Lehrstuhl für Steuerungs- und Regelungstechnik
Technische Universität München
Prof. Dr.-Ing./Univ. Tokio Martin Buss

Advancing Brain-Computer Interfaces for Robotic Embodiment Systems

Mohammad Abu-Alqumsan

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr.-Ing. Bernhard Seeber
Prüfer der Dissertation:	
1.	Prof. Dr.-Ing. Angelika Peer University of the West of England, Bristol, UK
2.	Prof. Dr.-Ing. Eckehard Steinbach

Die Dissertation wurde am 23.06.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 07.11.2016 angenommen.

Foreword

This work summarizes almost five years of my work at the Chair of Automatic Control Engineering (LSR) at the Technische Universität München (TUM). I would like to take this opportunity to express my gratitude to my supervisor Prof. Angelika Peer for the great support and for all the fruitful discussions and guidance that made a huge difference in every aspect of the hereby presented work. I am also grateful to Prof. Martin Buss and Prof. Dirk Wollherr for their great support and for giving me the opportunity to pursue my research path at the very exciting and enriching robotics laboratory of LSR .

I am indebted to thanking many people who made this thesis possible, but I remain responsible for all the shortcomings. Special thanks to my colleague in the VERE project, Robert Jenke, with whom I had the joy of working, sharing knowledge and traveling for business trips. I would like to thank my colleagues who made the time at LSR enjoyable and enriching. The list is long, but I would like particularly to thank Laith Alkurdi, Ken Friedl, Roderick de Nijs, Annemarie Turnwald, Sotiris Apostolopoulos, Muhammad Sheraz Khan, Milad Geravand, Stefan Klare, Philine Donner, Andreas Lawitzky, Shushu Ma, Alexander Pekarovskiy, Christian Landsiedel and Markus Schill. My thanks extend to Daniel Carton, Sebastian Erhart, Thomas Schauß and Moritz Stötter for the great support in all network-related matters. In terms of technical help, Wolfgang Jaschik, Thomas Lowitz, Tobias Stoeber, Domenik Weilbach and Kilian Weber could not have been more supportive or generous with their time, when I needed it the most. I thank them.

Sincere thanks to my students who devoted a great deal of their time and passion to the projects on which they worked and for helping in conducting the many experiments reported in this thesis. I thank in particular Felix Ebert, Mahmoud Kassar, Tailin Li, Sharat Embrandiri, Luis Vergara, Dullal Ghosh and Vincent Bobinski. Special thanks extend to all the volunteers who participated in these experiments for their patience.

Research efforts reported in this thesis served an integrative function in the collaborative EU project VERE (virtual embodiment and robotic re-embodiment). I was fortunate enough to collaborate and meet with remarkable and warmhearted people during the lifetime of the project. I would like to express my special appreciation to Emmanuele Tidoni, Prof. Salvatore M Aglioti (Sapienza University of Rome, Rome, Italy), Christoph Kapeller, Christoph Hintermüller, Christoph Guger (g.tec medical engineering, Shiedlberg, Austria) and Daniele Leonardis (Percro Laboratory, Scuola Superiore Sant'Anna, Pisa, Italy) for sharing their knowledge and for making it possible to test some of the VERE objectives with people with disabilities.

I am deeply grateful to my family for all the unconditional love, support and for always being the origin of my passion.

Munich, June 2016 Mohammad Abu-Alqumsan

To the memories of my father

Abstract

Thanks to the rapid development of brain-computer interfaces (BCIs), users can control and communicate with external devices without performing a single muscle movement. BCIs allow this by detecting the control and communication signals directly at their source, i.e. the brain. This renders BCIs highly applicable in different domains and of special importance to users with physical impairments. However, despite the great progress made to date, one should not expect BCIs to replace keyboards in the near future due to several factors. Firstly, the process of decoding brain activity remains error-prone due to the presence of noise and nonstationarity in the measured brain activity. Secondly, the bandwidth and bit rate of current BCIs compare poorly to conventional user interfaces. Furthermore, immersive robotic embodiment applications, which are the driving force behind this work, pose additional challenges to BCIs due to the varying time scales in which robots operate. To circumvent the current limitations, efforts were made firstly to improve the accuracy and reliability of the detection of the brain signals used in BCIs. On this account, we have focused on the detection of steady-state visual evoked potentials (SSVEPs) and proposed a new detection method that slightly outperforms state-of-the-art competing methods. Given the stereo-vision requirements of immersive applications, we have also explored and compared different stimuli presentation methods. Additionally, we have tackled the problem of detecting interaction error-related potentials due to their possible integration within BCI systems. Since reliable classification of these potentials typically requires long training sessions, special focus has been laid upon the classifier transferability problem. Application-specific improvements were also discussed for immersive robotic embodiment systems, where contextual and adaptive BCIs were proposed as a way to overcome the bandwidth limitations of current BCIs. Recognizing the importance of user intention recognition for effective interface self-adaptations, we have focused on developing a user-agnostic Bayesian framework to track and infer hidden user target goals within the context of navigation. A variety of experiments with human subjects were conducted to evaluate and numerically validate the proposed methodologies and algorithms.

Contents

Notations

Abbreviations

ACC	Anterior cingulate cortex
ALS	Amyotrophic lateral sclerosis
ANOVA	Analysis of Variance
AP	Action potential
AR	Augmented reality
AT	Assistive technology
BCI	Brain-computer interface
BP	Bereitschaftspotential
CCA	Canonical correlation analysis
CDF	Cumulative distribution function
CNS	Central nervous system
CNV	Contingent negative variation
CPD	Conditional Probability Distribution
CVARS	Canonical variates with autoregressive spectral analysis
DBN	Dynamic Bayesian Network
DFT	Discrete Fourier Transform
DoFs	Degrees of freedom
ECoG	Electrocorticography
EEG	Electroencephalography
EMG	Electromyography
ERD	Event-related desynchronization
ERP	Event-related potentials
ERSP	Event related spectral perturbation
ErrPs	Error-related potentials
ERS	Event-related synchronization
ERN	Error-related negativity
EPSP	Excitatory postsynaptic potentials
FFT	Fast Fourier Transform
FOV	Field of view
fMRI	functional magnetic resonance imaging
fNIRS	functional near-infrared spectroscopy
GAD	Grand average difference
GUI	Graphical user interface
HCI	Human-computer interaction
HMD	Head mounted display
ICA	Independent component analysis

IPSP	Inhibitory postsynaptic potentials
IR	Intention recognition
ITR	Information transfer rate
LCD	Liquid-crystal display
LDA	Linear discriminant analysis
LED	Light-emitting diode
LIS	Locked-in syndrome
LS	Least squares
MA	Moving average
MAP	Maximum A Posterior
MEC	Minimum energy combination
MEG	Magnetoencephalography
MI	Motor imagery
MI-BCI	Motor-imagery-based BCI
ML	Maximum likelihood
MMN	Mismatch negativity
MSI	Multivariate synchronization index
mVEP	Motion-onset visual evoked potential
MVFH	Minimum vector field histogram
P300-BCI	P300-based BCI
PCA	Principle component analysis
PET	Positron emission tomography
PNS	Peripheral nervous system
POMDP	Partially observable Markov decision process
PSD	Power spectral density
RBF	Radial basis function
RMS	Root mean square
ROS	Robot operating system
SC	Shared control
SCI	Spinal cord injury
SSVEP	Steady-state visual evoked potential
SSVEP-BCI	SSVEP-based BCI
SNR	Signal-to-noise ratio
SPECT	Single photon emission computed tomography
SPN	Stimulus-preceding negativity
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TVEP	Transient visual evoked potential
VEP	Visual-evoked potential
VFF	Vector force field

Conventions

Scalars, Vectors, and Matrices

Scalars are denoted by upper and lower case letters in italic type. We use bold uppercase letters (e.g. $\mathbf{Z}$) to denote two-dimensional *matrices* and bold lowercase letters (e.g. $\mathbf{z}$) to denote *column vectors*, and conversely $\mathbf{z}^T$ for *row vectors*. An arbitrary vector $\mathbf{z}$ is composed of elements z_i and an arbitrary matrix $\mathbf{Z}$ is composed of elements z_{ij} (i^{th} row, j^{th} column). A p-variate random variable is denoted by an uppercase letter, e.g. Z, where it refers to a vector of p random variables, e.g. $Z = [Z_1, Z_2, \dots Z_p]^T$. Realizations of Z are denoted the same way as vectors (e.g. $\mathbf{z}$). An arbitrary diagonal matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ can be defined with its diagonal elements as $\text{diag}(d_1, \dots, d_n)$.

x or X	Scalar
$\mathbf{z}$	Vector
$\mathbf{Z}$	Matrix
$\mathbf{Z}^T$	Transpose matrix of $\mathbf{Z}$
$\mathbf{Z}^{-1}$	Inverse matrix of $\mathbf{Z}$
$\mathbf{Z}^+$	Pseudoinverse matrix of $\mathbf{Z}$
$\|\mathbf{Z}\|_F$	Frobenius norm of matrix $\mathbf{Z}$
$\text{Tr}(\mathbf{Z})$	Trace of matrix $\mathbf{Z}$, i.e. the sum of all diagonal elements
$\lambda(\mathbf{Z})$	Ordered eigenvalues of $\mathbf{Z} \in \mathbb{R}^{n \times n}$, i.e. $\lambda(\mathbf{Z}) = (\lambda_1, \lambda_2, \dots \lambda_n)$
$\lambda_1(\mathbf{Z})$ or $\lambda_{\max}(\mathbf{Z})$	The dominant eigenvalue of matrix $\mathbf{Z}$
$\sigma(\mathbf{Z})$	Ordered singular values of $\mathbf{Z} \in \mathbb{R}^{n \times p}$, i.e. $\sigma(\mathbf{Z}) = (\sigma_1, \sigma_2, \dots \sigma_{\min(n,p)})$
$\lambda_1(\mathbf{Z})$ or $\lambda_{\max}(\mathbf{Z})$	The dominant singularvalue of matrix $\mathbf{Z}$
$\mathbf{0}$	Zero-vector or zero-matrix, i.e. $\mathbf{0} = [0, 0, \cdots 0]^T$
$E[x]$	expected value of a one dimensional random variable X
$E[\mathbf{x}]$	expected value of a p-variate random variable X
$f(\cdot)$	Scalar function
$\hat{x}$	Estimated or predicted value of x
$\tilde{x}$	Estimation error, i.e. $\tilde{x} = x - \hat{x}$
$\overline{x}$	Average value of x
$\|\cdot\|_p$	p-norm

Subscripts and Superscripts

x_k	Realized value of x at time k or the k^{th} element of a vector $\mathbf{x}$
$x_{0:k}$	Time sequence of variable x, i.e. $x_{0:k} = (x_1, x_2, \cdots x_k)$
$x^{(k)}$	The k^{th} realization of a variable x
${}^k x$	The k^{th} element in an enumerable set $\mathcal{X} = \{{}^1x, \cdots, {}^kx, \cdots, {}^nx\}$
${}^{1:j}x$	A subset or the complete set of $\mathcal{X}$
${}^j x_k$	The realization of a variable ${}^j x$ at time k
${}^{1:j}x_k$	The sequence of realized values for the variables ${}^{1:j}x$ at time k
${}^{1:j}x_{1:k}$	The sequence of realized values for the variables ${}^{1:j}x$ at times $\{1, \cdots, k\}$

Symbols

General

F_s	Sampling rate
t	Time
k	Discrete time
N_y	Number of EEG channels
N_x	Number of brain sources engaged for specific task/event
N_h	Number of harmonics in SSVEP response
N	Number of time samples in an EEG segment
O	Buffer overlap
T	Interface temporal resolution
b	Bit rate
$\mathbf{A}$	Mixing matrix
$\mathbf{Y}$	EEG data segment
$\mathbf{X}$	Source activity for a period of time
$\mathbf{X}_f$	SSVEP source model for frequency f
$\mathbf{\Psi}$	Additive noise term
$\mathbf{W}$	Spatial filter
$\mathbf{C}_{yy}$	True value of the auto-covariance matrix of random variable $\mathbf{y}$
$\mathbf{Q}_{yy}$	Sample spatial auto-covariance matrix of random variable $\mathbf{y}$
$\mathbf{Q}_{xy}$	Sample cross-covariance matrix between two random variables $\mathbf{x}$ and $\mathbf{y}$
$U[z]$	Z-transform of discrete sequence $u[k]$
P_{AR}	Spectral density of an autoregressive process
$F(f,t)$	Short-Time Fourier Transform (STFT) of a data segment
$\mathbf{T}$	Coherence matrix
$\mathbf{C}_{X,Y}$	Joint covariance matrix
E_m	Misclassification error
ρ_i	Canonical correlation coefficients

Adaptive BCIs for Robotic Embodiment

Θ	Head's yaw angle
Φ	Head's pitch angle
H	Gaze frame
W	Global or world frame
$\mathbf{P}_k$	Belief vector at time instant k
$s(\mathbf{P}_k)$	The non-uniformness of the belief vector at time instant k
u_k	User input at time instant k
$\mathbf{h}_k$	User gaze at time instant k
$\mathbf{x}_k$	Robot pose at time instant k
${}_W T^H$	Coordinate transformation from frame W to H
${}^k\mathbf{g}$	Pose of goal k
${}^m P_k$	Probability of goal m at time instant k
${}^m\mathbf{i}_k$	Mental path plan towards goal m from the robot's location at time k

1 Introduction

1.1 Background and Motivation

Brain-computer interfaces (BCIs) are *direct* communication and control channels between the brain and artificial devices like computers, prosthetic limbs or robots. The *directness* here refers to the fact that BCIs bypass the natural neural (outside the brain) and muscular pathways [1] which are required for all other kinds of human-human and human-machine interaction.

Due to different social, economic and scientific driving forces, BCIs have drawn the interest of a wide spectrum of scientists since its inception in 1970's by Vidal [2]. Most importantly, due to its directness, BCIs have a great potential use in assistive technology (AT) and rehabilitation for people with physical disabilities [3, 4]. Interest in AT, on the other hand, is driven by the emergence of ageing societies worldwide, and the social and economic implications this has in the future. The continuously increasing global population has been experiencing unprecedented shifts with respect to its age distribution throughout the past decades. According to estimates of the United Nations, the number of older persons has tripled throughout the past 50 years and is expected to more than triple in the next 50 years [5]. This will lead to a great imbalance between the number of people in need of health care due to age-related physical impairments, and the number of available caregivers. AT lends itself to counteract this foreseenable imbalance and, at the same time, it is expected to result in direct and indirect economic benefits [6]. In this regard, BCIs play an enabling role, especially when it comes to patients with locked-in syndrome and Amyotrophic Lateral Sclerosis (ALS) among others. Furthermore, BCIs have a restorative function as they can be combined with conventional therapy for rehabilitation in stroke people [7]. Additionally, according to different predictions, BCI-based gaming will probably be the first mass market of non-medical BCIs [8] and, albeit in its infancy, some promising examples already exist [9]. The field of BCIs is also driven by scientific enthusiasm and curiosity. Thanks to the integration of BCIs and assistive and immersive technologies, many questions like body ownership [10] and mind-brain dualism [11] are currently addressed in a systematic way.

Typically in BCIs, the users' brain signals are decoded into machine actions using a mapping that is known to both the users and the devices they control or communicate with. Due to its non-invasiveness, portability and relatively low cost, the Electroencephalography (EEG) technique is the most common way of measuring the brain activity in BCIs, and therefore we primarily discuss EEG-based BCIs throughout this work. One can also differentiate between two major categories of BCIs. Firstly, passive BCIs make use of spontaneous brain signals without demanding active user participation or voluntary mental effort. These include affective BCIs and workload monitoring systems [12]. On the other hand, active BCIs require user engagement in performing some sort of mental activity

or another, e.g. by attending to a specific stimulus among others or imagining a limb movement. For their relevance to communication and control applications, active BCIs are the main concern of this work.

From the continuous stream of EEG data, active BCIs make use of particular temporal and spectral patterns that occur in time about precisely timed events and at well localized brain sites. These include, but are not limited to, P300, steady-state visually evoked potentials (SSVEPs), error-related potentials (ErrPs), and event-related synchronization/desynchronization (ERS/ERD). SSVEPs are natural responses to repetitive visual stimulation patterns, whereby e.g. when the retina is excited by a flickering light in the range $4 - 100$ Hz [13], the observed brain activity over the visual cortex is typically characterized by a power increase at the same and higher harmonics of the driving frequency of the stimulus. While presented with concurrent stimuli with different flickering rates, users may choose to attend to one of them and thereby actively convey to the system their action of choice. The P300 signal, on the other hand, is elicited in the brain around 300 ms following rare stimuli presented randomly intermixed with more frequent standard stimuli. BCIs based on P300 often present, highlight or intensify all possible user commands in a random sequential order and users, similar to the SSVEP-based BCIs, are required to attend to the command they intend to choose. Error-related potentials (ErrPs) spontaneously appear time-locked to feedback presentations that indicate committed or observed errors. For ErrPs to be evoked, users need to be attentive when the feedback is presented. ERD/ERS refer to the case when events induce a reduction or an increase in the synchrony of the underlying brain activity in specific frequency bands [14]. To use ERD/ERS for control and communication purposes in BCIs, users need to learn by themselves a strategy how to modulate their brain signals, e.g. by imagining movements of different parts of the body.

Being observable under different conditions and with a certain level of repeatability and being characterized with relatively high signal-to-noise ratio (SNR), the aforementioned EEG patterns lend themselves relatively well to the decoding process. Therefore, they have been integrated within BCI applications spanning a wide spectrum of conventional human-machine interaction. These include, but are not limited to, controlling powered wheelchairs [15], telepresence robots [16], prosthetic limbs [17], gaming [9], web browsers [18] and spelling applications [19, 20]. Recently there has been an increasing interest, including this work, in using BCIs for physical embodiment systems and control of robotic avatars [21–23]. Such technology is beneficial to people with disabilities and healthy people alike, enabling them to be present in remote environments, and engage in physical interaction. Fig. 1.1 shows an overview of a typical BCI-based immersive physical embodiment system, highlighting the main components of state-of-the-art BCIs. The data acquisition step is commonly achieved with the EEG technique, but other methods were used as well [24, 25]. The signal processing step aims at improving the quality of the brain signals, by filtering out what is considered as noise, where the definition of noise varies according to the situation. The feature extraction and the classification steps are tightly connected, as the former step aims at providing the best features that can be used for reliable and robust classification. Stimulation (in active BCIs) and feedback (in general) are presented often through the visual modality, but again, different modalities can be and were frequently used, e.g. proprioceptive, auditory or even direct cortical stimulation.

