studies [Ferrez and Millán, 2008a, Schalk et al., 2000] demonstrated that the most discriminative brain regions for ErrP are fronto–central channels along the midline. We focused our analysis on channels Fz and Cz. The grand average ERPs of the eight subjects regarding to correct and incorrect feedback are shown in Fig. 4.8. There is an ERP after the presentation of both error and correct trials. The ErrP after the incorrect feedback has a positive peak around 200 ms followed by a negative peak around 300 ms (likely related to FRN). A second positive and negative peak appears around 450 ms (likely related to Pe) and 600 ms, respectively. The waveform of the ErrP response is similar to that observed in [Ferrez and Millán, 2008a, Zeyl et al., 2015, Margaux et al., 2012]. It seems that our experiments elicited an interaction ErrP, however the peaks appear later. It has been shown that the amplitude and latency of the ErrP are task-dependent [Iturrate et al., 2014]. The waveform of interaction ErrP is similar to the typical response ErrP, whereas the timing is similar to the feedback ErrP and to the

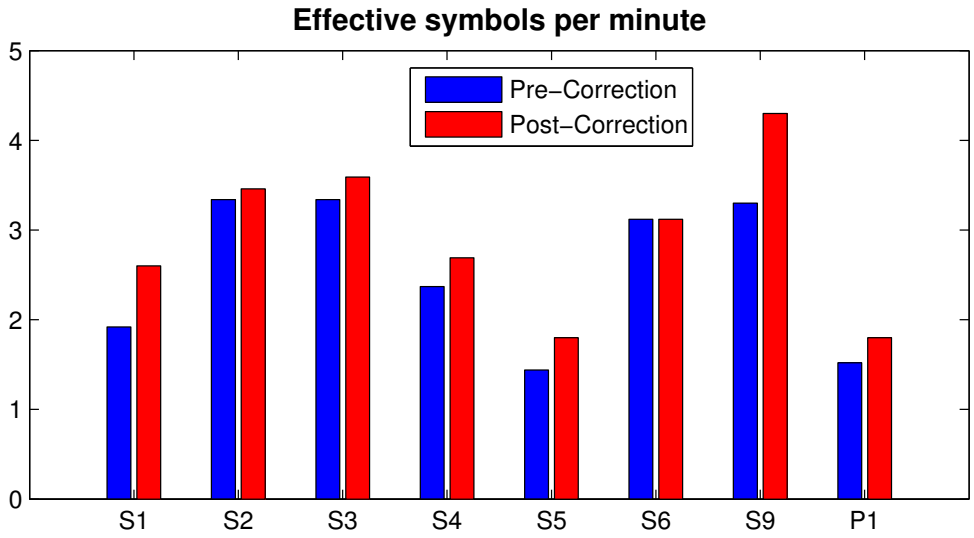


Figure 4.7: Effective symbols per minute for each subject in the two conditions: without error correction (blue) and with error correction (red).

observation ErrP, which are characterized by a negative deflection between 200 and 300 ms after the feedback/error. The ERP after correct feedback differs from the ErrP in latency, shape, and amplitude. It has only one positive peak that appears about 350 ms followed by a negative peak about 560 ms after the feedback. The correct ERPs were also reported in [Margaux et al., 2012, Zeyl et al., 2015]. However, in both studies, the waveforms between ErrP and correct ERP were very similar, except for a time lag between ERPs (larger in [Margaux et al., 2012]). So, our results differ from previous studies considering that distinct waveforms were evoked for wrong and correct feedback. The 1^{st} and 2^{nd} ErrP waveforms slightly differ. There is a latency difference mainly at the first positive peak and second negative peak (a difference of 70 ms and 100 ms respectively). The amplitude of the positive peak of the 1^{st} ErrP is slightly smaller than the 2^{nd} ErrP. On the other hand, the 1^{st} ErrP has a stronger first negative peak amplitude and slighter second negative peak amplitude.

The 2^{nd} correct ERP has a greater amplitude than the 1^{st} correct response. The differences between the 1^{st} and 2^{nd} correct ERPs may suggest that the significance of the feedback varies. For example, when the correct letter appears, it is an expected outcome, however, when the feedback is a correction of a wrong letter, it may represent greater relevance to the user since the system is correcting an error. The statistical R-square between ErrP and correct ERP identifies two clear discriminative components around 260-350 ms and 460-550 ms.

The waveform of the standards, true P300 targets, and false selected targets are also analyzed. Their grand-averages at channels Pz and PO7 are plotted in Fig. 4.9. The results show that the waveform associated with the wrong targets is different from that of the non-target events, approaching the P300 morphology, but still distinct from this one, as hypothesized from Fig. 4.5. This finding confirms our

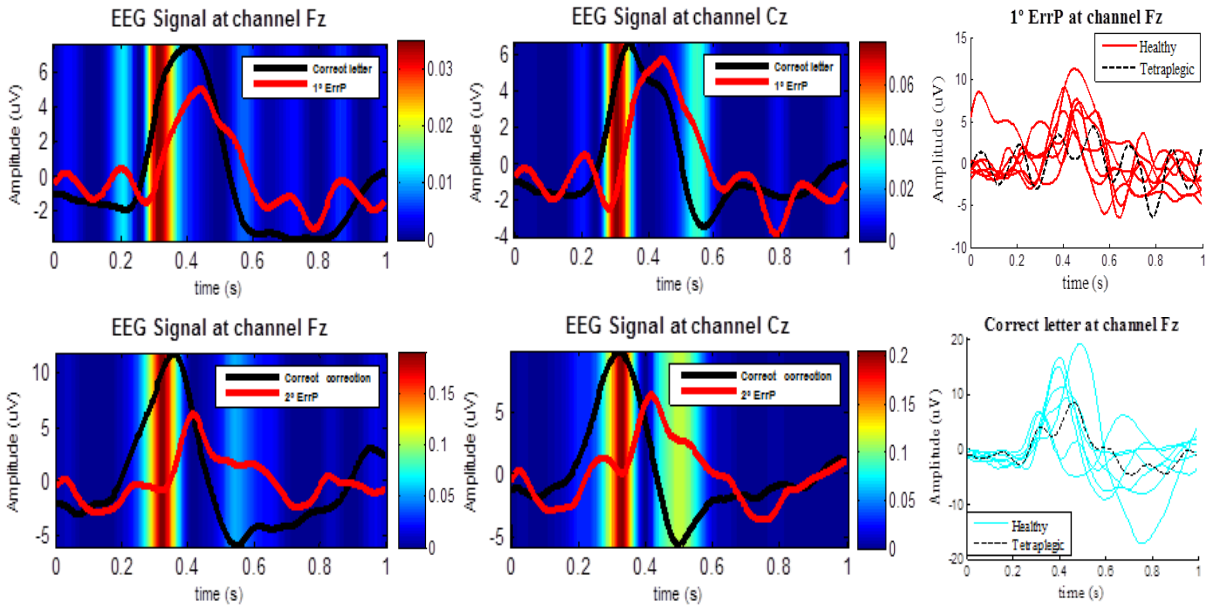


Figure 4.8: Grand average of first (top row: left and middle columns) and second (bottom row: left and middle columns) responses for error and correct feedback for the channels Fz and Cz. The background of these plots are the R-square values between correct and incorrect ERPs. The R-square identifies two discriminative components around 260-350 ms and 460-550 ms. Right column (top and bottom row): 1º ErrP and ERP after correct trials from Fz electrode for each healthy participant (solid line) and the tetraplegic participant (dashed line).

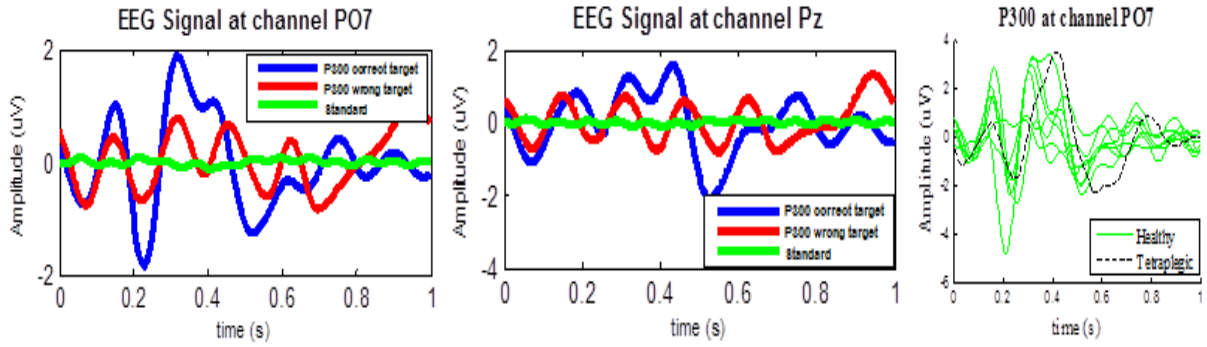


Figure 4.9: Grand average of EEG signals for electrodes PO7 (left column) and Pz (middle column) for three situations: correct P300 target, ERP incorrectly detected as target and non-target ERPs. Right column: P300 ERP from PO7 electrode for each healthy participant (solid line) and the tetraplegic participant (dashed line).

expectations and rationale to combine these features to build the error classifier. For comparison of individual waveforms, the last column of Fig. 4.4 and Fig. 4.9 shows respectively for each participant the waveforms of the ERPs associated with error and correct feedbacks at channel Fz and the waveforms of the P300 ERP at channel PO7.

4.2.8 Discussion

4.2.8.A Significance of the Online Results

The approach shows that is possible to use ErrPs for human-computer interaction, allowing the user to check or correct decisions of the system. Therefore, the proposed framework can be useful in many BCI applications beyond BCI spellers, in particular in human-machine collaborative systems, such as robotics, driving assistance, or emergency situations. The approach presented here extends our previous LSC paradigm [Pires et al., 2012] with this additional level of interaction, which can be used in different ways, namely as an ECS to increase the reliability of the system (the purpose of the current study); in the selection of samples to adapt the calibration model during online operation; or even as a secondary communication channel. To accommodate the double ErrP detection, the only change that was made in the LSC paradigm [Pires et al., 2012] was extending the ITI from 2.5 to 4 seconds, which slightly decreased the raw SPM. To our knowledge, this is the first study using a second ErrP to improve error correction. The online experiments with this approach showed an improvement of 5 percentage points, attaining 89.9%, which is considerable attending that the initial classification accuracy was already high, around 85% (typically, error correction systems have higher improvements in subjects with low BCI performance [Spüler et al., 2012]). The online error detector used the concatenation of the target ERP with the feedback response, an approach whose offline results showed a 7% improvement compared to the error detector using only the feedback response.

Table 4.4 compares the achieved online results with previous studies. We started with the highest initial spelling accuracy of all studies and yet there was an improvement for all subjects, except S6 who maintained the initial score. [Dal Seno et al., 2010] obtained no improvement, [Margaux et al., 2012] had a very small gain (only 0.5%). [Zeyl et al., 2015] had the highest online accuracy and the greatest gain in spelling accuracy, but the initial classification accuracy was much lower than in our study. We obtained the highest effective symbols per minute (2.92) and the greatest information transfer rate (14.19 bpm).

The tetraplegic participant (P1) had results similar to able-bodied subjects, and had an improvement of around 5% on spelling performance. Thus the system presented here might be a viable alternative for this participant and for other BCI target users.

As verified from the online results, the overall improvement of the double error approach relies on many factors, namely the rate of correction P_{cor} and the accuracy of error detection. The specificity of the 1^{st} error detection has an important impact as can be inferred from (4.4). Despite an average error detection of 88.4%, the levels of specificity varied across subjects. Those with lower specificity values had a performance degradation after the first feedback correction. The second feedback, allowed to overcome this degradation, by re-correcting the false positives, thus contributing to an overall improvement of 5.1% (Table 4.2).

Table 4.4: Online results obtained from previous BCI speller system using error detection and our proposed system.

Author	N^a	△ BECS	Dacc	Acc^b	BECS	ITR	eSPM	Pcor
Dal Seno et al., 2010	3	None	-	75.0^d	-	7.64^d	2.0^d	-
[Spüler et al., 2012]	23	0.52	-	-	2.34	-	-	-
[Margaux et al., 2012]	16	0.54d	0.5	62.5	1.37^d	-	-	34
[Zeyl et al., 2015]	11	-	13.7	92.6	-	8.79^d	2.54^d	-
This study	8^c	**0.29**	**5.1**	**89.9**	**3.6**	**14.19**	**2.92**	**44.2**

[a] N = number of participants.
[b] Acc is the final accuracy denoted as p in eq. (3.28) and (3.34) to compute respectively ITR and eSPM.
[-] not reported and unable to compute with provided data.
[c] 10 participants enrolled the study, but only 8 performed session 2.
[d] value computed from data provided in paper.

4.2.8.B ErrP Morphology

Both correct and wrong feedbacks evoked ERPs, in agreement with previous studies [Zeyl et al., 2015, Margaux et al., 2012], but here we have obtained a greater waveform distinction between them. The R-square analysis showed a strong discrimination (r^2 = 0.08 and r^2 = 0.26 for the first and second ErrP, respectively) between wrong and correct ERPs, which is of great importance for single trial classification. The 1^{st} and 2^{nd} feedback ErrP responses exhibited slight differences in terms of latency and amplitude. The amplitude of the positive peak of the 2^{nd} correct ERP is greater than that of the 1^{st}, which may be explained by the higher expectation of the user, due to the importance of correcting a wrong letter. The amplitude and latency of ErrPs have been related to user's motivation and workload [Hajcak et al., 2005, Iturrate et al., 2014], and it is also known that the amount of attentional resources have an effect in the morphology of ERPs [Polich, 2007].

4.2.8.C Limitations and Further Improvements

The implementation of a P300 BCI with error detection requires the calibration of the P300 and ErrP classifiers. Although the P300 calibration was quite fast (around 5 minutes), the ErrP calibration was longer so that a sufficient number of error-trials could be acquired under conditions similar to those expected during online operation, i.e., 10 to 20% error-rate and a long operation time. Performing a calibration for each session is something to avoid, since it limits the usability of the BCI system [Lotte, 2015, Wu, 2016]. Therefore we have validated the system using the calibration of the first session. The classifiers were trained in one day and the validation session was conducted on a different day, relying that the P300 and ErrP features remained almost the same between experiments. Although the classifiers presented a good generalization to ERPs variability across the two sessions, a small performance

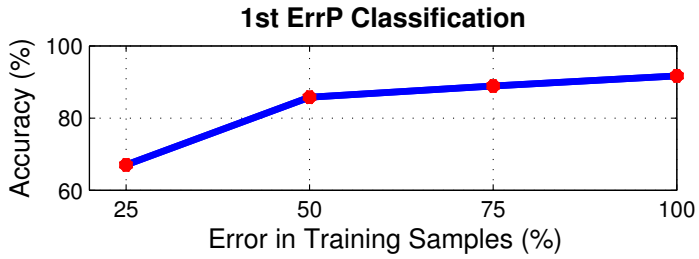


Figure 4.10: Classification of 1^{st} ErrP using datasets of calibration 2 with different percentages of error-trials (25, 50, 75 and 100%).

decay occurred, which was expected given that the classification is made at a single trial level (case of ErrP) and for a few number of repetitions (case of P300). The generalization of calibration models across sessions has a major impact on classification performance, particularly relevant when single trial classification is required, and it plays an important role in real-world applications to avoid recalibration procedures and performance loss [Iwane et al., 2016, Blankertz et al., 2011]. Doing the calibration only once mitigates the impact of a long calibration session. To analyze the effect of the amount of error-trials (directly associated to the calibration time), used for training the calibration model, on the classification performance, we split the calibration set in four subsets of consecutive samples, including respectively 25, 50, 75, and 100% of all error-samples, and we computed offline the classification accuracy using cross-validation. The results depicted in Fig. 4.10 show that from 50% of the total error-samples, the ErrP classification accuracy approximates the maximum. Therefore, it seems possible to combine several strategies to reduce calibration time. We explored another calibration approach that decreases the number of repetitions of target events to obtain a larger amount of error-trials in less time, the results are reported in the next section.

To further improve the proposed approach we need also to improve the correction rate (P_{cor}). Other methods beyond the use of the classifier's second best guess can be researched to determine the right target.

4.3 Generalization analysis of ErrP-calibration for Different Error-rates (case-study II)

As we verified in Section 4.2.3, the automatic recognition of error-related potentials requires a long calibration time in order to have enough error-samples to train the classifier. In this Section, we analyze whether it is possible to reduce the ErrP-calibration time in a P300-based brain-computer interface, by calibrating the BCI with a high rate of errors (wrong detections of user intent) [Cruz et al., 2018b]. We analyze if a high error-rate condition still produces a discriminable ErrP and if its classification model

generalizes well in sessions of different error-rates.

4.3.1 Introduction

The great variability of single-trial ErrPs poses difficulties in the generalization of the classification model across sessions. In [Iwane et al., 2016], authors evaluate the capability of the ErrP classifier to generalize across two different recording dates and across different inter-stimulus intervals (pace at which the stimuli are provided to the participants). The results showed that classification performance decrease in both conditions. Although the average waveform of the ErrPs is reported to be stable between sessions of different days, there is still a significant decrease in the classification rate between sessions, which can reach over 10% [Iwane et al., 2016, Cruz et al., 2018a]. In case-study I, we made a calibration with an error rate similar to that expected during online operation, i.e., with an error rate between 10 and 15%, and then we used this calibration model in a session of a different day. This calibration was quite long (about 2 hours) so that a significant number of error-trials could be gathered to train the classifier. The results of the validation test conducted some months after the calibration, showed a decrease of the classification accuracy that was more than 10% for some participants, leading to a significant impact on the online performance of the BCI. Other studies reported similar calibration approaches using real or sham errors [Spüler et al., 2012, Schmidt et al., 2012]. Such a long calibration session is suitable if used only once, but is impractical if used every time the user uses the system. On the other hand, the number of error-samples collected during a short calibration may be insufficient to train the classifier. One approach is to increase the error-rate during calibration to gather more error-trials in less time reducing the calibration time, however, it remains uncertain if a calibration obtained with a high error-rate generalizes well for a situation of an error-rate substantially lower.

In this study, we analyse whether the ErrPs elicited by a P300-speller with a high error-rate condition (around 40%) are still discriminable and classifiable, and whether a classification model obtained with this condition could generalize to a scenario in which it is expected an error-rate substantially lower (around 10-15%). This calibration with high error-rate would allow a greater amount of error-trials to be gathered in much less time.

4.3.2 P300-ErrP BCI

In this study, we used the same P300-ErrP BCI system as case-study I (section 4.2), that is, the integration of double ErrP detection into the lateral single character speller. We also used the same framework, namely preprocessing, features extraction, and classification methods (see a simplified representation of the classification pipeline in Fig. 4.1).

4.3.3 Participants

The experiments were carried out by five healthy participants (3 males, 2 females with age ranging 25-33 years old). All participants signed an informed consent to participate in the study. These participants took part also in the case-study I. Since there is a direct comparison of the results of that study with the present study, we kept the same identification of the participants, namely, S1, S2, S3, S5, and S6. During the experiments, participants sat in front of a computer screen at a distance of around 70 cm. They were asked to focus on the target stimulus and ignore the remaining non-target events, while mentally counting the times that the target flashes, helping to increase their attention level. Participants were instructed to be aware of the detected symbol to realize if it was the desired symbol. They were also informed that an automatic correction or re-correction could occur.

4.3.4 Calibrations and Online Session

The integration of error detection in the P300 speller requires two calibrations (calibration of the P300 classifier and calibration of the ErrP classifier). Therefore, the experiment consisted of three phases: P300 calibration (1^{st} phase), error-detector calibration (2^{nd} phase), and online session with error detection and correction (3rd phase). In this study, all three phases are performed on the same day, and in case-study I the first and second phases were held on the same day and the third phase was performed on a different day.

4.3.4.A P300-classifier calibration

In this calibration we gathered the EEG data associated with target and non-target events to train the models of the P300 classifier, which was used as the primary communication channel. The P300 calibration phase took about 5 minutes collecting 90 target epochs and 2430 non-target epochs.

4.3.4.B Error-detector calibration

Calibration1 - In this calibration procedure, participants spelled two 32-letter sentences (64 characters) without correction. It served to gather a dataset associated with positive and negative feedback responses. In order to obtain a high rate of wrong detections, the P300-LSC speller was set-up with a small number of stimuli rounds, namely $N_{rep} = 2$. The calibration took about 10 minutes, which included on average 26.4 wrong samples with a mean error-rate of 41.3%. The number of errors ranged between 21 and 40.

Calibration2 - Refers to the dataset of the error-detector calibration obtained in case-study I that will be used for comparison (see Table 4.5). The calibration lasted about two hours with a number of rounds averaging 4.2, yielding an error-rate of 18.7% (the number of errors ranged between 31 and 75).

Table 4.5: ErrP training and validation datasets used in the present study (Calibration1 and Test1) and in case-study I (Calibration2 and Test2).

	$Calibration1$	$Test1$	$Calibration2$ (case-study I)	$Test2$ (case-study I)
Error-rate	$\sim$ 40%	$\sim$ 15%	$\sim$ 20%	$\sim$ 15%
Duration	10 min	1 hour	2 hours	1 hour
N_{rep}(P300)	2	4.6	4.2	6

4.3.4.C Online Session

Test1 - This session used the classification models obtained in the calibration of the P300-classifier and in $Calibration1$. Participants were asked to spell the Portuguese sentence "ESTOU-A-ESCREVER-COM-UMA-INTERFACE-BCI" several times during approximately one hour. At the end of each sentence, the participants rested two minutes. The time that each participant took to write the sentence depended on the number of event repetitions. N_{rep} was settled individually for each user to reach $\approx 90\%$ accuracy (based on the outcome of the calibration of the P300-classifier). The BCI system performed automatic double-error detection and correction as described previously.

Test2 - Refers to the online session of case-study I, using the exact same conditions of $Test1$. The classification model was built from $Calibration2$.

4.3.5 Results and Discussion

4.3.5.A Evoked Potentials Morphology

Fig. 4.11 compares the grand averages of potentials elicited by wrong and correct feedback in channels Fz and Cz for 40% error-rate ($Calibration1$) and for 15% error-rate ($Test1$). The ErrP of $Calibration1$ has different waveforms of the ErrP of $Test1$ (Pearson's correlation coefficient $r = 0.64$ for both channels). The higher error-rate condition produced an ErrP with lower amplitude. This is consistent with other studies [Chavarriaga and Millán, 2010], which also reported a decrease of amplitude for higher error-rates, but here we also found changes in ErrP morphology, namely, the second positive peak of the ErrP elicited by the 40% error-rate condition had a lower latency (less than 80 and 90 ms for channels Fz and Cz, respectively). This resulted in an ErrP waveform very similar to the correct ERP signal (Pearson's correlation coefficient $r = 0.92$ and $r = 0.87$ for channel Fz and Cz respectively), which makes uncertain the possibility of a successful classification. This result may suggest an habituation effect that might reduce the impact of errors. On the other hand, for the lower error-rate condition the waveforms are clearly discriminable. There is also a consistency of the characteristics of the grand average of the 1^{st} ErrP and correct ERP in regard to those exhibited in $Test2$ (performed 22 months before). This stability across sessions was also reported in other studies [Chavarriaga and Millán, 2010, Iwane et al., 2016].

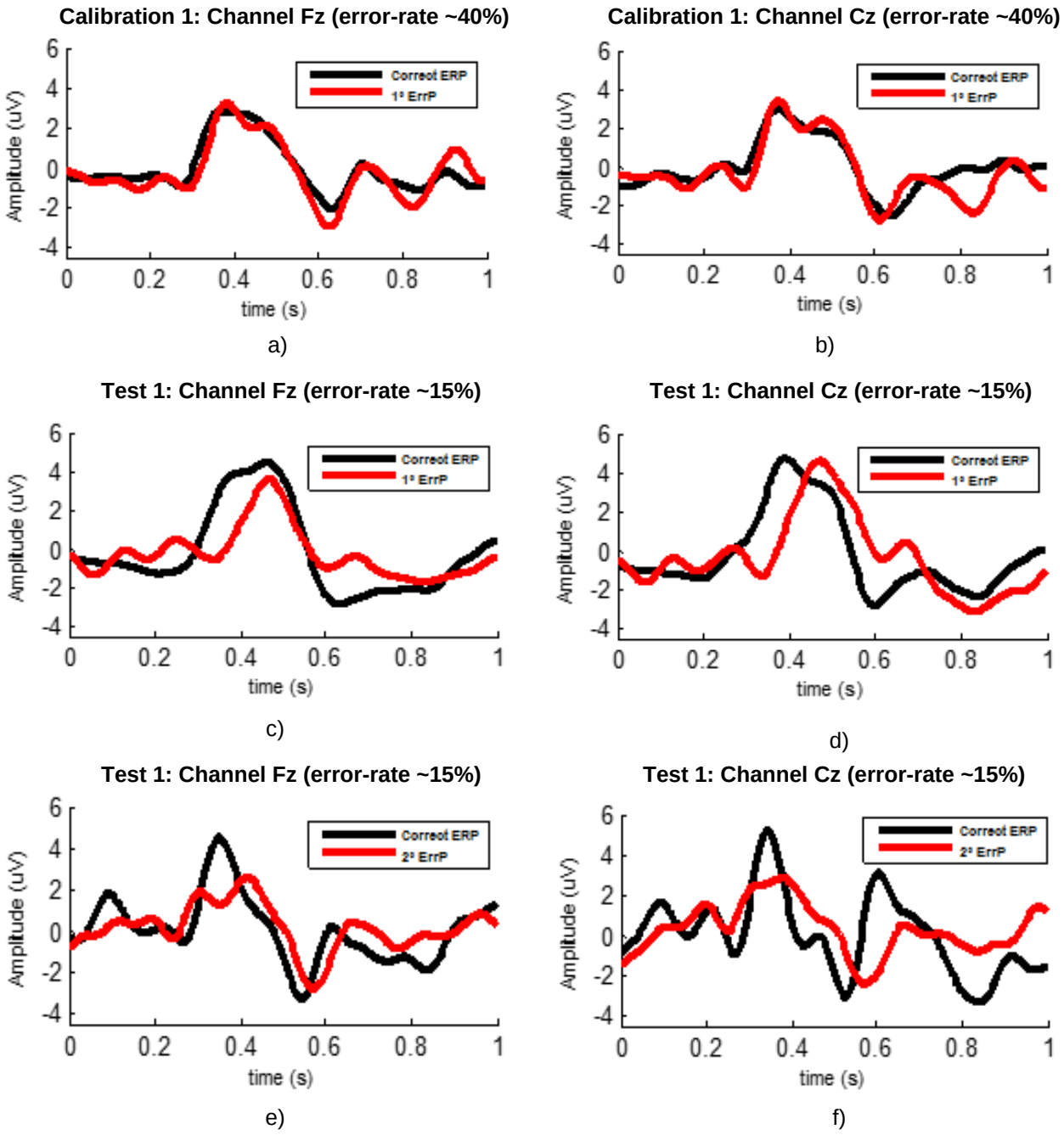


Figure 4.11: Grand average of evoked potentials after positive and negative feedback. $a)$ and $b)$ ERPs produced during the calibration of the ErrP detector with an error-rate of 40% for the channels Fz and Cz respectively. 1^{st} ErrP (c and d) and 2^{nd} ErrP (e and f) and ERP after correct feedback during the online session with error-rate of $\sim$ 15% for the channels Fz and Cz respectively.