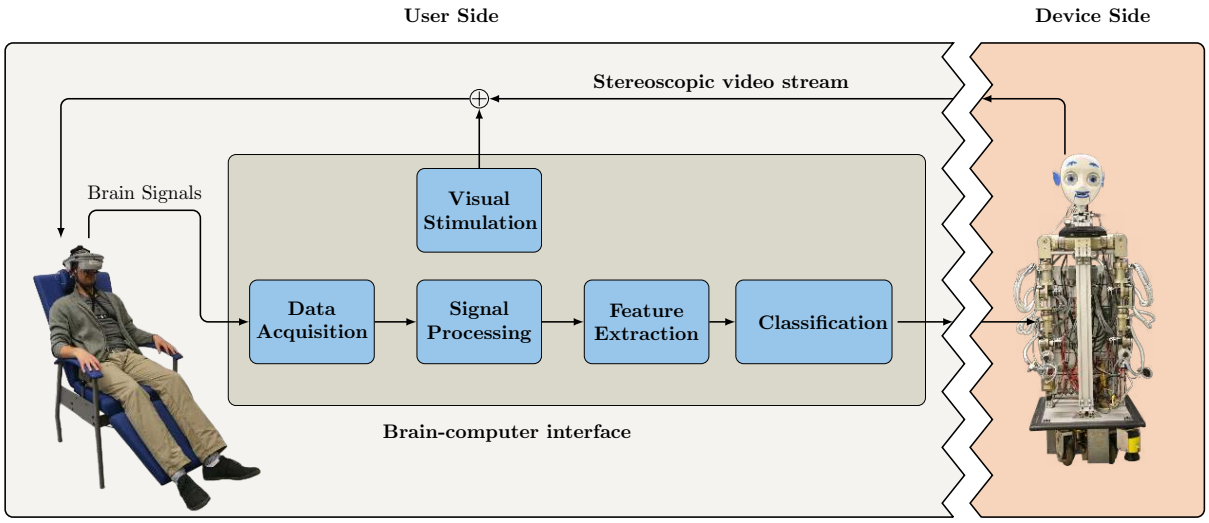


Figure 1.1: General overview of a BCI-based physical embodiment system. At the heart of the system is the BCI main modules typically found in almost all BCI systems. The device side here represents a robot avatar but can be replaced with any other external device, e.g. a wheelchair or prosthetic limb.

1.2 Problem Definitions, Challenges and Possible Remedies

Typically, the human user in immersive physical embodiment systems receives a continuous 3D stereoscopic video stream from the ego-perspective of his/her robot avatar, which he/she views through a head-mounted display (HMD), as depicted in Fig. 1.1. The locality of the robot is not relevant here, but we assume that the robot resides in a remote environment. BCI visual stimulation hereby is provided overlaid on the video stream and users selectively attend to one of the available commands. Ultimately, the goal is to enable users to interact with objects and other humans in the remote environment through intuitive and fluent BCI-based interaction, so that they act in the remote environment *as if* they were there, and move the robot's body *as if* it were theirs. This indeed requires a complex transformation matrix of the afferent and efferent pathways in the nervous system of the human user. The fact that BCIs do not require body movements at the user side allows e.g. to explore integrating multiple afferents in a congruent way to the robotic movements perceived through the visual modality. This is believed to be essential in order to embody an external object [26].

There are yet some limiting factors that make the realization of this objective a highly challenging task. These limitations are common to all BCI applications, but are more pronounced for robotic embodiment applications due to the varying time scales in which they, similar to own-body control, operate.

Firstly, decoding of brain signals is prone to errors due to artifacts and spontaneous background EEG which can be measured alongside the signals of interest. This "lost in translation" problem [11] has been addressed in the literature in many different ways, that literally investigated all aspects of the BCI pipeline, from the data acquisition to the

classification step. In the acquisition step, there is a continuous effort to come up with novel neuroimaging techniques that overcome the limited temporal and/or spatial resolution of the techniques currently available in the repertoire. In the signal processing step, novel methods are sought aiming at more powerful separation of the signals of interest from noise and other artifacts. Among these methods, spatial filters have proved to be powerful in many different BCI systems. On the other hand, the feature selection module highly benefits from advances in the field of neuroscience and findings that better identify the neural correlates of different cognitive functions. Such findings might guide the feature selection process in the temporal, spectral and spatial domains. The reliability and robustness of the classification step is highly dependent on the steps that precede it in the pipeline. Classification is done typically in a supervised fashion where a set of representative examples is obtained per subject from training sessions, and used to learn the discrimination boundaries between the different classes under consideration. Sometimes, as it is the case for interaction ErrPs classification, the required training sessions last for long time, and small changes in the experimental task typically require recording the lengthy training sessions anew. Classifier transferability is therefore highly desirable. On this account, there have been great efforts tackling classifier transferability across tasks and subjects and over time.

Secondly, the bandwidth and bit rate of BCIs are generally beyond the requirements of fluent interaction and is still far beyond than that of other classical user interfaces like keyboards and joysticks. This is partially due to the imperfect decoding of brain signals, but mainly due to limitations in the number of commands available for interaction, and the relatively long time required to issue a single command. What essentially contributes to the relatively poor temporal resolution of BCIs, is the trial averaging step which is often required to enhance the SNR, e.g. as in P300-based BCIs, and involves repeating the presentation of the different stimuli multiple times.

Thirdly, there is a great discrepancy between the temporal resolution and the degrees of freedom (DoFs) of the user interface and that of the robotic system, which is equipped with manipulation and locomotion capabilities. This implies a similar discrepancy between what the robot is able to do, and what the interface is able to mediate.

Taking uncertainty in user commands into consideration is therefore necessary to arrive at reasonable trade-offs between speed and accuracy of interaction. Adaptive BCIs, on the other hand, offer a way to overcome the limitations on the bit rate and the number of commands available for interaction. Additionally, adequate levels of robot autonomy are required to bridge the gap between the DoFs/temporal resolution of the robot avatar and the user interface, so that delayed user commands do not block the robot avatar from responding quickly to the rapid environmental changes in the remote environment. Controlling for the possible autonomy levels, that range from fully manual to goal-oriented or fully autonomous control, further allows for multilevel communication channels between the user and his/her robot avatar, with varying time scales. Together, these methods can push the limits of current BCI systems by making the best use of available contextual information, from the user, environment and history of interaction.

1.3 Organization and Main Contributions

This thesis deals with the problem of designing fluent and intuitive EEG-based BCI systems for immersive physical embodiment systems. In a general sense, BCIs mediate between the user (intentions) and the devices they communicate with or control (actions). It is this synergy between the intentions and actions and vice versa, from which we expect the fluency of interaction to emerge. Given the interdisciplinary nature of BCIs and their applications, we tried to approach the interaction problem from within different modules in its pipeline. Different methods and algorithms were developed to this end, and were evaluated and empirically validated with a series of experiments and studies with human subjects. All these studies are part of a larger project, which is approved by the Ethics Committee of the Faculty of Medicine of the Technical University of Munich (TUM).

The remaining of this thesis is organized as follows. Chapter 2 introduces the fundamentals and related topics required to understand the thesis. Chapters 3 and 4 address the problem of detecting EEG patterns typically used in BCIs aiming at reducing error rates and the implications of these errors. Chapter 5 presents adaptive, contextual and application-specific BCIs to overcome the bandwidth limitations of current BCIs, where a special focus is laid upon navigation tasks in robotic embodiment systems. Finally, this thesis concludes in chapter 6 with a summary of the main results in the core chapters and an outlook onto the future of BCIs and BCI-based robotic embodiment systems. The main contributions are outlined below.

Chapter 3 provides a genuine theoretical analysis to the foundations of the state-of-the-art detection methods of SSVEPs and the spatial filtering step involved in all of them. By considering the canonical correlation analysis (CCA) as a basis for this analysis, it is shown how these methods relate to each other and how they are essentially variants of the CCA. Empirical analysis shows that in low SNR regimes, most of these methods fail to provide reliable estimates as they ignore the presence of noise and interferences from background EEG activity, especially from the alpha brain waves, which might overlap with the driving frequencies of the different stimuli. The minimum energy combination (MEC), on the other hand, estimates the signal and noise powers from the spatially filtered signals, and therefore can handle different SNR regimes. Spatial filtering hereby involves a step of dimensionality reduction that is performed with discarding 90% of the noise power, where noise is defined here as the projection of the EEG signals onto the null space of the assumed SSVEP signals. Though this dimensionality reduction has proved to be effective, the criterion, on which it is based, seems somehow artificial. Based on this analysis, a new detection method is proposed, namely CVARS. The new method combines the power of the canonical variates and that of the autoregressive spectral analysis in estimating the signal and noise power levels, respectively. When evaluated with real EEG data in an unsupervised fashion, CVARS is shown to slightly outperform other state-of-the-art methods. We also show the tight relation between the standard procedures in the CVARS and MEC methods, and why the new method would be favored over MEC. Finally, we show that CVARS, when used in a supervised fashion, is able to provide reliable, accurate and rapid detection of SSVEPs, regardless of the SNR level in the raw EEG signals. On a parallel vein, comparative evaluation of different stimuli presentation/viewing methods shows that visual stimulation through HMDs is effective, and indeed results in higher

accuracies compared to monitors. Results also show that binocular stimuli viewing results in either binocular summation or inhibition, according to whether the stimulation is done with flickers that are completely in-phase or anti-phase, respectively. To our knowledge, the latter result has not been shown before.

Chapter 4 deals with the interaction error-related potentials (interaction ErrPs) and their classification. The importance of interaction ErrPs to BCIs stems from their potential use as a validation step for the decoding results obtained with other BCI paradigms. As classification of interaction ErrPs typically requires long training sessions to collect representative examples of EEG epochs that correspond to correct and incorrect trials, the problem of classifier transferability is examined in detail. Hereby, it is mainly discussed whether there are invariant features of interaction ErrPs and the consequences these invariants, if any, might have for their classification. Possible invariants are examined with respect to: (1) human mental processes that are required to assess interface actions (2) time (3) subjects (4) the way the continuous data is pre-processed. The three different experiments, designed in order to answer these questions, are presented. Results thereof show that interaction ErrPs are quite invariant with respect to (2) and (3) and exhibit high variability with respect to (1) and (4). These results translate directly to whether a trained linear discriminant analysis (LDA) classifier from one condition can be potentially used with other subjects, in other times, or for other tasks, and whether ErrPs obtained with slightly different processing pipelines can be compared with each other. Prior research on ErrPs have partially tackled some of these issues at separate occasions, which helped to formulate first hypotheses and guided the design of the three interfaces/experiments reported in this chapter. With this work, the aim is to ground irrelevant factors in the experimental design and in the preprocessing step so that concrete conclusions can be drawn with respect to the different sources of invariance and variability under consideration. A new task for P300 interaction is further proposed, which simulates adaptive BCIs in the sense that the interface elements dynamically change. Additionally, results from the analysis of interaction ErrPs, suggest that the novel inplace feedback presentation strategy, introduced with one of the experiments, is expected to circumvent the identified transferability problem among tasks as the mental processes needed to assess interface actions as correct or incorrect should be the same regardless of the nature of the underlying application. However, such claim needs to be verified with further experiments.

Chapter 5 addresses the specifics of robotic embodiment applications with emphasis on navigation tasks and proposes interface adaptations as one way, and perhaps the only way, out of the bandwidth bottleneck of current BCIs. Hereby, it is argued that in order for adaptive BCIs to be of any benefit, they should be able to reason about hidden user intentions, so that interface initiatives can be done in line with what the user has in mind. In particular, the problem of user intention recognition within the context of avatar navigation is treated in this chapter. Navigation is assumed to take place in structured environments with known maps and predefined and enumerable set of goal locations, whereas interaction between the robot avatar and the user is mediated with discrete interfaces, including SSVEP-based BCIs, and is limited to incremental commands that determine the direction of the robot's rotation or translation. The robot is assumed additionally to operate in the collision avoidance mode. To this end, a novel recursive Bayesian update rule is formulated so that it continuously tracks and infers users' end goals from user commands,

information about their gaze direction, and the relevant history of interaction (i.e. from the moment the robot starts to recede away from one of the defined goals). The output of the intention recognition module comes in the form of a belief vector, that sorts all goals according to how probable they are given the previously observed evidence. The intention recognition module additionally provides a measure of prospective confidence in its beliefs on the basis of the non-uniformness of these beliefs. A new metric is proposed to this end. The novelty of the Bayesian approach stems from the novel intuitive heuristics used in modeling user commands, user gaze and history of interaction. These heuristics are based on behavioral patterns observed in the general population during task execution, and thus the Bayesian inference system can be used in a plug-and-play fashion with the general population. Moreover, the proposed system allows to asynchronously respond to arriving evidence from different source (i.e. user commands and user gaze). Direct and indirect measures are adopted as performance metrics of the resulting Bayesian model. Direct measures reflect the accuracy of the inferred belief vectors, whereas indirect measures can be obtained only by making use of the belief vector in one way or another. To this end, we developed a simple probabilistic shared control scheme that builds upon the belief and the confidence of the intention recognition module, and silently modulates the magnitude of the translational and rotational steps. The number of commands needed to finish specific navigation tasks therefore is used as an indirect measure of the efficiency of the intention recognition module. This chapter additionally details the empirical evaluation of the intention recognition system with experiments with human subjects. To our knowledge, this the first time these experiments are done in BCI-based immersive physical embodiment settings. Most recruited subjects were able to complete the assigned tasks. Experimental results demonstrate the ability of the user-agnostic Bayesian intention recognition to infer the hidden user goals after a number of interactions. Additionally, the integration of the output of the intention recognition into shared control is shown to result in less user effort, measured as the number of commands required to accomplish the assigned tasks, indirectly proving the efficiency of the inference system. Several complementary implementation details guaranteed task completion.

Finally, chapter 6 summarizes the merits of the thesis and highlights future directions of possible improvements.

2 Foundations of Brain-computer Interfaces

Brain-computer interfaces bypass the natural muscular and neural pathways distributed over the human body by extracting useful control and communication information directly from the brain signals and directing this information towards external and artificial devices, e.g. prosthetic limbs. As has been shown in Fig. 1.1, BCIs (including the ones that will be encountered in this work) typically consist of data acquisition, processing, feature extraction, classification and stimulation modules. The following sections try to shed some light at each of these modules, and at other foundations necessary to understand this thesis. The limitations and the DoFs inherent in current EEG-based BCIs will be highlighted where appropriate.

This chapter proceeds as follows. Sec. 2.1 introduces the different neuroimaging techniques used in the data acquisition module and highlights, in particular, the fundamentals of electroencephalography (EEG) technique and the nature of the neural activity it measures. Sec. 2.2 outlines the main temporal and spectral patterns found in EEG signals. Sec. 2.3 provides a general overview of the signal processing methods that are typically used to enhance the SNR of these patterns. In Sec. 2.4, feature extraction and classification methods used in the thesis are presented. A brief description of popular types of BCI is provided in Sec. 2.5. The stimulation module is briefly described in Sec. 2.6 with emphasis on the visual stimulation modality and the visual neural pathways. A summary of this chapter is provided in Sec. 2.7.

2.1 Neuroimaging Data Acquisition

The data acquisition is the first step in any BCI and concerns itself with providing some sort of measure of the brain neural activity. There are several, direct and indirect, techniques that are typically used to this end. Direct methods either measure the electric field produced by this activity as in *Electroencephalography* (EEG) or the produced magnetic field as in *Magnetoencephalography* (MEG). On the other hand, indirect methods build on the premise that neural activity in the brain is accompanied by chemical and metabolic activity that can be measured e.g. with positron emission tomography (PET), functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS) or single photon emission computed tomography (SPECT), just to mention some. Each of these methods comes along with a unique set of advantages and disadvantages. Due to its absolute noninvasiveness, superior temporal resolution, mobility and relatively low cost, scalp EEG is the mostly used technique in BCIs and in monitoring brain activity in general [27]. For the same reasons, EEG-based BCIs will be the main focus of this work. Sec. 2.1.1 provides a general introduction to the neural activity in the brain, and Sec. 2.1.2 introduces the specific aspects of this activity that can be measured with EEG. Standards for EEG electrode placement are introduced in Sec. 2.1.3.

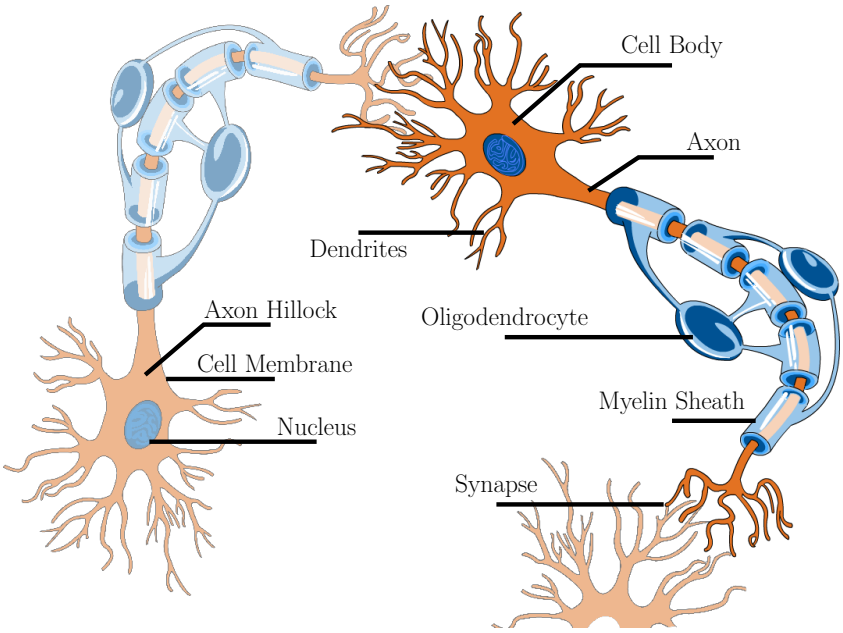


Figure 2.1: The structure of a typical neuron and its synaptic connections with adjacent neurons. Neurons vary in size and in structure, e.g. some neurons lack the dendrites or the myelin shealth.