Nevertheless, this waveform similarity did not avoid a significant performance decay between calibration and test sessions, as reported in [Iwane et al., 2016, Cruz et al., 2018a]. The 2^{nd} ErrP had a morphology different of the one obtained in case-study I. We hypothesize that this may be related to the different

Table 4.6: Offline results obtained from ErrP-calibration in $Calibration1$ ($\sim$ 40% error-rate) and in $Calibration2$ ($\sim$ 20% error-rate). Balanced accuracy was obtained from 10-fold cross validation.

	ErrP Calibration1 (error rate of $\sim$ 40%)			ErrP Calibration2 (error rate of $\sim$ 20%)			
Subjects	Error-rate (%)	Number of errors	Acc-ErrP1 (%)	Error-rate (%)	Number of errors	Acc-ErrP1 (%)	Acc-ErrP1 (%) (using half of error-samples)
S1	39.1	25.0	82.9	28.2	75.0	89.5	87.2
S2	32.8	21.0	72.8	14.5	55.0	93.9	86.6
S3	34.4	22.0	100.0	14.5	66.0	96.6	90.4
S5	40.6	26.0	81.9	24.8	66.0	87.4	70.8
S6	59.4	38.0	78.6	11.7	31.0	85.6	77.9
Mean	41.3	26.4	83.2	18.7	58.6	90.6	82.6

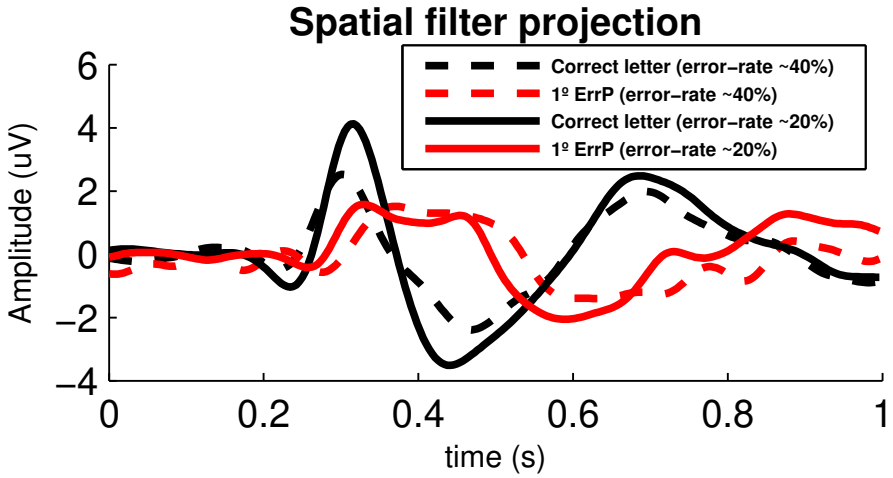


Figure 4.12: Grand average of the 1^{st} projection of the FCB spatial filter, using the EEG data of $Calibration1$ (dashed line) and $Test1$ (solid line).

information we provided to the participants. While in case-study I it was not explained what the double-error correction algorithm was doing, in the present study the participants were explained that if the corrected letter elicited a 2^{nd} ErrP, the first detected letter would be chosen as target letter. Therefore, letters corrected wrongly, owing to false positives, may have been less important to participants because they expected the correct letters would be re-selected.

4.3.5.B Offline Classification Accuracy

To evaluate the offline and online performance of the P300-ErrP BCI (Fig. 4.1), we computed the sensitivity, specificity, accuracy, and balanced accuracy metrics. Table 4.6 shows for each participant the error-rate, the total number of errors and the balanced accuracy of the 1^{st} error-detector obtained in $Calibration1$. For a direct comparison, we show the results for the same five subjects who participated in case-study I ($Calibration2$). For $Calibration1$, N_{rep} was 2 on average, which yielded an average

error-rate of 41.3%. However, participant S6 had a significantly higher error-rate (around 60%), i.e., more wrong samples than correct samples. The mean value of the balanced accuracy of the ErrP detector (obtained using 10-fold cross-validation) was 83.2%. Despite the high similarity between positive and negative feedback for the high error-rate condition, the accuracy of the classification was still high. Plotting the projection of the FCB spatial filter (Fig. 4.12) we observe that the spatial filter was able to maximize the discrimination of the features of the two classes, approaching the projection obtained with the lower error-rate condition, which explains the good accuracy. For $Calibration2$ the balanced accuracy of the ErrP detector was 90.6%. The average result of the ErrP-detector in $Calibration1$ was 7.4% lower, but the number of error-samples used for training was less than half of that of $Calibration2$, which influenced the classifier training. To verify whether this difference in classification accuracy is related to the high error-rate condition or it is due to the lack of error samples, we computed the results that would be achieved in $Calibration2$ if only half of the calibration dataset would have been used, which yielded an accuracy of 82.6%. Thus, for a similar number of errors, the two conditions have similar performance (around 83%). This suggests that the ErrP classifier is not strongly affected by the error-rate, but mainly by the number of training samples. Additionally, the FCB spatial filter seems to have an important contribution to confer robustness to changes of the ErrP waveforms between the two error-rate conditions. Authors in [Iwane et al., 2016] also highlighted the influence of spatial filtering in the classification accuracy across conditions.

To further analyze whether there is a direct correlation between ErrP accuracy, error-rate and number of error-samples we made a regression analysis, using data from $Calibration1$ and $Calibration2$ (Fig. 4.13). Joining the datasets of $Calibration1$ and $Calibration2$, we observe a linear relationship between ErrP accuracy and the number of error-samples (regression coefficient is 0.13), however we stress that this result is not statistically significant (t-test, $p = 0.22$). For the regression between ErrP accuracy and error-rate we analyzed the datasets separately. There is a negative relationship (coefficients of -0.24 and -0.27 respectively for $Calibration1$ and $Calibration2$), but the result is not statistically significant (t-test, $p = 0.68$ and $p = 0.19$ respectively). A two-way ANOVA was also performed to evaluate the combined effect of error-rate and number of error-samples in the classification performance. The p-value of 0.28 ($F = 1.32$) showed that the combination of these variables is not statistically significant. Despite the inconclusive analysis, the results point to the need for more error-samples for training, which could be obtained by increasing the duration of the calibration or pushing the error-rate to 50%, achieving a balanced number of samples with correct and wrong responses.

4.3.5.C Online Classification Accuracy

Table 4.7 presents the online BCI results obtained in $Test1$ using the classification model from $Calibration1$. The online average classification accuracy of the 1^{st} ErrP was 86.8% (error-rate of 14.6%), i.e., 3.6%

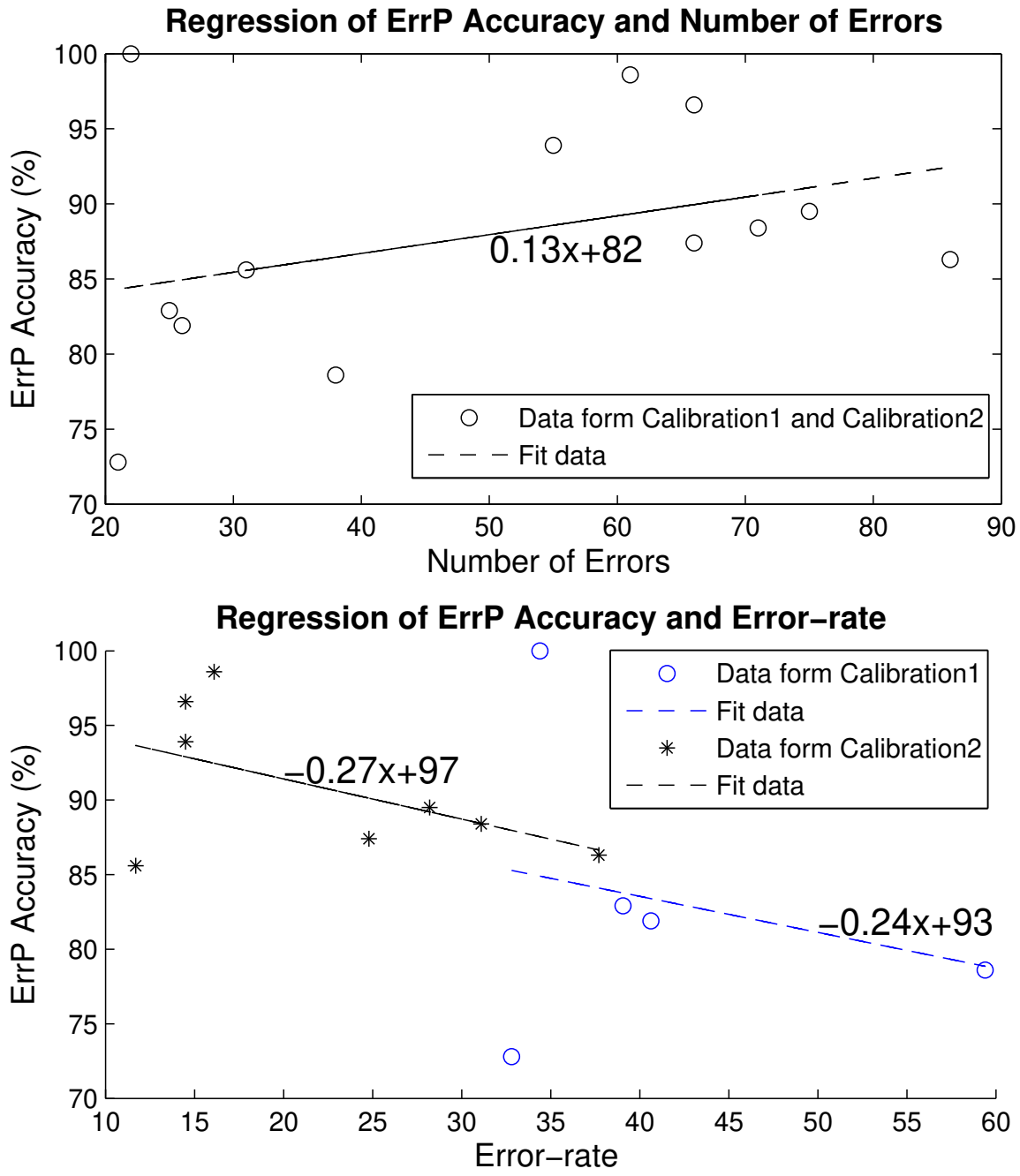


Figure 4.13: Linear regression between offline ErrP classification accuracy and number of errors (top) and error-rate (bottom), using datasets of error-detector training collected in the present and case-study I.

higher than the one obtained from cross-validation in $Calibration1$ (error-rate of 41.3%). This result showed that the calibration model obtained for the higher error-rate generalized well for the lower error-rate condition, the main hypothesis that we wanted to verify. Comparing the average accuracy achieved in $Test2$ (89.0%) there was only a small performance loss (2.2%).

Table 4.7: Online classification results in $Test1$ using the classification model from Calibration1 and accuracy of 1^{st} ErrP obtained in $Test2$ (case-study I).

	Pre-Acc ($Test1$)	Post-Acc ($Test1$)	Acc-ErrP1 ($Test1$)	Acc-ErrP2 ($Test1$)	Acc-ErrP1 ($Test2$)	N_{rep}
S1	79.6	90.5	91.4	75.0	90.8	5.0
S2	87.4	90.5	93.2	100.0	96.3	4.0
S3	90.5	94.7	99.5	89.5	89.5	3.0
S5	82.2	77.0	74.3	52.3	92.1	6.0
S6	87.4	90.5	75.7	85.7	76.3	5.0
Mean	85.4	88.7	86.8	80.5	89.0	4.6

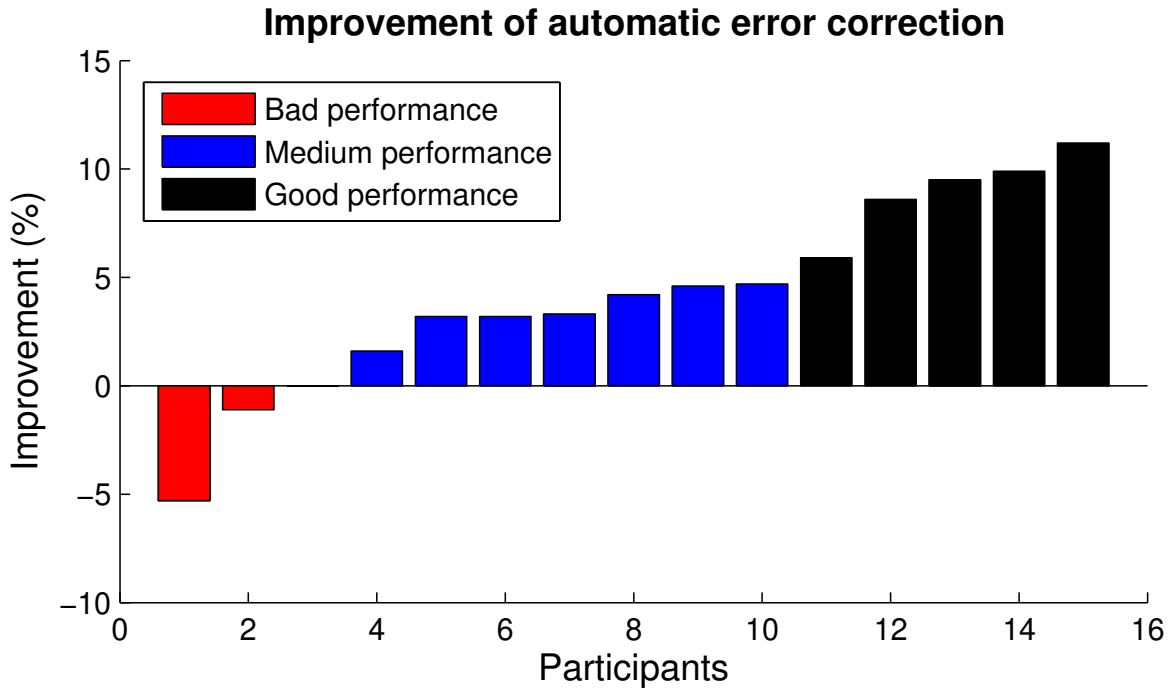


Figure 4.14: Improvement after automatic error correction for each participant using datasets of case-study I and case-study II.

Regarding the 2^{nd} ErrP (ErrP evoked when the automatic correction is wrong), the ErrP classification dropped to 80.5%. The classification of the 2^{nd} ErrP uses the model trained with the responses of the 1^{st} ErrP. This shows that the classification model of the 1^{st} ErrP does not generalize well to the 2^{nd} ErrP. This was also identified in case-study I, which consistently shows that the characteristics of the 1^{st} and 2^{nd} ErrPs are different.

4.3.5.D Analysis of Automatic Error Correction

The classification accuracy before the automatic error correction was 85.4%, and increased 3.2% after automatic error correction. However, the enhancement is not statistically significant (paired t-test, p =

0.14). Participant S1 had the greatest improvement, about 10%. On the other hand, participant S5 had a worse classification accuracy after correction. The 2^{nd} ErrP is effective in reducing the number of false positives by correcting the correct targets that are detected as errors in the 1^{st} error-detector, however when the classification accuracy of the 2^{nd} ErrP-detector is poor, the performance decreases (which was the case of this participant).

To analyze the relationship between the number of errors obtained in the calibration of the error-detector and the online improvement after automatic error correction, we grouped the data in three categories (Fig. 4.14): poor performance (the final classification accuracy is equal or lower than initial accuracy), medium performance (improvement less than 5%) and good performance (improvement greater than 5%). Averaging the number of errors got in $Calibration1$ and $Calibration2$ for participants of each category, the mean number of errors for poor, medium and high performance was 28.5, 51.3 and 56.8 respectively. This result may suggest that a calibration phase with at least 50 error-samples could always produce an improvement after error correction. However, it should be noted that the improvement also depends on the correction rate, i.e., the ability of replacing a wrong symbol by the correct one.

4.4 Conclusion

In this chapter, a new error detection and correction system based on double ErrP detection was proposed. Promising results were achieved with a significant increase of online accuracy, information transfer rate, and effective SPM validated with able-bodied participants and a tetraplegic participant. Therefore, the integration of error detection into the BCI system might provide a preponderant solution significantly improving BCI reliability to enable its use in clinical settings. Whit the approach used in case-study I, the ErrP calibration lasted about two hours which is too long. In case-study II we evaluated the generalization of the ErrP classifier over different error-rates. The online ErrP accuracy (trained with high error-rate and tested with low error-rate) was higher than the offline ErrP accuracy (trained and tested with high error-rate). This result is indicative that a classification model built from a calibration with high error-rate generalizes to conditions with lower error-rates. Nevertheless, the classification is still affected by the low number of samples gathered during training.

5

Towards User-friendly Brain-controlled Wheelchair (BCW)

Contents

Brain-controlled wheelchairs are a promising solution for people with severe motor disabilities who cannot use conventional interfaces. However, the low reliability of electroencephalographic signal decoding and the high user's workload imposed by continuous control of a wheelchair requires effective approaches. In this chapter, we propose a self-paced P300-based brain-computer interface combined with dynamic time-window commands and a collaborative-controller [Cruz et al., 2021]. The self-paced approach allows users to switch between control and non-control states without requiring any additional task or mental strategy, while the dynamic time-window commands allow balancing the reliability and speed of the BCI. The collaborative controller, combining user's intentions and navigation information, offers the possibility to navigate in complex environments and to improve the overall system reliability. The feasibility of the proposed approach and the impact of each system component (self-paced, dynamic time-window, and collaborative controller) were systematically validated in a set of experiments conducted with seven able-bodied participants and six physically disabled participants steering a robotic wheelchair in real office-like environments (Case-study III). In this chapter, we also investigate whether detection of user's mental state (e.g. stress) using GSR can be used to adapt a human-machine collaborative controller (Case-study IV).

5.1 A self-paced BCI with a collaborative controller for highly reliable wheelchair driving (case-study III)

5.1.1 Introduction

People suffering from conditions that affect neuromuscular structures and functions tend to lose a significant degree of autonomy in daily living activities. Powered wheelchairs may help them to increase their levels of mobility and quality-of-life [Davies et al., 2003]. However, many of them become unable to use conventional interfaces, as a result of impairment severity or physical ability deterioration [Frank et al., 2000]. For those with severe motor impairments, brain-computer interfaces may be an alternative solution as it is possible to send commands through brain signals without requiring muscle activity [Wolpaw et al., 2002, Lopes et al., 2013, Carlson and Millan, 2013, Lopes et al., 2016]. Yet, using a BCI to control a robotic wheelchair is a very challenging task because BCI has low transfer rates and limited accuracy [Kübler et al., 2015]. Controlling a BCI system requires continuous and high levels of attention and focus, which imposes a high mental and physical workload that limits its usability. In turn, this workload can cause attention shifts and fatigue, resulting in even greater uncertainty in decoded brain commands. In the context of brain-controlled wheelchairs steered in real-world scenarios, the low reliability and rate of BCI commands can lead to disastrous safety consequences for the user and the system [Bi et al., 2013]. For this reason, when compared to other BCI applications such as spellers or

games, BCWs require much higher reliability and general usability, which is only possible if they integrate an assistive navigation system that perceives the wheelchair's surroundings and performs suitable and smooth trajectories, considering the user intents. This can be accomplished by combining user and machine outputs in a so-called collaborative controller [Carlson and Demiris, 2012, Carlson and Millan, 2013, Duan et al., 2014, Lopes et al., 2016, Li et al., 2016], allowing BCI commands, which encode high-level goals, to be provided at sparse intervals without the need for precise, low-level continuous steering (see Figure 5.1).

The aforementioned collaborative approach may not yet be sufficient for effective use of a BCI because the user still needs to provide BCI commands in regular time-windows and has to be continuously focused, which is a mentally demanding task [Iturrate et al., 2009, Diez et al., 2013, Duan et al., 2014]. Self-paced control (also known as asynchronous control) provides the possibility for users to send BCI commands only when they wish to, at their own pace. This is therefore a very desirable feature, which can lead to less mental effort and more natural driving interaction [Rebsamen et al., 2010, Carlson and Millan, 2013, Wang et al., 2014, Lopes et al., 2016, Zhang et al., 2015b, Yu et al., 2017, Yu et al., 2018]. To implement a self-paced BCI, the system must automatically recognize control and non-control states. In a state of non-control, it is understood that the user does not want to select any command. The main goal of this study is to research ways to increase both the reliability and usability of BCWs and the main contributions are: 1) proposal and validation of a new dynamic time-window approach for BCI commands based on the degree of the classifier's confidence, and its combination with the self-paced approach, that adjusts the BCI speed to the user's performance over time. To the best of our knowledge, very few works have used dynamic time-window methods in brain-actuated wheelchairs (and those used different approaches) and none have done so in a non-simulated environment [Kaufmann et al., 2014]. This automatic adjustment decreases the users' performance fluctuations that may arise from changes in the users' attention, thus maintaining the most stable reliability naturally; 2) validation of a self-paced approach that frees the user from being continuously focused on the BCI. This is achieved through a non-control state that does not involve any additional task for the user, as he/she only has to be relaxed in a state of inattention. At the same time, the approach also tunes the rate of false positives, which is different from other P300-based brain-actuated wheelchair approaches. The proposed self-paced detector contributes to a natural BCI operation increasing the usability of the system; 3) combination of three impactful features in a single framework: self-paced control, dynamic adjustment of time-window commands and collaborative control, aiming at high overall reliability. This led to an overall performance greater than 99% without decreasing the BCI speed, which shows the feasibility of the approach in this application but which can be extended to other different contexts; 4) validation of the BCW in a realistic office-like environment with severe physically disabled participants. This represents a contribution to the effective validation of BCI approaches, as most studies have validated their approaches only with

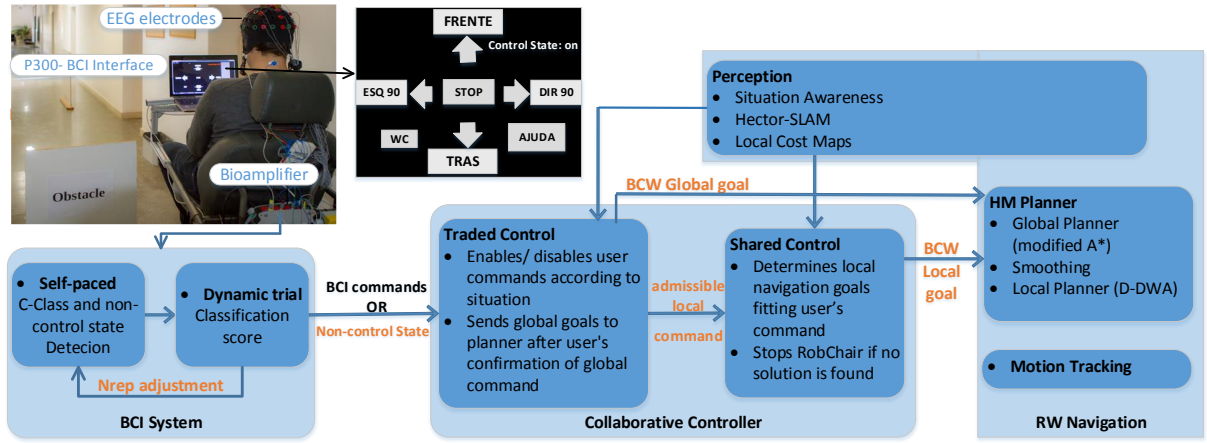


Figure 5.1: Schematic representation of the overall BCW system, which is composed of three main modules: BCI System, Collaborative Controller, and Robotic Wheelchair Navigation.

Table 5.1: Gender, age and BCI experience of able-bodied participants.

Subjects	S1	S2	S3	S4	S5	S6	S7
Age	32	25	22	22	21	21	23
Sex	F	F	M	M	M	M	M
BCI experience	YES	NO	NO	NO	NO	NO	NO

healthy participants.

5.1.2 Experimental Design

5.1.2.A Participants

This study comprises two groups of participants: 7 able-bodied users, referred to as Group I, and 6 participants with severe motor disabilities referred to as Group II (see Tables 5.1 and 5.2). The study was ethically assessed and approved by the board of the Cerebral Palsy Association of Coimbra (APCC) and was conducted complying with the code of Ethics of the Declaration of Helsinki. Informed consent was obtained from all participants, explaining the aims of the study, their role as participants (e.g., voluntary participation), and the ethical commitments of the research team (e.g., data anonymization, guarantee of confidentiality). The sample of able-bodied users (S1-S7) was composed of students and researchers with ages between 21 and 32 years old, with a mean age of 23.7 years. Only one participant had previous experience with P300-based BCI and none had experience in driving a wheelchair. Table 5.1 presents their ages, gender, and their previous BCI experience. Group II (P1-P6) included outpatients from APCC with ages between 21 and 50 years old, averaging 37.5 years old. Primary clinical diagnoses for the latter group included Cerebral palsy (CP) (2 cases), spinal cord injury (1 case), agenesis of the four members (1 case), limb-girdle muscular dystrophy (1 case), and Duchenne muscular dystrophy (1

case). Table 5.2 contains a more detailed description of the disabled participants, as well as a summary of their clinical and functional data. They were all dependent on human assistance in daily activities but still able to use powered wheelchairs independently with customized interfaces (even though some of them with great difficulty).

Table 5.2: Characterization of motor disabled participants: gender, age, BCI experience, and a summary of their clinical and functional data.

Subjects	Age	Gender	BCI experience	Diagnosis	Type of disability	Level of motor functionality	Interface used to steer powered wheelchair
P1	49	M	NO	CP (and sensory impairment)	congenital	Moderate autonomy in day-to-day activities. Upper and lower limbs with spasticity. Slight movements.	Joystick controlled by hand; Good efficiency.
P2	35	F	NO	Agenesis of the four members (and sensory impairment)	congenital	Moderate autonomy in day-to-day activities. Very small limbs; Slight movement disturbance.	Joystick controlled by stump; Good efficiency.
P3	50	M	NO	CP (and cognitive impairment)	congenital	Moderate autonomy in day-to-day activities. Upper and lower limbs with spasticity. Slight movements.	Joystick controlled by hand; Good efficiency.
P4	45	M	NO	SCI (and sensory impairment)	acquired	Head and upper limbs control(tetraparesis). Moderate autonomy in day-to-day activities.	Joystick controlled by hand; Good efficiency.
P5	25	M	NO	Limb-girdle muscular dystrophy	congenital	Low autonomy in day-to-day activities. Proximal muscle weakness.	Joystick controlled by hand; Good efficiency.
P6	21	M	NO	Duchenne muscular dystrophy	congenital	Low autonomy in day-to-day activities. Muscle weakness.	Joystick controlled by hand; Good efficiency.

5.1.2.B BCI Graphical User Interface and Commands

The BCI uses the visual paradigm described in section 3.1.2, which comprises 7 steering commands. These symbols flash randomly with a highlight time of 100 ms and an inter-stimulus interval of 75 ms, resulting in a stimuli onset asynchrony (SOA) of 175 ms. For every round of the oddball paradigm, each symbol is flashed once. An EEG data segment (epoch) with 256 time samples (one second) is extracted for each stimulus onset. Because of the low signal-to-noise ratio of P300 ERPs, several rounds are usually required to collect several target epochs, to improve the target classification.

For the conventional approach (here called a static trial) the number of repetitions/rounds necessary to select a target is fixed, settled according to the performance of users in the calibration phase. For the dynamic trial approach (dynamic-time commands), the number of repetitions per trial during the online operation varies according to the target classification score. The overall trial time (TT) needed for symbol classification is computed from

$$TT = N_{rep} \times N_s \times SOA + CT \tag{5.1}$$

where $\mathbf{N}_s = 7$ is the number of symbols, and $CT = 1$ is the time associated with the last flash of the trial. Inter-Trial Interval, i.e. the time between each set of rounds, was set at one second.