2.1.1 Neural Activity in the Human Brain

Most of the brain functions, like memory, motor control, or supervision of glandular secretion, just to mention some, are carried out by the *neurons* and their complex interconnections, i.e. *neural networks*. Around 85 billion neurons are found in the human brain [28] immersed in a fluid known as the *interstitial fluid*. Neurons are typically made of a *cell body (soma)*, *dendrites* and an *axon*, all enclosed by the cell membrane that separates the interiors of the cell from the outside world. The cell body is composed of a nucleus and other typical cellular metabolic machinery. Dendrites are the receiving terminals of neurons, and often form a forked array of tree-shaped branches extending from the cell body. The axon, on the other side, is a tapering cylindrical projection that joins to the cell body at a cone-shaped elevation called the *axon hillock* [29]. Fig. 2.1 depicts the structure of a typical neuron.

The different neurons communicate with each other through two types of special junctions, referred to as *electrical* and *chemical synapses*. Electrical synapses are faster than their chemical counterpart as they facilitate direct ion flow between neighboring neurons in both directions. They are however a distinct minority [30]. On the other hand, the flow of electrical nerve impulses in chemical synapses is strictly unidirectional and requires the release of special chemical neurotransmitters from the presynaptic neuron (i.e. the sending neuron) into the interstitial fluid that separates it from the postsynaptic neuron (i.e. the receiving neuron). The released neurotransmitters are received by the postsynaptic neurons through specialized receptors.

Neurons, similar to other cells in the human body, maintain a specific potential difference

across their membranes, often called the *membrane potential* [1]. In resting state, the value of the membrane potential is typically around -70 mV. Deviations of the membrane potential from its resting level is the source of all electrical neural activity. Such deviations are triggered by internal (e.g. blood pressure or activity of other neurons) or external stimuli (e.g. touch or light). When such stimuli bring the membrane potential at the axon hillock to a level that is larger than the firing threshold (typically -55 mV), a nerve impulse or *action potential* (AP) is fired and propagates downstream across the axon towards the axon terminals and from there to the postsynaptic cells (i.e. downstream neurons, muscular or glandular cells). An AP lasts for a few milliseconds only. Should the stimulus not be strong enough to bring the membrane potential to a level that surpasses the firing threshold, an AP does not form, and the neuron returns shortly back to its resting state. Similarly, an AP does not fire when the incoming stimuli make the membrane potential more negative (i.e. hyperpolarization). With respect to its effect on the postsynaptic AP, a neurotransmitter can be inhibitory (i.e. leading to hyperpolarization) or excitatory (i.e. brings the membrane potential towards the firing threshold). The resulting membrane potential at the postsynaptic neuron for the two cases are respectively called the *inhibitory postsynaptic potential* (IPSP) and *excitatory postsynaptic potential* (EPSP). A postsynaptic neuron might be subjected to multiple IPSPs and EPSPs, in which case, the net postsynaptic potential determines whether an AP is going to be initiated or not.

2.1.2 Electroencephalography (EEG)

Electric field potentials recorded from multiple electrodes placed at different locations on the scalp is called scalp EEG. When electrodes are surgically implanted under the skull (but not within the brain) the technique is called intracranial EEG (iEEG) or electrocorticography (ECoG) [31]. When electrodes are placed inside the brain, the recording is referred to as local field potentials (LFPs). Magnetoencephalography (MEG) is closely related to EEG, and alternatively measures the weak magnetic field produced by the underlying neural activity using superconducting sensors placed at short distances around the head, and often such setup is done inside a magnetically shielded room, rendering MEG of inferior portability and higher cost compared to EEG.

Typically, large populations of spatially aligned neurons with temporally overlapping activity (not necessarily initiating APs) are required in order for EEG and MEG to be observable at the recording sites [32, 33]. As argued in [34, 35], EEG (and MEG) mainly reflects the activity of IPSPs and EPSPs rather than the stronger action potentials. This is due to their longer duration which allows for their superposition to take place compared to the briefly lasting APs. The net activity of neurons is often modeled as *current dipoles* [36]. This way, fluctuations in EEG and MEG can be thought of as manifestations of changes in the magnitude, position and orientation of these dipoles. Source localization concerns itself with reconstructing these dipoles from the observed potentials or magnetic field. A major difference between EEG and MEG is that EEG is sensitive to neural sources of arbitrary orientations but mostly sensitive to activity oriented perpendicular to the surface of the head, whereas MEG is mainly sensitive to tangential sources [37]. The presence of such

[1] The membrane potential results from concentration of potassium (K^-) and chloride (Cl^-) ions inside the cell and the concentration of sodium (Na^+) and calcium ions (Ca^+2) outside.

tangential and radial sources in the head results from the folding of the cerebral cortex. Cortical pyramidal cells in the cerebral cortex are believed to be the main contributors to the measured scalp EEG. These neurons, making up around 70% of the neocortex [38], have elongated apical dendrites that are aligned perpendicular to the scalp [35].

Due to the remoteness of the recording sites in scalp EEG from the sites of the underlying brain activity, neural currents need to pass through the surrounding tissue and through different mediums e.g. cranial meninges, skull and scalp, with different conductivities and complex geometries [39]. Consequently, the recorded EEG activity is a spatially smoothed version of the underlying source activity as these mediums act mostly as spatial filters. This is often referred to as the *volume conduction* of the head.

2.1.3 EEG Electrode Placement

In order to reduce measurement variability in repeated EEG recordings on different times and with different subjects, EEG electrodes are typically placed over the scalp according to the international 10-20 system. Hereby, measurement sites are defined with respect to each other and relative to four anatomical references, i.e. the nasion, inion, left pre-auricular point (LPA), and right pre-auricular point (RPA). The distance between the nasion and inion defines the total posterior-interior distance, and the distance between LPA and RPA defines the total right-left distance. The distance between adjacent electrodes is set to either 10% or 20% of the total posterior-anterior distance or the right-left distance, and hence the 10-20 name of the standard. Recording sites are typically identified with a combination of letters and numbers, that are derived from the different lobes and sides of the cerebral cortex that underly them. Frontal, parietal, temporal and occipital sites thereby identified with the letters “F”, “P”, “T” and “O” respectively. Frontal pole and central electrodes are identified with the letter “C” and the letter combination “Fp”. Left and right hemisphere sites are given respectively odd and even numbers, whereas midline sites are appended with the letter “z”.

There are 21 sites defined with the 10-20 system in total. Since this is not large enough for most purposes, extensions to the system are available such that extra electrodes are introduced between the already defined ones and are identified accordingly. For instance, sites that lie between frontal and central sites are identified with the letter combination “FC” and those that lie in between parietal and occipital sites are identified with “PO”. Fig. 2.2 shows the distribution of the electrodes in the extended 10-20 standard (sometimes referred to as the 10-10 standard). Electrode placement throughout this work is done according to the defined sites in this system. State-of-the-art commercial EEG acquisition systems often come with elastic EEG caps where recording sites are accurately marked.

Often, electrical activity at each site is measured with respect to one or more reference electrodes. Additionally, a ground electrode is used for common mode rejection in the differential amplifiers used in the acquisition apparatus [40].

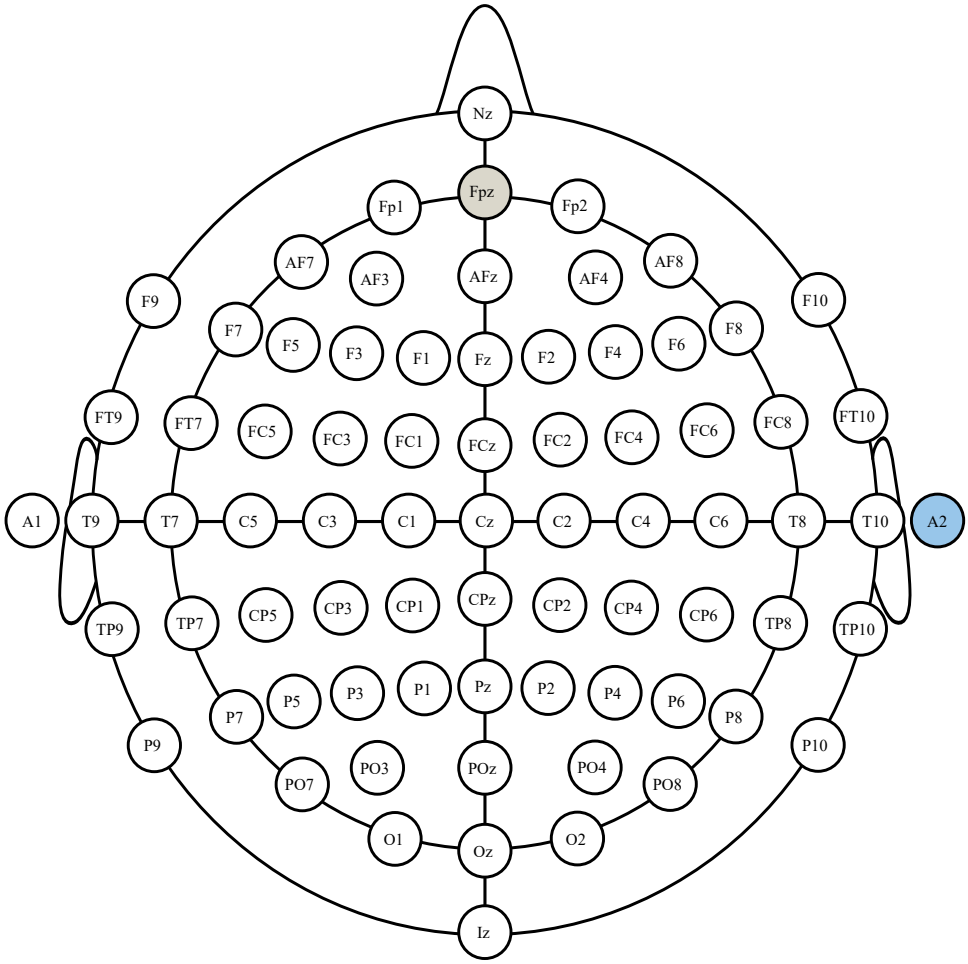


Figure 2.2: Electrode placement according to the extended 10-20 system. The reference electrode is placed at the right earlobe throughout this work and the ground electrode at Fpz. Illustration modified from original by Marius 't Hart licensed under CC BY-NC-SA 3.0.

2.2 Patterns in the EEG Signals

Scalp EEG signals can be fully represented either in the time or the frequency domain, where each representation can also be fully reconstructed from the other. In different BCI applications, however, one representation or the other might be more convenient for analyzing the EEG patterns under consideration. The following subsections outline the main spectral and temporal patterns in EEG signals.

2.2.1 EEG Patterns in the Time Domain

EEG data in the time domain is characterized by continuous fluctuations that instantaneously reflect the underlying brain activity. These fluctuations lie normally in the range $[-100, 100]$ μV as they unfold over time [35]. Due to the high complexity of interpreting continuous data, and attributing them to the underlying source activity, EEG is primarily investigated in controlled experiments about precisely timed events, that can be paced externally (e.g. presentation of auditory, tactile or visual stimuli) or internally (e.g. mental thought or activation of muscles). In the latter case, however, the time of the event onset can be revealed only in an indirect way (e.g. with EMG).

Reproducible brain activity observed in EEG about the onset of specific experimental events is referred to as *event-related potentials* (ERPs). The amplitude of typical ERPs ranges from 1 to 20 μV [41], which is clearly below the noise level. Noise in EEG signals can be of encephalic origin (e.g. cerebral background activity) and non-encephalic sources (e.g. movement of the eye or the head). *During experiments*, non-encephalic noise can be reduced by restricting subject movements, but these cannot be completely eliminated. A standard *post-experiment* noise reduction practice in ERP research is to average all epochs extracted about the events of interest (after manually or automatically rejecting artifact-contaminated epochs). This leaves ERPs reflecting mainly time-and-phase-locked brain activity to the onset of events under consideration. Averaging epochs per experimental condition and across trials and groups (i.e. different subjects) is typically referred to as *grand averaging*. Traditionally, ERPs were referred to as evoked potentials since it was believed that they were primarily evoked by the presentation of stimuli. The neutral term currently used, that is ERPs, reflects the fact that these potentials might as well arise from the cognitive, affective and motor demands of the situation [35].

Observed ERP *peaks*, i.e. local minima and maxima, should be carefully distinguished from the underlying cortical and subcortical activity that is associated with particular neural or psychological processes, referred to as *components* [42]. More on this in Sec. 2.3. ERP components are often differentiated and named according to their polarity, latency, duration and spatial distribution.

The following presents some of the commonly encountered ERPs and their component(s), and some of the experimental conditions in which they are typically observed. The interested reader might refer to [42] for more details.

- *P300* is a positive deflection that is typically observed in oddball experimental paradigms around 300 ms following infrequent and irregular target stimuli presented intermixed with other standard and frequent nontarget stimuli. P300 is mostly

visible at the parietal and central electrodes and its peak amplitude was found to be negatively correlated with the frequency of the target stimuli. In three-stimuli experimental paradigms (target, nontarget and distractor stimuli), it is often differentiated between two subcomponents of P300, namely the frontally distributed P3a and the centroparietal P3b. Active attention to stimuli is required for P300 to be elicited.

- *The Mismatch Negativity (MMN)* is typically observed in oddball paradigm experiments after an auditory (or somatosensory) stimulus that deviates from repeated standard stimuli. The deviant stimulus might differ in terms of its frequency, duration, intensity or locus of origin from the standard stimulus [42]. MMN is characterized with a negativity that spans the post-stimulus time range from 100 ms to 250 ms and is largest at central midline sites. MMN can be elicited even when subjects are instructed to pay no attention to the stimuli.

- *N400 and P600* are language-related ERPs, where N400 appears in reading tasks as a response to contextually inappropriate words and P600 appears after contextually correct but physically deviant words (e.g. for words shown with larger font than other words) [42].

- *Visual-evoked potentials (VEPs)* are responses to brief visual stimulation e.g. by a flash of light or pattern reversal. For instance, a brief stimulation with flashing light gives rise to several components that vary in their locus and latency. Short-latency components appear mostly at periocular sites, the midrange components appear mostly at the vertex (Cz) and parietal sites (e.g. P3), and the long-latency components appear in the vertex as a peak around 50-70 ms (C1), a positive peak around 100 ms (P100), a negative peak around 130 ms (N130) and another negative peak at about 200 ms (N200) [42]. These components can be written as C1-P100-N130-N200 pattern. Similarly, brief pattern reversal (e.g. checkerboard pattern reversal without luminance change) gives rise to a N70-P100-N135 pattern that appears mostly at the occipital sites referenced to the anterior scalp [42, 43]. Furthermore, recordings from experiments where stimulation was performed within individual hemifields, a reversal is observed such that the ipsilateral occipital sites showed NPN pattern where the contralateral sites showed a PNP pattern [42]. These VEPs are often referred to as *transient* VEPs or TVEPs. On the other hand, when flashing or pattern reversal stimulation is performed for an extended period of time and with a high rate, the TVEPs will overlap as a manifestation of the overlap in the activity of the underlying neural circuitry. This gives rise to what is often referred to as *steady-state* VEPs or SSVEPs. Chapter 3 provides in-detail description of the SSVEPs and their detection methods. Furhtermore, motion-onset VEPs (mVEPs) are characterized with P1-N2-P2 pattern and appear as a response to visual motion.

- *Error-related potentials (ErrPs)* is a family of ERPs that appear in human EEG time-locked to the occurrence of errors committed by the subjects themselves, or by other humans or interfaces. Due to the relevance of ErrPs to BCIs as a potential validation step, chapter 4 discusses this kind of ERPs in details.

- *Anticipation potentials.* In ERP research, one comes across three special ERPs characterized by negative slow waves and are typically linked to anticipation. These are: the *Bereitschaftspotential* (BP), the *contingent negative variation* (CNV) and the *stimulus-preceding negativity* (SPN). BP and CNV are sometimes considered as response-related potentials [44]. BP or the readiness potential can be observed mainly prior to voluntary and self-paced movements. Usually, one can differentiate between an early slow negative wave or *early BP* (2000 to 1500 ms pre-movement) and a steeper negative-going wave or *late BP* (400 to 500 ms pre-movement) [42, 45]. CNV, on the other hand, is observed in between two stimuli presented in sequence with predictable inter-stimulus interval, where the first stimulus is a warning of the second one which demands a quick response (e.g. button press) [35, 46]. The CNV was observed to reach its maximum at the onset of the second stimulus. The SPN is another slow negative shift preceding anticipated stimuli that provide significant information [42].

2.2.2 EEG Patterns in the Frequency Domain

The first meaningful piece of information that was extracted from EEG recordings is related to the presence of alpha and beta waves, and hence the names, as the inventor [2] of EEG discovered these waves. Ever since, many other regular waves have been observed in the brain activity recorded with EEG (or MEG). All these waves/bands are defined by a band-limited frequency range. The major waves are listed below in the order they appear on the spectral axis.

- *delta band (0-4 Hz)* appears mostly during sleep, and often is considered as an index of the deepness of sleep.

- *theta band (4-7.5 Hz)* is associated with access to unconscious material, deep meditation and is probably related to the level of arousal [47].

- *alpha band (8-13 Hz)* is most visible in the occipital lobe when eyes are closed and declines with background illumination [48]. The tau waves appear in the same frequency band over the midtemporal cortical regions and are blocked by auditory, rather than visual, stimulation. Likewise, the so called Rolandic mu waves (or upper alpha) are observed over the sensorimotor cortical areas and decline with simple motor movements like moving a finger or a toe [34].

- *beta band (13 - 30 Hz)* is most visible in the frontal and central regions in waking state. It is typically associated with active thinking and attention, and declines over central areas with motor activity or tactile stimulation [47].

- *gamma band (>30 Hz)* is modulated by sensory input and associated with working memory and attention [49].

[2] Hans Berger (1873 - 1941)

In addition to the presence of these rhythms, EEG exhibits a $1/f$ spectral profile, resembling the profile of pink noise [3]. According to the logic in [34], due limitations of the axon conductivity and synaptic delays, the $1/f$ profile can be partially explained by the fact that the slower the oscillation in the brain, the greater the number of neurons that can get engaged, the wider the cortical areas being recruited and the larger the resulting mean power will be.