5.1.2.C Calibration sessions

Before starting the driving tasks, each participant performed a calibration session to obtain the classification models. Participants were seated in the RobChair with the computer screen positioned in front of them at a distance of approximately 30 cm, in the same conditions they have while driving the wheelchair. Participants were instructed to focus on the pre-defined target commands, successively provided at the top of the screen, and to mentally count whenever a target command flashes. Calibration consisted of a sequence of 9 symbols, and 9 rounds per symbol, collecting 81 target epochs and 486 non-target epochs, taking about 2 minutes. Participants performed only one calibration, from which all classification models and parameters were obtained to control the 3 performed tasks.

5.1.2.D RobChair system

The navigation module which includes perception, planning, and collaborative controller was developed by other elements of our team [Lopes et al., 2016]. RobChair is a robotic wheelchair (RW) with differential actuation, equipped with optical encoders coupled to each motorized wheel and an hokuyo UTM-30LX scan laser. Its navigation architecture is implemented in ROS and is composed of three main modules: perception, planning, and motion tracking. Currently, the perception module is composed of: situation awareness; Simultaneous Localization and Mapping (SLAM), performed with Hector-SLAM [Kohlbrecher et al., 2011], and multi-resolution local cost maps, as described in [Lopes et al., 2016]. The planning

module is based on the hybrid motion (HM) planner, which is composed of a global planner based on a modified version of the A* algorithm, a smoother and a Double-Dynamic Window (D-DWA) approach for local planning. The collaborative controller is a decision-making module composed of two layers: traded and shared controllers (see Fig. 5.1). It receives sparsely issued high-level commands from the P300-based BCI that can either be global or local commands. Global commands consist of a set of target goals belonging to the navigational space (e.g. WC). As soon as a user selects a global command through the P300-based BCI, the traded controller is in charge of validating the command after user confirmation, sending it directly to the global planner. On the other hand, local commands allow the RW to navigate between local goals previously defined on the topological map (also referred to as decision points). A decision point is defined as an ambiguous place in the map because it allows several directions to be taken from there (e.g. intersections or bifurcations). In these situations, the user provides a direction through the P300-based BCI (e.g., FORWARD, BACK, etc.), and the collaborative controller determines the closest local goal in the direction provided by the user. Although these commands might be understood as low-level, they are, in fact, high-level commands because they provide a way to choose a pre-defined local goal. If the user issues a local command the traded controller enables/disables that command depending on the information provided by the situation awareness module (e.g. if a user issues the LEFT command but it is only possible to move forward, the command is disabled). The shared controller is in charge of determining the appropriate local navigation goal according to the admissible local command provided by the traded controller, situation, place being navigated, and topological map information (i.e. predefined subgoals or decision points). More details on the collaborative control algorithm can be found in [Lopes et al., 2016]. RobChair was ergonomically adapted with help of APCC staff, to be used by severe motor impaired participants.

5.1.2.E Navigation scenarios

The experiments consisted of steering the RobChair in a real indoor office environment. Participants performed three navigation tasks as described below. The first task used the self-paced BCI with Static trial time (STT), the second one used the self-paced BCI with dynamic trial time, and the third one used a non self-paced approach. The order of the tasks was the same for all participants, Task1-Task2 for the physically disabled participants, and Task1-Task2-Task3 for the able-bodied group. Each participant performed the experiments on the same day. Before starting the task, each participant went through the designated path seated in the wheelchair, while an external operator was driving the wheelchair using a joystick, and the decision points were shown. Then, at the starting point of the route, after the calibration, each participant was enabled to become familiar with the interface, selecting commands, but with the wheelchair stopped. The familiarization time was variable between participants, ensuring that each one understood the task.

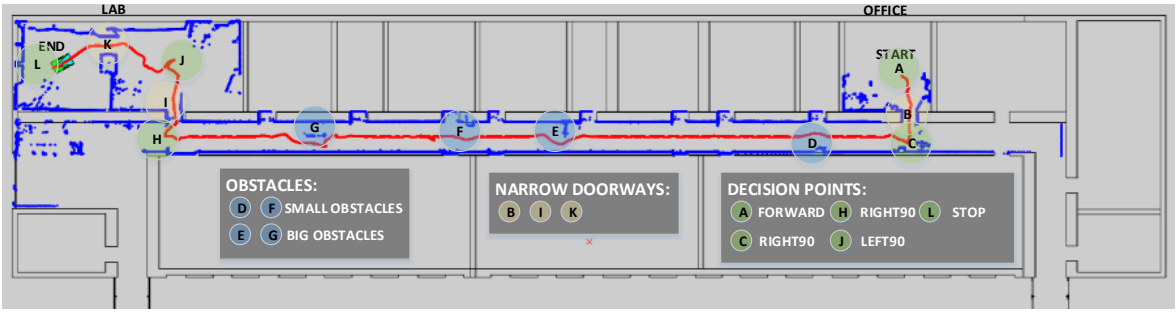


Figure 5.2: Map with scenarios in which the participants performed Task1 and Task3 with 4 obstacles (D, F, E and G), 3 narrow passages (B, I and K) and 5 decision points (A, C, H, J and L).

Task1 - Collaborative and self-paced control with STT. Users steered the RobChair following the map route depicted in Fig. 5.2, from an office, represented as START, to a lab, represented as END using the self-paced P300-based BCI with STT, whose implementation is explained in section 5.1.3.B. Before starting the tests, participants were instructed about the route, which included three narrow doorways (B, I, and K), two small obstacles (D and F), and two large obstacles (E and G). The minimum number of decisions to reach the final destination was 5, i.e. participants had to provide commands at each decision point (A, C, H, J, and L). However, the BCI is always outputting a command (target symbol or non-control state) at every trial. To detect the occurrence of false positives and false negatives, users were instructed to press an adapted switch whenever the BCI system selected an erroneous command, and someone was always behind for double checking.

Task2 - Collaborative and self-paced control with DTT: The navigation task consisted of moving from the lab signed by START to the hall near the ELEVATOR, as shown on the map in Fig. 5.3, using the self-paced P300-based BCI with the DTT approach. This route included a room, passage through two doors, and navigation in a corridor. To perform this navigation task, users were required to provide commands to start it and to choose the appropriate local goal in decision points (A, C, E, F, and H). However, the BCI is always outputting a command (target symbol or non-control state) at every trial.

Task3 - Collaborative and non self-paced control with STT: In this task, we have evaluated the non-self-paced control in which the user had to provide a target selection at every trial, as in this mode the BCI could not detect the non-control state. The route was the same as in $Task1$, which consisted of going from OFFICE to LAB. RobChair's speed was programmed to slow down at every decision point, that is, points in which the robot could not make a decision without an appropriate user command.

5.1.2.F EEG Data Acquisition and Preprocessing

The electroencephalographic and GSR signals were recorded with a 16-channel g.USBamp bioamplifier. The EEG signals were recorded with active Ag/AgCl electrodes at positions Fz, Cz, C3, C4, CPz, Pz,

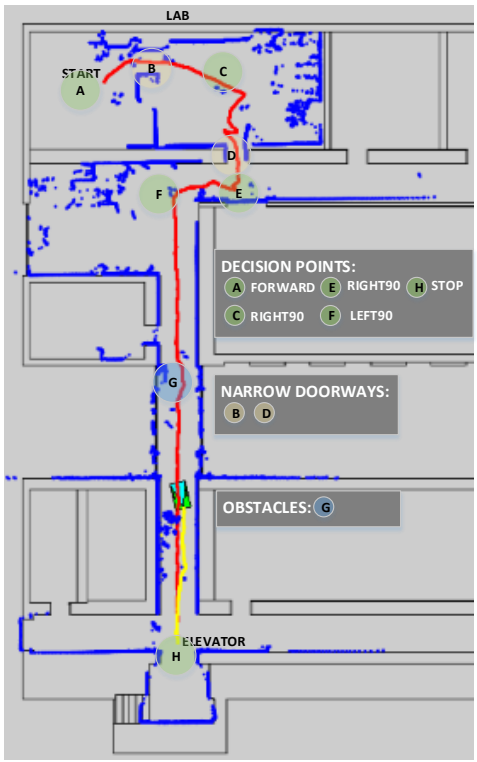


Figure 5.3: Map with scenarios in which the participants performed Task2 with one obstacle (G), 2 narrow passages (B and D), and 5 decision points (A, C, E, F, and H).

P3, P4, PO7, PO8, POz and Oz according to the international extended 10-20 standard system (see a photo of the system setup in Fig. 5.1). The electrodes were referenced to the right or left earlobe and the ground was the AFz electrode. The GSR signals were captured from the index and middle fingers of the non-dominant hand using the g.GSRsensor2. All signals were sampled at 256 Hz and filtered using a 50 Hz notch filter. EEG signals were filtered by a 0.5-30 Hz band-pass filter and GSR signals were filtered by a low-pass filter with 1Hz cutoff frequency.

5.1.3 Methods: Self-paced P300-based BCI and Dynamic Trial

5.1.3.A Online Classification Pipeline

The online classification pipeline of the self-paced P300 BCI system is schematically represented in Fig. 5.4. After preprocessing, the data is segmented into epochs of 1 second, and then the epochs are normalized to zero mean and unitary standard deviation. The number of repetitions per trial is selected from the calibration session. Then, the normalized epochs of the N_{rep} repetitions are averaged for each channel and a feature extractor is applied, namely, a statistical spatial filter (C-FMSB) that uses

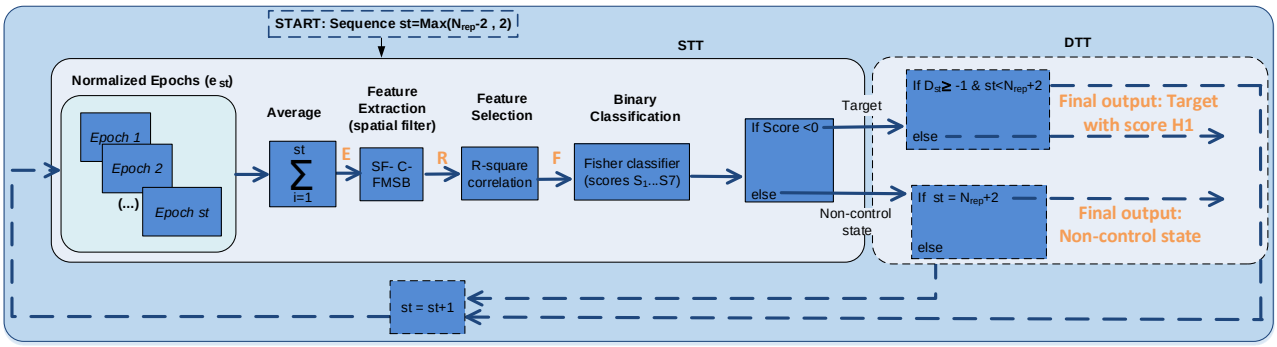


Figure 5.4: Schematic representation of the overall pipeline of the self-paced P300-based BCI. Symbols E, R, F, and H1 refer to the averaged epochs, extracted features, selected features, and symbols with the highest score, respectively. The main solid line block represents the self-paced approach with STT. The dashed line blocks complement the STT block to implement the DTT approach (according to Algorithm 1).

a suboptimum approach combining two criteria, the Fisher criterion and the SNR maximization (see details in [Pires et al., 2011a]). Considering the averaged epochs $E_{N_c \times L}$, where $N_c = 12$ is the number of electrodes and $L = 256$ is the number of samples, the spatial filter projection is obtained from

$$Z_{1:2} = V_{1:2}^T E \tag{5.2}$$

where $W_{1:2}$ are the 2 optimal filters that correspond to the eigenvectors associated with the largest eigenvalues obtained from the solution of the Generalized eigenvalue decomposition (GED) using both FC criteria and Max-SNR. The resulting feature vector is the concatenation of the two projections, $R_{1\times 2L} = [z_1\ z_2]$, corresponding to 512 features. From these, the 120 most relevant features are selected using the R-square correlation method, leading to a feature vector $F_{1\times 120}$. The feature vector is then classified by a FLD classifier, as described in sections 5.1.3.B and 5.1.3.C.

5.1.3.B Self-Paced Mode

In self-paced mode, the BCI system needs to detect control and non-control states. In the control state, participants were asked to focus on the target symbols, whilst in the non-control state (user has no intention of selecting a target), participants were asked to keep looking at the screen, but being relaxed without specifically attending to any of the symbols. This was thought of as the most realistic scenario in those situations where the BCI is controlled by users unable to perform any motor movement. This scenario is different from most of the proposed P300-based BCWs that require a mental task to switch to the non-control state, for example, closing the eyes, performing mental tasks (e.g., reading a newspaper), selecting an extra symbol, or combining different neural mechanisms (e.g. P300 with motor imagery) [Rebsamen et al., 2010, Wang et al., 2014, Zhang et al., 2015b, Yu et al., 2017, He et al.,

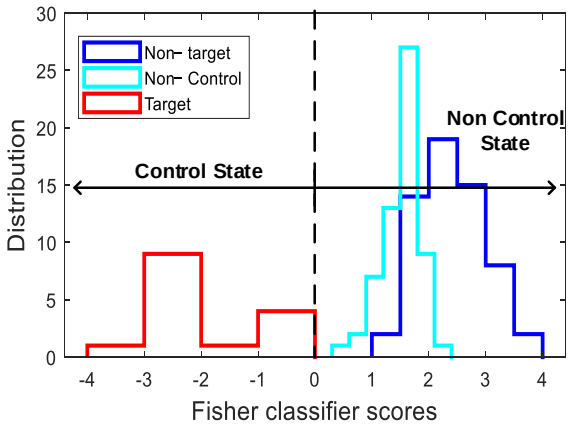


Figure 5.5: Distribution (histogram) of FLD classifier scores of target, non-target, and non-control epochs obtained from a representative participant.

2016]. The self-paced mode comprises 3 classes: target, non-target, and non-control state. Preliminary analyses, during which the calibration was performed with these three classes, showed that the epochs from the non-control class are very similar to non-target epochs, producing a similar classification score distribution as shown in the histogram of Fig. 5.5. Thus, the 3-class classification problem could be then transformed into a binary classification, and it was not necessary to consider the non-control state epochs to detect this class, nor to collect these epochs during calibration (we only need to collect target and non-target epochs). The Fisher classifier scores are positive for non-target and non-control epochs and negative for target epochs (the boundary is set to 0). However, the values of the scores depend on the threshold α, set for each participant, which adjusts the false positive rate as explained ahead, corresponding to virtually move the decision boundary (i.e., instead of moving the boundary, the scores are moved left or right according to the threshold).

Considering the most discriminative features F_i from each class, where $i \in \{+, -\}$ (target $(+)$ and non-target $(-)$), the FLD projection is obtained as

$$\mathbf{y_i} = w\mathbf{F_i} + \mathbf{b} \tag{5.3}$$

where w is the linear discriminative vector, and b is defined as

$$\mathrm{b} = -\frac{(b_1 + b_2)}{2} \tag{5.4}$$

with b_1 and b_2 computed from

$$\mathbf{b}_1 = \frac{1}{K_+}\sum_{k=1}^{K_+}\mathbf{Y}_+ + \sigma(\mathbf{Y}_+) \tag{5.5}$$

$$\mathbf{b}_2 = \frac{1}{K_-}\sum_{k=1}^{K_-}\mathbf{Y}_- - \alpha \times \sigma(\mathbf{Y}_-) \tag{5.6}$$

where K_i is the number of training samples in class i, σ is the standard deviation of target events and non-control state and α is a threshold that adjusts the false positive rate (FPR). False Positives (FP) occur when the user does not intend to convey a command (non-control state) but the system detects a command. False Negatives (FN) occur when the user wants to provide a command but the systems detects a non-control state. The classifier is tuned to minimize the false positive rate (FPR) as it is considered that the impact of a false non-control state is better than a wrong command. FPs can lead to unwanted trajectories of the wheelchair which may render complicated to return to the desired destination goal (e.g., a unwanted 'BACK' command). Yet, when passing through decision points, FNs can also lead to unwanted navigation paths, but as the the wheelchair speed slows down at decisions points, the user has more than one chance to provide the desired command. The threshold α was set experimentally for each participant, with an increment of 0.25 within the interval $[1 : 0.25 : 3]$. Using the calibration data, the FPR and FNR are computed for each increment of α, and then the α that produces the lowest FPR and an FNR less than 10% is selected.

5.1.3.C Dynamic Trial Approach

The DTT approach adjusts dynamically the number of repetitions to user's performance, balancing BCI speed and performance. Throughout the online operation, the P300 classification is computed for each sub-trial of index st. This value varies between $N_{rep} - 2$ and $N_{rep} + 2$ and its minimum value is limited to 2. N_{rep} is set in the calibration session matching a 90% offline classification accuracy (N_{rep} is the same for STT and DTT). The overall DTT approach is described in Algorithm 1 and schematically represented in Fig. 5.4. Starting with $st = Max(N_{rep} - 2, 2)$, the EEG signal is segmented into epochs, pre-processed and averaged. Features are then extracted, selected, and classified using the models trained in the calibration session. A decision parameter D_{st}, which defines the desired degree of confidence to recognize the target command, is computed as the difference between the symbol with the highest score (H_1) and second-highest score (H_2). If the D_{st} value is less than -1 (an empirical value set experimentally and the same for all participants), a valid prediction is identified, the classification output is the target with score H_1, and the system proceeds to the next detection. Otherwise, there is a null prediction, that is, no target symbol is identified during the sequence st, and the vector with epoch (E_{st}) is updated by adding the next epoch ($st = st + 1$). This procedure is repeated until a valid prediction is obtained or the number of sequences is equal to $N_{rep} + 2$.

Algorithm 1 Dynamic trial time algorithm.

E defines segmented epochs
N_{rep} is set according to calibration data (P300 accuracy around 90%)
Start with $st = Max(N_{rep} - 2, 2)$
while $st \leq N_{rep} + 2$ **do**
$E = 0$
for $k = 1$ **to** $k = st$ **do**
    $E = E + E(k)$
end for
$E = \frac{E}{st}$
$Z_{1:2} = W_{1:2}^T E$ (spatial filter projections)
Apply feature selector
Compute classification score (H_i) for each event applying FLD (eq. 5.3)
Select the highest score (H_1) and the second highest score (H_2)
$H_1 \equiv \max S_j \quad , j\epsilon\{1, \cdots, N_s\}$
$H_2 \equiv \max S_j \quad , j\epsilon\{1, \cdots, N_s\} \backslash H_1$
Compute the normalized difference between H_1 and H_2: $D_{st} = \frac{H_2 - H_1}{H_1}$
if $D_{st} < -1$ **then**
    'Valid' prediction and the classification output is the target with score H_1. The system is ready to proceed to the next detection
    break
else
    'Null' prediction, that is, no target symbol is identified during this sequence, so
    $st = st + 1$
end if
end while

5.1.3.D Online performance metrics

We evaluate the feasibility of the BCI system through the accuracy (Acc_{BCI}), number of FP and FN (see definition in section 5.1.3.B), and the number of Wrong target (WT), that is, the user is willing to send a command but the BCI detects a wrong target. For example, a user wants to provide a FORWARD command and the BCI detects left. The BCI accuracy is defined as

$$Acc_{BCI} = \frac{Total_{com} - (FP + FN + WT)}{Total_{com}} \tag{5.7}$$

where $Total_{com}$ is the total number of selections, i.e., the sum of the number of Control commands (CC), i.e., target selections, and the number of Non-control commands (NCC), i.e., trials in which the user does not want to select any target.

The global accuracy of the BCW is referred to as Acc_{BCW} and was computed taking into account the performance of the collaborative controller:

$$Acc_{BCW} = 1 - \frac{BCW_{err}}{CC + NCC} \tag{5.8}$$

where BCW_{err} is the number of the overall BCW errors, that is, the number of wrong commands at the

output of the collaborative controller.

5.1.4 Experimental Results

The experiments were conducted with 7 able-bodied participants ($S1$ to $S7$) and 6 motor impaired participants ($P1$ to $P6$) described in Tables 5.1 and 5.2. Participants of Group I performed the three navigation tasks, namely, $Task1$, $Task2$, $Task3$, and participants of Group II performed only $Task1$ and $Task2$. In $Task1$, it was used the self-paced control with STT approach (fixed number of repetitions). In $Task2$, it was used the self-paced control with the DTT approach (number of repetitions was automatically adjusted online to user's performance). The navigation time (from starting point to the final destination) took on average 11 and 8 minutes respectively for $Task1$ and $Task2$. $Task3$ ran in non self-paced mode, taking on average 11 minutes. The overall experiment lasted between 2 hours and a half and 3 hours for group I and between 3 hours and a half for group II, including setup, calibration, familiarization, navigation times, and questionnaires. Disabled participants did not perform $Task3$ because the time allowed by their institution to carry out the experiments was not enough to accomplish all three tasks. On the other hand, the results of $Task3$ obtained with Group I were very conclusive about the high difficulty and workload in using a non-self paced approach, so it was considered that this task would be unnecessary and unsuitable for the disabled participants. It should be stressed out that participants coming from APCC suffer from severe motor disabilities and required complicated transportation logistics to travel to the site of the experiments. Additionally, it was required that each participant was accompanied by a therapist or caregiver and a psychologist during the whole experimental process.

5.1.4.A BCI performance

All commands received by the Hybrid Motion Planner of the RobChair result from the combination of the detected BCI command with the collaborative controller. Therefore, we need to assess both BCI accuracy (Acc_{BCI}) and "BCI + collaborative controller" accuracy (Acc_{BCW}) to analyze the impact of each module. The online results obtained for $Task1$, $Task2$ and $Task3$ are presented in Table 5.3, Table 5.4 and Table 5.5. For Group I, the average BCI classification accuracies, calculated according to (5.7), are 97.1%, 94.5% and 89.1% for $Task1$, $Task2$ and $Task3$, respectively. The number of commands provided by the users has been also calculated as it is one of the most important quantitative metrics to assess user effort and continuous workload. To accomplish $Task1$, participants in Group I needed to issue on average 10 control commands (target selections), while in $Task3$ the same group issued on average 73 control commands (same path of $Task1$). Although in $Task1$ only 5 decisions were necessary to reach the final destination, participants provided 5 extra commands on average due to wrong BCW commands or due to localization problems. For example, sometimes RobChair misidentified local deadlocks stopping, thereby requiring new commands from the user that led to a trajectory replanning.

The self-paced mode used in $Task1$ has clearly shown its effectiveness in considerably decreasing the number of control commands required to drive the RobChair. It is also possible to conclude that participants spent on average 86.3% of the time in a state of non-control, which undoubtedly greatly reduced the time that users were focused on target selection, with an expected positive impact on users' workload. The collaborative controller increased the overall accuracy of $Task1$, $Task2$, and $Task3$ by 2.9%, 4.5%, and 5.5% respectively, reaching 100%, 99.1%, and 94.6%, leading to very high reliability of the overall system. None of the BCI errors made by Group I in $Task1$ had an impact on the navigation as the collaborative controller rejected them all. For Group II, BCI results were just slightly lower than for Group I, but the BCW accuracy was almost the same, as the collaborative controller corrected most of the wrong BCI commands. These results show the effectiveness of the "Self-paced + collaborative" control approach, with performance remaining stable across patients with varied levels of physical disability.

Table 5.3: Online performance for both groups in Task1: self-paced mode with static trial time

Subjects	CC	NCC	WT	FP	FN	BCW $_{err}$	TT	Acc$_{BCI}$ (%)	Acc$_{BCW}$ (%)
S1	8	61	0	1	2	0	7.1	95.7	100.0
S2	12	72	0	0	2	0	5.9	97.6	100.0
S3	11	58	0	0	0	0	7.1	100.0	100.0
S4	10	78	0	0	0	0	7.1	100.0	100.0
S5	10	66	0	0	2	0	7.1	97.4	100.0
S6	7	57	0	0	2	0	8.4	96.9	100.0
S7	14	49	1	4	0	0	7.1	92.1	100.0
Average	**10.3**	**63.0**	**0.1**	**0.7**	**1.1**	**0.0**	**7.1**	**97.1**	**100.0**
P1	14	63	1	3	2	1	7.1	92.2	98.7
P2	23	104	1	5	1	0	7.1	94.5	100.0
P3	13	85	0	2	4	0	7.1	93.9	100.0
P4	13	68	0	3	4	1	7.1	91.4	98.8
P5	17	77	0	0	1	0	5.9	98.9	100.0
P6	12	66	0	2	0	0	7.1	97.4	100.0
Average	**15.3**	**77.2**	**0.3**	**2.5**	**2.0**	**03**	**6.9**	**94.7**	**99.6**

CC = number of control commands, NCC = number of non-control commands, WT = wrong targets, FP = False Positives, FN = False Negatives, BCW$_{err}$ = number the BCW errors, TT = overall trial time, Acc$_{BCI}$ is the BCI accuracy, Acc$_{BCW}$ is the BCW accuracy.

The average number of FP and FN showed in Tables 5.3 and 5.4 are respectively 1.1 and 0.7 for Group I, and 1.9 and 2.0 for Group II, showing that the control vs. non-control state detection is very effective. The comparison between the BCI accuracy obtained for $Task1$ and $Task2$ gives a measure of the impact of the DTT approach. The BCI classification accuracy was high but lower than using the STT ($Task1$), and the time to select a command was reduced in about 1 s for both groups (*paired t-test, p= 0.003 and p = 0.03*). This shows that the dynamic trial time can be used to adjust the BCI speed vs accuracy. In order to favor accuracy, the DTT method should be more restrictive in the degree of confidence of the command (given by D_{st} in Algorithm 1).

Table 5.4: Online performance for both groups in Task2: self-paced mode with dynamic trial time

Subjects	CC	NCC	WT	FP	FN	BCW$_{err}$	TT$_{Max}$	TT$_{Min}$	TT$_{Mean}$	Acc$_{BCI}$ (%)	Acc$_{BCW}$ (%)
S1	7	27	0	1	0	1	5.9	4.7	5.5	97.1	97.1
S2	11	31	0	0	0	0	8.4	3.5	4.9	100.0	100.0
S3	14	27	0	0	0	0	9.6	5.9	6.7	100.0	100.0
S4	10	31	0	1	0	0	5.9	4.7	5.8	97.6	100.0
S5	19	30	0	4	1	0	9.6	5.9	6.4	89.8	100.0
S6	10	23	0	1	1	0	10.8	5.9	8.0	93.9	100.0
S7	41	23	7	4	0	2	8.4	4.7	6.3	82.8	96.9
Average	**16.0**	**27.4**	**1.0**	**1.6**	**0.3**	**0.4**	**8.4**	**5.0**	**6.2**	**94.5**	**99.1**
P1	9	33	0	0	2	0	7.1	7.1	7.1	95.2	100.0
P2	36	36	5	6	2	1	9.6	4.7	6.8	81.9	98.6
P3	9	31	0	0	1	0	9.6	5.9	7.1	97.5	100.0
P4	12	79	0	1	4	1	5.9	4.7	5.7	94.5	98.9
P5	13	50	0	0	0	0	4.7	3.5	4.4	100.0	100.0
P6	8	31	1	1	3	0	7.4	5.2	6.2	87.2	100.0
Average	**14.5**	**43.3**	**1.0**	**1.3**	**2.0**	**0.3**	**7.4**	**5.2**	**6.2**	**92.7**	**99.6**

CC = number of control commands, NCC = number of non-control commands, WT = wrong targets, FP = False Positives, FN = False Negatives, BCW$_{err}$ = number of BCW errors, TT$_{Max}$ = maximum trial time, TT$_{Min}$ = minimum overall trial time, TT$_{Mean}$ = mean of trial time, Acc$_{BCI}$ is the BCI accuracy, Acc$_{BCW}$ is the BCW accuracy.