Different peaks representing strong activity from one or more of the aforementioned waves are typically observed superimposed on the $1/f$ profile [41]. In most subjects, a peak within the alpha band can be observed. These characteristics of the EEG signals in the frequency domain have serious implications on the SSVEPs and their detection. This will be explained in detail in chapter 3.

2.2.3 EEG Patterns in the Time-Frequency Domain

Trial averaging, albeit helpful in making most of the *event-unrelated potentials* die away, fails to extract the non-phase-locked activity related to special events that induce a reduction or an increase in the synchrony of the underlying brain activity in a specific frequency band [14]. This decrease/increase in synchrony are referred to as *event-related desynchronization* (ERD) and *event-related synchronization* (ERS), respectively. Voluntary movements, for instance, are characterized by pre-movement contralateral ERD in the mu (upper alpha) and beta bands over the sensorimotor areas. The ERD in these frequency bands becomes bilaterally symmetrical shortly before the movement. Movements are often followed by ERS in the beta band. Quantitatively, ERD/ERS can be assessed with *event-related spectral perturbation* (ERSP) analysis or the band power method, just to mention some [14].

2.3 EEG Signal Processing

The main objective of EEG signal processing is to increase the SNR of typical patterns found in EEG signals, like the ones presented in previous section, contaminated with noise. The following subsections highlight the main methods used for this purpose throughout this work.

2.3.1 Signal Processing in the Time Domain

Based on the nature of EEG signals that was described earlier, the continuous EEG data can be modeled linearly as a superposition of a mixture of source activity and additive noise. Formally, we can write that

$$\mathbf{y}^T(t) = \mathbf{x}^T(t) \cdot \mathbf{A} + \boldsymbol{\psi}^T(t), \tag{2.1}$$

where $\mathbf{y}(t) \in \mathbb{R}^{N_y}$ denotes the reading from the multichannel EEG system at time instant t, N_y denotes the number of recording electrodes, $\mathbf{x}(t) \in \mathbb{R}^{N_x}$ denotes the activity of N_x cortical or subcortical sources, and $\boldsymbol{\psi}(t) \in \mathbb{R}^{N_y}$ denotes the additive noise at each electrode

[3] This type of frequency profile is frequently observed in nature, as it is the case in e.g. many biological (from cellular level to behavioral level) signals [50] and speech and music signals [51].

site. The *mixing* or *weighting matrix* is denoted with $\mathbf{A} \in \mathbb{R}^{N_x \times N_y}$ and is assumed to be unknown constant over a short time segment. For digitized EEG data, the time index k is used instead in (2.1). It is worth repeating here that the noise includes the background EEG activity and possible artifacts and typically is defined according to the situation under consideration. In ERPs, for instance, the signal is defined as the activity of the population of neurons which get recruited in response to stimulus presentation and the noise corresponds to all other background activity plus possible artifacts. Since ERPs by definition are of limited length around the event of stimulus presentation, the model for ERPs can be rewritten in matrix form as

$$\mathbf{Y} = \mathbf{X}\mathbf{A} + \mathbf{\Psi}, \tag{2.2}$$

where $\mathbf{Y} \in \mathbb{R}^{N \times N_y}$ denotes the EEG epoch of N samples, $\mathbf{X} \in \mathbb{R}^{N \times N_x}$ denotes the sources activity for the same N samples, $\mathbf{\Psi} \in \mathbb{R}^{N \times N_y}$ denotes the noise and $\mathbf{A}$ is defined as before. Typically, $N \gg N_x$ and $N \gg N_y$.

The ultimate goal in ERP research is to reveal the source activity $\mathbf{X}$ from multiple observations of EEG epochs, i.e. $\mathbf{Y}^{(i)}$, where $i = \{1, 2, \cdots M\}$, time-locked to the same events or experimental conditions. The most straightforward method on this account has been already described, which is epoch averaging. Hereby, the average signal or the observed ERPs is computed with $\overline{\mathbf{Y}} = \frac{1}{M} \sum_{i=0}^{i=M} \mathbf{Y}^{(i)}$, which asymptotically converges to the true mean value $\mathbf{XA}$, since according to the central limit theorem, the noise will decay with $\sqrt{M}$. Obviously, what is being revealed with averaging is not the underlying source activity, but rather a mixture thereof. To illustrate this, we consider a simulated situation similar to [42] in Fig. 2.3. Hereby, we assume, without loss of generality, three cortical sources that respond to a specific stimulus (i.e. three ERP components). The activity of these sources within a specific period of time is denoted by $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t)]^T$, giving rise to the observed signals $\mathbf{y}(t) = [y_1(t), y_2(t), y_3(t)]^T$. For the moment, we ignore the presence of noise, so that epoch averaging is not necessary to enhance the SNR. For arbitrary choice of the mixing matrix in Fig. 2.3, it can be easily seen that the different peaks in the observed ERP signals do not map one-to-one to the underlying source activity or the ERP components. Most importantly, the signal $y_2(t)$ has no trace of the ERP components originating at $x_2(t)$ and $x_3(t)$. The inevitable presence of noise in observed epochs is expected to complicate the story even further.

Noise is often considered to be uncorrelated white noise and independent of the signal under consideration. Both assumptions, ignore the fact that the background EEG (a typical ingredient in the overall noise) is influenced by the experimental events. For instance, it has been shown that the EEG activity undergoes a reduction with visual stimulation [52]. Additionally, the well-observed reduction in alpha waves with eyes open strengthens our belief that this model does not hold in general.

2.3.2 Signal Processing in the Spatial Domain

From our hypothetical example in Fig. 2.3, we note that if the mixing matrix is known and it happens to be nonsingular, the ERP components can be recovered with $\hat{\mathbf{X}} = \overline{\mathbf{Y}}\mathbf{A}^{-1}$. In general, the inverse only exists if $N_y = N_x = \text{rank}(\mathbf{A})$, in which case recovery happens

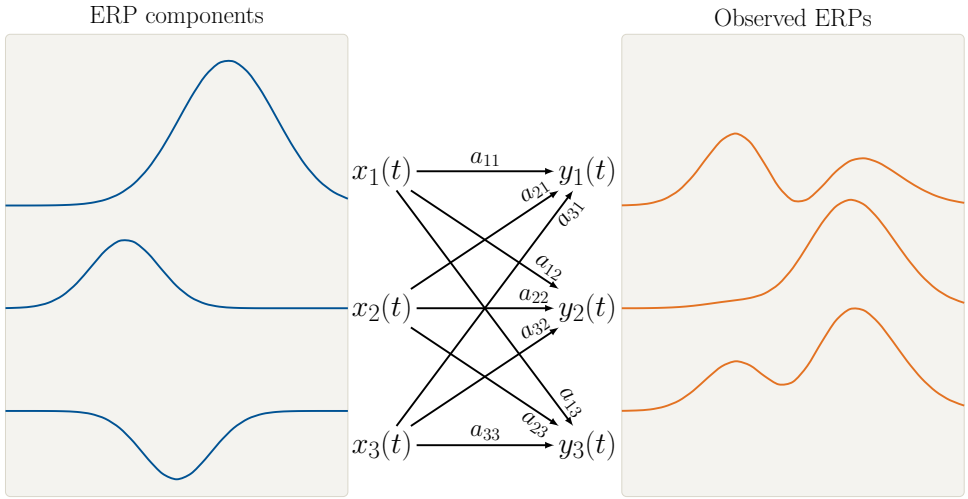


Figure 2.3: A simulated example, similar to [42], highlighting the effect of volume conduction of the head on the observed ERPs with respect to the ERP components. This example clearly shows that, even in absence of noise, the peaks in the observed ERPs do not necessarily map one-to-one to the underlying source activity

with certainty. For the case when the number of recording sites is larger than the number of sources, i.e. $N_y > N_x$, which is the most common case, we will still have an exact solution $\hat{\mathbf{X}} = \overline{\mathbf{Y}}\mathbf{A}^{+}$. For completeness, the case $N_y < N_x$ has infinitely many solutions. This is another reason why it is desirable to record as many electrodes as necessary to have $N_y > N_x$. However, the matrix $\mathbf{A}$ and the underlying source activity are often unknown, not to mention the noise which might survive the averaging process.

Spatial filtering concerns itself with the problem of estimating the matrix $\mathbf{W}$, such that we can compute $\mathbf{S} = \mathbf{YW}$ as a plausible estimate of what the underlying source activity might be. The matrix $\mathbf{W} \in \mathbb{R}^{N_y \times N_s}$ denotes the matrix of N_s spatial filters, $[\mathbf{w}_1, \mathbf{w}_2 \cdots, \mathbf{w}_{N_s}]$. Note here that we used N_s instead of N_x as the number of the underlying sources is also unknown.

Obviously, estimating $\mathbf{W}$ is only possible if additional assumptions can be made about the data. To this end, spatial *principal component analysis* (PCA) and *independent component analysis* (ICA) offer a solution based on assumptions about the statistical properties of the underlying components, i.e. statistical independence and non-Gaussianity in ICA and linear uncorrelatedness in PCA. PCA and ICA have been used often in ERP research [53, 54]. However, when other assumptions can be made about the nature of the source signals, whatever these are, this extra knowledge can be used to guide the derivation of $\mathbf{W}$. This is the case for SSVEPs, and will be thoroughly discussed in chapter 3.

The observed EEG signals are spatially correlated with each other (due to volume conduction) with a true (but unknown) auto-covariance matrix $\mathbf{C}_{yy} \in \mathbb{R}^{N_y \times N_y}$. Note that EEG signals are non-stationary in general but can be assumed stationary in the wide sense for short time segments only, i.e. the covariance matrix $\mathbf{C}_{yy}$ is a function of time. The time index is dropped for simplicity. If the forward model in (2.2) is known, then the auto-covariance matrix can be computed with $\mathbf{C}_{yy} = \mathbf{A}^T\mathbf{C}_{xx}\mathbf{A} + \mathbf{C}_{\psi\psi}$, assuming the noise and the source activity are uncorrelated. Since neither the forward model nor the true

value of the covariance matrix is known a priori, the (unbiased) estimate of the covariance matrix can be computed from the sample data with

$$\mathbf{Q}_{yy} = \frac{1}{N-1} \sum_{i=1}^{i=N} (\mathbf{y}_i - \overline{\mathbf{y}}) (\mathbf{y}_i - \overline{\mathbf{y}})^T . \tag{2.3}$$

This equation is written throughout this work in the form $\mathbf{Q}_{yy} = \frac{1}{N}\mathbf{Y}^T\mathbf{Y}$, where the observation matrix $\mathbf{Y}$ is centered around the mean. Using N in the formula instead of $N-1$ is an approximation with negligible effect on the computation since $N \gg N_y$. For completeness, the cross-covariance matrix between two random vectors X and Y, of dimensions p and q respectively, is denoted by $\mathbf{C}_{xy} \in \mathbb{R}^{p \times q}$. The sample cross-covariance matrix can be similarly estimated from N samples of X and Y with $\mathbf{Q}_{xy} = \frac{1}{N}\mathbf{X}^T\mathbf{Y}$. Unless explicitly stated otherwise, all sample matrices encountered in this work will be centered or made centered by subtracting the sample mean from all observations.

A common practice in a wide range of spatial filtering techniques involves a step of spatial pre-whitening on the observed signals. Spatial pre-whitening concerns itself with finding the matrix $\mathbf{W}$, such that $\mathbf{Y}_w = \mathbf{YW}$ has the identity auto-covariance matrix. An obvious and straightforward solution is $\mathbf{W} = \mathbf{Q}_{yy}^{-1/2}$. Another feasible and straightforward solution is $\mathbf{W} = \mathbf{R}^{-1}$ which can be computed from the QR-decomposition of the sample data $\mathbf{Y} = \mathbf{QR}$.

2.3.3 Signal Processing in the Frequency Domain

EEG data at each recording site in $\mathbf{Y}$ can be transformed into its Discrete Fourier Transform (DFT) representation using

$$Y_i[k] = \frac{1}{N} \sum_{n=0}^{N-1} y_i[n] e^{-jk(2\pi/N)n}, \quad \text{where } j = \sqrt{-1}, k \in \{0, 1, \cdots N-1\}. \tag{2.4}$$

Fast Fourier Transform (FFT) is the golden method used in DFT computations. The power spectral density (PSD) can be estimated with the FFT of the autocorrelation function of the EEG signal, where the result is often referred to as the *periodogram*.

Alternatively, autoregressive (AR) models can be used to estimate the PSD for individual channels in EEG data segments. Generally speaking, any time series $s[n]$ can be represented by an AR model of the form

$$\sum_{m=0}^{m=p} \alpha_m s[n-m] = u[n], u[n] \sim \mathcal{N}(0, \sigma_u^2) \tag{2.5}$$

where p defines the order of the process, which is often denoted as AR(p), $\alpha_i \in \mathbb{R}, \forall i \in \{1, 2, \cdots p\}$ define the p-parameters of the model, and $u[n]$ are sampled from a white noise process of zero mean and σ_u^2 power. By definition, $\alpha_0 = 1$. The spectral representation of AR(p) can be easily obtained with the Z-transform. Let $S[z]$ and $U[z]$ be the Z-transform of the sequences $s[n]$ and $u[n]$ respectively, which results in $\sum_{m=0}^{m=p} \alpha_m z^{-m} S[z] = U[z]$. This

yields the transfer function of the AR model, i.e. $H[z]$ to be

$$H[z] = \frac{S[z]}{U[z]} = \frac{1}{\sum_{m=0}^{m=p} \alpha_m z^{-m}}. \tag{2.6}$$

Straightforwardly, we can compute the power spectral density of the process with

$$P_{\mathrm{AR}}(\omega) = \sigma_u^2 |H(e^{j\omega})|^2 = \frac{\sigma_u^2}{|\sum_{m=0}^{m=p} \alpha_m e^{-j\omega m}|^2} \tag{2.7}$$

$$= \frac{\sigma_u^2}{|1 + \alpha_1 e^{-j\omega} + \alpha_2 e^{-j2\omega} \cdots \alpha_p e^{-jp\omega}|^2}, \tag{2.8}$$

where $j = \sqrt{-1}$. We note here that the set of model parameters completely defines the process in its temporal and spectral representations. AR model parameters can be estimated from data by solving Yule-Walker equations [55] or using multiple linear regression models [56]. By the latter method, all the α_i parameters can be estimated at once with $[\hat{\alpha}_1, \hat{\alpha}_2, \cdots \hat{\alpha}_p]^T = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{s})$ from the vector of the observed data $\mathbf{s}$ and the matrix $\mathbf{X}$ that contains p-lagged versions of the same data. The power of the driving white noise $\hat{\sigma}_u$ can be estimated from the power of the error term in the fitted model. The biggest advantage of the AR power density estimation is the continuity of its estimated power profile. Yet, the adequateness of a fitted model is highly dependent on its assumed order [57].

2.3.4 Signal Processing in the Time-Frequency Domain

To quantify non-phase-locked brain dynamics around experimental events, i.e. ERD/ERS, Makeig [58] has introduced the *event related spectral perturbation analysis* (ERSP), that is a localized and slightly modified version of Short-Time Fourier Transform (STFT) around the events of interest. Hereby, ERSPs are computed from the amplitude spectra of short and overlapping data segments extracted with respect to the experimental events and smoothed with a moving average filter. These spectra are normalized by a baseline computed from the EEG data segments immediately preceding the events of interest. The computation of average ERSP from multiple epochs follows straightforwardly, with

$$\mathrm{ERSP}(f,t) = \frac{1}{M} \sum_{i=1}^{i=M} |F_i(f,t)|^2,$$

where $F_i(f,t)$ is the spectral representation of trial i at time t computed from STFT as previously described [59]. Relevantly, the inter-trial coherence (ITC) provides a measure for phase synchronization in multiple EEG segments time-locked to the same event. It can be computed with

$$\mathrm{ITC}(f,t) = \frac{1}{M} \sum_{i=1}^{i=M} \frac{F_i(f,t)}{|F_i(f,t)|}.$$

The ERSP and ITC measures will be primarily used in chapter 4 to quantify the spectral dynamics of interaction ErrPs.

2.4 Feature Extraction and Classification

As has been mentioned earlier, the two problems of feature extraction and classifications are tightly related. The goal of classification in general is to find a mapping function $h : \mathcal{X} \to \mathcal{Y}$, that maps from the domain of the d-dimensional feature space $\mathcal{X} = \mathbb{R}^d$ to the range of class labels $\mathcal{Y} = \{\omega_1, \omega_2, \cdots \omega_l\}$, where l is the number of classes available. Feature extraction, on the other hand, concerns itself with finding the features that can represent the different classes under consideration in a way that facilitates classification. Features used to discriminate different ERPs are usually extracted temporally from EEG segments time-locked to the events of interest and spatially over the brain regions that are believed to give rise to them. Classification of ERPs, on the other hand, is typically achieved in a supervised fashion, whereby representative examples for each class are collected from one or more training sessions. Hereby, $\hat{h}$ is learned from a training dataset ($\mathcal{D}$) containing $n = n_1 + n_2 + \cdots n_l$ tuples of observations and their labels, i.e.

$$\mathcal{D} = \left\{ \left(\mathbf{x}^{(1)}, h\left(\mathbf{x}^{(1)}\right)\right), \left(\mathbf{x}^{(2)}, h\left(\mathbf{x}^{(2)}\right)\right), \cdots \left(\mathbf{x}^{(n)}, h\left(\mathbf{x}^{(n)}\right)\right) \right\},$$

where $n_1, n_2, \cdots n_l$ are the number of available examples for class $\omega_1, \omega_2, \cdots \omega_l$, respectively.

2.4.1 Multi-class Linear Discriminant Analysis (LDA)

Gaussian-based linear discriminant analysis (LDA) assumes a normal distribution for each class, such that $\mathbf{x}|\omega_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $i \in \{1, 2, \cdots l\}$, or in long form

$$f(\mathbf{x}|\omega_\mathbf{i}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_\mathbf{i}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}. \tag{2.9}$$

LDA additionally assumes that all classes share a common covariance matrix ($\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \forall i$). The mapping function $h_{LDA}(\mathbf{x})$ is defined hereby with

$$\begin{aligned}
\hat{h}_{LDA}(\mathbf{x}) &= \underset{i\in\{1,2\cdots l\}}{\operatorname{argmax}} \ \log\left(f\left(\mathbf{x}|\omega_i\right)\right) && (2.10)\\
&= \underset{i\in\{1,2\cdots l\}}{\operatorname{argmax}} \ -\log\left((2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}\right) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) && (2.11)\\
&= \underset{i\in\{1,2\cdots l\}}{\operatorname{argmax}} \ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) && (2.12)\\
&= \underset{i\in\{1,2\cdots l\}}{\operatorname{argmax}} \ \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i && (2.13)\\
&= \underset{i\in\{1,2\cdots l\}}{\operatorname{argmax}} \ \mathbf{w}_i^T\mathbf{x} + b_i, && (2.14)
\end{aligned}$$

where we substituted $b_i = -\frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i$ and $\mathbf{w}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i$. The multi-class LDA will be used in supervised SSVEP detection in chapter 3.

2.4.2 Two-class LDA

The two-class LDA will be encountered in the classification of P300 and ErrPs signals in chapter 4. Hereby, the class labels are defined as $\omega_1 = 1$, $\omega_2 = -1$ and the mapping function can be simplified to $h_{LDA}(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$, where $\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ and $b = \frac{-1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T\mathbf{w}$.

Since the true means and covariance matrices for each class are unknown, estimates thereof are substituted for the computations of $\mathbf{w}$ and b. The sample means are computed with $\hat{\boldsymbol{\mu}}_1 = \frac{1}{n_1}\sum_{i=1}^{i=n_1}\mathbf{x}^{(i)}$ and $\hat{\boldsymbol{\mu}}_2 = \frac{1}{n_2}\sum_{i=n_1+1}^{i=n}\mathbf{x}^{(i)}$, and the pooled sample covariance matrix is computed with $\hat{\boldsymbol{\Sigma}} = \frac{1}{n-2}\Big[(n_1-1)\hat{\boldsymbol{\Sigma}}_1 + (n_2-1)\hat{\boldsymbol{\Sigma}}_2\Big]$, where $\hat{\boldsymbol{\Sigma}}_1$ and $\hat{\boldsymbol{\Sigma}}_2$ are the within-class sample covariance matrices, which can be estimated with $\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{n_1-1}\sum_{i=1}^{i=n_1}(\mathbf{x}^{(i)} - \hat{\boldsymbol{\mu}}_1)(\mathbf{x}^{(i)} - \hat{\boldsymbol{\mu}}_1)^T$ and similarly for $\hat{\boldsymbol{\Sigma}}_2$. These estimates of the covariance matrices are known as the maximum likelihood (ML) estimates, which fail to provide invertible $\hat{\boldsymbol{\Sigma}}$ when $n < d$ [60]. As a remedy, we adopted the analytical shrinkage covariance estimator proposed in [60] (using function `cov_shrink()` from BCILAB [61]). Throughout this work, we will refer to the LDA with ML covariance estimator as ML-LDA and with the shrinkage estimator as shrinkage-LDA.

2.4.3 Discrimination Power of Features

In order to quantitatively assess the discrimination power of the spatial and temporal samples that contribute to the feature vectors in $\mathcal{D}$, often the signed r^2 discrimination test is used [62, 63]. Hereby, the signed r^2 is computed with

$$r_k^2 = \frac{\text{sign}\left(\text{cov}\left(\mathbf{x}_k, \boldsymbol{\omega}\right)\right)\text{cov}(\mathbf{x}_k, \boldsymbol{\omega})^2}{\text{Var}(\mathbf{x}_k)\text{Var}(\boldsymbol{\omega})}, \forall k \in \{1, 2, \cdots d\}, \tag{2.15}$$

where the vector $\boldsymbol{\omega}$ is constructed from all the sample labels, i.e. $\boldsymbol{\omega} = [h(\mathbf{x}^{(1)}), h(\mathbf{x}^{(2)}), \cdots h(\mathbf{x}^{(n)})]^T$ and $\mathbf{x}_k$ is constructed by concatenating the k^{th} element of all sample feature vectors. The signed r^s will be encountered in chapter 4 when we discuss the main temporal and spatial features of interaction ErrPs.

2.5 Types of EEG-based BCIs

There are several types of BCIs that provide relatively reliable communication and control channels. For the sake of consistency, we will refer to all possible commands which a user can pass to the interface as *interface elements*. ERPs and ERD/ERS are so far the mostly used pieces of information about the underlying brain activity that is used in BCIs. For this purpose, certain ERPs lend themselves better than others, e.g. due to high SNR and/or the nature of the tasks in which they can be elicited. Of these, P300, SSVEP and MI-based BCIs are the primary types found in state-of-the-art BCIs. This is not to say that other patterns were not already or have the potential to be used reliably in BCIs. Single-trial CNV was proposed to be integrated as a master switch within BCI systems in [64]. CNV [46] and/or BP [45] were proposed to be exploited as an evidence for control

commands. Moreover, mVEPs were used to communicate with a 5 and 6-button virtual keyboard in [65, 66].

P300-based BCIs (or P300-BCIs), in its most recent versions, adopt a visual (or tactile) matrix-like presentation of the interface elements, e.g. alphanumerals in spelling tasks. Users continuously attend to a target element of their choice in the visual matrix which undergoes random flashing (or intensification) of its rows and columns. Flashing of the target elements is expected to elicit P300 whereas flashing of nontarget elements is not. Therefore, careful segmentation and classification of the continuously acquired EEG can reveal the hidden interface element, at which the user was attending. Classification of acquired trials as target/nontarget is done in a supervised fashion, whereby per subject and per condition representative trials are collected during training sessions. In training sessions, users attend to cued target elements and flashing is done as usual. The mode, in which P300 systems run with cued targets, is often referred to as *copy spelling mode*, whereas when targets are chosen freely by subjects, e.g. in online sessions, is referred to as *free spelling mode*.

In SSVEP-based BCIs (or SSVEP-BCIs), users are continuously presented with different interface elements that concurrently flicker but with different frequencies. Similar to P300-based interfaces, users can select their element of interest simply by attending to it. P300 and SSVEP-based BCIs are often referred to as *selective-attention-based BCIs* as they require attending to one out of several stimuli.

MI-BCIs are completely asynchronous and do not require external stimulation, but typically require intensive training so that users can modulate their brain waves. Most of motor-imagery-based BCIs (or MI-based BCI) are realized by discriminating between ERD/ERS spatial patterns associated with different motor imagery tasks, e.g. imagining the movement of the right hand vs. the left hand. MI-BCIs typically offer few interface elements (in this case the interface is not visible to the user except for the feedback, if any). The mostly popular MI-BCIs typically discriminate between two classes of MI (left hand vs. right hand). Yet, systems that discriminate between 3-5 classes have been already realized [15, 67–69]. Some evidence is available and hints that discriminating features exist between simple limb motor imagery and compound limb motor imagery [70], which potentially allows to discriminate between 7 different classes. However, increasing the number of interface elements in MI-BCIs comes at the expense of reduced accuracies. On the other hand, P300 and SSVEP-BCIs can offer much more interface elements, without sacrificing accuracy that much. Yet, this might come at the expense of poorer temporal resolution of interaction. Different BCI types are usually compared with respect to their *information transfer rate* (ITR).

2.5.1 Information Transfer Rate (ITR)

Wolpaw et al. [71] have suggested a formula based on mutual information to compute bit rate for BCI interfaces. Generally speaking, for an interface with (M)-equiprobable elements, where elements can be correctly selected with a probability (P) and wrongly with the remaining probability, i.e. $(1 - P)$, where the probability $(1 - P)$ is uniformly distributed over the $M - 1$ undesired elements, the bit rate (b) in bits/selection can be

computed with

$$b = \log_2 M + P \log_2 P + (1 - P) \log_2 \left(\frac{1 - P}{M - 1} \right). \tag{2.16}$$

The bit rate in bits/second can be easily computed with $r = b/T$, where T is the interaction resolution of the interface in seconds, i.e. the mean time in seconds required to make one selection.

Due to the high bandwidth demands of immersive robotic applications, we rely primarily in this work on SSVEP and P300-BCIs since they typically offer relatively high values of M and P and thus for b. Chapter 5 proposes adaptive and contextual BCIs to overcome the current limitations of b in SSVEP and P300-BCIs.

2.6 Stimulation

In BCIs that rely on external stimuli, like P300-BCIs, stimuli can be presented to users through one of the main sensory modalities, i.e. visual, auditory [72] or tactile [73, 74]. Of these, visual stimulation is the mostly used in general and the only used modality throughout this work. The next subsection details the human visual system aiming at providing some insights into the basic neural pathways of the visual stimulation. This information can be regarded as a reference when we discuss the different viewing conditions of SSVEP stimuli in chapter 3.

2.6.1 Human Visual System

The human visual system is composed of the eye, optic nerve and the brain regions that are recruited for multilevel image processing in the occipital lobe. Visual signals need to travel all way back from the eyes at the anterior side of the head to the visual cortex at the posterior side. When the retina in each eye gets excited by light, neural signals are initiated at its photoreceptive cells, referred to as *rods* and *cones*, and pass through bipolar neurons to special neurons called *ganglion cells* . The axons of all these ganglion neurons form a bundle that is called the *optic nerve* [29]. The two optic nerves meet at the *optic chiasm*, where axons from each hemifield form two contralateral groups called *optic tracts*. Signals that travel through the axons of the ganglion cells are forwarded to neurons in the *lateral geniculate nucleus* (LGN) of the thalamus, which relay them to the visual cortex. The visual pathways are depicted in Fig. 2.4.

In BCI research, there are continuous efforts to identify possible ways, by which the visual pathways can be utilized for control and communication. The most successful example in this regard is the SSVEP-based BCI. As previously mentioned, when the human eye is stimulated by flickering light with specific frequencies, EEG recorded over the visual cortex, i.e. at the very end of the visual pathways, can reveal the frequency of that stimulation. Other successful examples exist, e.g. mVEPs as reported in [65, 66].

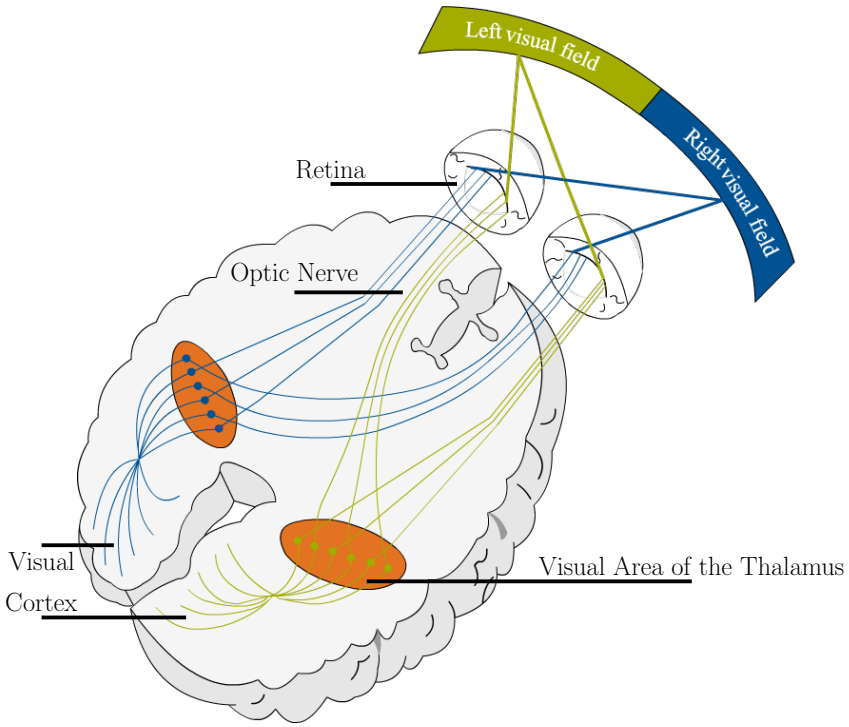


Figure 2.4: The human visual system and visual pathways.

2.7 Summary

This chapter has detailed the different modules that make up the BCI pipeline. In particular, from the discussion of the EEG neuroimaging technique, it became clear that the electrical potentials measured with EEG are manifestations of the activity of temporally overlapping and spatially aligned large populations of neurons. However, only a spatially filtered version of the source activity is measured with EEG due to volume conduction of the head. One main objective in EEG signal processing is to reverse this process, i.e. to estimate the source activity through spatial filtering of the EEG signals. This is the main topic in chapter 3. Additionally, the processing of EEG signals in time domain has been shown to rely mainly on epoch averaging. Despite its simplicity, the averaging technique has uncovered many ERPs, of which the most commonly observed were described alongside the tasks in which they can be observed. A brief treatment to the problems of feature extraction and classification has been given. In particular, the presented LDA classifiers will find application in chapters 3 and 4. Furthermore, the most popular types of BCIs were introduced, where it was shown that BCIs based on selective attention can offer larger number of interface elements than other popular alternatives and thus they are selected in this work as appropriate means for control and communication in immersive robotic embodiment systems. Since these types of BCIs rely heavily on the visual modality to deliver stimulation, the visual neural pathways have been detailed. This aimed at highlighting the role these pathways play in SSVEP-BCIs in particular and the ongoing research efforts to come up with new EEG-based communication paradigms.

3 Steady-State Visual Evoked Potentials

Steady-state visual evoked potentials (SSVEPs) refer to the involuntary brain response to repetitive visual stimulation of the eye in humans and some non-human primates [75]. The measured scalp EEG potentials are characterized in the frequency domain by constant amplitude and phase at the Fourier components of the stimulation frequency and its higher harmonics. SSVEPs can be observed when the driving frequency of the stimulus is in the range $4 - 100$ Hz [13], below which the electrical excitations of the visual system are able to abate before the new stimuli are presented, and in this case, it is the transient VEPs that can be observed [76]. Capilla et al. [77] showed that steady-state VEPs can be accurately predicted from the linear summation of appropriately constructed transient responses and concluded that both can be attributed to the same underlying neural mechanism.

Due to several factors, SSVEPs have been a core concept in non-invasive EEG-based BCI applications. Among these are its robustness and relatively high SNR, the high information transfer rate (ITR) it delivers and the short training time required, if any, before it can be used for online applications. Typically, in SSVEP-based BCIs, different frequency-tagged stimuli are displayed simultaneously with each stimulus given a predefined mapping to a system command. This mapping is also known to the user, who can control or communicate with the system, simply by attending to the stimulus corresponding to the command of interest. Relatively easy discrimination between the different frequencies is facilitated by the fact that selective attention to stimulus location modulates SSVEP [78]. In one study, 48 SSVEP-based interface elements were used for a generic remote control application, allowing for a transfer rate of around 60-90 bit/min [79]. Other applications in which SSVEPs were integrated include, but are not limited to, telephone dialing applications [80], spelling applications [81, 82], video games [83] and robot control [84]. Additionally, SSVEP-BCIs were proposed to be used as a possible means to communicate breathing conditions in patients suffering respiratory disorders [85].

Though these applications might differ in terms of the interaction temporal resolution they ask for, they all strive for higher SSVEP detection accuracies. To this end, spatial filtering has proved to be a powerful pre-processing step for SSVEP detection and boosted typical detection rates both in offline analysis and online SSVEP-BCI applications. What essentially makes spatial filtering very effective in SSVEP detection compared to other scenarios, is that some concrete assumptions about the source signals can be made and therefore unmixing matrices (i.e. the matrix of spatial filters) can be searched for in a constrained space. This is a common feature in most, if not all, state-of-the-art detection methods, which additionally build upon the second order statistics of the acquired EEG data, that is, its spatial auto-covariance and cross-covariance with what is assumed to be a pure SSVEP response. In this chapter, these methods are analyzed theoretically and empirically and the similarities/discrepancies between them are examined, by considering canonical correlation analysis (CCA) as a basis for analysis. This analysis shows that most

of these methods fail to provide reliable detection in low SNR regimes, mainly as they ignore the contribution of noise at each possible driving frequency. MEC includes the estimation of noise in the computation of its scores, and therefore it has proved to outperform other methods. However, MEC incorporates additionally a dimensionality reduction step, which despite its proven effectiveness, is based on an artificial criterion. On the basis of these results, a new detection method, namely CVARS, is proposed. CVARS gets its detection power from combining the power of the canonical variates and that of the autoregressive spectral analysis in estimating the signal and noise power levels. The new method shows superior results to other state-of-the-art methods.

Since our interest in SSVEPs stems from their potential use in immersive robotic embodiment applications, appropriate presentation/viewing methods of the visual stimuli using HMDs are also considered. The work in [86] compared stimulation with an LCD monitor to HMD stimulation for a virtual reality (VR) game. Results have shown an increase of 10% in ITR with HMDs and improved user engagement. This is taken a bit further with a study that compares monocular and binocular viewing with an LCD monitor and an HMD. Results show that monocular viewing of the SSVEP stimuli provides reliable detection results, both with the LCD monitor and the HMD device, but higher for the HMD. Importantly, we found that anti-phase dichoptic stimulation results in binocular inhibition, most likely due to binocular rivalry, whereas in-phase stimulation of the two eyes leads to binocular summation.

This chapter proceeds as follows. Sec. 3.1 presents the basic source model for SSVEPs. Sec. 3.3 highlights the foundations of the state-of-the-art detection methods and emphasizes the points where they converge or diverge and proposes the CVARS method as an alternative. Sec. 3.4 provides in-detail information about the materials and methods used to collect real EEG data from healthy subjects for different viewing methods. Results concerning the different detection and stimuli viewing methods are presented respectively in Sec. 3.5 and Sec. 3.6. The chapter concludes in Sec. 3.7.

3.1 SSVEP Source Model

We assume that after the retina is excited by flickering light, and after the transient VEPs vanish [87], pure SSVEP responses appear as multiple phase-shifted sinusoidal waves whose frequencies are integer multiple (up to N_h) of the driving frequency. The assumed pure waves propagate to the scalp where EEG signals are measured. Due to volume conduction of the head, a linear combination of these source signals corrupted with noise will be observed at each measurement location, i.e. electrode. The additive noise might have encephalic or non-encephalic sources, and is generally non-stationary. However, for short EEG segments, it is often assumed to be stationary in the wide sense [88]. No statistical knowledge about the noise is additionally assumed here.