Table 5.5: Online Performance for healthy participants (Group I) in Task3: non self-paced mode with static trial-time

Subjects	CC	WT	BCW$_{err}$	TT	Acc$_{BCI}$ (%)	Acc$_{BCW}$ (%)
S1	85	21	10	7.1	75.3	88.2
S2	73	2	2	5.9	97.3	97.3
S3	65	4	4	7.1	93.8	93.8
S4	71	4	4	7.1	94.4	94.4
S5	85	7	3	7.1	91.8	96.5
S6	62	11	3	8.4	82.3	95.2
S7	71	8	2	7.1	88.7	97.2
Average	**73.1**	**8.1**	**4.0**	**7.1**	**89.1**	**94.6**

CC = number of control commands, NCC = number of non-control commands, WT = wrong targets, TT = overall trial time, Acc$_{BCI}$ is the BCI accuracy, Acc$_{BCW}$ is the BCW accuracy.

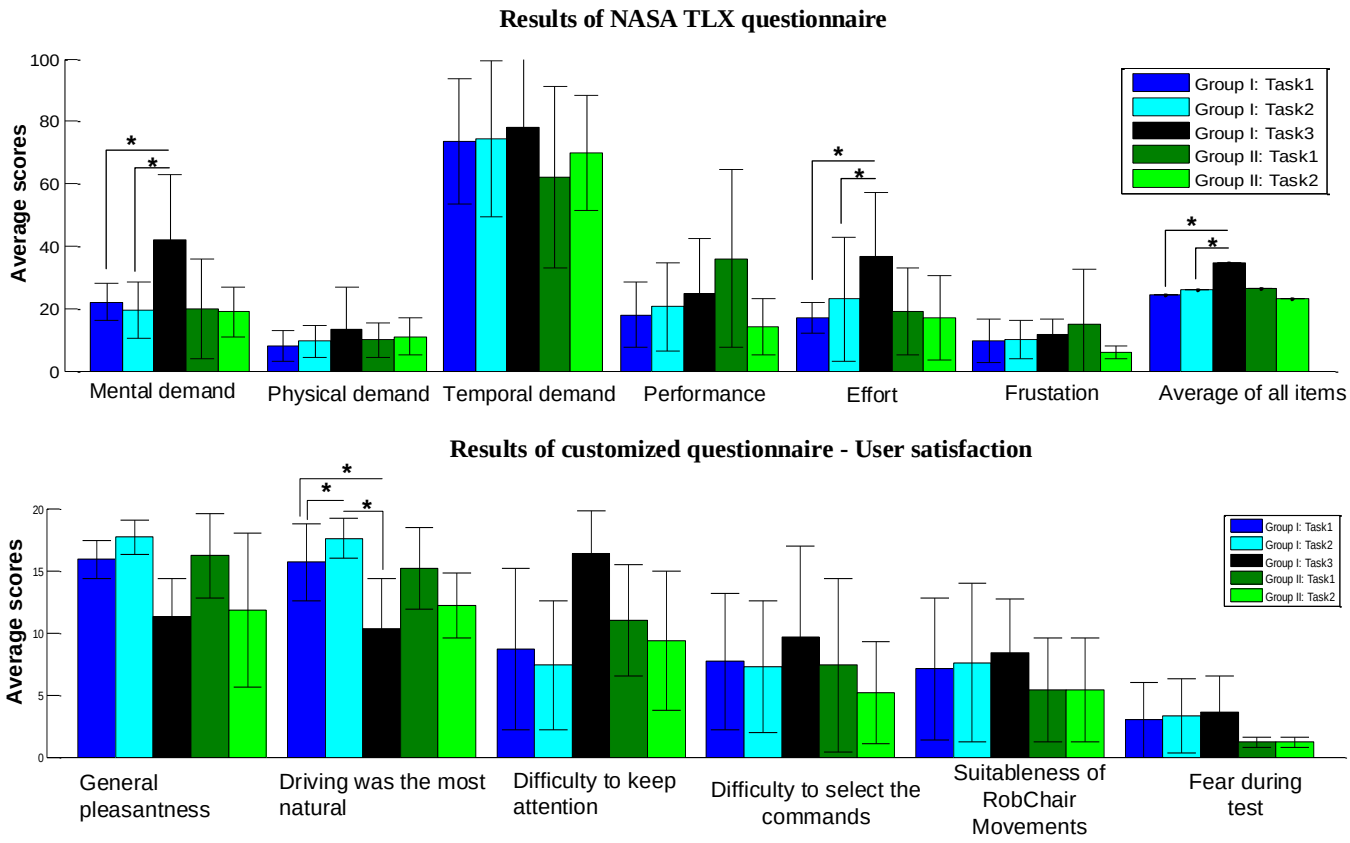


Figure 5.6: Results of the questionnaires for each group. Top: NASA TLX raw rating scores and unweighted average of all items (scale 0-100); and Bottom: results of user satisfaction customized questionnaire (scale 1-20)). (*) indicates items that were statistically significant.

5.1.4.B Subjective questionnaires

Participants were asked to answer a two-part questionnaire assessing their subjective perception of the performed tasks. The first part of the questionnaire was based on the NASA-TLX [NASA, 1986] to assess mental demand, physical demand, temporal demand, performance, effort, and frustration of the three tasks. Only one part of the NASA-TLX has been applied, that is, the participants rated each subscale but did not evaluate the contribution of each factor (weight). The overall workload for each subject is therefore an unweighted average of these six subscales. The second part is a customized questionnaire that compares the 3 tasks regarding the degree of user satisfaction, with questions directed to the specific tasks. The NASA-TLX workload scores range between 0 and 100 and participants rated the customized questionnaire on a 20-point scale, ranging from 1 (very low) to 20 (very high). For the performance parameter, the scale is inverted, i.e., 1 is perfect and 20 is a failure. In addition, psychometric questionnaires were applied to Group II to assess the emotional state according to State-Trait Anxiety Inventory [Spielberg et al., 1970] and the Beck Depression Inventory [Beck et al., 1961], in order

to ensure that the participants were emotionally capable of performing the tasks.

The average results of the two subjective questionnaires are in Fig. 5.6. Group I reported mental demand and effort as significantly higher in $Task3$, compared to $Task1$ (diff=20.0, paired t-test, p=0.01, and diff=19.5, p=0.02, respectively), and compared to $Task2$ (diff=22.5, paired t-test, p=0.006, and diff=13.5, p=0.002, respectively), as expected from the quantitative results. $Task2$ was scored as slightly less mental demanding than $Task1$ for both groups (diff=2.5 and diff=1), but the difference is not statistically significant. The perceived performance of Group I in $Task1$ was slightly greater than in $Task2$ and $Task3$, although BCI classification accuracies of the 3 tasks were very similar. Group I classified $Task3$ as more temporal demanding, without significant differences between tasks (paired t-test, *p=0.26* and *p=0.35*). For Group I, temporal demand was similar for $Task1$ and $Task2$ (*diff=1*), and for Group II $Task2$ was considered slightly less demanding (*diff=8*). The scores for physical demand and frustration of both groups are low (less than 3). These results suggest that participants felt comfortable and fearless. The average of Group I for NASA-TLX items shows that $Task3$ was the task with the highest taskload (paired t-test, *p=0.03* and *p=0.009*). For Group II, $Task2$ presents the lower taskload, but the difference is not statistically significant. Comparative results between Task1 and Task2 could have been influenced by the difference between tasks. Although both tasks have the same number of decision points, Task2 has a shorter path and fewer obstacles.

In the second questionnaire, $Task3$ was reported as the less pleasant, less natural, more difficult to keep the attention, and the more difficult to select commands, which again corroborates the results of the quantitative metrics. Participants of Group I classified $Task2$ as more natural than $Task1$ (*p=0.03*). For Group I, $Task2$ is more pleasant than $Task1$ (*p=0.06*), while Group II considered $Task1$ more pleasant (*p=0.07*), only near statistical significance.

Regarding the perception of "Suitableness of RobChair Movements", Group II considered the movements more appropriate than Group I.

5.1.4.C Discussion

A brain-controlled wheelchair is a complex system that requires a high level of reliability and safety and involves intelligent navigation systems. The goal of this study was to assess the impact of the combination of a collaborative controller with a self-paced control (using STT and DTT approaches) on users' effort, the naturalness of interaction, and system reliability when driving a robotic wheelchair with a BCI. Able-bodied and motor impaired participants used the proposed self-paced BCI with a mean accuracy of 95.8% and 93.7% (average of $Task1$ and $Task2$), respectively. These results show an effective implementation of the BCI classifier and in particular of the control vs. non-control state detection. Still, the average number of FP was higher than the number of FN, which was not what was intended. A posterior offline analysis made after the experiments showed that a better tuning of the threshold α

could have decreased the number of false positives. The collaborative controller increased the overall system accuracy to above 99% for both groups, clearly showing its importance for the reliability of the BCW. Even using the non self-paced approach (which yielded a 89% BCI accuracy) the collaborative controller increased the overall BCW accuracy to 94.6%. The collaborative controller proved to have the desired effect, by discarding wrong BCI commands and replacing them by the intended ones, thereby reducing the impact of lower BCI performances. The self-paced control enormously reduced the number of the required commands, specifically from 73 to 10 on average. This reduction had a significant impact on the perceived overall task workload as shown in Fig. 5.6, in particular on mental demand and effort. Accordingly, the greater workload of the non self-paced operation was reflected in a decrease of the BCI accuracy in 8.0%, when compared to the self-paced operation. Analysing $Task1$ vs. $Task2$ it was found that the DTT increased the BCI speed by reducing the time per trial in about 1 sec, but it slightly decreased the BCI performance, which was not the desired outcome.

Based on these results, we can state that the self-paced approach had a very high impact on the reliability, naturalness, and workload demand of the BCW, and the collaborative controller had a high impact on the reliability of the BCW, with increased relevance when the BCI performance was lower. Although with a lower impact on the entire system, the DTT approach showed the possibility of adjusting BCI speed vs. user's performance. This will be a subject of future research. For example, the D_{st} threshold should be individually tuned for each participant to ensure an improvement of the BCI accuracy. Moreover, the impact of the DTT approach may have been diminished by the impact of the self-paced control, since the user is less susceptible to lack of attention and fatigue. Overall, participants scored $Task1$ and $Task2$ very similarly. As regards participants' subjective preference in performing the designated tasks, there were no significant differences between the two groups. The subjective results show a very positive user experience feedback regarding workload demand and the naturalness of control of the overall system.

Table 5.6 shows a comparison between different brain-actuated wheelchair architectures that are closely related to our system, i.e., that use a control scheme combining user and machine commands, a self-paced paradigm, and experiments with real wheelchairs. Only one of the studies reported experiments with motor disabled participants [He et al., 2016], which emphasizes the need for more studies involving the potential target users, in a perspective of human-centred design. Additionally, most of the experiments of the proposed works were performed in highly structured environments set up in the lab. Our work presents the most complex navigation scenario including both healthy and severely motor disabled participants. To the best of our knowledge, our proposal is the only one achieving an overall accuracy greater than 99%, which validates the proposed BCI and navigation approaches. Moreover, from this group of studies, our study is the only one assessing user experience through subjective questionnaires.

Table 5.6: Summary of related BCW works that use a collaborative control, a self-paced paradigm and experiments with real wheelchair.

Study	BCI approach	Self-paced paradigm	Environment	Subjects
Rebsamen et al. [Rebsamen et al., 2010]	P300 and MI	Static trial	Structured environment based on corridors and rooms without obstacles.	Healthy: 5
Carlson and Millan [Carlson and Millan, 2013]	MI	Static trial	Unstructured environment based on an office room with static obstacles.	Healthy: 4
Wang al. [Wang et al., 2014]	MI and P300 and blink	Static trial	Semi-structured environment based on corridors with static obstacles.	Healthy: 4
Zhang al. [Zhang et al., 2015b]	MI or P300	Static trial	Semi-structured environment based on a domestic room with static obstacles.	Healthy: 9
He al. [He et al., 2016]	P300	Static trial	Real. Trajectory with two destinations in a room.	Healthy: 8; Disabled: 5
Our study	**P300**	**Dynamic trial**	**Semi-structured environment based on office rooms and corridors with static obstacles and narrow passages.**	**Healthy: 7; Disabled: 6**

The good results achieved by motor disabled participants suggests that the proposed BCI may represent an effective solution for wheelchair control. Still, more extensive experiments with a wider group of participants and in more natural living contexts are needed to validate the approaches. The overall results have been very promising and motivate new further research already under way, namely the integration of vision sensors to recognize semantic features, such as doors, tables, chairs, which will be incorporated dynamically as target goals in the interface. Error related potentials (ErrP) that we have already used in a different context [Cruz et al., 2018a] are also being integrated to improve the reliability of the BCI commands.

5.2 Detection of Stressful Situations While Steering a BCW (case-study IV)

Using the dataset obtained in case-study III we analyze whether the galvanic skin response, recorded in participants steering a robotic wheelchair with a BCI, is elicited in particular contexts susceptible to triggering emotional arousals such as stress. Additionally, it is researched whether GSR can be automatically detected to adapt a human-machine collaborative controller. Healthy and motor impaired participants were asked to steer the RobChair ISR-UC prototype [Lopes et al., 2016] in indoor office

environments, including complex scenarios such as passing narrow doors, avoiding obstacles, and with situations of unexpected trajectories of the wheelchair (controlled by an operator without users knowledge). Emotional state recognition can be applied in several contexts. For example, if a high level of stress is detected, the BCI can be automatically adjusted, by increasing the time for command detection; or RobChair's navigation system can take full control, disregarding user's input commands. Thus, the system could be able to adapt according to both navigation context and user's state, increasing the overall system reliability.

5.2.1 Introduction

Users who drive a robotic wheelchair through a brain-computer interface provide only sparse commands encoding direction decisions or destinations, while the navigation system is responsible for perceiving the environment, planning, and executing the trajectories. Therefore, users know that they do not have full control of the wheelchair, strongly relying on the machine. This can make them experience stressful reactions such as fear and anxiety, when placed in difficult or unexpected situations (e.g., narrow-door passages or unexpected obstacles). These emotional reactions may eventually contribute to a degradation of the BCI classification performance leading to unwanted navigation commands, which can worsen the stressing situations. Changes in emotional states such as anxiety or fear produce bodily reactions not controlled by the person, which are driven by the sympathetic chain of the Autonomic nervous system (ANS) [Dawson et al., 2007]. One such reaction is the increase of the sweat glands activity, resulting in increased skin conductance. Thus, galvanic skin response, which measures the electrical conductance of the skin, can be used as a measurable parameter reflecting stress and other emotional states [Cornelia et al., 2013, Setz et al., 2010]. Automatic detection of user's emotional state can be useful for applications in which humans and machines interact, particularly in human-machine collaborative systems, where systems can adapt to human's emotional state. GSR signal has two components, the tonic level of skin conductance referred to as Skin conductance level (SCL), and rapid phasic responses referred to as skin conductance responses [Dawson et al., 2007]. An example of a typical GSR signal is presented in Fig. 5.7. SCL varies slowly and it is continuously changing. SCRs appear as accentuated peaks and may or may not be related to stimulus events. They are characterized by a minimum amplitude of 0.01 μS (micro Siemens) and a latency value between 1-3 s, a rise time value between 1-3 s and half recovery time value between 2-10 s with an exponential decay [Dawson et al., 2007, Boucsein, 2012]. The skin conductance response can be elicited by a specific stimulus (S-SCR) or elicited spontaneously, that is, without any identifiable stimulus, referred to as non-specific SCR (NS-SCRs). The number of NS-SCRs peaks elicited in a given time period (frequency) is typically 1-3 per minute in rest periods and over 20 per minute in high arousal situations [Boucsein, 2012]. GSR and other physiological signals such as body temperature and heart rate, also controlled by the ANS, have been used in real-

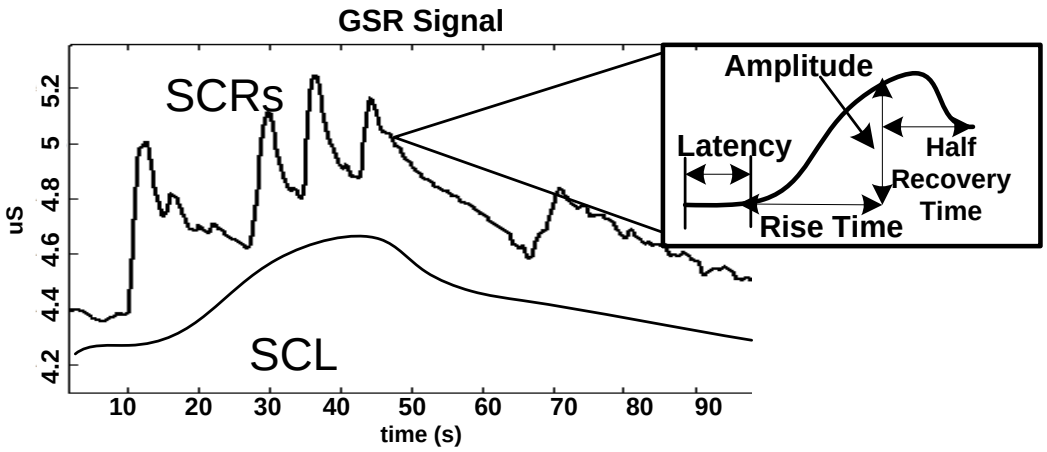


Figure 5.7: Skin conductance responses (SCRs) and skin conductance level (SCL) of a typical GSR signal (filtered by a low-pass filter at 1 Hz).

world applications to detect emotional arousal. Healey and Picard [Healey and Picard, 2005] explored different biosignals (e.g. electrocardiogram, electromyogram, GSR) during real-world driving tasks in order to detect driver's stress levels (low, medium, high). They recorded data during rest periods and while driving on the highway and in the busy main street. These data were divided into 5-minute segments and then classified in the 3 stress levels. They showed that the most relevant information to distinguish these stress levels is obtained through skin conductivity and heart rate metrics. GSR was also used in computer-based arithmetic and reading tasks [Mundell et al., 2016, Nourbakhsh et al., 2012], to measure user's performance and cognitive load. Most of the works analyze the GSR signal in laboratory conditions, pointing out possible applications, but without actually using GSR to influence or change the behaviour of the system.

5.2.2 SCR Detection

Several approaches have been used to extract SCRs from GSR. Greco et al. [Greco et al., 2016] consider the GSR signal as the sum of three components: phasic, tonic, and noise and use a convex optimization approach (cvxEDA) to split these 3 components. Then, the area of the phasic component is computed and a signal with an area exceeding a 0.5 threshold is considered a SCR. In [iMotions, 2017] a median filter is applied and subtracted from the signal to remove the SCL component (baseline). In the phasic data, the SCRs are detected based on a threshold for the peak onset and a threshold for the peak offset ($< 0\mu S$). Given that SCL is continuously moving, methods based on baseline correction are the most effective for SCR detection.

We propose a baseline correction inspired in [Golotvin and Williams, 2000] (usually used for correction of baseline distortion in spectra), and then we fit specific characteristics of typical SCRs signals

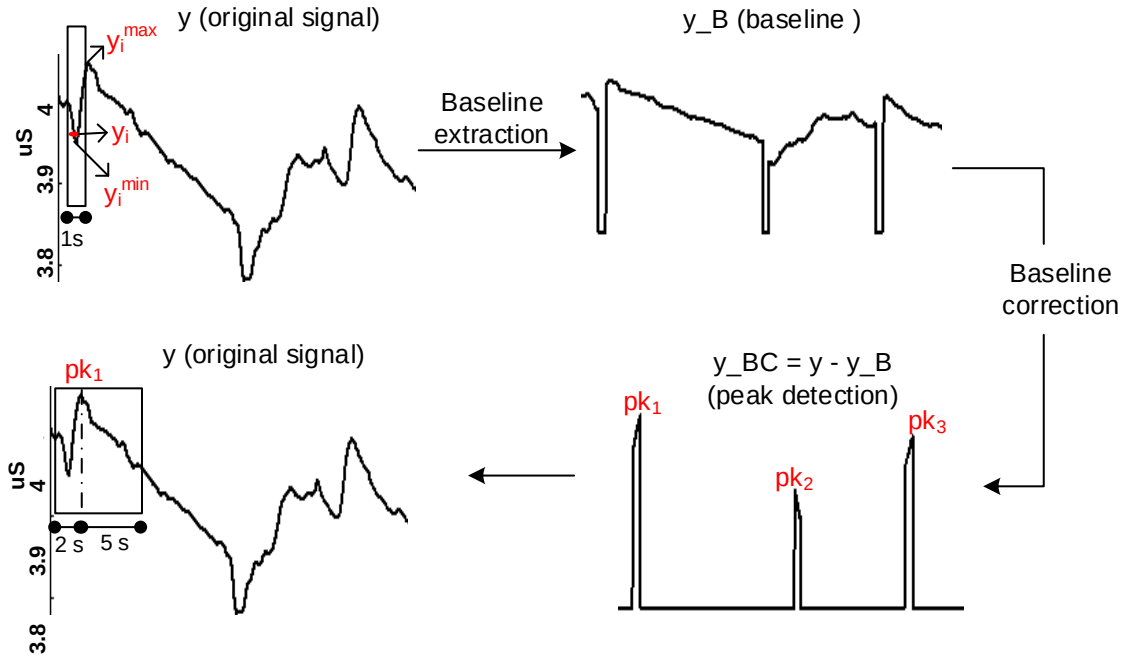


Figure 5.8: SCRs peaks detection. First, a sliding window of 1-second is used to detect baseline points on signal y, then peaks are detected from baseline corrected signal (y_BC). A sliding window of 7-seconds is used to verify if peak pk in signal y has an amplitude higher than 0.01 μS and if it has an exponential decay.

(amplitude and exponential decay) to each possible SCR peak. We call the proposed method feature-based peak detection. Let us consider the GSR signal represented by y (see Fig. 5.8). To decide if a point $y(i)$ belongs to the baseline, a sliding window of 1-second-width is centred at each time sample i. Then, the minimum and maximum values (y_i^{min} and y_i^{max}) are computed for each window. If their difference is less than a threshold value (TH), the point $y(i)$ is considered part of the baseline. The baseline signal (y_B) is constructed from

$$y_B = \begin{cases} y(i), & \text{if } y_i^{max} - y_i^{min} < TH \\ c, & \text{otherwise} \end{cases} \tag{5.9}$$

The threshold is $TH = n \times \sigma_{noise}$, where σ_{noise} is the noise standard deviation and n is usually set between 2 and 4. In our experiments it was empirically adjusted to 4. To compute the σ_{noise}, the signal y is divided into 32 equal regions. Then, the standard deviation σ of each region is calculated and σ_{noise} is the one with minimum value. The constant c is the minimum value of GSR signal for each participant obtained during the rest period (before the onset of the experiment).

Then, the baseline curve is subtracted from the GSR signal resulting in the baseline corrected signal $y_{BC} = y - y_B$, clearly revealing the peaks (pk) associated with SCRs. Matching the instants of the detected peaks on signal y and using a rise time value of 2 s and a half recovery time of 5 s we compute the amplitude of each peak as the difference between maximum and minimum of the window. We fit the

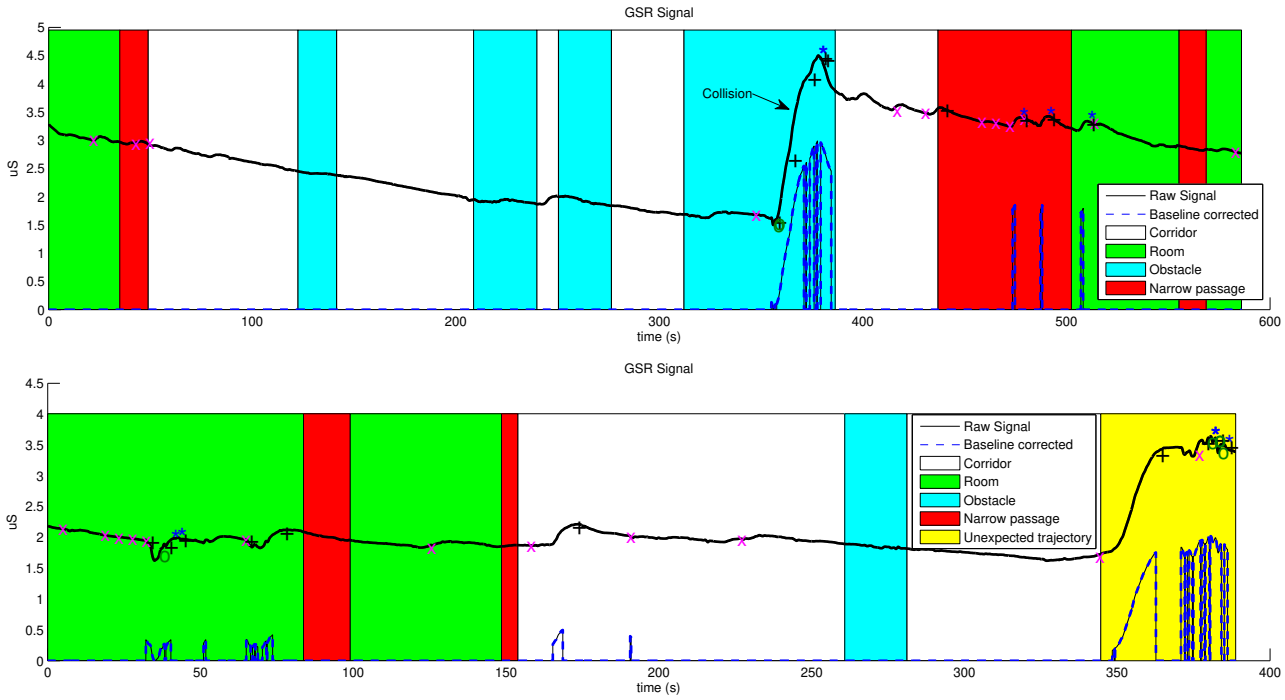


Figure 5.9: GSR signal during first (top) and second (bottom) navigation task. Plots contain low-pass-filtered GSR signal (solid line), GSR signal after baseline correction (dashed line), peaks detected by our algorithm (marked with $*$), peaks detected by the median filter (marked with $+$), peaks detected by cvxEDA methods (marked with o) and BCI commands (marked with $\times$).

data window with an exponential $a \times e^{-bx}$. If $a > 0$, $0 < b < 1$ and the amplitude is greater than 0.01 μS, the peak is considered a SCR, otherwise it is discarded. The algorithm is flexible since it uses two characteristics of the SCR (amplitude and exponential decay), which can be adjusted.

5.2.3 Experimental Results and Discussion

5.2.3.A Analysis of GSR Signal in Stressful Situations

The experiments were carried out by seven healthy participants (Group I - GI) and six motor disabled participants (Group II - GII), however the results of two motor disabled participants were discarded, one because there was no electrodermal response, and the other because of strong artifact contamination. Participants steered RobChair with the self-paced P300-based BCI on two distinct routes. The BCI accuracy was 97.1% and 96.4% in group GI, respectively for the first and the second route, with an average time per command $TT = 6.7s$. Group GII achieved a BCI accuracy of 95.6% and 94.9% for the first and second routes with a $TT = 6.2s$.