Formally, we state that in response to flickering light with driving frequency f, the values recorded over time at each electrode i can be written as

$$y_i(t) = \sum_{h=1}^{N_h} a_{i,h} \sin\left(2\pi h f t + \phi_{i,h}\right) + \psi_i(t), \tag{3.1}$$

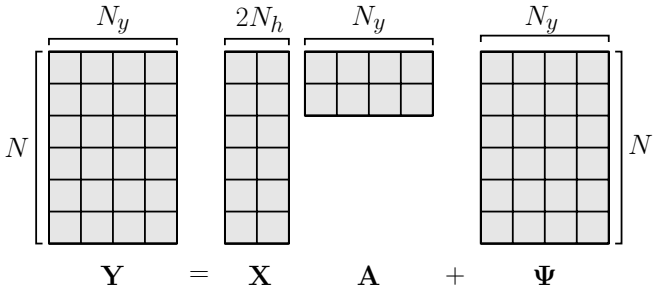


Figure 3.1: Matrix representation of the EEG model containing SSVEPs.

where $i \in \{1, \ldots, N_y\}$ and N_y is the total number of recording electrodes. The terms $a_{i,h}$ and $\phi_{i,h} \in \,]-2\pi, 0]$ respectively denote the harmonic-specific amplitude and phase lag at electrode i. The additive noise contributing to $y_i(t)$ is denoted by $\psi_i(t)$. For the digitized EEG with sampling rate of F_s (in Hz), (3.1) becomes

$$\begin{aligned} y_i[k] &= \sum_{h=1}^{N_h} a_{i,h} x_{i,h}[k] + \psi_i[k], \text{ where} \\ x_{i,h}[k] &= \sin\left(2\pi h f \frac{k}{F_s} + \phi_{i,h}\right) \\ &= \sin\left(2\pi h f \frac{k}{F_s}\right)\cos\left(\phi_{i,h}\right) + \cos\left(2\pi h f \frac{k}{F_s}\right)\sin\left(\phi_{i,h}\right). \end{aligned}$$

With conformable transformation of the amplitude values $a_{i,h}$, the above notation can be rewritten for a collection of N samples from all N_y electrodes (see Fig. 3.1) in matrix form as

$$\mathbf{Y} = \mathbf{X}\mathbf{A} + \mathbf{\Psi}. \tag{3.2}$$

Here, $\mathbf{Y} = \begin{bmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_{N_y}\end{bmatrix}, \mathbf{\Psi} \in \mathbb{R}^{N \times N_y}$ and the mixing (or propagation) matrix $\mathbf{A} \in \mathbb{R}^{2N_h \times N_y}$. The source model matrix $\mathbf{X} \in \mathbb{R}^{N \times 2N_h}$ is defined now as

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_h & \cdots & \mathbf{X}_{N_h}\end{bmatrix}, \tag{3.3}$$

where $\mathbf{X}_h = [\sin(2\pi h f k') \;\; \cos(2\pi h f k')] \in \mathbb{R}^{N \times 2}$, and $k' = \frac{1}{F_s}, \frac{2}{F_s}, \ldots, \frac{N}{F_s}$.

In absence of any statistical knowledge about the noise, $\mathbf{A}$ can be estimated from the source model and the acquired EEG data by $\mathbf{A}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ which minimizes the total least squares error, i.e. $\|\mathbf{Y} - \mathbf{X}\mathbf{A}\|_F^2$. Equivalently, the LS estimate of $\mathbf{A}_{LS}$ can be written as $\mathbf{A}_{LS} = \mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}$, where $\mathbf{Q}_{xx}$ and $\mathbf{Q}_{xy}$ respectively denote the sample spatial auto-covariance matrix of the pure SSVEP response and the sample cross-covariance between the acquired EEG data and the assumed source model, both are estimated from N observations. Despite the fact that the source signals are defined deterministically, they can be modeled as stochastic with empirical means computed on segments of length N, which can be approximated with the zero-vector $\mathbf{0}$ for large N. Due to the orthogonality of the basis vectors in $\mathbf{X}$, the sample covariance matrix $\mathbf{Q}_{xx}$, for centered and normalized $\mathbf{X}$, can also be approximated for relatively large N with $\mathbf{I}_{N_x}$, where $N_x = 2N_h$. In this formulation, the number of harmonics in the SSVEP response, i.e. N_h, is not a random

variable but rather an unknown deterministic value.

For completeness, if the noise is known to be spatially correlated and its non-singular covariance matrix $C_{\psi\psi}$ is known, the estimate of $\mathbf{A}$ can be improved with $\mathbf{A}_{GLS} = \left(\mathbf{X}^T C_{\psi\psi}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T C_{\psi\psi}^{-1}\mathbf{Y}$, where $\mathbf{A}_{GLS}$ is the generalized least squares estimate that minimizes the squared Mahalanobis distance, i.e. $(\mathbf{Y}-\mathbf{X}\mathbf{A})^T C_{\psi\psi}^{-1}(\mathbf{Y}-\mathbf{X}\mathbf{A})$. It might be tempting to try estimating the noise from training sessions with no visual stimulation and plug that into the formula for the GLS estimate. However, we refrain from doing that for two simple reasons. First, EEG is in general non-stationary (we only assume weak stationarity for short time segments). Second, there is evidence in the literature that visual stimulation reduces background EEG activity [52].

Where it is necessary, in order to avoid ambiguity when we refer to the source model for the different driving frequencies, the subscript f_l will be added to matrix $\mathbf{X}_{f_l}$ to indicate the source model of the driving frequency under consideration.

The previously assumed source model can generalize with little modification to different visual stimulation patterns, other than the flickering light. For instance, visual stimulation can be done with pattern (e.g. checkerboards) reversal stimuli. The only difference between the two cases, is that SSVEPs are modulated at the fundamental stimulus frequency for the flicker, and at the stimulus reversal rate (i.e. twice the stimulus frequency) for pattern reversal stimuli [89, 90].

In addition to the stimuli's driving frequency, the amplitude of the SSVEPs is highly sensitive to other parameters of the stimulator, like the spatial frequency, contrast, hue, orientation and the used equipment [78]. There exists in the literature a plethora of work which tried to estimate the correlation between these parameters and the evoked potentials. For instance, Teng et al. [91] have found that square waves with 50% duty cycle produced significantly higher discriminative power than sinusoidal and triangular waves displayed with a light-emitting diode (LED). Campbell and Maffei [92] found that a linear relation governs the amplitude of the SSVEP and the log of contrast of a phase-alternating grating stimulus. Additionally, it was found that it is necessary to keep the contrast level ratio as $\sqrt{2}$ between the monocular and binocular viewing conditions in order to obtain evoked potentials that have comparable amplitudes among the two conditions. Clearly, such information can be integrated to provide more realistic models for SSVEPs. This, however, remains to be solved with more research in the field.

3.2 Problem Statements and Challenges

During concurrent repetitive visual stimulation, an SSVEP detector aims at finding the stimulus, at which the user is attending, based on multi-channel EEG data segments $\mathbf{Y} \in \mathbb{R}^{N\times N_y}$, obtained online from continuous scalp EEG data by means of buffering (with buffer length N and buffer overlap O where $0 \leq O < N$). Based on the requirements of the application at hand, the buffer length and the interaction temporal resolution (T) are determined. T is defined as the shortest time (in samples) between two consecutive commands. The overlap can be computed from $O = \max(0, N-T)$.

The problem of stimulus identification from EEG data thus can be formulated as having M spatially-distributed flickering lights, driven by different frequencies $f_1, f_2, \ldots, f_M$, and

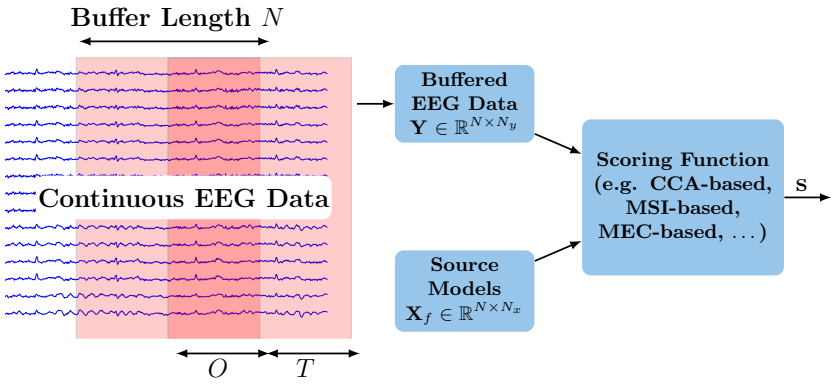


Figure 3.2: Schematic of a general SSVEP detector from continuous EEG data. Different scoring functions can be used to provide the score vector $\mathbf{s} \in \mathbb{R}^{M+1}$, whereby a decision about the user intention can be made every T samples.

a mapping function $\hat{g}$ is sought, where $\hat{g} : \mathbb{R}^{N \times N_y} \mapsto \{f_0, f_1, \ldots, f_M\}$, and f_0 denotes the idle state, i.e. when the user does not attend to any of the stimuli. Often, $\hat{g}(\mathbf{Y})$ is defined as the argument which maximizes a score function or a test statistic s. SSVEP detection can be formally written as

$$\hat{f}_l = \hat{g}(\mathbf{Y}) = \underset{f_l \in \{f_0, f_1, \ldots, f_M\}}{\operatorname{argmax}} \; s(\mathbf{Y}, \mathbf{X}_{f_l}). \tag{3.4}$$

We denote by $g(\mathbf{Y})$ the ground truth frequency of the stimulus, to which the user attends while $\mathbf{Y}$ is being acquired. The score $s(\mathbf{Y}, \mathbf{X}_{f_0})$ is considered here as to test whether or not a given response is statistically significant and not due to noise fluctuations and background EEG. More often than not, it is defined as a constant threshold which is either computed from the EEG data segments themselves or a priori computed from training data. For completeness, let $\mathbf{s} \in \mathbb{R}^{M+1}$ denote the vector containing the score value for all frequencies plus the idle state.

Fig. 3.2 depicts the schematic model of SSVEP detection in continuous EEG data. Available scoring functions used in (3.4) will be discussed in section 3.3.

The user in SSVEP-mediated applications continuously shifts his/her gaze between the M spatially distributed stimuli for active control and towards the stimuli-free areas of the display for the idle control state. Obviously, such interaction is asynchronous and completely paced with the user actions. However, and regardless of the SSVEP detection method used, recognition of these asynchronous spatial attention shifts does not happen usually on the spot due to inherent limiting factors of the buffering stage. The shorter the buffer size (small N and T), the faster is the response of the system. With larger buffer sizes, however, the temporal random fluctuations in the score function $s(\mathbf{Y}, \mathbf{X}_{f_l})$ are made less severe as the noise (the assumed source of variability in the evoked potentials) attenuates typically in proportion to the square root of the number of time averages done on the data. On the other hand, larger buffer sizes introduce delays into the system and reduce the achievable bit rate. Consequently, finding a trade-off between interaction accuracy and speed is of high importance for practical systems.

3.3 SSVEP Detection Methods

Early investigated methods for SSVEP analysis and detection have relied mainly on Power Spectral Density Analysis (PSDA) with Fast Fourier Transform (FFT) applied on single-channel EEG data. Herewith, the temporal EEG signal is transformed into its Fourier representations where test statistics can be derived from obtained information about the power (or amplitude) [76, 93–96], phase [97] or both [98], at all considered driving frequencies. FFT-based methods require relatively long data segments to give reasonable detection results since the frequency resolution (Δf) in the Fourier domain is determined by the reciprocal of the temporal data length available (e.g. it is required to have a 4 s data segment to get a frequency resolution of $\Delta f = 0.25$ Hz). Known issues with FFT like the grid effect (Fourier components cannot be computed for frequencies that are not an integral multiple of Δf) and spectral leakage (energy spillover from one frequency bin to adjacent ones due to rectangular windowing) highly affect the calculated amplitude and phase precision and should be accounted for by choosing suitable window functions and segment lengths [78, 99, 100]. In [101], authors argue that for an arbitrary recording, there might not be enough information to benefit from phase information in SSVEP detection.

Recently, there has been a great tendency towards different methods that rely on spatial filtering of multi-channel EEG data, which proved to be more efficient and stable than FFT-based methods. The basic idea here is to find a spatial filter that transforms the original multi-channel EEG signals into single or multi-channels with desirable characteristics. Friman et al. [102] proposed the minimum energy combination (MEC) spatial filter, which aims at minimizing the noise energy, with the noise defined as the projection of the EEG signals onto the null space of the pure SSVEP vectors. Alternatively, the canonical correlation analysis (CCA) aims at finding a pair of spatial linear combinations for both, the EEG signals and the assumed pure SSVEP responses, which jointly maximizes the correlation between the resulting canonical variates [103]. The maximum contrast combination (MCC) filter maximizes the SNR, defined with the generalized Rayleigh quotient [102, 104]. Finally, the multivariate synchronization index (MSI), through spatial whitening, extracts a single metric that reflects the synchronization level between the EEG signals and assumed pure SSVEPs [105]. Additionally, parametric spectral density analysis, as in autoregressive methods, can be used to provide an estimate for the noise power levels after removing the total energy of the driving frequencies from the spatially filtered signals. This was shown to provide reliable test statistics for SNR, on which SSVEP detection can be based [102, 106].

In the next subsections, we theoretically show the similarities between the spatial filters and the scoring functions used by the standard unsupervised CCA, MSI and MCC approaches to SSVEP detection, where we conclude also that their detection accuracies should not differ significantly. Based on the theoretical analysis of the state-of-the-art methods, we propose a new method, namely the *canonical variates with autoregressive spectral analysis* (CVARS) that estimates the signal and noise power levels from the canonical variates which leads to slight improvement in correct detection rates. We additionally show the conditions on which the MEC and CVARS provide similar results.

3.3.1 Canonical Correlation Analysis (CCA)

Lin et al. [103] used canonical correlation analysis (CCA) to recognize the narrow-band driving frequency of SSVEPs from EEG data. The CCA-based method was found to outperform the FFT-based spectrum estimation method in terms of classification accuracy. This result has been repeatedly reported in [95, 107]. The superior performance can be attributed to the ability of CCA to reveal spatial coherence in data contaminated by either white Gaussian noise or colored noise fields, should the data have high SNR [108].

CCA [109] does that by finding the maximally correlated pairs among all possible linear combinations of two zero-mean multivariate random variables X and Y, where $\mathbf{x} \in \mathbb{R}^{N_x \times 1}$ and $\mathbf{y} \in \mathbb{R}^{N_y \times 1}$. Without loss of generality, we assume in the following that $N_x \leq N_y$.

Formally, we look for the canonical weight vectors $\mathbf{w}_x$ and $\mathbf{w}_y$ where $x = \mathbf{w}_x^T\mathbf{x}$ and $y = \mathbf{w}_y^T\mathbf{y}$, such that the correlation coefficient between the canonical variates x and y, $\rho_1(x, y)$ is maximized. By definition,

$$\begin{aligned}
\rho_1(x,y) &= \frac{E[xy]}{\sqrt{(E[x^2]E[y^2])}} \\
&= \frac{E[\mathbf{w}_x^T\mathbf{x}\mathbf{y}^T\mathbf{w}_y]}{\sqrt{E[\mathbf{w}_x^T\mathbf{x}\mathbf{x}^T\mathbf{w}_x]E[\mathbf{w}_y^T\mathbf{y}\mathbf{y}^T\mathbf{w}_y]}} \\
&= \frac{\mathbf{w}_x^T E[\mathbf{x}\mathbf{y}^T]\mathbf{w}_y}{\sqrt{\mathbf{w}_x^T E[\mathbf{x}\mathbf{x}^T]\mathbf{w}_x \mathbf{w}_y^T E[\mathbf{y}\mathbf{y}^T]\mathbf{w}_y}} \\
&= \frac{\mathbf{w}_x^T\mathbf{C}_{xy}\mathbf{w}_y}{\sqrt{\mathbf{w}_x^T\mathbf{C}_{xx}\mathbf{w}_x\mathbf{w}_y^T\mathbf{C}_{yy}\mathbf{w}_y}}.
\end{aligned} \tag{3.5}$$

Since scaling of $\mathbf{w}_x$ and $\mathbf{w}_y$ doesn't affect the objective function, the search space is limited by constraining the variance of the variates x and y to be 1 [110]. This leads to the new optimization problem

$$\begin{aligned}
\mathbf{w}_x, \mathbf{w}_y = \ & \underset{\mathbf{w}_x, \mathbf{w}_y}{\operatorname{argmax}} \ \mathbf{w}_x^T\mathbf{C}_{xy}\mathbf{w}_y \\
& \text{subject to } \ \mathbf{w}_x^T\mathbf{C}_{xx}\mathbf{w}_x = \mathbf{w}_y^T\mathbf{C}_{yy}\mathbf{w}_y = 1
\end{aligned}$$

By introducing Lagrange multipliers, one can easily obtain the following generalized eigenvalue problems

$$\mathbf{C}_{xy}\mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\mathbf{w}_x = \rho_1^2\mathbf{C}_{xx}\mathbf{w}_x \tag{3.6}$$

$$\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\mathbf{w}_y = \rho_1^2\mathbf{C}_{yy}\mathbf{w}_y. \tag{3.7}$$

Due to the fact that $\mathbf{C}_{xx}$ and $\mathbf{C}_{yy}$ denote covariance matrices, which are symmetric positive semi-definite, (3.6) and (3.7) can be rearranged into two standard symmetric

eigenvalue problems,

$$\mathbf{T}\mathbf{T}^T\mathbf{w}'_x = \rho_1^2\mathbf{w}'_x, \tag{3.8}$$
$$\mathbf{T}^T\mathbf{T}\mathbf{w}'_y = \rho_1^2\mathbf{w}'_y, \tag{3.9}$$

where $\mathbf{T} = \mathbf{C}_{xx}^{-1/2}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-1/2}$, $\mathbf{w}'_x = \mathbf{C}_{xx}^{1/2}\mathbf{w}_x$ and similarly $\mathbf{w}'_y = \mathbf{C}_{yy}^{1/2}\mathbf{w}_y$. The matrix $\mathbf{T}$ is referred to as the *coherence matrix* and denotes the cross-covariance between the whitened vectors $\mathbf{C}_{xx}^{-1/2}\mathbf{x}$ and $\mathbf{C}_{yy}^{-1/2}\mathbf{y}$. This yields that $\rho_1(x,y) = \sigma_{\max}(\mathbf{T}) = \sqrt{\lambda_{\max}(\mathbf{T}\mathbf{T}^T)}$, i.e. respectively the maximum singular value and the square root of the maximum eigenvalue. The matrix product $\mathbf{T}\mathbf{T}^T$ is often referred to as *squared coherence matrix* [111]. Other uncorrelated canonical variates can be found using the remaining eigenvectors and eigenvalues [110].