Examples of RobChair routes and relevant points are marked in maps of Fig. 5.2 and Fig. 5.3. Routes, decision points, and events were recorded synchronously with GSR and EEG. In Fig. 5.9, we show the complete GSR recording for a representative able-bodied participant during the first and

the second navigation tasks, as well as the navigation events labelled as follows: corridors (white), rooms (green), obstacles (blue), narrow passages (red), and the unexpected RobChair trajectory (yellow). There were also some slight lateral scrapings on obstacles, which are marked in Fig. 5.9 as collisions. The black solid line represents the preprocessed lowpass-filtered GSR signal and the blue dashed line represents the GSR signal after baseline correction. Baseline correction makes significant peaks evident, enabling the detection of S-SCRs and NS-SCRs. Fig. 5.9 shows that a strong SCR was elicited when a lateral collision with an obstacle G occurred, while no SCRs were elicited when RobChair passed previous obstacles without collisions. The first and third narrow passages did not elicit any SCRs and the second narrow passage elicited some SCRs. In the second navigation task (bottom of Fig. 5.9) there is a well-defined SCR with high amplitude originated by an unexpected trajectory, resulting from the remote operation of RobChair without participant knowledge.

The automatic detection of GSR was evaluated through the proposed FBPD method, the median filter approach [iMotions, 2017], and the convex optimization approach [Greco et al., 2016], considering the visual detection of a human expert as ground truth. The human detection, although susceptible of misclassification, provides the most reliable reference for the validation of the automatic methods. Fig. 5.9 shows the SCRs detected by the 3 automatic detectors. We assumed as S-SCRs all peaks that occurred in narrow passages, obstacles, collisions and unexpected RobChair trajectories, and as NS-SCRs all peaks that occurred in rooms and corridors. The 3 algorithms are user-independent, that is, the model parameters are the same for all participants. The median filter has only one adjustable parameter, the width of the window, which was set to 4 s centered on the current sample. We evaluated the cvxEDA method with the same parameters defined in [Greco et al., 2016].

Table 5.7 shows the number of S-SCRs and NS-SCRs detected manually and by the automatic algorithms. The classification accuracy was respectively 82.4%, 69.3% and 33.3% for GI and 87.6%, 32.7% and 78.9% for GII. The FBPD method performed better than the other two with a classification accuracy above 82% for both groups. These results show that detecting SCRs based on exponential fitting as proposed in FBPD is more effective than the approaches based on peak and area thresholds. The median filter and cvxEDA methods [Greco et al., 2016], [iMotions, 2017] had irregular performance, with a tendency to overestimate the number of detected peaks in one group and underestimate in the other group. The amplitude and frequency of S-SCRs and NS-SCRs are known to be related to stress level. Table 5.7 shows these two parameters, taking the SCRs detected with FBPD. The frequency of S-SCRs elicited in collisions was not calculated since collisions occurred during very brief periods of time. The SCRs elicited in narrow passages and obstacle contour had lower amplitudes than the SCRs that occurred in rooms and corridors, and their frequencies were typical of NS-SCRs (1-3 per minute) [Boucsein, 2012]. Therefore, we can conclude that narrow passages and obstacles did not affect participants. Collision and unexpected trajectory were the events eliciting the SCRs with the

Table 5.7: Number of SCRs elicited during the experiments

	Narrow passages		Obstacles		Unexpected trajectories		Rooms and corridors		Automatic detection accuracy (%)		Collisions[a]
	GI (35)	GII (20)	GI (35)	GII (20)	GI (7)	GII (4)	GI	GII	GI	GII	GI (3)
SCRs detected manually	33	18	43	27	22	17	143	112			3
SCRs detected using proposed FBPD	28	19	34	22	25	16	172	128	82.4	87.6	3
SCRs detected using median filter	38	11	52	19	33	14	226	60	69.3	32.7	3
SCRs detected using cvxEDA	15	16	34	11	30	27	76	121	33.3	78.9	1
Amplitude (μS)	0.3	0.3	0.2	0.3	1.0	0.9	0.7	0.5			1.7
Frequency (per min)	2.1	2.4	1.9	2.0	4.2	4.9	1.6	2.0			

[a]During the navigation tasks performed by GI, 3 collisions (slight lateral scraping) occurred and all of them elicited one SCR. For GII there were no collisions.

highest amplitude and frequency. These two situations show clear evidence of users' arousal, i.e., the participants were very sensitive to sudden events. Figure 5.10 illustrates well these conclusions. The unexpected trajectories event is discriminated at the top right corner. Collision events are not represented, because they are very brief and therefore impossible to measure the SCR frequency.

Table 5.8 shows the percentage of events that elicited at least one SCR. For GI, we found that 17 narrow passages, 19 obstacles, 3 collisions and 6 unexpected trajectories elicited at least one SCR. Collisions elicited SCRs in 100% of the times, and the unexpected trajectories of RobChair resulted in SCRs in 87.5% of the times. However, SCRs were elicited in only 48.6% of narrow passages and 54.3% of obstacles. For GII we verified similar results, SCRs were elicited in 100.0% of unexpected trajectories and in only 45.0% of narrow passages and 60.0% of obstacles. For GI, 76.5% of SCRs elicited in narrow passages and 100.0% of SCRs elicited in obstacles occurred on the first task. For GII 78.8% and 75.0% of SCRs elicited in narrow passages and bypassing obstacles occurred on the first task. These results could mean that the participants gained confidence in RobChair in the first task, reducing their emotional arousal in the second task.

The presence of NS-SCRs increases the difficulty in associating SCRs to stressful events. Most state-of-the-art applications use SCRs features (e.g., peak rate, peak height, response durations) disregarding the discrimination between SCRs and NS-SCRs. For a system that aims to self-adjust to user's emotional arousals the non-discrimination between S-SCRs and NS-SCRs may be critical. As shown in

Table 5.8: Number of events that elicited at least one SCR

	Narrow passages		Obstacles		Unexpected trajectories		Collisions	
	GI (35)	GII (20)	GI (35)	GII (20)	GI (7)	GII (4)	GI (3)	GII (0)
SCRs detected manually	17	9	19	12	6	4	3	-
SCRs detected using proposed FBPD	14	11	16	9	7	4	3	-
SCRs detected using median filter	20	7	19	6	7	3	3	-
SCRs detected using cvxEDA	4	5	7	4	4	3	3	-
SCRs elicited (%)	48.6	45.0	54.3	60.0	85.7	100.0	100.0	-

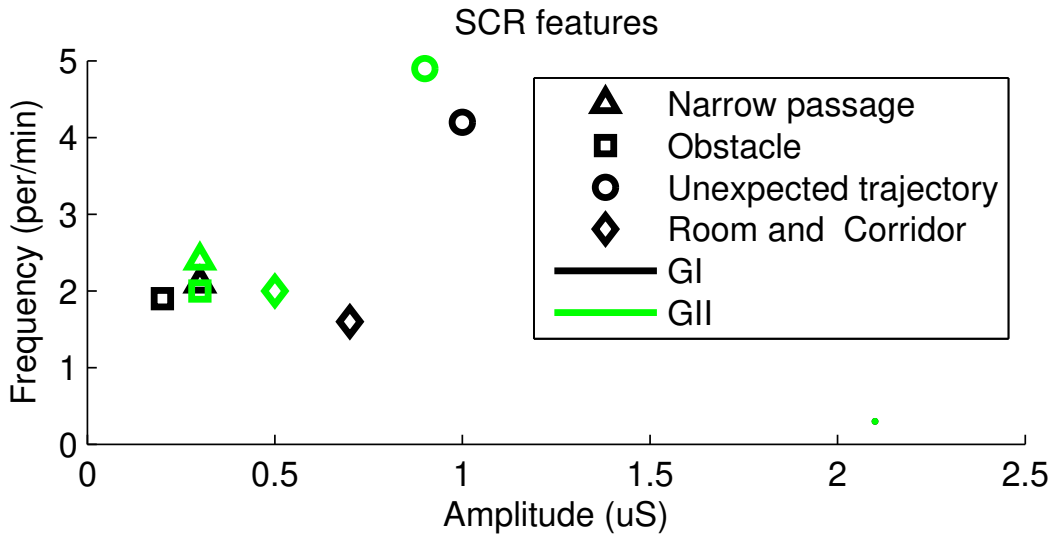


Figure 5.10: Scatter plot of average frequency vs average amplitude of SCRs in different events.

Fig. 5.10, amplitude, and frequency of peaks can be used to adjust a threshold for distinguishing both. In the future, other SCR features such as rise time and area under the responses could be explored in order to increase the discrimination between S-SCR and NS-SCR.

5.2.3.B Relationship Between SCRs and BCI Performance

An analysis was carried out to investigate whether information of SCRs can be used to anticipate wrong BCI commands and if wrong commands led to an increase of SCRs. We analyzed the last 4 s and 60 s before each BCI command and the 60 s after the commands. The participants of GI and GII performed the two navigation tasks with 143 and 91 correct BCI commands and 13 and 10 wrong BCI commands, respectively (Table 5.9 and Table 5.10). The wrong commands were preceded by a percentage of SCRs higher than the correct commands for group GI. However, for group GII the percentage of SCRs

Table 5.9: Number of correct BCI commands that elicited at least one SCR

	Correct BCI commands - GI: 143 and GII: 91					
	Period of 4 s before a BCI command		Period of 60 s before a BCI command		Period of 60 s after a BCI command	
	GI	GII	GI	GII	GI	GII
SCRs detected using proposed FBPD	26	11	91	75	88	77
Percentage of commands that elicited SCRs	18.2	12.1	63.6	82.4	61.5	84.6

Table 5.10: Number of Wrong BCI Commands that Elicited at Least one SCR

	Wrong BCI commands - GI: 13 and GII: 10					
	Period of 4 s before a BCI command		Period of 60 s before a BCI command		Period of 60 s after a BCI command	
	GI	GII	GI	GII	GI	GII
SCRs detected using proposed FBPD	5	0	11	8	11	9
Percentage of commands that elicited SCRs	38.5	0.0	84.6	80.0	84.6	90.0

for correct and wrong commands was similar. Therefore, we could not establish a clear relationship between SCRs and BCI performance that could be used to anticipate a BCI error. Comparing the percentage of commands that elicited SCRs before and after correct and wrong commands, we verify that the percentage is similar for correct commands and that it is higher after wrong commands in group GII. The results suggest that BCI errors could have caused emotional arousal in group GII.

The correlation between SCRs and BCI performance was inconclusive and a set of experiments is being prepared to further research this correlation. The information of SCR may be used to detect stressful situations while users were driving a brain-actuated RobChair. The method FBPD proposed for automatic detection of SCRs proved to be very effective comparing to the state of the art methods. Despite the encouraging results, the lack of clear discrimination between S-SCR and NS-SCR hampers the identification of stressful situations.

5.3 Conclusion

In this chapter, we assessed the impact of the integration of collaborative control, self-paced control, and dynamic-time commands into a BCW system. The system was validated by 7 healthy participants and 6 motor disabled patients in real office-environment navigation tasks. Both able-bodied and motor disabled participants successfully controlled the BCI system with an average BCI accuracy of 95.8% and

93.7% respectively and the collaborative controller corrected most of wrong commands increasing the accuracy to more than 99% for both groups. In case-study II, the GSR was analyzed to infer whether it could be reliably correlated to stressful situations, and automatically detected while users were driving a BCW. The results demonstrated that information of SCR may be used to detect stressful situations. The method FBPD proposed for automatic detection of SCRs proved to be very effective comparing to state of the art methods. Despite the encouraging results, the lack of clear discrimination between S-SCR and NS-SCR hampers the identification of stressful situations.

6

Spatial filtering: Riemannian Geometry Approach

Contents

Cross-session and cross-subject classification problem in BCI applications is very challenging due to the many changes that can happen between sessions and between subjects, such as electrode positions and electrode impedance, level of attention or stress. Invariance properties of the Riemannian manifold of symmetric positive definite (SPD) matrices can be used to deal with this issue since EEG data can be represented through spatial covariance matrices which are SPD. In this chapter, we propose and describe a spatial filter based on Riemannian geometry able to use the invariance properties of Riemannian distance to handle cross-session and cross-subject generalization. The proposed method is validated with the dataset BCI-Double-ErrP-Dataset [Cruz et al., 2020] obtained in [Cruz et al., 2018a] and benchmark datasets [BCI-Competition, 2015]. The results evidence that the proposed filter improves the generalization across sessions and across subjects, and that it is robust to the amount of error training samples used to build the classification model.

6.1 Introduction

BCI is being increasingly used for a variety of applications including clinical and non-medical applications [Nicolas-Alonso and Gomez-Gil, 2012], [Nam et al., 2018, pages 89-247], however, it is still rarely used outside laboratories. One of the major issues is the need to calibrate the BCI system each time it is used, which is annoying and unpractical. This re-calibration procedure is due to non-stationarity of EEG signals, which is maybe caused by: 1) the variations of users' mental states, such as stress and fatigue, and 2) technical interferences, such as changes in electrode positioning and electrode impedance. Therefore, EEG signal typically changes across sessions and it changes dramatically across subjects. Thus, the classification model built with data collected from one session/subject usually works poorly on a new session/subject. This issue can be tackled by TL techniques. Transfer learning can be defined as "the ability of a system to recognize and apply knowledge and skills learned from previous tasks to novel tasks" [Pan and Yang, 2009]. In the context of BCIs, the previous tasks can represent the previous sessions or subjects, and the novel tasks can correspond to the current session or subject. Therefore, in a cross-session problem, the data of a new session is classified by learning from the past data recorded in previous sessions of the user. In the cross-subject problem, the data of a new subject is predicted by learning from the data of other subjects. In this study, we propose a new spatial filter method called the Riemannian Fisher criterion beamformer that extends our previous FCB filter introduced in [Pires et al., 2011a] to a Riemannian manifold of SPD matrices. $RFCB$ uses the invariance properties of Riemannian distance to obtain a robust filter for extracting discriminative and invariant features that are not strongly affected by cross-session and cross-subject variability.

6.2 Preliminaries

Lets consider a spatio-temporal matrix $X \in \Re^{N_c \times L}$ representing the EEG epoch $X = [x(t_1) \quad x(t_2) \quad ... \quad x(t_L)]$ of $N_c = 12$ channels with $L = 256$ time samples. The spatial covariance matrix $COV_i \in \Re^{N_c \times N_c}$ of i-th trial is computed from

$$COV_i = X_i X_i^T \tag{6.1}$$

Definition 1 (SPD matrix). [Barachant et al., 2010] The matrix A is said to be Symmetric Positive Definite if $A = A^T, x^T Ax \neq 0, \forall x = 0$ and its eigenvalues are positives.

Definition 2 (Domain). [Pan and Yang, 2009] A domain $\mathcal{D} = \{\mathcal{X}, P(x)\}$ comprises a d-dimensional feature space $\mathcal{X}$ and a marginal probability distribution $P(x)$. Usually, different domains may have different feature spaces or different marginal probability distributions.

We consider the EEG data from previous sessions or subjects in the source domain $\mathcal{D}_S = \{(x_{S1}, y_{S1}), \quad ... \quad (x_{SN}, y_{SN})\}$, where $x_{Si} \in \mathcal{X}_S$ is the data instance and $y_{Si} \in \mathcal{Y}_S$ is the associated class label. Similarly, the current sessions or subjects are in the target domain $\mathcal{D}_T = \{(x_{T1}, y_{T1}), \quad ... \quad (x_{TN}, y_{TN})\}$, where $x_{Ti} \in \mathcal{X}_T$ and $y_{Ti} \in \mathcal{Y}_T$.

6.3 Riemannian Fisher Criterion Beamformer

This section describes the proposed Riemannian Fisher criterion beamformer approach that uses the invariance properties of Riemannian metrics to deal with the variability between session/subject.

6.3.1 Robustness of Riemannian structure using EEG signal

The EEG signal $X(t) \in \Re^{N_c \times L}$ generated by the brain sources $s(t) \in \Re^d$ can be represented by

$$X(t) = As(t) \tag{6.2}$$

where $d \leq N_c$ is the number of EEG dipolar fields and $A \in \Re^{N_c \times d}$ is a mixing matrix [Congedo et al., 2017]. The mixing matrix varies according to the physical properties of the head, the impedance of the electrodes, and electrode positioning. Thus, it strongly depends on the conditions of each session. Therefore, it is not surprising that cross-session BCI learning is a very challenging task and cross-subject learning is even more difficult. In order to analyze the variability between sessions let us consider two trials (i and j) of source domain (session 1) with the mixing matrix A_S and with the source covariance matrices C_i and C_j, and two trials of target domain (session 2) with the mixing matrix A_T, and the same source covariance matrices C_i and C_j. The sensor covariance matrices of the two trials from session 1 can be expressed as $P1_i = A_S C_i A_S^T$ and $P1_j = A_S C_j A_S^T$, and the sensor covariance matrices of the

two trials from session 2 are given by $P2_i = A_T C_i A_T^T$ and $P2_j = A_T C_j A_T^T$. The difference between A_S and A_T results in covariance matrices that are different from each other, however the congruence invariance of the AIRM guarantees that

$$\delta(P1_i, P1_j) = \delta(C_i, C_j) \quad and \quad \delta(P2_i, P2_j) = \delta(C_i, C_j) \tag{6.3}$$

and hence, it yields

$$\delta(P1_i, P1_j) = \delta(P2_i, P2_j) \tag{6.4}$$

Therefore, the Riemannian distance between covariance matrices of session 1 is similar to the Riemannian distance between covariance matrices of session 2, thereby it is expected that a spatial filter that maximizes the Riemannian distance between classes will be identical either calculated using the data recorded in session 1 or the data from session 2. Thus, we propose to extend the FCB to the Riemannian manifold of the SPD matrix taking advantage of the congruence invariance property of the AIRM to minimize the effect of data variability in cross-sessions and cross-subjects.

6.3.2 The RFCB algorithm

The proposed RFCB algorithm computes the spatial filters $W^{N_c \times m}$ that projects the EEG signal to a lower m-dimensional subspace ($m < N_c$) more discriminative, which simultaneously maximize the Riemannian distance between classes and minimizes the Riemannian distance within classes. The proposed spatial filter is inspired in [Harandi et al., 2017] that uses a discriminant analysis approach on Riemannian manifold for dimensionality reduction in image and video classification. Consider $f(W, P) : \mathcal{S}_{++}^{N_c} \mapsto \mathcal{S}_{++}^{m}$ a generic mapping that maps a SPD matrix P from the original dimension N_c to a lower dimension m defined as $f(W, P) = W^T P W$. The spatial filters W are computed to preserve as much as possible the intra-class compactness (intra-class distances) while maximizing the inter-class distances, by solving the following optimization problem

$$W = \arg \min_{W^T W = I_m} \mathcal{L}(W) \quad where \quad \mathcal{L}(W) = \frac{\widetilde{S}_w(W)}{\widetilde{S}_b(W)} \tag{6.5}$$

where $\widetilde{S}_w(W)$ and $\widetilde{S}_b(W)$ are the within-class and between-class cost functions formulated as follows [Nguyen and Artemiadis, 2018]

$$\widetilde{S}_w(W) = \frac{1}{2} \sum_{i,j \neq i}^{K} g_w(P_i, P_j) \delta^2(f(W, P_i), f(W, P_j)) \tag{6.6}$$

$$\widetilde{S}_b(W) = \frac{1}{2} \sum_{i,j \neq i}^{K} g_b(P_i, P_j) \delta^2(f(W, P_i), f(W, P_j)) \tag{6.7}$$

where $\delta(f(W,P_i), f(W,P_j))$ denotes the Riemannian distance, i and j are the trials, K is the number of all trials, $g_w(P_i,P_j)$ and $g_b(P_i,P_j)$ are the within and between class similarity to compute the affinity between SPD matrices. g_w and g_b can be defined using several methods such as Fisher Discriminant Analysis, Local scaling (LS), and k-Nearest Neighbor. Here we used the LS approach [Sugiyama, 2007], given by

$$g_w(P_i,P_j) = \begin{cases} \frac{A(P_i,P_j)}{K_c}, & \text{if } P_i \ and \ P_j \in same \ class \\ 0, & \text{if } P_i \ and \ P_j \notin same \ class \end{cases} \tag{6.8}$$

$$g_b(P_i,P_j) = \begin{cases} A(P_i,P_j) \times (\frac{1}{K} - \frac{1}{K_c}), & \text{if } P_i \ and \ P_j \in same \ class \\ \frac{1}{K}, & \text{if } P_i \ and \ P_j \notin same \ class \end{cases} \tag{6.9}$$

where K_c is the number of trials in class c, $A(P_i,P_j)$ is the affinity matrix defined as

$$A(P_i,P_j) = exp(-\frac{\delta^2(P_i,P_j)}{\rho_i \times \rho_j}) \quad with \quad \rho_i = \delta(P_i, P_i^{(k)}) \tag{6.10}$$

where $P_i^{(k)}$ is the k-th nearest neighbor of P_i, and $k = 7$ as recommended in [Sugiyama, 2007]. The AIRM is used to compute the similarity between SPD matrices and to define the nearest neighbours.

The minimization problem (6.5) is an optimization problem on a Grassmann manifold, and as proposed in [Harandi et al., 2017], it can be solved by the conjugate gradient descent method. We need to compute the gradient $\nabla_W \mathcal{L}(W)$ on the manifold

$$\nabla_W \mathcal{L}(W) = (I_{Nc} - WW^T) \sum_{i,j \neq i}^{K} \frac{\partial \mathcal{L}(W)}{\partial \delta^2(P_i,P_j)} \frac{\partial \delta^2(P_i,P_j)}{\partial W} \tag{6.11}$$

From [Harandi et al., 2017] the Jacobian $\frac{\partial \delta^2}{\partial W}$ of the squared distance of the AIRM metric is given by

$$\frac{\partial \delta^2(W^T P_i W, W^T P_j W)}{\partial W} = 4\Big(P_i W (W^T P_i W)^{-1} - P_j W (W^T P_j W)^{-1}\Big) \times log\Big(W^T P_i W P_j W (W^T P_j W)^{-1}\Big) \tag{6.12}$$

This optimization problem was solved through the Manopt toolbox [Boumal et al., 2014], which provides several techniques for solving the optimization over various Riemannian manifolds. The solution W finds the lower-dimensional subspace, but it does not identify the individual directions of each spatial filter. Thus, to obtain the directions of each spatial filter, in the last step of the algorithm, we solve the generalized eigenvalue decomposition problem within the detected subspace. That is, we calculate the $S_b' = W^T S_b W$ and $S_w' = W^T S_w W$ and compute the eigenvectors $V' = eig(S_b', S_w')$ sorted according to the largest eigenvalue. The final optimum spatial filters are defined as

$$V_f^{Nc \times m} = W \times V' \tag{6.13}$$

Algorithm 2 Riemannian Fisher Criterion Beamformer algorithm.

Segment EEG into epochs, perform the pre-processing and define the label of epochs

1: **Begin**
2: X_i defines segmented epoch $i = 1 \ ... \ K$, where K is the number of all trials
3: Compute the covariance matrices P_i of each trials using eq. (3.8)
4: Compute the affinity matrix $A(P_i, P_j)$ using eq. (6.10)
5: Compute the the within and between class similarity ($g_w(P_i, P_j)$ and $g_b(P_i, P_j)$) using eq. (6.8 and 6.9)
6: Compute the within-class and between-class cost functions ($\widetilde{S}_w(W)$ and $\widetilde{S}_b(W)$) using eq. (6.6 and 6.7)
7: Solve the minimization problem $W = \arg\min_{W^T W = I_m} \mathcal{L}(W) = \frac{\widetilde{S}_w(W)}{\widetilde{S}_b(W)}$ using the conjugate gradient descent method
8: Compute the $S_b^{'} = W^T S_b W$ and $S_w^{'} = W^T S_w W$
9: Solve the generalized eigenvalue decomposition problem $V'^{m \times m} = eig(S_b^{'}, S_w^{'})$
10: Compute the final spatial filters $V_f = WV'$
11: **End**

The main steps of RFCB method are described in Algorithm 2.

6.3.3 Feature vector

After computing the optimal spatial filters (V_f), the spatial filter projection is obtained from

$$Y = V_f^T X \tag{6.14}$$

The first two projections are concatenated resulting in a vector of 512 features. Then, using the R-square correlation method the 200 most discriminative features are selected.

6.4 Datasets

We evaluate the proposed approach using an In-house ErrP dataset (dataset I) [Cruz et al., 2020] and a benchmarking ErrP dataset (dataset II) [Chavarriaga and Millán, 2010], described below.

6.4.1 In-house dataset (dataset I)

The dataset was acquired during a P300-speller task performed by seven able-bodied participants, and one tetraplegic participant (P1) with medullar injury (C4/C5 level). Twelve EEG channels were recorded at 256Hz. This dataset is available online in [Cruz et al., 2020], for details see appendix A. The experiment comprised two sessions with three phases. Session 1 included P300 calibration (phase 1) and ErrP calibration (phase 2), and session 2 was the final P300-ErrP system (phase 3). These two

sessions were held on different days, except for participant S9, who made the two sessions on the same day. Here, we only used the ErrP calibration data (session 1) and the final P300-ErrP data (session 2). Session 1 has on average 304 trials and the number of errors varied across participants with the minimum value of 31 and the maximum of 86. Session 2 has on average 176 trials and the minimum number of errors was 8 and the maximum was 40. The EEG data was preprocessed with the same framework used in [Cruz et al., 2018a]. More precisely, the signals were filtered using a 50 Hz notch filter and a band-pass filter with a lower cutoff frequency of 1 Hz and a higher cutoff frequency of 10 Hz.

6.4.2 Benchmark dataset (dataset II)

This dataset was gathered while users were observing the performance of an external agent, but without having any control over the agent [Chavarriaga and Millán, 2010]. The EEG signal was recorded from 6 different subjects with 64 electrodes, and at a sampling rate of 512 Hz. The experiment contains two sessions conducted several weeks apart. Participants monitored a moving cursor to reach the target location. Each session consisted of 10 blocks of 3 minutes. More details about this dataset can be found in [Chavarriaga and Millán, 2010]. Here, the data were downsampled at 256 Hz and the same twelve EEG channels of our in-house dataset were used, which are Fz, Cz, C3, C4, CPz, Pz, P3, P4, PO7, PO8, POz, and Oz.

6.5 Results: dataset I

In this section, we present the results of BCI-Double-ErrP-Dataset [Cruz et al., 2020]. For a comparative analysis of generalization, three methods are evaluated: 1) the proposed RFCB; 2) FCB used in [Cruz et al., 2018a]; and 3) the Tangent space spatial filter (TSSF) proposed in [Xu et al., 2020]. In contrast with our spatial filter approach that obtains the optimal filter matrix from the lower-dimensional SPD manifold, the authors in [Xu et al., 2020] obtain the optimal filter on the tangent space and mapped back onto the manifold (see implementation details in appendix B). We evaluate the methods in three scenarios:

- **Intra-session cross-validation.** For each session the classification is done through cross-validation.
- **Cross-session generalization.** Session 1 was used as training data and session 2 as test data (TrS1-TeS2), and vice versa (TrS2-TeS1), session 1 and session 2 were recorded in different days.
- **Cross-subject generalization.** The empirical results showed that cross-subject generalization can be affected by both variabilities between subjects and bad performance of the training subject. Thus, to avoid the effect of bad performance we used as a training set, subjects with great cross-validation classification, and good generalization between sessions. In order to evaluate the

Table 6.1: Cross-validation classification accuracy using session 1 with different values of m

	Lower SPD Dimension (parameter m)									
Subjects	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
S1	86.8	86.5	85.8	86.2	88.5	87.1	88.5	89.1	88.5	88.5
S2	81.4	73.9	75.8	77.4	78.4	78.2	80.5	76.8	76.2	77.5
S3	93.1	90.8	91.4	92.7	92.7	93.9	93.0	94.6	94.4	94.0
S4	80.7	80.0	81.7	82.2	80.2	80.3	80.0	82.8	82.3	84.1
S5	67.0	69.3	74.6	73.6	71.8	78.1	77.9	80.4	79.9	79.9
S6	68.0	78.5	74.4	80.0	75.7	76.8	78.8	76.4	76.2	76.4
S9	92.3	94.2	91.2	87.7	93.2	92.7	93.6	95.2	94.5	97.3
P1	52.8	72.5	59.0	67.2	62.5	68.1	68.0	70.6	68.2	68.7
Mean	**77.8**	**80.7**	**79.2**	**80.9**	**80.4**	**81.9**	**82.5**	**83.2**	**82.5**	**83.3**

consistency of the results, we selected two subjects with these properties (S3 and S9). Cross-session results showed that the best results are obtained when session 1 was used as the training set and session 2 as the test set, therefore, we used the data of session 1 of subject S3 or subject S9 as a training set and the data of session 2 of the other subjects as a test set.

For RFCB the within and between class similarity (eq. 6.8 and eq. 6.9) was defined with $K_w = 3$ and $K_b = 1$. For cross-session generalization, the dimensionality of the lower SPD manifold ($\mathcal{S}_{++}^{m}$) was obtained by cross-validation. For cross-subject generalization, we selected the parameter m that produced the best results (excluding the test subject). In all methods, the feature vector was classified with a Bayes classifier.

6.5.1 Dimensionality selection of the lower SPD manifold

The $RFCB$ algorithm finds the optimal filter $V_f^{N_c \times m}$ with the lower dimension (more discriminative) than the original manifold ($N_c = 12$), that preserves the original SPD structure. To analyze the effect of the dimensionality of the lower SPD manifold (parameter m), we performed the classification of session 1 using cross-validation with a different value of m. The parameter m was varied within the interval [2 : $N_c - 1$]. The results presented in Table 6.1 show that the maximum mean accuracy is 83.3% obtained for $m = 11$. The performance is very sensitive to parameter m. For example, in participants S5, S6, and P1 the difference between the maximum accuracy and minimum accuracy is greater than 10%.

6.5.2 Intra-session Performance

In this section, we present the results of intra-session analyses using dataset I [Cruz et al., 2020].

The classification performance of the $RFCB$, FCB, and $TSSF$, using intra-session data (the data of each session is classified through cross-validation) are presented in Table 6.2. The results showed that the mean classification accuracy was respectively 84.9%, 83.6%, and 72.3% for session 1, and 79.2%,

Table 6.2: Offline classification accuracy for intra-session cross-validation data

	RFCB		FCB		TSSF	
Subjects	**Session1**	**Session2**	**Session1**	**Session2**	**Session1**	**Session2**
S1	89.1	94.8	88.5	93.8	80.2	78.3
S2	81.4	72.2	77.1	68.4	69.1	75.5
S3	94.6	95.6	93.0	95.9	81.3	76.8
S4	84.1	80.5	84.9	78.8	69.7	60.8
S5	80.4	72.9	82.1	68.2	72.6	56.8
S6	80.0	54.3	78.9	49.2	60.7	52.7
S9	97.3	98.0	97.1	98.0	62.8	83.9
P1	72.5	65.4	67.2	65.4	81.8	56.1
Mean	**84.9**	**79.2**	**83.6**	**77.2**	**72.3**	**67.6**

77.2%, and 67.6% for session 2. The mean classification accuracy obtained for data of session 2 was lower by 5.7%, 6.4%, and 4.7% respectively. This drop is likely due to the difference in the number of error samples (Session 2 has a much less number of error training samples). When session 1 was used as a training set the $RFCB$ and FCB had similar performance and $TSSF$ had a lower performance. When session 2 was used as training set the mean accuracy of $RFCB$ was 2.0% and 11.6% higher compared with FCB, and $TSSF$ respectively (paired t-test, p = 0.019, and p = 0.006), i.e. $RFCB$ was less sensitive to the amount of training samples.

6.5.3 Cross-session Performance

As verified in section 6.5.1 the performance of the $RFCB$ algorithm is affected by the value of the lower SPD manifold dimension. Therefore, for cross-session we selected m following two approaches: 1) for each participant we select the dimension m that produces the best performance, referred to as $RFCB_{Best}$, and 2) parameter m was set automatically by cross-validation, referred to as $RFCB_{Aut}$. $RFCB_{Best}$ is a biased approach as it uses the tests results to tune the parameter. It was used as a reference to evaluate the performance of $RFCB_{Aut}$.

For cross-session data, the $RFCB$ method performed better than the other two methods as shown in Table 6.3. The mean classification accuracy of $RFCB_{Best}$ is very similar for both sessions with a value around 85%, meaning that the variability across sessions and the lack of error samples do not affect the performance. The $RFCB_{Aut}$ approach has the mean classification accuracy value of 82.3%, and 78.6% for session 2 and session 1, respectively. There is a decrease of about 4%, however, this value is less than the decreases obtained using FCB and $TSSF$, which is about 8% and 5%, respectively. The FCB method presented better results than $TSSF$ with a mean classification accuracy of 80.6% for session 2 and 72.7% for session 1. These results show that the proposed $RFCB_{Best}$ approach has better generalization across sessions and it is less sensitive to the variation of the number of error training samples than FCB and $TSSF$ methods. The paired t-test show that the difference for session

Table 6.3: Offline classification accuracy for cross-session data using session 1 as training data and session 2 as test data (TrS1-TeS2), and using session 2 as training data and session 1 as test data (TrS2-TeS1)

	RFCB_Best		RFCB_Aut		FCB		TSSF	
	TrS1-TeS2	TrS2-TeS1	TrS1-TeS2	TrS2-TeS1	TrS1-TeS2	TrS2-TeS1	TrS1-TeS2	TrS2-TeS1
S1	93.4	81.1	92.0	81.1	93.4	85.4	79.6	82.5
S2	93.2	83.4	85.5	77.5	89.9	57.7	69.9	62.3
S3	96.6	83.8	96.3	83.8	95.8	87.6	78.2	75.8
S4	79.3	84.4	78.5	84.2	80.1	77.3	65.0	62.9
S5	72.2	85.2	70.6	81.4	66.3	70.3	57.1	51.4
S6	71.2	86.2	71.2	82.7	56.1	58.2	45.9	48.9
S9	94.9	85.6	93.1	79.6	93.1	82.7	85.9	61.7
P1	76.8	86.6	71.1	58.1	69.7	62.6	58.8	53.9
Mean	**84.7**	**84.5**	**82.3**	**78.6**	**80.6**	**72.7**	**67.6**	**62.4**

Table 6.4: Classification accuracy for cross-subject data using subjects S9 and S3 as training data

	Subjects	RFCB_Best	RFCB_Aut	FCB	TSSF
Train S9	S1	88.4	88.4	84.7	66.0
	S2	81.6	74.6	68.3	53.3
	S3	93.2	92.1	89.8	60.2
	S4	82.4	82.4	78.8	58.3
	S5	63.6	63.6	58.1	55.2
	S6	66.9	63.4	66.3	56.9
	P1	60.5	60.5	60.1	50.0
	Mean	**76.7**	**75.0**	**72.3**	**57.1**
Train S3	S1	87.8	87.8	70.0	60.1
	S2	72.8	72.8	65.9	55.2
	S4	78.5	78.5	73.0	44.9
	S5	70.6	70.6	64.1	54.6
	S6	76.2	76.2	58.6	45.0
	S9	82.9	76.9	79.8	38.9
	P1	60.1	60.1	60.1	50.9
	Mean	**75.6**	**74.7**	**67.4**	**49.9**

2 (paired t-test, p = 0.029, and p = 0.032) and session 1 (paired t-test, p = 0.020, and p = 0.019) are statistically significant.

6.5.4 Cross-subject Performance

For cross-subject data, in $RFCB_{Aut}$ approach, the dimension m was the one that produced the best results without the test subject.

The results reported in Table 6.4 demonstrated that for $RFCB_{Best}$ and $RFCB_{Aut}$ approaches the mean accuracies are almost the same when subject S3 or subject S9 is used as a training set, attaining values of 76.7%, 75.6%, 75.0%, and 74.7% respectively. For FCB using subject S9 as training set

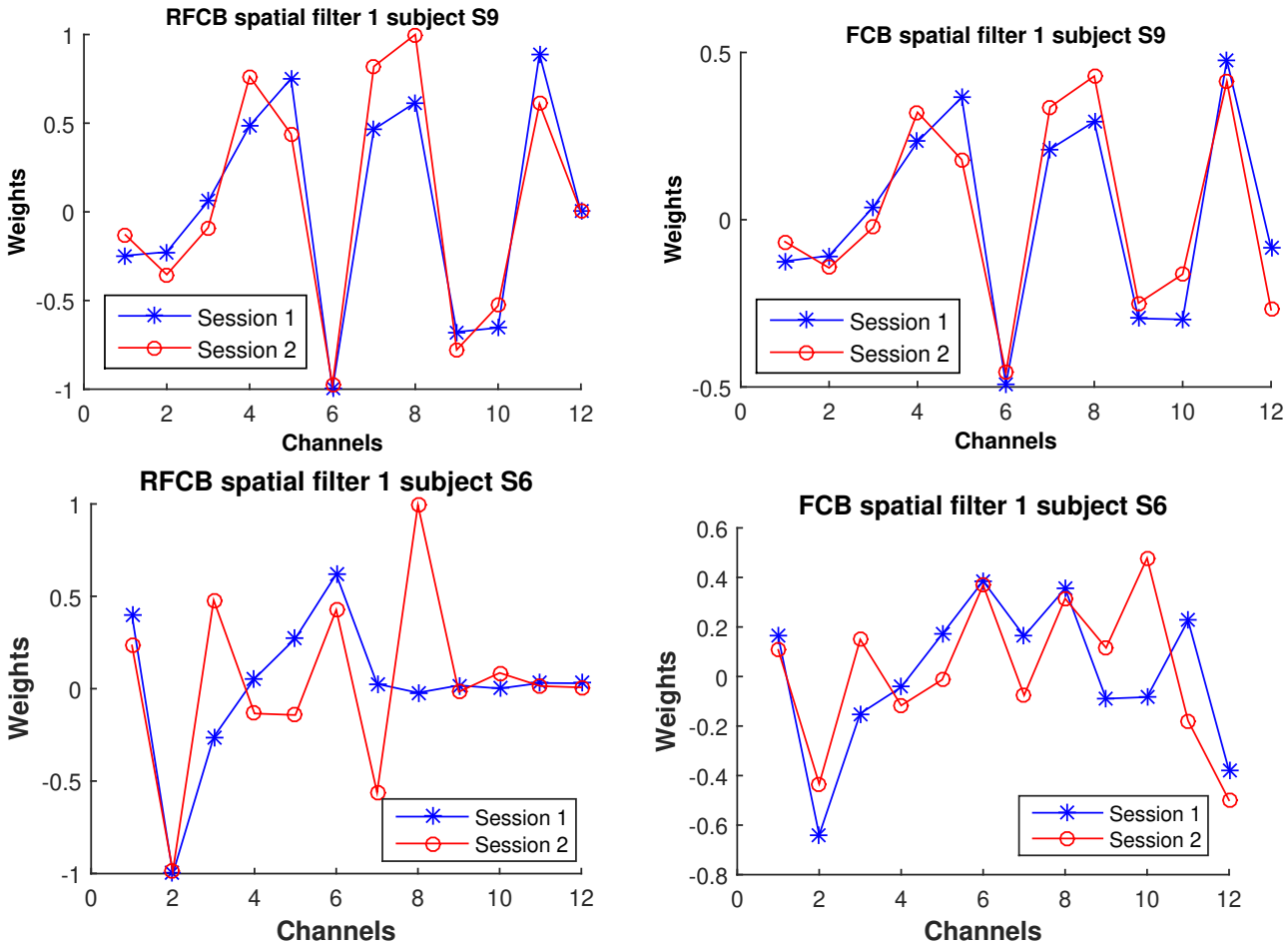


Figure 6.1: The coefficients of the first spatial filters estimated using RFCB method (left), and FCB method (right) for subject S9 (with the best generalization), and subject S6 (with the worse generalization) using session 1 and session 2 data from all 12 channels.

produced an accuracy about 5% higher than subject S3. $TSSF$ method has poor generalization across subjects. Using the data of subject S9 or subject S3 as training set $RFCB_{Aut}$ presented statistically better results than FCB and $TSSF$ methods (paired t-test, p = 0.033, p = 0.025, p = 0.001, and p = 0.0004). When compared to user dependent model, $RFCB_{Aut}$ showed a decrease in classification accuracy of about 6%.

6.5.5 Robustness of the spatial filter coefficients

To analyze the spatial filter coefficients robustness across sessions and across subjects, we compared the coefficient index obtained using $RFCB$ and FCB methods. For cross-session generalization the analysis was made for two subjects, subject S9 who had the best generalization, and subject S6 who had the worse generalization, using the data gathered in session 1 and the data from session 2. Figure

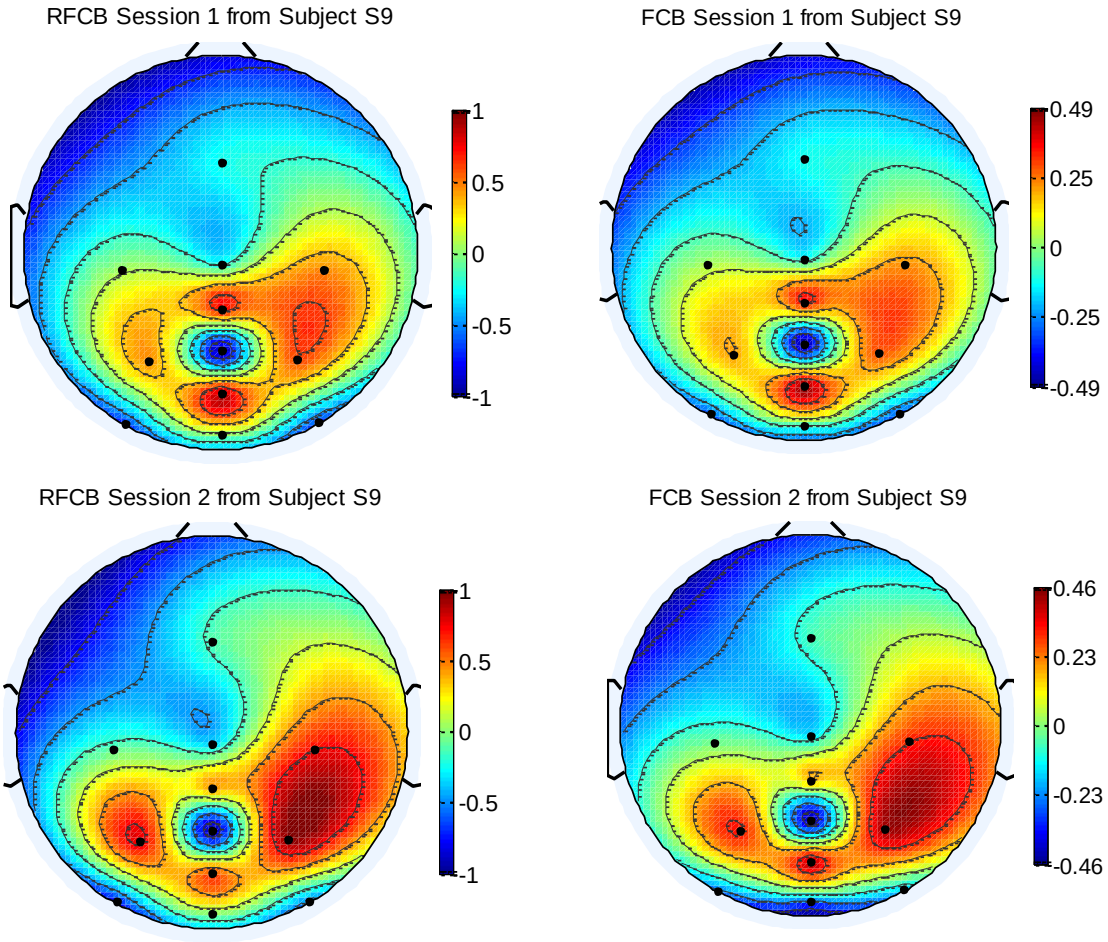


Figure 6.2: Scalp topography of the first spatial filters estimated using $RFCB$ method (left), and FCB method (right) for subject S9 with data from session 1 and session 2.

6.1 shows the weights of the first spatial filters estimated using the data of each session, and Figure 6.2 presents the interpolation of the coefficients indexes at electrode positions. When using subject S9, the spatial filter coefficients obtained from the two sessions are very similar for both $RFCB$ and FCB methods (correlation coefficient, r=0.935, and r=0.927, for $RFCB_{Best}$, and FCB respectively). Although the spatial filter coefficients between sessions are similar, there is a variability in the classification accuracy between sessions of about 9% and 15% for $RFCB_{Best}$, and FCB respectively (Table 6.3). For subject S6, the correlation between the coefficient indexes decreased for both methods (correlation coefficient, r=0.621, and r=0.616, for $RFCB_{Best}$, and FCB respectively), and there is again a variability in the classification accuracy between sessions of 15% and 27% for $RFCB_{Best}$, and FCB respectively (Table 6.3). Despite the great variability of classification between sessions for subject S6 (table 6.3), $RFCB_{Best}$ proved to be much more effective than FCB. Scalp topography shows that coefficients indexes with the greatest values are at the centroparietal region for both methods.

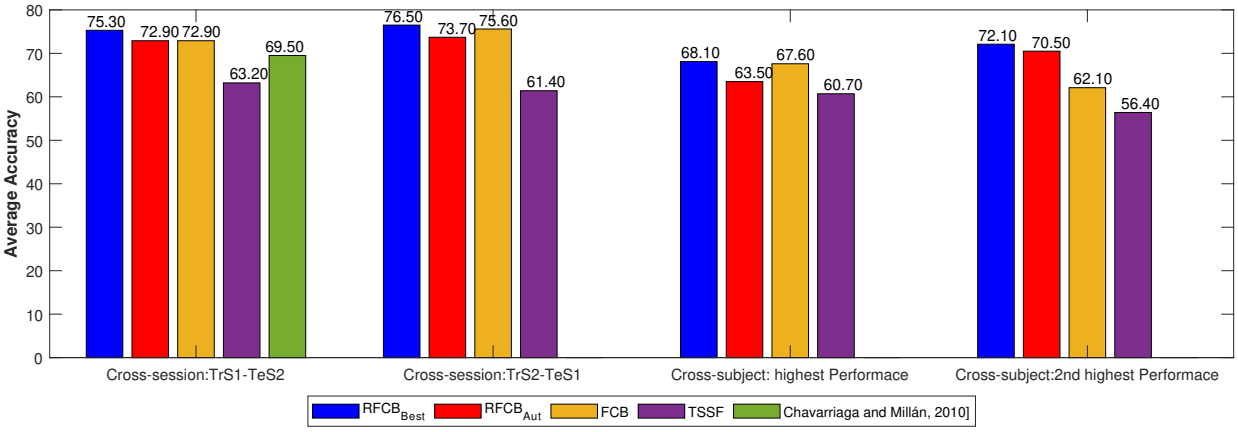


Figure 6.3: The classification accuracy using a benchmark dataset [Chavarriaga and Millán, 2010].

6.6 Results: dataset II

In this section, we present the results obtained with dataset II (monitoring error-related potentials dataset [Chavarriaga and Millán, 2010]). We evaluated the cross-session and cross-subject generalization. Figure 6.3 shows the mean accuracy of all subjects. Regarding to cross-session data, the mean classification accuracy of both sessions were 75.9%, 73.3%, 74.3%, and 62.3% for $RFCB_{Best}$, $RFCB_{Aut}$, FCB, and $TSSF$ respectively. There was a low performance variability across-session with average values of 1.3%, 0.6%, 2.7%, and 1.8% respectively. These results showed that the $RFCB_{Best}$ method was less sensitive to the variation between sessions and it had a slightly higher classification accuracy.

For cross-subject generalization, results demonstrated high classification variability when the subject with highest or second-highest performance (combining accuracy and generalization) was used as the training set, and $RFCB$ approaches achieved higher classification accuracy when subject with second-highest performance was used as the training set. In order to assess whether the defined criteria for selecting the subject to be used as training set produced the best results, we used the data of other subjects as training set. The best results were obtained using the subject with highest performance, second-highest performance, and fourth performance for $TSSF$, $RFCB$, and FCB methods. Therefore, in the future other criterion to select the training set should be explored.

6.7 Discussion and conclusion

In this study, we proposed a new spatial filter method based on Riemannian geometry that uses the invariance properties of Riemannian distance to deal with cross-session and cross-subject variability. The robustness of the $RFCB$ method was assessed in a cross-session and cross-subject classification

problem using an in-house dataset and a benchmark dataset. $RFCB_{Best}$ approach achieved a stable classification accuracy across-session for both datasets with values about 85% and 76% respectively.

Regarding the in-house dataset results, when session1 is used as a training set the cross-session average classification accuracy is similar to the one obtained from cross-validation and when session2 is used as a training set the cross-session accuracy is 5.4% higher than the one obtained from cross-validation. These results show the robustness of the $RFCB_{Best}$ algorithm to the changes occurring between sessions and to the lack of error training samples. For the $RFCB_{Aut}$ approach we verified similar results with a small decrease in performance. Comparing cross-subject vs cross-validation, for $RFCB_{Best}$ and $RFCB_{Aut}$ approaches we verified that the average classification accuracy is about 9% and less 6% and they have similar performance when using either subject S3 or subject S9 as a training set.

The $RFCB$ method also produced better results in the benchmark dataset, and the classification accuracy is higher (5.8%) than the one achieved in the original work [Chavarriaga and Millán, 2010], this strengthens the effectiveness of the proposed method.

$RFCB_{Best}$ had higher classification accuracy than $RFCB_{Aut}$ in both datasets, and the performance was statistically better in cross-session data, except when using session 2 as training set and session 1 as test set (paired t-test, p = 0.058) using the in-house dataset, thus in future research, we will explore others ways to automatically define the parameter m.

The results showed that the session used as a training set had an impact on classification performance, therefore, further analysis would be done to understand which characteristics should have a training dataset to achieve good performance.

To explore the impact of methods used to define the distance between two SPD matrices we made a further analysis comparing several divergence methods. Figure 6.4 compares the results obtained

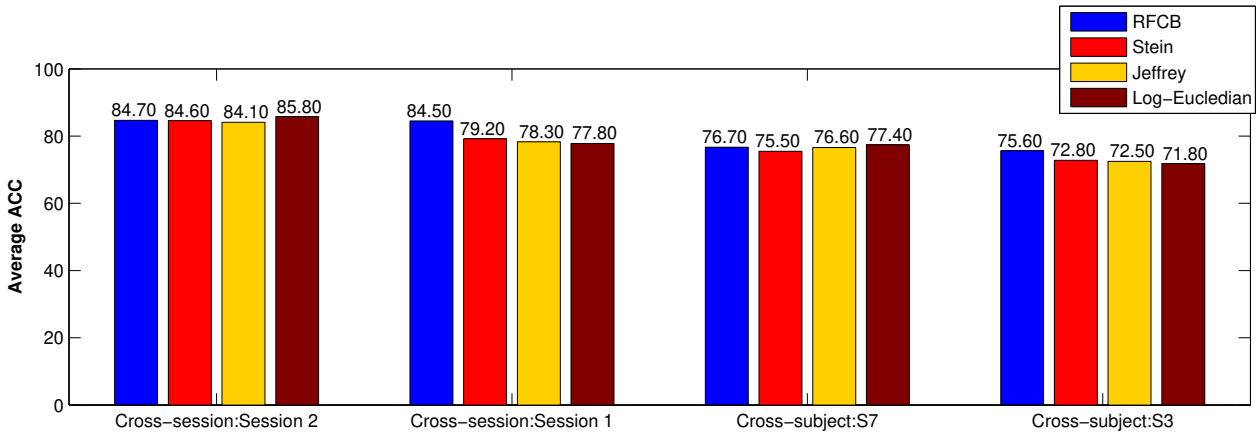


Figure 6.4: The classification accuracy of $RFCB$ algorithm using Riemannian metric, Stein divergence, Jeffrey divergence, and log-euclidean distance.

with Riemaniann metric, Stein divergence [Sra, 2011], Jeffrey divergence [Cherian et al., 2012] and log-euclidean distance [Arsigny et al., 2006]. Results showed that the $RFCB$ algorithm with Riemannian metric had higher performance than the other methods in cross-session analysis when session 2 is used as a training set, and in cross-subject analysis when subject S3 is used as a training set. The difference between AIRM and the log-euclidean distance is statistically significant (paired t-test, p = 0.039, and p = 0.027) and for Stein and Jeffrey divergence the difference approach the significance (paired t-test, p = 0.083, p = 0.068, p = 0.077, and p = 0.083). This means that, $RFCB$ using Riemannian distance is more robust than when used other divergence methods with respect to the variability of the number of error samples in the data training and the variability of the performance of the subject used as a training set.

7

Conclusion

Contents

7.1 Summary of Thesis Contributions

The development of a reliable and practical BCI system is the subject of considerable interest for many BCI researchers. In this thesis, we proposed approaches for improving the reliability and usability of BCI systems. More precisely:

- **Double ErrP detection.** A new error detection and correction system based on double ErrP detection was proposed. Moreover, the ERP elicited by the feedback and the ERP detected as a target were concatenated at the feature level leading to a simple data-augmentation method which proved to be effective. The approach combining the double-ErrP detection and data augmentation was validated with able-bodied participants and a tetraplegic participant, showing a significant increase in online accuracy, information transfer rate, and effective symbols per minute. The results achieved in chapter 4 were above the state-of-the-art, and they showed that the error monitoring process is crucial to enhance the human-machine interaction, for example using BCI systems.

The effective use of BCI in real-world scenarios still depends on key issues related to the usability aspect. In this thesis two key aspects were addressed:

- **Self-paced control.** We proposed a new approach that integrates collaborative control, self-paced control and dynamic-time commands into a BCW system. The proposed self-paced control is natural without requiring any additional tasks (mental or physical) to detect the control state. At the same time, the approach also tunes the rate of false positives. The dynamic time-window approach balances the performance and speed of the BCI, and the collaborative controller validated the user's intentions. The proposed approaches proved to be extremely effective with overall accuracy near 100% for both able-bodied and motor disabled participants.

- **The calibration process.** Usually, before each session a calibration phase is performed to build the classification model, limiting the use of BCI. A new spatial filter method based on Riemannian geometry was proposed that allows to re-use the data from previous sessions or other subjects to predict new samples. The results evidenced that, the proposed method is not strongly affected by the cross-session variability and by the low number of samples in the training dataset. This is an important contribution to make the BCI system easier to use and more practical.

7.2 Future Improvements

The approaches described in this thesis presented encouraging results, however, there are still several ways to improve them. Regarding ErrP applications, we applied ErrP during a P300-speller task which has discrete events and we knew when the error occurs, future work could be the use of ErrP in a

continuous task, where we do not know the precise moment that the error occurs. The use of ErrP to improve the systems' reliability depends on the accuracy of ErrP detection and correction rate. We achieved a high ErrP detection accuracy, but the correction rate is low. Thus, further improvements in the current system could be obtained by researching new approaches to increase the correction rate.

Users steered the wheelchair in office-like environments requiring challenging tasks, such as narrow door passages and obstacle (static and dynamic) avoidance. Although complex, the scenarios were still very controlled and different from users' daily home settings. Therefore, further experiments in more natural living contexts are needed to validate the approaches.