For later use, we define the decomposition $\mathbf{T}^T\mathbf{T} = \mathbf{W}'_y\mathbf{P}^2\mathbf{W}'^T_y$, where $\mathbf{W}'_y \in \mathbb{R}^{N_y \times N_x}$ and $\mathbf{P}$ has all the canonical correlations on its diagonal, i.e. $\mathbf{P} = \mathrm{diag}(\rho_1, \rho_2, \ldots, \rho_{N_x}) = \mathrm{diag}(\sqrt{\lambda(\mathbf{T}\mathbf{T}^T)})$. Similarly, $\mathbf{T}\mathbf{T}^T = \mathbf{W}'_x\mathbf{P}^2\mathbf{W}'^T_x$, where $\mathbf{W}'_x \in \mathbb{R}^{N_x \times N_x}$.

The score function in (3.4) can thus be defined for the standard CCA method as $s_{\mathrm{cca}} = \rho_1$. Alternatively, other score functions can be derived as an arbitrary function of $(\rho_1, \ldots, \rho_{N_x})$, in the form $s_{\mathrm{fcca}} = f(\sigma(\mathbf{T}))$.

Since true covariance matrices are not known a priori, the coherence matrix is defined with the empirical estimates thereof. Recall that $\mathbf{Q}_{xx} \approx \mathbf{I}_{N_x}$. Given enough samples for $\mathbf{Y}$, i.e. $N \gg N_y$, then $\mathbf{Q}_{yy}$ will be full rank and invertible.

Equivalently, the canonical correlations [112] can be found by first applying the QR decomposition of $\mathbf{Y} = \mathbf{Q}_y\mathbf{R}_y$ and $\mathbf{X} = \mathbf{Q}_x\mathbf{R}_x$, to obtain the orthonormal matrices $\mathbf{Q}_x$ and $\mathbf{Q}_y$ and the full rank $\mathbf{R}_x$ and $\mathbf{R}_y$ matrices. The second step involves the singular value decomposition of $\mathbf{Q}_x^T\mathbf{Q}_y$ as $\mathbf{U}\mathbf{P}\mathbf{V}^T$.

This formulation is computationally more efficient and provides more insights on the geometric interpretation of CCA. Hereby, the canonical correlations correspond to the cosine of the principal angles between the two subspaces spanned by the column spaces of $\mathbf{Q}_x$ and $\mathbf{Q}_y$ or formally, $\sigma(\mathbf{Q}_x^T\mathbf{Q}_y) = \big(\cos(\theta_1), \ldots, \cos(\theta_{\min(N_y,N_x)})\big)$. Geometrically, the maximum canonical correlation is the cosine of the smallest angle possible between any two vectors in the subspaces spanned by $\mathbf{Q}_x$ and $\mathbf{Q}_y$. Fig. 3.3 illustrates this relation.

3.3.2 Multivariate Synchronization Index (MSI)

More recently, Zhang et al. [105] introduced the multivariate synchronization index (MSI) for online SSVEP detection, where the synchronization level between the source model and the acquired EEG is measured based on the S-estimator [113]. The joint covariance matrix $\mathbf{C}_{X,Y}$ which includes the auto and cross-covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$ can be written in block form as

$$\mathbf{C}_{X,Y} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix}. \tag{3.10}$$

The transform $\mathbf{U}$ that orthogonalizes the diagonal block matrices, i.e. whitens the

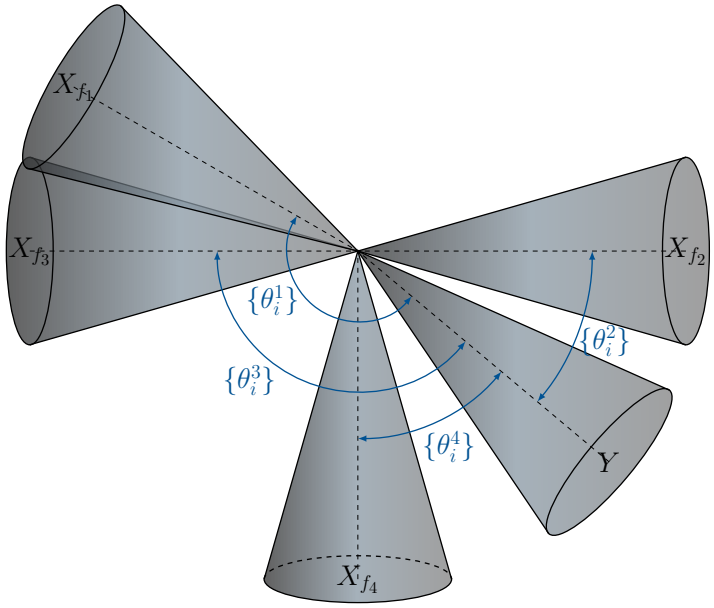


Figure 3.3: Principal angles between the subspaces spanned by the different source models and the acquired EEG data. Each is a subspace in $\mathbb{R}^N$ and is represented by a convex cone. Note also that the column spaces of the different $\mathbf{X}_{f_l}$ may intersect if they share, at least, one basis vector, as it is the case with $\mathbf{X}_{f_1}$ and $\mathbf{X}_{f_3}$, where $f_1 = 15$ and $f_3 = 10$, and $N_h = 3$.

original data matrices $\mathbf{X}$ and $\mathbf{Y}$, was applied such that $\mathbf{R} = \mathbf{U}\mathbf{C}_{X,Y}\mathbf{U}^T$ and

$$\mathbf{U} = \begin{bmatrix} \mathbf{C}_{xx}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{yy}^{-1/2} \end{bmatrix}, \mathbf{R} = \begin{bmatrix} \mathbf{I}_{N_x} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{I}_{N_y} \end{bmatrix}, \tag{3.11}$$

where $\mathbf{T} = \mathbf{C}_{xx}^{-1/2}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-1/2}$ is the coherence matrix already encountered. Let $P = 2N_h + N_y$, $\lambda(\mathbf{R}) = (\lambda_1, \ldots, \lambda_P)$, and the normalized eigenvalues to be defined as $\lambda_i' = \frac{\lambda_i}{P}$. The synchronization index then can be obtained from the entropy-like quantity

$$s_{\text{msi}} = 1 + \frac{\sum_{i=1}^{P} \lambda_i' \log(\lambda_i')}{\log(P)}. \tag{3.12}$$

The MSI-based score is tightly related to s_{cca}. In order to find the exact relationship, we need to find the eigenvalues of the matrix $\mathbf{R}$ defined in (3.11), which involves solving the characteristic equation $\det(\mathbf{R} - \lambda\mathbf{I}) = 0$. Hereby,

$$\det\left(\begin{bmatrix} \mathbf{I}_{N_x} - \lambda\mathbf{I}_{N_x} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{I}_{N_y} - \lambda\mathbf{I}_{N_y} \end{bmatrix}\right) = 0 \tag{3.13}$$

Without loss of generality, we assume in the following that $N_y \geq N_x$. Using the Schur

determinant identity we can rewrite (3.13) as

$$\begin{aligned}
0 &= \det\left(\mathbf{I}_{N_y} - \lambda\mathbf{I}_{N_y}\right) \cdot \det\left((1-\lambda)\mathbf{I}_{N_x} - \frac{1}{1-\lambda}\mathbf{T}\mathbf{T}^T\right) \\
&= (1-\lambda)^{N_y} \cdot \det\left(\frac{1}{1-\lambda}\left((1-\lambda)^2\mathbf{I}_{N_x} - \mathbf{T}\mathbf{T}^T\right)\right) \\
&= (1-\lambda)^{N_y} \cdot \frac{1}{(1-\lambda)^{N_x}} \cdot \det\left((1-\lambda)^2\mathbf{I}_{N_x} - \mathbf{T}\mathbf{T}^T\right) \\
&= (1-\lambda)^{N_y-N_x} \cdot \det\left((1-\lambda)^2\mathbf{I}_{N_x} - \mathbf{T}\mathbf{T}^T\right),
\end{aligned}$$

which means that either $\lambda = 1$ or $(1-\lambda)^2$ is one of the eigenvalues of $\mathbf{T}\mathbf{T}^T$. Consequently, there are exactly $(N_y - N_x)$ eigenvalues of $\mathbf{R}$ that take the value 1. The remaining $2N_x$ eigenvalues can be related to the canonical correlations (Sec. 3.3.1) with $\lambda = 1 \mp \rho_i$, where $i = 1, \ldots, N_x$. This renders s_{msi} as a mere nonlinear function of all canonical correlations, and as a special case of the score s_{fcca}. Additionally, the filtering step (i.e. whitening) involved is similar to that in the CCA method.

3.3.3 Minimum Energy Combination (MEC)

Friman et al. [102] proposed to apply the spatial filter $\mathbf{W}_{\text{MEC}} \in \mathbb{R}^{N_y \times N_s}$, which minimizes the noise energy in $\mathbf{S} = \mathbf{Y}\mathbf{W}_{\text{MEC}} \in \mathbb{R}^{N \times N_s}$, where $N_s \leq N_y$. The noise here is defined as the difference between the original EEG signal and its best LS approximation in the subspace spanned by the SSVEP sinusoids, and thus estimated as

$$\tilde{\mathbf{\Psi}} = \mathbf{Y} - \mathbf{X}\mathbf{A}_{LS} = \mathbf{Y} - \mathbf{X}\mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}. \tag{3.14}$$

Note that $s_{\text{cca}}(\tilde{\mathbf{\Psi}}, \mathbf{X}) = 0$. In fact this is another way to say that the noise estimate $\tilde{\mathbf{\Psi}}$ is orthogonal to the subspace spanned by the columns of $\mathbf{X}$. The sample noise covariance matrix can be written as $\mathbf{Q}_{\psi\psi} = \mathbf{Q}_{yy} - \mathbf{Q}_{yx}\mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}$ with the eigendecomposition $\mathbf{Q}_{\psi}\mathbf{D}_{\psi}\mathbf{Q}_{\psi}^T$. The spatial filter $\mathbf{W}_{\text{MEC}}$ is then obtained by concatenating the last N_s vectors in $\mathbf{Q}_{\psi}$ which correspond to the least proportion of energy in $\tilde{\mathbf{\Psi}}$, e.g. correspond to the eigenvalues whose sum does not exceed $\text{Tr}(\mathbf{Q}_{\psi\psi})/10$ [102].

A test statistic can be derived from the filtered signals, to which we will refer as $s_{\text{mec}}(\mathbf{Y}, \mathbf{X})$ and is obtained with [102]

$$s(\mathbf{Y}, \mathbf{X}) = \frac{1}{N_s N_h}\sum_{l=1}^{N_s}\sum_{k=1}^{N_h}\frac{\hat{P}_{kl}}{\hat{\sigma}_{kl}^2}, \tag{3.15}$$

where $\hat{P}_{kl} = \|\mathbf{X}_k^T\mathbf{s}_l\|^2/N$ estimates the signal power and $\hat{\sigma}_{kl}$ provides an estimate to the noise power at the k^{th} harmonic in the l^{th} spatially filtered signal. $\hat{\sigma}_{kl}$ is obtained by fitting a p-order autoregressive AR(p) model, with parameters $\{\hat{\alpha}_{l1}, \ldots, \hat{\alpha}_{lp}, \hat{\sigma}_l^2\}$ to the data of each column (l) in the matrix $\tilde{\mathbf{S}} = \mathbf{S} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{S} = \tilde{\mathbf{\Psi}}\mathbf{W}_{\text{MEC}} \in \mathbb{R}^{N \times N_s}$ [15], where

$l \in \{1, \dots N_s\}$. Noise power can be computed by substituting $\omega = 2\pi f/F_s$ to (2.8) as

$$\hat{\sigma}_{kl}^2 = \frac{\hat{\sigma}_l^2}{|1 + \sum_{m=1}^{p} \hat{\alpha}_{lm} \exp(-j2\pi mkf/F_s)|^2}, \tag{3.16}$$

where F_s is the sampling frequency and f is the stimulation frequency.

The discrimination power of the statistic in (3.15) stems from its ability to incorporate the noise power estimate at the frequencies under consideration. So far, the score functions in CCA and MSI reflected the signal power only.

Furthermore, we can write the noise covariance matrix as

$$\begin{aligned}
\mathbf{C}_{\psi\psi} &= \mathbf{C}_{yy} - \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy} \\
&= \mathbf{C}_{yy}\left(\mathbf{I}_{N_y} - \mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\right) \\
&= \mathbf{C}_{yy}\left(\mathbf{I}_{N_y} - \mathbf{C}_{yy}^{-1/2}\mathbf{T}^T\mathbf{T}\mathbf{C}_{yy}^{1/2}\right) \\
&= \mathbf{C}_{yy}\left(\mathbf{I}_{N_y} - \mathbf{C}_{yy}^{-1/2}\mathbf{W}'_y\mathbf{P}^2\mathbf{W}'^T_y\mathbf{C}_{yy}^{1/2}\right) \\
&= \mathbf{C}_{yy}\left(\mathbf{C}_{yy}^{-1/2}\mathbf{W}'_y\left(\mathbf{I}_{N_y} - \mathbf{P}^2\right)\mathbf{W}'^T_y\mathbf{C}_{yy}^{1/2}\right) \\
&= \mathbf{C}_{yy}^{1/2}\mathbf{W}'_y\left(\mathbf{I}_{N_y} - \mathbf{P}^2\right)\mathbf{W}'^T_y\mathbf{C}_{yy}^{1/2} \qquad (3.17) \\
&= \mathbf{C}_{yy}\mathbf{W}_y\left(\mathbf{I}_{N_y} - \mathbf{P}^2\right)\mathbf{W}_y^T\mathbf{C}_{yy}. \qquad (3.18)
\end{aligned}$$

Multiplying both sides with $\mathbf{W}_y^T$ from the left and $\mathbf{W}_y$ from the right, yields the following relation

$$\mathbf{W}_y^T\mathbf{C}_{\psi\psi}\mathbf{W}_y = \mathbf{I}_{N_y} - \mathbf{P}^2. \tag{3.19}$$

Therefore it is possible to diagonalize the noise covariance matrix with the canonical weights matrix $\mathbf{W}_y$, which is generally not orthogonal, as by definition $\mathbf{W}_y^T\mathbf{C}_{yy}\mathbf{W}_y = \mathbf{I}_{N_x}$. Recall that the diagonalization of $\mathbf{C}_{\psi\psi} = \mathbf{Q}_\psi\mathbf{D}_\psi\mathbf{Q}_\psi^T$ in the original paper of the MEC method was done with eigendecomposition [102]. Should the EEG data be spatially white or pre-whitened, i.e. $\mathbf{C}_{yy} = \mathbf{I}_{N_y}$ before running the MEC procedure, and N_s is fixed to $2N_h$, then $\mathbf{W}_y = \mathbf{W}'_y = \mathbf{W}_{\text{MEC}}$. This result provides another intuitive insight on the MEC filtering. When the original EEG signals are spatially pre-whitened, MEC maximizes the canonical correlation coefficients while aiming at minimizing the noise energy since the smallest diagonal elements in $\mathbf{D}_\psi$ and $\left(\mathbf{I}_{N_y} - \mathbf{P}^2\right)$ correspond to the largest canonical correlations.

3.3.4 Maximum Contrast Combination (MCC)

The goal of the maximum contrast combination (MCC) method is to find the linear spatial filter that maximizes the generalized Rayleigh quotient [102, 104]. Formally, we are after $\mathbf{w}$ which maximizes $\lambda = \frac{\mathbf{w}^T\mathbf{C}_{yy}\mathbf{w}}{\mathbf{w}^T\mathbf{C}_{\psi\psi}\mathbf{w}}$ subject to $\|\mathbf{w}\|_2 = 1$. The true covariance matrices are not known and thus they are substituted with their sample estimates. The sample noise covariance matrix $\mathbf{Q}_{\psi\psi}$ can be found the same way as in section 3.3.3. With the help of Lagrangian multipliers, one can show that λ attains its maximum with the dominant

eigenvector of the matrix $\mathbf{C}_{\psi\psi}^{-1}\mathbf{C}_{yy}$, which can be rewritten with the result in (3.17) as

$$\mathbf{C}_{\psi\psi}^{-1}\mathbf{C}_{yy} = \mathbf{C}_{yy}^{-1/2}\mathbf{W}_y' \left(\mathbf{I}_{N_y} - \mathbf{P}^2\right)^{-1} \mathbf{W}_y'^T \mathbf{C}_{yy}^{1/2}. \tag{3.20}$$

Thus, $\lambda_i(\mathbf{C}_{\psi\psi}^{-1}\mathbf{C}_{yy}) = 1/(1-\rho_i^2) = f_{\text{mcc}}(\rho_i)$, $\forall i \in \{1, \ldots, \min(N_x, N_y)\}$, and consequentially, $\lambda_{\max} = 1/(1-\rho_{\max}^2)$. The spatial filter $\mathbf{w}$ which attains the maximum quotient can be found by normalizing the columns $\mathbf{C}_{yy}^{-1/2}\mathbf{W}_y' = \mathbf{W}_y$ with respect to the Euclidean norm, and picking the one corresponding to $\rho_{\max}$. This proves that MCC and the CCA methods have exactly the same discrimination power since the function $f_{\text{mcc}}(\rho_1)$ is monotonically increasing in ρ_1.

3.3.5 Canonical Variates with Autoregressive Spectral Estimation of Noise (CVARS)

As stated earlier, the standard CCA method is able to reveal spatial coherence in high SNR regimes. When SNR $\ll 1$, however, the canonical coefficients mainly reflect the correlation between the noise and the assumed source signals, which often leads to erroneous detection as the separate contribution of the signal and the noise to the total values of ρ_i cannot be determined.