In the $RFCB$ method, we obtain the spatial filter with the data from previous sessions or other subjects. We suggest that in the future explore the unsupervised $RFCB$ algorithm that the filter is adapted with the data of the current session/subject. The low dimensionality of the new SDP matrix is an important parameter that needs to be tuned. Therefore, it would be worth exploring other methods to choose the best dimensionality.

Bibliography

[Absil et al., 2009] Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

[Ahn et al., 2014] Ahn, M., Lee, M., Choi, J., and Jun, S. C. (2014). A review of brain-computer interface games and an opinion survey from researchers, developers and users. *Sensors*, 14(8):14601–14633.

[Alamgir et al., 2010] Alamgir, M., Grosse-Wentrup, M., and Altun, Y. (2010). Multitask learning for brain-computer interfaces. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 17–24.

[Alexander and Brown, 2010] Alexander, W. H. and Brown, J. W. (2010). Computational models of performance monitoring and cognitive control. *Topics in cognitive science*, 2(4):658–677.

[Amaral et al., 2018] Amaral, C., Mouga, S., Simões, M., Pereira, H. C., Bernardino, I., Quental, H., Playle, R., McNamara, R., Oliveira, G., and Castelo-Branco, M. (2018). A feasibility clinical trial to improve social attention in autistic spectrum disorder (asd) using a brain computer interface. *Frontiers in neuroscience*, 12:477.

[Arsigny et al., 2006] Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2006). Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 56(2):411–421.

[Arvaneh et al., 2014] Arvaneh, M., Robertson, I., and Ward, T. E. (2014). Subject-to-subject adaptation to reduce calibration time in motor imagery-based brain-computer interface. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6501–6504. IEEE.

[Barachant et al., 2010] Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2010). Riemannian geometry applied to bci classification. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 629–636. Springer.

[Barachant et al., 2011] Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2011). Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928.

[Barbosa et al., 2016] Barbosa, S., Pires, G., and Nunes, U. (2016). Toward a reliable gaze-independent hybrid bci combining visual and natural auditory stimuli. *Journal of neuroscience methods*, 261:47–61.

[BCI-Competition, 2015] BCI-Competition (2015). Bci challenge @ ner 2015. https://www.kaggle.com/c/inria-bci-challenge.

[Beck et al., 1961] Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. (1961). An inventory for measuring depression. *Archives of general psychiatry*, 4(6):561–571.

[Berger, 1929] Berger, H. (1929). Über das elektroenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, 87(1):527–570.

[Bernstein et al., 1995] Bernstein, P. S., Scheffers, M. K., and Coles, M. G. (1995). " where did i go wrong?" a psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1312.

[Bi et al., 2013] Bi, L., Fan, X. A., and Liu, Y. (2013). EEG-based brain-controlled mobile robots: A survey. *IEEE Transactions on Human-Machine Systems*, 43(2):161–176.

[Birbaumer et al., 2000] Birbaumer, N., Kübler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser, J., Iversen, I., Kotchoubey, B., Neumann, N., and Flor, H. (2000). The thought translation device (TTD) for completely paralyzed patients. *IEEE Transactions on Rehabilitation Engineering*.

[Bishop, 2006] Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

[Blankertz et al., 2011] Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of erp components—a tutorial. *NeuroImage*, 56(2):814–825.

[Blankertz et al., 2010] Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L. E., Sturm, I., Curio, G., et al. (2010). The berlin brain–computer interface: non-medical uses of bci technology. *Frontiers in neuroscience*, 4:198.

[Botvinick et al., 1999] Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., and Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, 402(6758):179.

[Boucsein, 2012] Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

[Boumal et al., 2014] Boumal, N., Mishra, B., Absil, P.-A., and Sepulchre, R. (2014). Manopt, a matlab toolbox for optimization on manifolds. *The Journal of Machine Learning Research*, 15(1):1455–1459.

[Brunner et al., 2015] Brunner, C., Birbaumer, N., Blankertz, B., Guger, C., Kübler, A., Mattia, D., Millán, J. d. R., Miralles, F., Nijholt, A., Opisso, E., et al. (2015). Bnci horizon 2020: towards a roadmap for the bci community. *Brain-computer interfaces*, 2(1):1–10.

[Brunner et al., 2006] Brunner, C., Scherer, R., Graimann, B., Supp, G., and Pfurtscheller, G. (2006). Online control of a brain-computer interface using phase synchronization. *IEEE Transactions on Biomedical Engineering*, 53(12):2501–2506.

[Carlson and Demiris, 2012] Carlson, T. and Demiris, Y. (2012). Collaborative control for a robotic wheelchair: evaluation of performance, attention, and workload. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(3):876–888.

[Carlson and Millan, 2013] Carlson, T. and Millan, J. d. R. (2013). Brain-controlled wheelchairs: a robotic architecture. *IEEE Robot Autom Mag*, 20(1):65–73.

[Carter et al., 1998] Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., and Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280(5364):747–749.

[Carter et al., 2000] Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., and Cohen, J. D. (2000). Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, 97(4):1944–1948.

[Castermans et al., 2014] Castermans, T., Duvinage, M., Cheron, G., and Dutoit, T. (2014). Towards effective non-invasive brain-computer interfaces dedicated to gait rehabilitation systems. *Brain sciences*, 4(1):1–48.

[Chapin et al., 1999] Chapin, J. K., Moxon, K. A., Markowitz, R. S., and Nicolelis, M. A. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature neuroscience*, 2(7):664.

[Chavarriaga and Millán, 2010] Chavarriaga, R. and Millán, J. d. R. (2010). Learning from EEG error-related potentials in noninvasive brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(4):381–388.

[Chavarriaga et al., 2014] Chavarriaga, R., Sobolewski, A., and Millán, J. d. R. (2014). Errare machinale est: The use of error-related potentials in brain-machine interfaces. *Frontiers in Neuroscience*, 8(8 JUL):1–13.

[Chen et al., 2015] Chen, X., Wang, Y., Gao, S., Jung, T.-P., and Gao, X. (2015). Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *Journal of Neural Engineering*.

[Cherian et al., 2011] 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.

[Cherian et al., 2012] Cherian, A., Sra, S., Banerjee, A., and Papanikolopoulos, N. (2012). Jensen-bregman logdet divergence with application to efficient similarity search for covariance matrices. *IEEE transactions on pattern analysis and machine intelligence*, 35(9):2161–2174.

[Choi, 2012] Choi, K. (2012). Control of a vehicle with eeg signals in real-time and system evaluation. *European journal of applied physiology*, 112(2):755–766.

[Christ et al., 2000] Christ, S., Falkenstein, M., Heuer, H., and Hohnsbein, J. (2000). Different error types and error processing in spatial stimulus-response-compatibility tasks: behavioural and electrophysiological data. *Biological psychology*, 51(2-3):129–150.

[Cincotti et al., 2008] Cincotti, F., Mattia, D., Aloise, F., Bufalari, S., Schalk, G., Oriolo, G., Cherubini, A., Marciani, M. G., and Babiloni, F. (2008). Non-invasive brain–computer interface system: towards its application as assistive technology. *Brain research bulletin*, 75(6):796–803.

[Coles et al., 2001] Coles, M. G., Scheffers, M. K., and Holroyd, C. B. (2001). Why is there an ern/ne on correct trials? response representations, stimulus-related components, and the theory of error-processing. *Biological psychology*, 56(3):173–189.

[Combaz et al., 2012] Combaz, A., Chumerin, N., Manyakov, N., Robben, A., Suykens, J., and Van Hulle, M. (2012). Towards the detection of error-related potentials and its integration in the context of a P300 speller brain–computer interface. *Neurocomputing*, 80:73–82.

[Comerchero and Polich, 1999] Comerchero, M. D. and Polich, J. (1999). P3a and p3b from typical auditory and visual stimuli. *Clinical neurophysiology*, 110(1):24–30.

[Congedo et al., 2017] Congedo, M., Barachant, A., and Bhatia, R. (2017). Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4(3):155–174.

[Congedo et al., 2011] Congedo, M., Goyat, M., Tarrin, N., Ionescu, G., Varnet, L., Rivet, B., Phlypo, R., Jrad, N., Acquadro, M., and Jutten, C. (2011). " brain invaders": a prototype of an open-source p300-based video game working with the openvibe platform.

[Cornelia et al., 2013] Cornelia, S., Gravenhorst, F., Schumm, J., Arnrich, B., and Tröster, G. (2013). Towards long term monitoring of electrodermal activity in daily life. *Personal and Ubiq Comp Journ*.

[Corralejo et al., 2014] Corralejo, R., Nicolás-Alonso, L. F., Álvarez, D., and Hornero, R. (2014). A p300-based brain–computer interface aimed at operating electronic devices at home for severely disabled people. *Medical & biological engineering & computing*, 52(10):861–872.

[Coyle et al., 2010] Coyle, D., McGinnity, T. M., and Prasad, G. (2010). Improving the separability of multiple eeg features for a bci by neural-time-series-prediction-preprocessing. *Biomedical Signal Processing and Control*, 5(3):196–204.

[Cruz et al., 2015] Cruz, A., Garcia, D., Pires, G., and Nunes, U. (2015). Facial expression recognition based on eog toward emotion detection for human-robot interaction. In *BIOSIGNALS*, pages 31–37.

[Cruz et al., 2021] Cruz, A., Pires, G., Lopes, A., Carona, C., and Nunes, U. J. (2021). A self-paced bci with a collaborative controller for highly reliable wheelchair driving: Experimental tests with physically disabled individuals. *IEEE Transactions on Human-Machine Systems*, 51(2):109–119.

[Cruz et al., 2019] Cruz, A., Pires, G., Lopes, A. C., and Nunes, U. J. (2019). Detection of stressful situations using gsr while driving a bci-controlled wheelchair. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1651–1656. IEEE.

[Cruz et al., 2018a] Cruz, A., Pires, G., and Nunes, U. J. (2018a). Double errp detection for automatic error correction in an erp-based bci speller. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(1):26–36.

[Cruz et al., 2018b] Cruz, A., Pires, G., and Nunes, U. J. (2018b). Generalization of errp-calibration for different error-rates in p300-based bcis. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 644–649. IEEE.

[Cruz et al., 2020] Cruz, A., Pires, G., and Nunes, U. J. (2020). BCI DOUBLE ERRP DATASET. https://ieee-dataport.org/open-access/error-related-potentials-primary-and-secondary-errp-and-p300-event-related-potentials-. [Online; accessed July-2020].

[Dal Seno et al., 2010] Dal Seno, B., Matteucci, M., and Mainardi, L. (2010). Online detection of P300 and error potentials in a BCI speller. *Computational Intelligence and Neuroscience*, 2010.

[Davies et al., 2003] Davies, A., Souza, L. D., and Frank, A. (2003). Changes in the quality of life in severely disabled people following provision of powered indoor/outdoor chairs. *Disability and Rehabilitation*, 25(6):286–290.

[Dawson et al., 2007] Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). The electrodermal system. *Handbook of psychophysiology*, 2:200–223.

[De Venuto and Mezzina, 2018] De Venuto, D. and Mezzina, G. (2018). User-centered ambient assisted living: Brain environment interface. In *2018 7th Mediterranean Conference on Embedded Computing (MECO)*, pages 1–4. IEEE.

[Dehaene, 2018] Dehaene, S. (2018). The error-related negativity, self-monitoring, and consciousness. *Perspectives on Psychological Science*, 13(2):161–165.

[Devlaminck et al., 2011] Devlaminck, D., Wyns, B., Grosse-Wentrup, M., Otte, G., and Santens, P. (2011). Multisubject learning for common spatial patterns in motor-imagery bci. *Computational intelligence and neuroscience*, 2011:8.

[Di Gregorio et al., 2018] Di Gregorio, F., Maier, M. E., and Steinhauser, M. (2018). Errors can elicit an error positivity in the absence of an error negativity: Evidence for independent systems of human error monitoring. *NeuroImage*, 172:427–436.

[Diez et al., 2013] Diez, P. F., Müller, S. M. T., Mut, V. A., Laciar, E., Avila, E., Bastos-Filho, T. F., and Sarcinelli-Filho, M. (2013). Commanding a robotic wheelchair with a high-frequency steady-state visual evoked potential based brain–computer interface. *Medical engineering & physics*, 35(8):1155–1164.

[Ding and Peng, 2005] Ding, C. and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205.

[Duan et al., 2014] Duan, J., Li, Z., Yang, C., and Xu, P. (2014). Shared control of a brain-actuated intelligent wheelchair. In *Proceeding of the 11th World Congress on Intelligent Control and Automation*, pages 341–346. IEEE.

[Eaton and Miranda, 2014] Eaton, J. and Miranda, E. R. (2014). On mapping eeg information into music. In *Guide to Brain-Computer Music Interfacing*, pages 221–254. Springer.

[Ehlis et al., 2014] Ehlis, A.-C., Schneider, S., Dresler, T., and Fallgatter, A. J. (2014). Application of functional near-infrared spectroscopy in psychiatry. *Neuroimage*, 85:478–488.

[Endrass et al., 2007] Endrass, T., Reuter, B., and Kathmann, N. (2007). Erp correlates of conscious error recognition: aware and unaware errors in an antisaccade task. *European Journal of Neuroscience*, 26(6):1714–1720.

[Fabiani et al., 2004] Fabiani, G. E., McFarland, D. J., Wolpaw, J. R., and Pfurtscheller, G. (2004). Conversion of eeg activity into cursor movement by a brain-computer interface (bci). *IEEE transactions on neural systems and rehabilitation engineering*, 12(3):331–338.

[Falkenstein et al., 1996] Falkenstein, M., Hohnsbein, J., and Hoormann, J. (1996). Differential processing of motor errors. *Recent advances in event-related brain potential research*, pages 579–585.

[Falkenstein et al., 1989] Falkenstein, M., Hohnsbein, J., Hoormann, J., and Blanke, L. (1989). Error processing in choice reaction tasks with focused and crossmodal divided attention. an erp study. In *Ninth International Conference on Event-Related Potentials of the Brain, Noordwijk, The Netherlands*.

[Falkenstein et al., 1991] Falkenstein, M., Hohnsbein, J., Hoormann, J., and Blanke, L. (1991). Effects of crossmodal divided attention on late erp components. ii. error processing in choice reaction tasks. *Electroencephalography and clinical neurophysiology*, 78(6):447–455.

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

[Faller et al., 2017] Faller, J., Allison, B. Z., Brunner, C., Scherer, R., Schmalstieg, D., Pfurtscheller, G., and Neuper, C. (2017). A feasibility study on ssvep-based interaction with motivating and immersive virtual and augmented reality. *arXiv preprint arXiv:1701.03981*.

[Farwell and Donchin, 1988] 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(6):510–523.

[Farwell et al., 2014] Farwell, L. A., Richardson, D. C., Richardson, G. M., and Furedy, J. J. (2014). Brain fingerprinting classification concealed information test detects us navy military medical information with p300. *Frontiers in neuroscience*, 8:410.

[Fernandez-Rodriguez et al., 2016] Fernandez-Rodriguez, A., Velasco-Alvarez, F., and Ron-Angevin, R. (2016). Review of real brain-controlled wheelchairs. *Journal of neural engineering*, 13(6):061001.

[Ferrez and Millán, 2005] Ferrez, P. W. and Millán, J. d. R. (2005). You are wrong!—automatic detection of interaction errors from brain waves. In *Proceedings of the 19th international joint conference on Artificial intelligence*, number CONF.

[Ferrez and Millán, 2008a] Ferrez, P. W. and Millán, J. d. R. (2008a). Error-related eeg potentials generated during simulated brain–computer interaction. *IEEE transactions on biomedical engineering*, 55(3):923–929.

[Ferrez and Millán, 2008b] Ferrez, P. W. and Millán, J. d. R. (2008b). Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy. In *Proceedings of the 4th international brain-computer interface workshop and training course*, number CNBI-CONF-2008-004, pages 197–202.

[Forster and Pavone, 2008] Forster, B. and Pavone, E. F. (2008). Electrophysiological correlates of crossmodal visual distractor congruency effects: evidence for response conflict. *Cognitive, Affective, & Behavioral Neuroscience*, 8(1):65–73.

[Frank et al., 2000] Frank, A., Ward, J., Orwell, N., McCullagh, C., and Belcher, M. (2000). Introduction of a new nhs electric-powered indoor/outdoor chair (epioc) service: benefits, risks and implications for prescribers. *Clinical Rehab.*, 14(6):665–673.

[Friedman et al., 2001] Friedman, D., Cycowicz, Y. M., and Gaeta, H. (2001). The novelty p3: an event-related brain potential (erp) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, 25(4):355–373.

[Friedman et al., 1997] Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine learning*, 29(2-3):131–163.

[Gao et al., 2014] Gao, S., Wang, Y., Gao, X., and Hong, B. (2014). Visual and auditory brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 61(5):1436–1447.

[Gehring and Fencsik, 2001] Gehring, W. J. and Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *Journal of Neuroscience*, 21(23):9430–9437.

[Gehring et al., 1993] 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.

[Gibbons and Beneteau, 2010] Gibbons, C. and Beneteau, E. (2010). Functional performance using eye control and single switch scanning by people with als. *Perspectives on Augmentative and Alternative Communication*, 19(3):64–69.

[Golotvin and Williams, 2000] Golotvin, S. and Williams, A. (2000). Improved baseline recognition and modeling of ft nmr spectra. *J of Magn Reson*, 146(1):122–125.

[Graimann et al., 2009] Graimann, B., Allison, B., and Pfurtscheller, G. (2009). Brain–computer interfaces: A gentle introduction. In *Brain-computer interfaces*, pages 1–27. Springer.

[Greco et al., 2016] Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., and Citi, L. (2016). cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Trans Biomed Eng*, 63(4):797–804.

[Gutberlet et al., 2009] Gutberlet, I., Debener, S., Jung, T., and Makeig, S. (2009). Techniques of eeg recording and preprocessing. *Quantative EEG Analysis Methods and Clinical Applications*, pages 23–49.

[Hajcak et al., 2005] Hajcak, G., Moser, J. S., Yeung, N., and Simons, R. F. (2005). On the ern and the significance of errors. *Psychophysiology*, 42(2):151–160.

[Harandi et al., 2017] Harandi, M., Salzmann, M., and Hartley, R. (2017). Dimensionality reduction on spd manifolds: The emergence of geometry-aware methods. *IEEE transactions on pattern analysis and machine intelligence*, 40(1):48–62.

[Hazrati and Erfanian, 2010] Hazrati, M. K. and Erfanian, A. (2010). An online eeg-based brain–computer interface for controlling hand grasp using an adaptive probabilistic neural network. *Medical engineering & physics*, 32(7):730–739.

[He and Wu, 2019] He, H. and Wu, D. (2019). Transfer learning for brain-computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*.

[He et al., 2016] He, S., Zhang, R., Wang, Q., Chen, Y., Yang, T., Feng, Z., Zhang, Y., Shao, M., and Li, Y. (2016). A p300-based threshold-free brain switch and its application in wheelchair control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(6):715–725.

[Healey and Picard, 2005] Healey, J. A. and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE trans Intell Transp Syst*, 6(2):156–166.

[Hoffmann et al., 2005] 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. In *Conference Proceedings. 2nd International IEEE EMBS Conference on Neural Engineering, 2005.*, pages 97–100. IEEE.

[Hoffmann et al., 2008] Hoffmann, U., Vesin, J.-M., Ebrahimi, T., and Diserens, K. (2008). An efficient p300-based brain–computer interface for disabled subjects. *Journal of Neuroscience methods*, 167(1):115–125.

[Holroyd and Coles, 2002] Holroyd, C. B. and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, 109(4):679.

[Holz et al., 2015] Holz, E. M., Botrel, L., and Kübler, A. (2015). Independent home use of brain painting improves quality of life of two artists in the locked-in state diagnosed with amyotrophic lateral sclerosis. *Brain-Computer Interfaces*, 2(2-3):117–134.

[Huettel and McCarthy, 2004] Huettel, S. A. and McCarthy, G. (2004). What is odd in the oddball task?: Prefrontal cortex is activated by dynamic changes in response strategy. *Neuropsychologia*, 42(3):379–386.

[Hwang et al., 2013] Hwang, H.-J., Kim, S., Choi, S., and Im, C.-H. (2013). Eeg-based brain-computer interfaces: a thorough literature survey. *International Journal of Human-Computer Interaction*, 29(12):814–826.

[Hyvärinen and Oja, 2000] Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430.

[iMotions, 2017] iMotions (2017). Galvanic skin response the complete pocket guide. *iMotions – Biometric Research*.

[Iturrate et al., 2009] Iturrate, I., Antelis, J. M., Kubler, A., and Minguez, J. (2009). A noninvasive brain-actuated wheelchair based on a p300 neurophysiological protocol and automated navigation. *IEEE transactions on robotics*, 25(3):614–627.

[Iturrate et al., 2014] Iturrate, I., Chavarriaga, R., Montesano, L., Minguez, J., and Millán, J. (2014). Latency correction of event-related potentials between different experimental protocols. *Journal of neural engineering*, 11(3):036005.

[Iwane et al., 2016] Iwane, F., Chavarriaga, R., Iturrate, I., and Millán, J. d. R. (2016). Spatial filters yield stable features for error-related potentials across conditions. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 000661–000666. IEEE.

[Jasper, 1958] Jasper, H. (1958). The 10/20 international electrode system. *EEG and Clinical Neurophysiology*, 10(2):370–375.

[Jayaram et al., 2016] Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Computational Intelligence Magazine*, 11(1):20–31.

[Jeuris et al., 2012] Jeuris, B., Vandebril, R., and Vandereycken, B. (2012). A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39(ARTICLE).

[Jost and Jost, 2008] Jost, J. and Jost, J. (2008). *Riemannian geometry and geometric analysis*, volume 42005. Springer.

[Jung and Berger, 1979] Jung, R. and Berger, W. (1979). Hans bergers entdeckung des elektrenkephalogramms und seine ersten befunde 1924–1931. *Archiv fuer Psychiatrie und Nervenkrankheiten*, 227(4):279–300.

[Kang and Choi, 2014] Kang, H. and Choi, S. (2014). Bayesian common spatial patterns for multi-subject eeg classification. *Neural Networks*, 57:39–50.

[Kang et al., 2009] Kang, H., Nam, Y., and Choi, S. (2009). Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Processing Letters*, 16(8):683–686.

[Karim et al., 2006] Karim, A. A., Hinterberger, T., Richter, J., Mellinger, J., Neumann, N., Flor, H., Kübler, A., and Birbaumer, N. (2006). Neural internet: Web surfing with brain potentials for the completely paralyzed. *Neurorehabilitation and Neural Repair*, 20(4):508–515.

[Kaufmann et al., 2014] Kaufmann, T., Herweg, A., and Kübler, A. (2014). Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials. *Journal of neuroengineering and rehabilitation*, 11(1):7.

[Kevric and Subasi, 2017] Kevric, J. and Subasi, A. (2017). Comparison of signal decomposition methods in classification of eeg signals for motor-imagery bci system. *Biomedical Signal Processing and Control*, 31:398–406.

[Kindermans et al., 2014] Kindermans, P.-J., Schreuder, M., Schrauwen, B., Müller, K.-R., and Tangermann, M. (2014). True zero-training brain-computer interfacing–an online study. *PloS one*, 9(7):e102504.

[Kohlbrecher et al., 2011] Kohlbrecher, S., Von Stryk, O., Meyer, J., and Klingauf, U. (2011). A flexible and scalable slam system with full 3d motion estimation. In *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 155–160. IEEE.

[Kohlmorgen and Blankertz, 2004] Kohlmorgen, J. and Blankertz, B. (2004). Bayesian classification of single-trial event-related potentials in eeg. *International Journal of Bifurcation and Chaos*, 14(02):719–726.

[Krauledat et al., 2008] Krauledat, M., Tangermann, M., Blankertz, B., and Müller, K.-R. (2008). Towards zero training for brain-computer interfacing. *PloS one*, 3(8):e2967.

[Kübler et al., 2015] Kübler, A., Holz, E. M., Sellers, E. W., and Vaughan, T. M. (2015). Toward independent home use of brain-computer interfaces: a decision algorithm for selection of potential end-users. *Archives of physical medicine and rehabilitation*, 96(3):S27–S32.

[Kumar et al., 2019] Kumar, S., Yger, F., and Lotte, F. (2019). Towards adaptive classification using riemannian geometry approaches in brain-computer interfaces. In *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–6. IEEE.

[Lee et al., 2006] Lee, H., Cichocki, A., and Choi, S. (2006). Nonnegative matrix factorization for motor imagery eeg classification. In *International Conference on Artificial Neural Networks*, pages 250–259. Springer.

[Lemm et al., 2011] Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *Neuroimage*, 56(2):387–399.

[Li et al., 2020] Li, F., He, F., Wang, F., Zhang, D., Xia, Y., and Li, X. (2020). A novel simplified convolutional neural network classification algorithm of motor imagery eeg signals based on deep learning. *Applied Sciences*, 10(5):1605.

[Li et al., 2016] Li, Z., Zhao, S., Duan, J., Su, C.-Y., Yang, C., and Zhao, X. (2016). Human cooperative wheelchair with brain–machine interaction based on shared control strategy. *IEEE/ASME Transactions on Mechatronics*, 22(1):185–195.

[Lim et al., 2019] Lim, C. G., Poh, X. W. W., Fung, S. S. D., Guan, C., Bautista, D., Cheung, Y. B., Zhang, H., Yeo, S. N., Krishnan, R., and Lee, T. S. (2019). A randomized controlled trial of a brain-computer interface based attention training program for adhd. *PloS one*, 14(5).

[Liu et al., 2018] Liu, Y., Wei, Q., and Lu, Z. (2018). A multi-target brain-computer interface based on code modulated visual evoked potentials. *PloS one*, 13(8).

[Lopes et al., 2016] Lopes, A., Rodrigues, J., Perdigao, J., Pires, G., and Nunes, U. (2016). A new hybrid motion planner: Applied in a brain-actuated robotic wheelchair. *IEEE Robot Autom Mag*, 23(4):82–93.

[Lopes et al., 2013] Lopes, A. C., Pires, G., and Nunes, U. (2013). Assisted navigation for a brain-actuated intelligent wheelchair. *Robotics and Autonomous Systems*, 61(3):245–258.

[Lotte, 2015] Lotte, F. (2015). Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. *Proceedings of the IEEE*, 103(6):871–890.

[Lotte et al., 2018] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F. (2018). A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005.

[Lotte et al., 2007] 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 and Guan, 2010] Lotte, F. and Guan, C. (2010). Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms. *IEEE Transactions on biomedical Engineering*, 58(2):355–362.

[Luck and Kappenman, 2011] Luck, S. J. and Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components*. Oxford university press.

[Luu et al., 2003] Luu, P., Tucker, D. M., Derryberry, D., Reed, M., and Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science*, 14(1):47–53.

[Machado and Balbinot, 2014] Machado, J. and Balbinot, A. (2014). Executed movement using eeg signals through a naive bayes classifier. *Micromachines*, 5(4):1082–1105.

[Malmivuo et al., 1995] Malmivuo, P., Malmivuo, J., and Plonsey, R. (1995). *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA.

[Margaux et al., 2012] Margaux, P., Emmanuel, M., Sébastien, D., Olivier, B., and Jérémie, M. (2012). Objective and subjective evaluation of online error correction during p300-based spelling. *Advances in Human-Computer Interaction*, 2012.

[McFarland et al., 1997] McFarland, D. J., McCane, L. M., David, S. V., and Wolpaw, J. R. (1997). Spatial filter selection for eeg-based communication. *Electroencephalography and clinical Neurophysiology*, 103(3):386–394.

[McFarland et al., 2015] McFarland, D. J., Sarnacki, W. A., and Wolpaw, J. R. (2015). Effects of training pre-movement sensorimotor rhythms on behavioral performance. *Journal of neural engineering*, 12(6):066021.

[McMillan et al., 1995] McMillan, G. R., Calhoun, G., Middendorf, M., Schnurer, J., Ingle, D., and Nasman, V. (1995). Direct brain interface utilizing self-regulation of steady-state visual evoked response (ssver). In *Proc. RESNA'95 Annu. Conf*, volume 15, pages 693–695.

[Millán et al., 2004] Millán, J. R., Renkens, F., Mourino, J., and Gerstner, W. (2004). Noninvasive brain-actuated control of a mobile robot by human eeg. *IEEE Transactions on biomedical Engineering*, 51(6):1026–1033.

[Moakher, 2005] 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.

[Morioka et al., 2015] Morioka, H., Kanemura, A., Hirayama, J.-i., Shikauchi, M., Ogawa, T., Ikeda, S., Kawanabe, M., and Ishii, S. (2015). Learning a common dictionary for subject-transfer decoding with resting calibration. *NeuroImage*, 111:167–178.

[Moses et al., 2018] Moses, D. A., Leonard, M. K., and Chang, E. F. (2018). Real-time classification of auditory sentences using evoked cortical activity in humans. *Journal of neural engineering*, 15(3):036005.

[Mueller et al., 2008] Mueller, V., Brehmer, Y., Von Oertzen, T., Li, S.-C., and Lindenberger, U. (2008). Electrophysiological correlates of selective attention: a lifespan comparison. *BMC neuroscience*, 9(1):18.

[Muglerab et al., 2008] Muglerab, E., Benschc, M., Haldera, S., Rosenstielc, W., Bogdancd, M., Birbaumerae, N., and Kübleraf, A. (2008). Control of an internet browser using the p300 event-related potential. *International Journal of Bioelectromagnetism*, 10(1):56–63.

[Mundell et al., 2016] Mundell, C., Vielma, J. P., and Zaman, T. (2016). Predicting performance under stressful conditions using galvanic skin response. *arXiv preprint arXiv:1606.01836*.

[Nam et al., 2018] Nam, C. S., Nijholt, A., and Lotte, F. (2018). *Brain–computer interfaces handbook: technological and theoretical advances*. CRC Press.

[NASA, 1986] NASA (1986). Nasa task load index (tlx) v. 1.0 manual.

[Ng et al., 2014] Ng, D. W.-K., Soh, Y.-W., and Goh, S.-Y. (2014). Development of an autonomous bci wheelchair. In *2014 IEEE Symposium on Computational Intelligence in Brain Computer Interfaces (CIBCI)*, pages 1–4. IEEE.

[Nguyen and Artemiadis, 2018] Nguyen, C. H. and Artemiadis, P. (2018). Eeg feature descriptors and discriminant analysis under riemannian manifold perspective. *Neurocomputing*, 275:1871–1883.

[Nicolas-Alonso and Gomez-Gil, 2012] Nicolas-Alonso, L. F. and Gomez-Gil, J. (2012). Brain computer interfaces, a review. *sensors*, 12(2):1211–1279.

[Nourbakhsh et al., 2012] Nourbakhsh, N., Wang, Y., Chen, F., and Calvo, R. A. (2012). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proc of the 24th Australian Computer-Human Interaction Conf*, pages 420–423. ACM.

[O'Connell et al., 2007] O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., Robertson, I. H., and Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. *European Journal of Neuroscience*, 25(8):2571–2579.

[Overbeek et al., 2005] Overbeek, T. J., Nieuwenhuis, S., and Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the pe vis-à-vis the ern/ne. *Journal of Psychophysiology*, 19(4):319–329.

[Pailing and Segalowitz, 2004] Pailing, P. E. and Segalowitz, S. J. (2004). The error-related negativity as a state and trait measure: Motivation, personality, and erps in response to errors. *Psychophysiology*, 41(1):84–95.

[Pan and Yang, 2009] Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

[Parafita et al., 2013] Parafita, R., Pires, G., Nunes, U., and Castelo-Branco, M. (2013). A spacecraft game controlled with a brain-computer interface using SSVEP with phase tagging. In *SeGAH 2013 - IEEE 2nd International Conference on Serious Games and Applications for Health, Book of Proceedings*.

[Perdiz et al., 2019] Perdiz, J., Cruz, A., Nunes, U. J., and Pires, G. (2019). A hybrid brain-computer interface fusing p300 erp and electrooculography. In *Mediterranean Conference on Medical and Biological Engineering and Computing*, pages 1755–1766. Springer.

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

[Pichiorri et al., 2015] Pichiorri, F., Morone, G., Petti, M., Toppi, J., Pisotta, I., Molinari, M., Paolucci, S., Inghilleri, M., Astolfi, L., Cincotti, F., et al. (2015). Brain–computer interface boosts motor imagery practice during stroke recovery. *Annals of neurology*, 77(5):851–865.

[Picton, 1992] Picton, T. W. (1992). The p300 wave of the human event-related potential. *Journal of clinical neurophysiology*, 9(4):456–479.

[Pinti et al., 2020] Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., and Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fnirs) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, 1464(1):5.

[Pires, 2018] Pires, G. (2018). LSC:P300-ErrP Video. https://home.isr.uc.pt/~gpires/ErrP/video_ErrP.mp4. [Online; accessed July-2018].

[Pires et al., 2011a] Pires, G., Nunes, U., and Castelo-Branco, M. (2011a). Statistical spatial filtering for a P300-based BCI: tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis. *Journal of neuroscience methods*, 195(2):270–281.

[Pires et al., 2012] Pires, G., Nunes, U., and Castelo-Branco, M. (2012). Comparison of a row-column speller vs. a novel lateral single-character speller: assessment of BCI for severe motor disabled patients. *Clinical Neurophysiology*, 123(6):1168–1181.

[Pires et al., 2011b] Pires, G., Torres, M., Casaleiro, N., Nunes, U., and Castelo-Branco, M. (2011b). Playing tetris with non-invasive bci. In *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–6. IEEE.

[Pires, 2011] Pires, G. P. (2011). *Biosignal Classification for Human Interface with Devices and Surrounding Environment*. PhD thesis.

[Polich, 2007] Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118(10):2128–48.

[Qiu et al., 2017] Qiu, Z., Allison, B. Z., Jin, J., Zhang, Y., Wang, X., Li, W., and Cichocki, A. (2017). Optimized motor imagery paradigm based on imagining chinese characters writing movement. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(7):1009–1017.

[Rakotomamonjy and Guigue, 2008] Rakotomamonjy, A. and Guigue, V. (2008). Bci competition iii: dataset ii-ensemble of svms for bci p300 speller. *IEEE transactions on biomedical engineering*, 55(3):1147–1154.

[Raza et al., 2016] Raza, H., Cecotti, H., Li, Y., and Prasad, G. (2016). Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface. *Soft Computing*, 20(8):3085–3096.

[Rebsamen et al., 2010] Rebsamen, B., Guan, C., Zhang, H., Wang, C., Teo, C., Ang, M. H., and Burdet, E. (2010). A brain controlled wheelchair to navigate in familiar environments. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(6):590–598.

[Reichert et al., 2016] Reichert, J. L., Kober, S. E., Schweiger, D., Grieshofer, P., Neuper, C., and Wood, G. (2016). Shutting down sensorimotor interferences after stroke: A proof-of-principle smr neurofeedback study. *Frontiers in human neuroscience*, 10:348.

[Rivet et al., 2009] Rivet, B., Souloumiac, A., Attina, V., and Gibert, G. (2009). xdawn algorithm to enhance evoked potentials: application to brain–computer interface. *IEEE Transactions on Biomedical Engineering*, 56(8):2035–2043.

[Saichoo and Boobrahm, 2019] Saichoo, T. and Boobrahm, P. (2019). Brain computer interface for real-time driver drowsiness detection. *Thai Journal of Physics*, 36(1).

[Salazar-Gomez et al., 2017] Salazar-Gomez, A. F., DelPreto, J., Gil, S., Guenther, F. H., and Rus, D. (2017). Correcting robot mistakes in real time using eeg signals. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 6570–6577. IEEE.

[Samek et al., 2012] Samek, W., Vidaurre, C., Müller, K.-R., and Kawanabe, M. (2012). Stationary common spatial patterns for brain–computer interfacing. *Journal of neural engineering*, 9(2):026013.

[Schalk et al., 2008] Schalk, G., Miller, K. J., Anderson, N. R., Wilson, J. A., Smyth, M. D., Ojemann, J. G., Moran, D. W., Wolpaw, J. R., and Leuthardt, E. C. (2008). Two-dimensional movement control using electrocorticographic signals in humans. *Journal of neural engineering*, 5(1):75.

[Schalk et al., 2000] Schalk, G., Wolpaw, J. R., McFarland, D. J., and Pfurtscheller, G. (2000). EEG-based communication: Presence of an error potential. *Clinical Neurophysiology*, 111(12):2138–2144.

[Schlögl et al., 1997] Schlögl, A., Lugger, K., and Pfurtscheller, G. (1997). Using adaptive autoregressive parameters for a brain-computer-interface experiment. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136)*, volume 4, pages 1533–1535. Ieee.

[Schmidt et al., 2012] Schmidt, N. M., Blankertz, B., and Treder, M. S. (2012). Online detection of error-related potentials boosts the performance of mental typewriters. *BMC neuroscience*, 13(1):19.

[Schönleitner et al., 2019] Schönleitner, F. M., Otter, L., Ehrlich, S. K., and Cheng, G. (2019). A comparative study on adaptive subject-independent classification models for zero-calibration error-potential decoding. In *IEEE International Conference on Cyborg and Bionic Systems 2019*.

[Schröder et al., 2005] Schröder, M., Lal, T. N., Hinterberger, T., Bogdan, M., Hill, N. J., Birbaumer, N., Rosenstiel, W., and Schölkopf, B. (2005). Robust eeg channel selection across subjects for brain-computer interfaces. *EURASIP Journal on Applied Signal Processing*, 2005:3103–3112.

[Schultz, 2002] Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 36(2):241–263.

[Seno et al., 2010] Seno, B. D., Matteucci, M., and Mainardi, L. T. (2010). The Utility Metric : A Novel Method to Assess the Overall Performance of Discrete Brain – Computer Interfaces. 18(1):20–28.

[Serby et al., 2005] Serby, H., Yom-Tov, E., and Inbar, G. F. (2005). An improved p300-based brain-computer interface. *IEEE Transactions on neural systems and rehabilitation engineering*, 13(1):89–98.

[Setz et al., 2010] Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., and Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable eda device. *IEEE Trans Inf Technol Biomed*, 14(2):410–417.

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

[Sharma et al., 2018] Sharma, N., Verma, S., and Nagpal, P. B. (2018). A review of brain fingerprinting process. *A Review of Brain Fingerprinting Process*, 12(1):5–5.

[Song and Sepulveda, 2020] Song, Y. and Sepulveda, F. (2020). Comparison between covert sound-production task (sound-imagery) vs. motor-imagery for onset detection in real-life online self-paced bcis. *Journal of NeuroEngineering and Rehabilitation*, 17(1):1–11.

[Spielberg et al., 1970] Spielberg, C., Gorsuch, R., and Lushene, R. (1970). Test manual for the state-trait anxiety inventory.

[Spüler et al., 2012] Spüler, M., Bensch, M., Kleih, S., Rosenstiel, W., Bogdan, M., and Kübler, A. (2012). Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. *Clinical Neurophysiology*, 123(7):1328–1337.

[Sra, 2011] Sra, S. (2011). Positive definite matrices and the symmetric stein divergence. *arXiv preprint arXiv:1110.1773*.

[Sugiyama, 2007] Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of machine learning research*, 8(May):1027–1061.

[Sun et al., 2016] Sun, B., Feng, J., and Saenko, K. (2016). Return of frustratingly easy domain adaptation. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[Tang et al., 2018] Tang, J., Liu, Y., Hu, D., and Zhou, Z. (2018). Towards bci-actuated smart wheelchair system. *Biomedical engineering online*, 17(1):111.

[Teeuw, 2010] Teeuw, J. (2010). Comparison of error-related eeg potentials. In *13th Twente Student Conference on IT*.

[Treder et al., 2011] Treder, M. S., Schmidt, N. M., and Blankertz, B. (2011). Gaze-independent brain-computer interfaces based on covert attention and feature attention. *Journal of neural engineering*, 8(6):066003.

[Vamvakousis and Ramirez, 2014] Vamvakousis, Z. and Ramirez, R. (2014). P300 harmonies: A brain-computer musical interface. In *ICMC*.

[van Schie et al., 2004] van Schie, H. T., Mars, R. B., Coles, M. G., and Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature neuroscience*, 7(5):549.

[Van Veen and Carter, 2005] Van Veen, V. and Carter, C. S. (2005). Separating semantic conflict and response conflict in the stroop task: a functional mri study. *Neuroimage*, 27(3):497–504.

[Veen and Carter, 2002] Veen, V. v. and Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of cognitive neuroscience*, 14(4):593–602.

[Vidal et al., 2000] Vidal, F., Hasbroucq, T., Grapperon, J., and Bonnet, M. (2000). Is the 'error negativity'specific to errors? *Biological psychology*, 51(2-3):109–128.

[Vidal, 1973] Vidal, J. J. (1973). Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1):157–180.

[Vilela and Hochberg, 2020] Vilela, M. and Hochberg, L. R. (2020). Applications of brain-computer interfaces to the control of robotic and prosthetic arms. In *Handbook of Clinical Neurology*, volume 168, pages 87–99. Elsevier.

[Volosyak, 2011] Volosyak, I. (2011). Ssvep-based bremen–bci interface—boosting information transfer rates. *Journal of neural engineering*, 8(3):036020.

[Wang et al., 2014] Wang, H., Li, Y., Long, J., Yu, T., and Gu, Z. (2014). An asynchronous wheelchair control by hybrid eeg–eog brain–computer interface. *Cognitive neurodynamics*, 8(5):399–409.

[Wolpaw et al., 2000] Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., Donchin, E., Quatrano, L. A., Robinson, C. J., and Vaughan, T. M. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering*, 8(2):164–173.

[Wolpaw et al., 2002] 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 et al., 1991] Wolpaw, J. R., McFarland, D. J., Neat, G. W., and Forneris, C. A. (1991). An eeg-based brain-computer interface for cursor control. *Electroencephalography and clinical neurophysiology*, 78(3):252–259.

[Wolpaw and Wolpaw, 2012] Wolpaw, J. R. and Wolpaw, E. W. (2012). Brain-computer interfaces: something new under the sun. *Brain-computer interfaces: principles and practice*, pages 3–12.

[Wu, 2016] Wu, D. (2016). Online and Offline Domain Adaptation for Reducing BCI Calibration Effort. *IEEE Transactions on Human-Machine Systems*, pages 1–14.

[Xu et al., 2020] Xu, J., Grosse-Wentrup, M., and Jayaram, V. (2020). Tangent space spatial filters for interpretable and efficient riemannian classification. *Journal of Neural Engineering*, 17(2):026043.

[Yamamoto et al., 2020] Yamamoto, M., Sadatnejad, K., Tanaka, T., Islam, M. R., Tanaka, Y., and Lotte, F. (2020). Detecting eeg outliers for bci on the riemannian manifold using spectral clustering. In *42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society (EMBC'2020)*.

[Yeung et al., 2004] Yeung, N., Botvinick, M. M., and Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological review*, 111(4):931.

[Yger et al., 2016] Yger, F., Berar, M., and Lotte, F. (2016). Riemannian approaches in brain-computer interfaces: a review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1753–1762.

[Yu et al., 2018] Yu, Y., Liu, Y., Jiang, J., Yin, E., Zhou, Z., and Hu, D. (2018). An asynchronous control paradigm based on sequential motor imagery and its application in wheelchair navigation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(12):2367–2375.

[Yu et al., 2017] Yu, Y., Zhou, Z., Liu, Y., Jiang, J., Yin, E., Zhang, N., Wang, Z., Liu, Y., Wu, X., and Hu, D. (2017). Self-paced operation of a wheelchair based on a hybrid brain-computer interface combining motor imagery and p300 potential. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12):2516–2526.

[Zanini et al., 2017] Zanini, P., Congedo, M., Jutten, C., Said, S., and Berthoumieu, Y. (2017). Transfer learning: a riemannian geometry framework with applications to brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116.

[Zeyl et al., 2015] Zeyl, T., Yin, E., Keightley, M., and Chau, T. (2015). Adding Real-Time Bayesian Ranks to Error-RelatedPotential Scores Improves Error Detection and Auto-Correction in aP300 Speller. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 4320(c):1–1.

[Zeyl et al., 2016] Zeyl, T., Yin, E., Keightley, M., and Chau, T. (2016). Adding real-time Bayesian ranks to error-related potential scores improves error detection and auto-correction in a P300 speller. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(1):46–56.

[Zhang et al., 2015a] Zhang, H., Chavarriaga, R., Khaliliardali, Z., Gheorghe, L., Iturrate, I., and Millán, J. d. R. (2015a). EEG-based decoding of error-related brain activity in a real-world driving task. *Journal of Neural Engineering*.

[Zhang et al., 2015b] Zhang, R., Li, Y., Yan, Y., Zhang, H., Wu, S., Yu, T., and Gu, Z. (2015b). Control of a wheelchair in an indoor environment based on a brain–computer interface and automated navigation. *IEEE transactions on neural systems and rehabilitation engineering*, 24(1):128–139.

A

BCI-Double-ErrP-Dataset

A.1 Experiment associated with dataset

BCI-Double-ErrP-Dataset is an EEG dataset recorded while participants used a P300-based BCI speller (Fig. A.1). This speller uses a P300 post-detection based on Error-related potentials (ErrPs) to detect and correct errors (i.e. when the detected symbol does not match the user's intention). After the P300 detection, an automatic correction is made when an ErrP is detected (this is called a "Primary ErrP"). The correction proposed by the system is also evaluated, eventually eliciting a "Secondary ErrP" if the correction is wrong. The overall approach is called "Double-ErrP detection and correction" [Cruz et al., 2018a]. The recorded datasets are useful to research the use of ErrPs to increase BCI reliability and to research how "Secondary ErrPs" can be used to increase the naturalness of Human-machine interaction. The datasets are also useful to research ErrP variability across subjects and across sessions in different days.

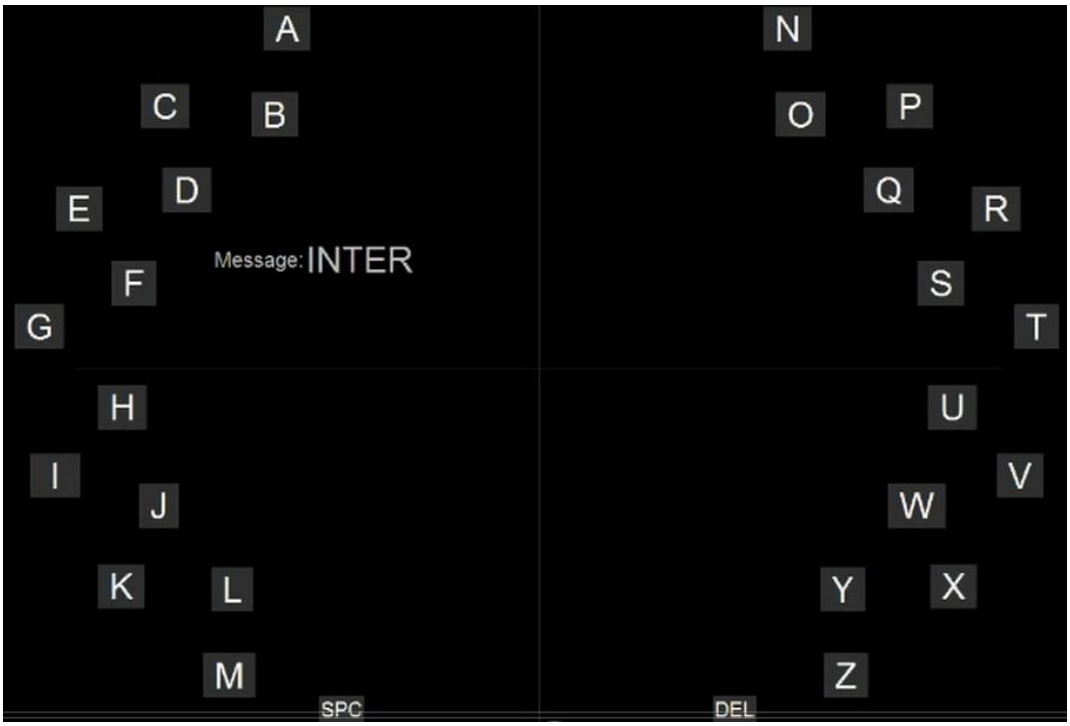


Figure A.1: Lateral single character speller paradigm [Cruz et al., 2018a]. Each symbol is highlighted once in each round. The highlight time of the stimuli is 75 ms and the inter-trial interval is 4s.

A.2 EEG recording and Participants

The EEG signals were recorded from nine able-bodied participants (S1-S9) and one tetraplegic participant (P1) with medullar injury (C4/C5 level) with ages between 24 and 43 years old. However, the data set comprises only 8 subjects because two participants (S7 and S8) took part only in the first session of the experiments. Scalp EEG was recorded using 12 Ag/AgCl electrodes, except for participant S2 who used passive electrodes (g.USBamp bioamplifier, g.tec, Austria). The right earlobe and AFz were used as a reference and ground electrodes respectively. The EEG signals were sampled at 256 Hz and pre-processed using a notch filter at 50 Hz and a band-pass filter at [0.1 - 100] Hz.

A.3 Sessions

The experiment comprised two sessions with three phases:

- **Session 1:** P300 calibration (1st phase) and ErrP calibration (2nd phase);
- **Session 2:** final P300-ErrP system evaluation (3rd phase), that is, the detection of targets events

(P300 classification) and the detection (and correction) of wrong selections (ErrP detection – Primary and Secondary ErrPs).

P300 Calibration/Training set (session 1): participants focused attention on the 10 letters of the word “INTERFACES”. For each letter, all symbols flashed 9 times (collecting 90 targets and 2430 non-targets). This calibration was required to train the P300 classifier to operate the P300-speller.

ErrP Calibration/Training set (session 1): participants had to write several times the Portuguese sentence “ESTOU-A-ESCREVER-COM-UMA-INTERFACE-BCI” (38 characters) without either interruption or correction. The number of repetitions within a trial was selected based on the user’s performance in the P300 calibration (for an offline accuracy around 90%). This calibration was required to collect ErrPs in order to train the ErrP classifier to detect errors. The EEG data associated with each spelled sentence are stored in Matlab files (e.g., file ”S2_Sess1_ sentence3.mat” for Subject 2, session 1, sentence 3).

P300-ErrP Test set (session 2): participants spelled the same sentence of the ErrP calibration. This session was held on a different day of session 1 for all participants except S9. These datasets are de most relevant as they correspond to the final task where we detect both 1st and 2nd ErrPs.

A.4 Instructions to use datasets

There are 4 folders:

- **P300˙Training˙Session1:** dataset of the P300 calibration (session 1);
- **Session 2:** dataset of the ErrP calibration (session 1);
- **P300˙Training˙Session1:** dataset of the testing session (session 2);
- **P300˙Training˙Session1:** Matlab code to extract P300, 1st ErrP and 2nd ErrP. **Run example.m**.

Each file (e.g., S*_Sess*_ sentence*.mat) contains a big matrix signal with the following fields:
Line1: Timestamp of each sample.
Line2: EEG samples recorded from Fz.
Line3: EEG samples recorded from Cz.
Line4: EEG samples recorded from C3.
Line5: EEG samples recorded from C4.
Line6: EEG samples recorded from CPz.
Line7: EEG samples recorded from Pz.
Line8: EEG samples recorded from P3.
Line9: EEG samples recorded from P4.

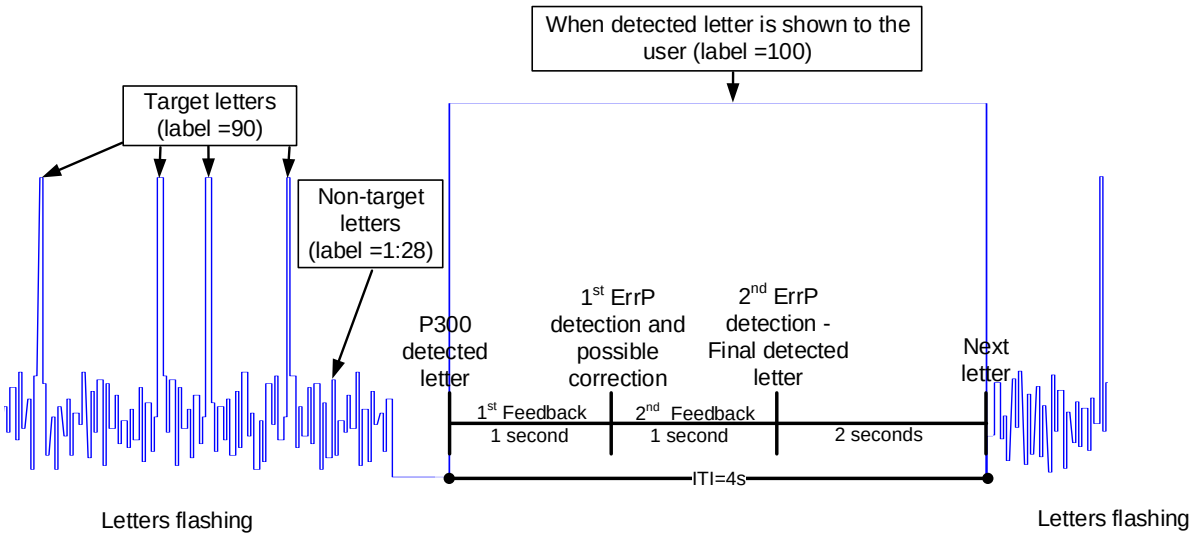


Figure A.2: Description of events saved on line 14.

Line10: EEG samples recorded from PO7.

Line11: EEG samples recorded from PO8.

Line12: EEG samples recorded from POz.

Line13: EEG samples recorded from Oz.

Line14: Contains the information of symbols. It is labelled as (see details in Fig. A.2):

- **'90'**, code of the target symbol when hilighted;
- **'1:28'**, codes of the non-targets when highlighted; and
- **'100'**, code indicating that the detected letter is shown to the user.

Line15: A zero vector except for the timestamps when feedbacks occur:

- **'1'** for wrong feedback (FIRST FEEDBACK); and
- **'2'** for correct feedback (FIRST FEEDBACK), Session 2 also has the information of the second feedback, with value:
- **'3'** for wrong feedback (SECOND FEEDBACK); and
- **'4'** for correct feedback (SECOND FEEDBACK).

B

Tangent space spatial filter

The tangent space spatial filter algorithm [Xu et al., 2020] obtains the filter on the tangent space and projects it to the manifold. We first compute the covariance matrices using (eq. 6.1), then the Riemannian mean (C^m) is calculated unsin eq. 3.10 . It is used as the reference point for the tangent space (eq. 3.13). The tangent vector (S) is computed, and the filters can be obtained using different criterion. Here we used the Fisher Linear Discriminant criterion. Therefore, we compute the within-class and between-class scatter matrix and then we solve the generalized eigenvalue decomposition problem. The filter is the eigenvector sorted according to the largest eigenvalue and it is mapped back onto the manifold (C^w). After we solve the generalized eigenvalue decomposition problem: GED(C^w,C^m). The optimum spatial filter (V) is the eigenvector sorted according to the largest absolute value of the logarithm of the eigenvalues.