Therefore, we propose to estimate the noise at each frequency for each signal after spatial filtering with parametric spectral density estimation. That is, the noise power is estimated by fitting an autoregressive model to canonical variates after cleaning them from all energy at the SSVEP driving frequencies and their higher harmonics. The test statistic s_{cvars} is computed exactly as in (3.15) with the AR(p) models fitted on $\tilde{\mathbf{S}} = \mathbf{S} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{S}$, where $\mathbf{S} = \mathbf{Y}\mathbf{W}_y$ and $\mathbf{W}_y \in \mathbb{R}^{N_y \times N_x}$ is the canonical weighting matrix. Additionally, we observe that $\mathbf{X}^T\mathbf{S} = N\mathbf{W}_x\mathbf{P}$ and therefore $\hat{P}_{kl} = \rho_l^2(w_{kl,1}^2 + w_{kl,2}^2)$, where $w_{kl,1}$ and $w_{kl,2}$ denote the respective weight of the sine and cosine signals of the k^{th} harmonic in the l^{th} canonical variate.

3.3.6 Discussion

In the light of the previous analysis and findings, it is obvious that the scores of the CCA, MCC and MSI methods correspond to different functions of the canonical correlations (ρ_i), and their scores can all be considered special cases of s_{fcca}. Common to all of them is the spatial pre-whitening step of both, the EEG signals and the assumed SSVEP pure response, which is spatially white by construction. Though it is not necessary to do so to obtain the scoring functions s_{cca} and s_{mcc}, these methods additionally involve applying the transform $\mathbf{C}_{yy}^{-1/2}$ to the left singular vectors of the coherence matrix in order to obtain the optimal spatial filter. The spatial filtering of the MEC and CVARS methods is accomplished with the eigenvectors of the noise covariance matrix and the canonical weighting matrix $\mathbf{W}_y$, respectively. Additionally, both involve spectral analysis of AR(p) models fitted on the spatially filtered data cleaned from the energy at the driving frequencies of the SSVEP. The rationale behind using (3.16) in [102, 106] is that the temporally colored noise in each column $\tilde{\mathbf{s}}_l$ in the matrix $\tilde{\mathbf{S}}$, can be modeled as discrete-time autoregressive random process of order p. As a result, this modeling allows to whiten the temporally colored noise,

and produce an unbiased (or with small bias [106]) estimate for the noise power at each stimulation frequency and its higher harmonics. The CVARS method, therefore, involves whitening the data, both spatially and temporally, before it can provide the scoring function s_{cvars}. If the data is spatially pre-whitened before applying the MEC method, then results will be very similar to those of the CVARS method, but not exactly the same as N_s used to compute s_{mec} is governed by a fraction of the total energy in the noise signal and in case of CVARS, $N_s = 2N_h$ is fixed.

3.4 Material and Methods

Several experiments with volunteer subjects were conducted. The objective of these experiments is twofold. First, to provide a means to evaluate the different detection methods. Second, to apply the newly proposed CVARS method for investigating the adequateness of HMD in viewing SSVEP visual stimuli, and whether binocular stimulation is necessary in this case to provide reliable detection of SSVEPs.

3.4.1 Subjects

A total of 10 healthy adults (1 female) aged 29.3 ± 5.5 (range $22 - 39$) with normal or corrected-to-normal vision served as paid volunteer subjects in this study. During the experiments, the participants were seated 0.65 m away from an LCD monitor on a comfortable armchair in a slightly dimmed room. All participants gave their written informed consent. Participants were additionally asked to fill in pre- and post-questionnaires, that were meant to collect data about the level of tiredness before and after the experiment in addition to some demographical data.

Scalp EEG signals were recorded from 16 electrodes positioned according to the international extended 10-20 electrode system over the parieto-occipital scalp areas at P3, Pz, P4, PO9, PO7, PO3, POz, PO4, PO8, PO10, O9, O1,Oz, O2, O10 and Iz. Electrodes were referenced to the right earlobe and the ground electrode was positioned at FPz. The signals were acquired with sampling rate of 512 Hz using g.USBamp acquisition system (g.tec medical engineering GmbH, Schiedlberg, Austria) and band-pass filtered at $0.5 - 60$ Hz. The power line interference at 50 Hz was removed with a 4^{th} order butterworth notch-filter with $48 - 52$ Hz stop band. All electrodes were filled with highly conductive gel in order to reduce impedance.

3.4.2 Experimental Paradigm

A $22''$ liquid-crystal display (LCD) monitor and a head-mounted display (HMD) from Oculus VR, United States, both having 60 Hz refresh rate, were used to view the stimuli which consisted of four spatially distributed flickering rectangles presented simultaneously to participants. The driving frequencies of the stimuli were chosen as integer divisors of the display refresh rate, namely 15, 12, 10 and 8.57 Hz. The chosen and *fixed* driving frequencies are known to evoke moderate to high SSVEP's amplitude strength [114]. The spatial distribution of the stimuli is shown in Fig. 3.4. The EEG acquisition and visual

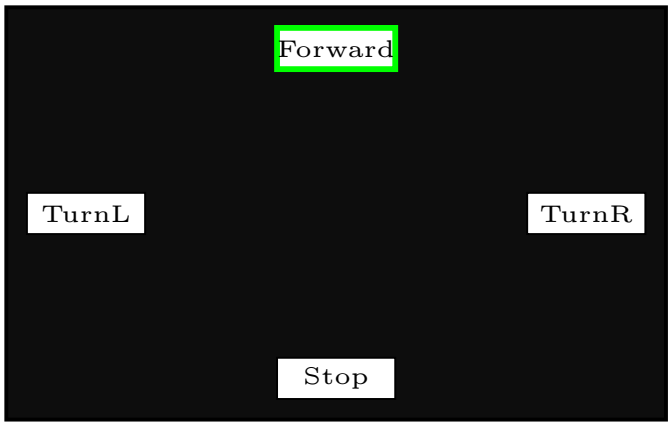


Figure 3.4: Stimulus presentation. In synchronization with the display refresh rate, four white rectangles were flickered on and off against a black screen at rates of 15 Hz (Forward), 12 Hz (TurnL), 10 Hz (TurnR) and 8.57 Hz (Stop). The labeling of the different rectangles serves robotic control applications. The size proportions between the flicker and the display are preserved in the figure.

stimulation were running on two different computers, and synchronized with the screen overlay control interface (SOCI) [115, 116].

Each subject sequentially viewed the stimuli on both displays (i.e. the monitor and the HMD) with either their left eye only, right eye only, or both eyes. This resulted in a total of 2×3 different viewing conditions which were pseudo-randomized across subjects. Subjects were assigned either to finish all the HMD or the LCD conditions at first to minimize electrode displacement that might take place after mounting/unmounting the HMD. Two consecutive sessions per condition were recorded for later offline analysis. In total, each subject underwent 12 sessions, each of around 5 min duration. The whole experiment lasted around two hours including preparations and rest breaks between sessions. Participants used an eye patch to view the LCD monitor in the monocular viewing conditions. This is not necessary in case of the HMD, since stimulating individual eyes can be achieved with software by rendering (or not) the stimuli separately on each of the two displays that make up the HMD. Importantly, when stimuli were rendered for binocular vision with the HMD, the right and left eyes were stimulated differently as the on-off flickering of each stimulus on the right display was the complement version to that on the left display (i.e. anti-phase dichoptic stimulation). During the experiments, no head stabilizer was used and subjects were free to move their head. This means in the monitor case, that the subjects were able to place the stimuli of interest in the fovea by jointly moving their head and the eyes (the contribution of the eye movement to the gaze is likely to be minimal). On the other hand, in the HMD conditions, only eye movement could bring the stimulus of interest to the fovea.

Each recording session started with a blank screen for around 15 seconds followed by the presentation of the flickering stimuli. During stimulation, subjects were instructed to overtly sustain the spatial attention on the cued stimulus. Stimuli were highlighted in turn with a green rectangle as shown in Fig. 3.4 for 10 seconds followed by a rest period of

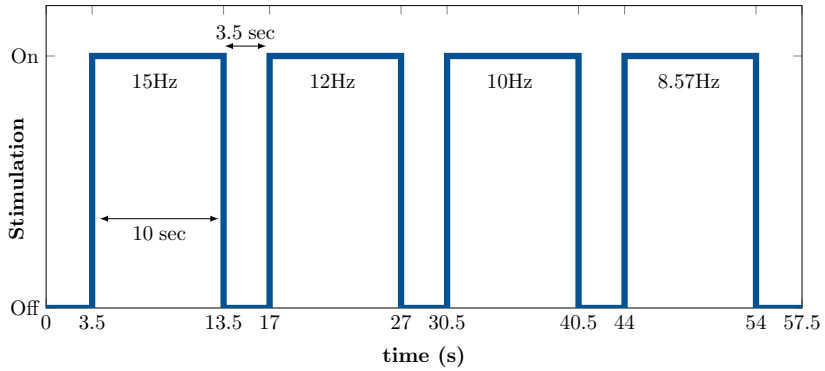


Figure 3.5: Trigger timing for one complete stimulation sequence. When stimulation is on, all flickering stimuli are presented concurrently. Each stimulus is highlighted for 10 seconds as shown with the stimulation train signal.

3.5 seconds, in which the screen went blank. The stimuli were cued in the descending order of their driving frequencies (i.e. the sequence will be top, left, right and bottom according to the stimuli constellation shown in Fig. 3.4). Detailed timing of one complete sequence is shown in Fig. 3.5. Each recording session consisted of five such full sequences.

3.5 Experimental Results - Detection Methods

The different detection methods are compared with regard to their average accuracy ($\bar{P}_D$), which is computed from labeled EEG segments as the ratio of the correctly classified segments to the total number of available segments. Average misclassification error can be easily computed with $E_m = 1 - \bar{P}_D$. In order to provide results that can be compared with those in the literature, we will base our evaluation throughout this section solely on the two recording sessions obtained with binocular viewing in the LCD monitor condition. The results for other viewing conditions are reported comparatively in Sec. 3.6

The accuracy for each of the different detection methods is highly influenced by the choice of the key system parameters, e.g N, T, N_h and F_s. In the following, we will firstly highlight the effect of each of these parameters individually on unsupervised CCA detection accuracy. We chose the CCA method here as the other scoring functions can be obtained from the canonical coefficients, canonical variates and the canonical weights, and thus it can serve as an indicator of the information gain/loss that accompanies parameter change. By fixing these parameters in the light of the empirical evaluation of CCA, the unsupervised detection accuracy of the CVARS is then compared to the state-of-the-art methods.

3.5.1 CCA Results with Varying Key System Parameters

Labeled non-overlapping EEG data segments (i.e. $O = 0$) are extracted from the two available recording sessions per subject. The segment size was varied between $0.5-5$ seconds with steps of 0.5 seconds, respectively yielding 200, 100, 60, 50, 40, 30, 20, 20, 20 and 20 segments, per stimulation frequency. Segments which were obtained during the idle state were not included in the evaluation since we consider unsupervised CCA at this stage.

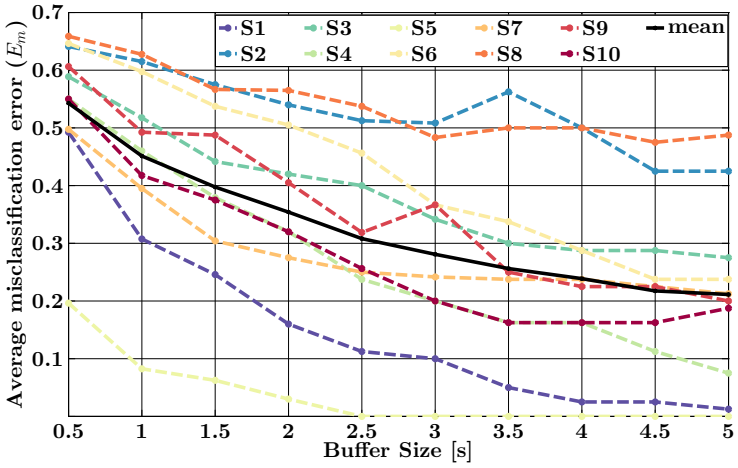


Figure 3.6: Average misclassification error for CCA computed as $E_m = 1 - \bar{P}_D$ for varying buffer length, $N_h = 1$.

Fig. 3.6 shows the average misclassification error (E_m) for all subjects as a function of the buffer size, when the CCA method was used with $N_h = 1$ and $F_s = 512$ Hz. With larger buffer sizes, one can observe that misclassification errors for all subjects get suppressed due to enhanced SNR and more accurate estimates of $\mathbf{Q}_{yy}$ and $\mathbf{Q}_{xy}$.

Fig. 3.7(a) shows the average misclassification error of the CCA method (averaged over all subjects) for different number of harmonics N_h that ranges from 1 to 6 as a function of buffer length. For $N_h > 3$, no further improvement is observed, which cannot be explained by the fact that EEG data itself was lowpass-filtered with a cutoff frequency 60 Hz since the fourth harmonic of the maximum stimulation frequency is not rejected thereby. However, the fourth harmonic of the driving frequency $f_2 = 12$ Hz, lies within the stop band of the notch filter. In order to fully isolate the influence of the notch filter, we reevaluated the average misclassification error of the unsupervised CCA method, excluding f_2 from the analysis. Again, the results shown in Fig. 3.7(b), suggest that no further improvement for $N_h > 3$. The increased accuracy for $N_h > 1$ is however not statistically significant as it was also reported in [107]. The value of N_h is set to 3 throughout the remaining of this work.

As has been mentioned earlier, one can derive arbitrary scoring functions s_{fcca} from the canonical correlations ρ_i. Fig. 3.8 shows the results for $s_{\text{fcca}} = \sum_1^k \rho_i$, while fixing $N_h = 3$, and $k \in \{1, 2, \ldots 2N_h\}$. The performance degradation of s_{fcca} for $k > 1$ suggests that the fluctuations in $\rho_i, \forall i > 1$ over time cannot be used in winner takes all (WTA) assignment as ρ_1 for SSVEP detection.

The effect of changing the sampling rate is shown in Fig. 3.9. Downsampling with a factor of 2 or 4, which respectively resembles sampling frequency of 256 and 128 Hz leads to accuracies that are comparable with the full data segments (with $F_s = 512$). Estimation of covariance matrices $\mathbf{Q}_{yy}$ and $\mathbf{Q}_{xy}$ is not, therefore, significantly affected by downsampling. This behavior suggests that adjacent samples in $\mathbf{Y}$ are correlated (and they are) and that allowing for Δt between samples that is larger than the expected maximum correlation lag would not affect the obtained results. Going to downsampling factor of 8 ($F_s = 64$) deteriorates the performance significantly, as this would introduce aliasing and loss of information of the higher harmonics (with frequencies larger than $F_s/2$). Throughout the

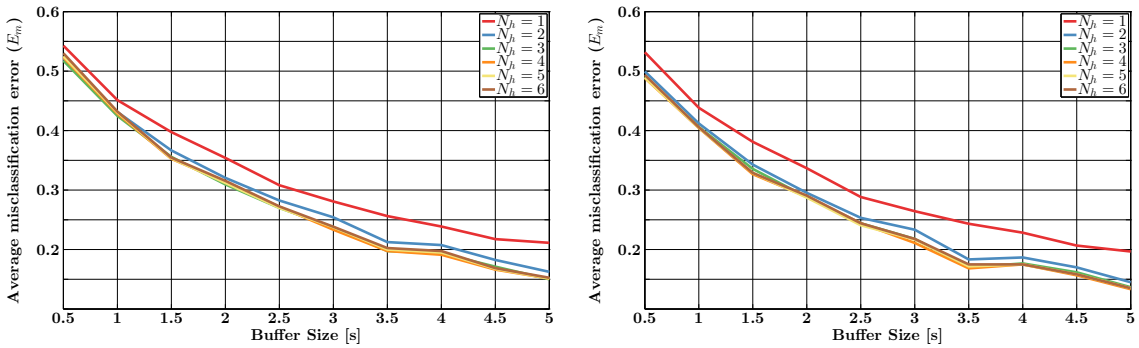


Figure 3.7: Average misclassification error for CCA computed with different number of harmonics. For $N_h > 3$, no further improvement in accuracy is observed. Evaluation is based on (a) all stimulation frequencies (b) and excluding $f_2 = 12$ Hz.

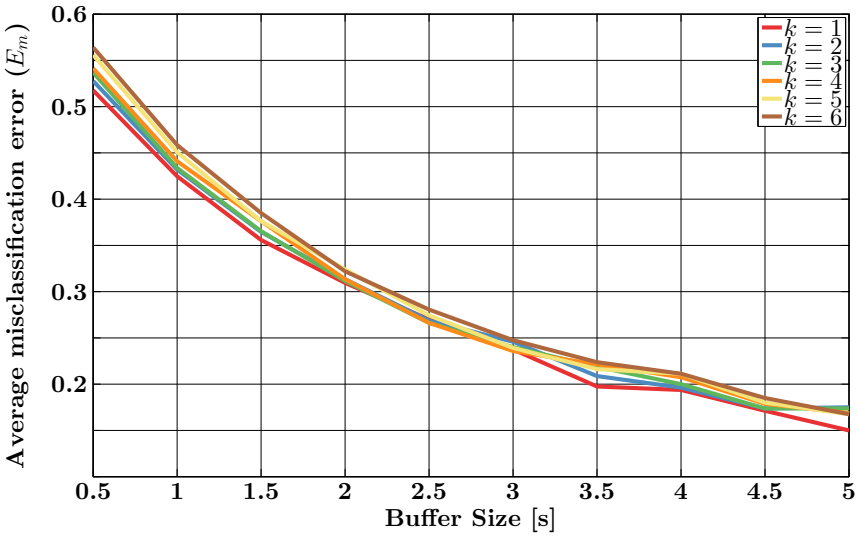


Figure 3.8: Average misclassification error for CCA averaged over all frequencies and subjects computed for different scoring functions $s_{\text{fcca}} = \sum_i^k$, with $k \in \{1, 2, \ldots, 2N_h\}$.

remaining of this chapter, we use $F_s = 512$ Hz. Sampling rates $F_s \in \{256, 128\}$ are expected to produce similar results in terms of the reported scores and the reported misclassification rates.

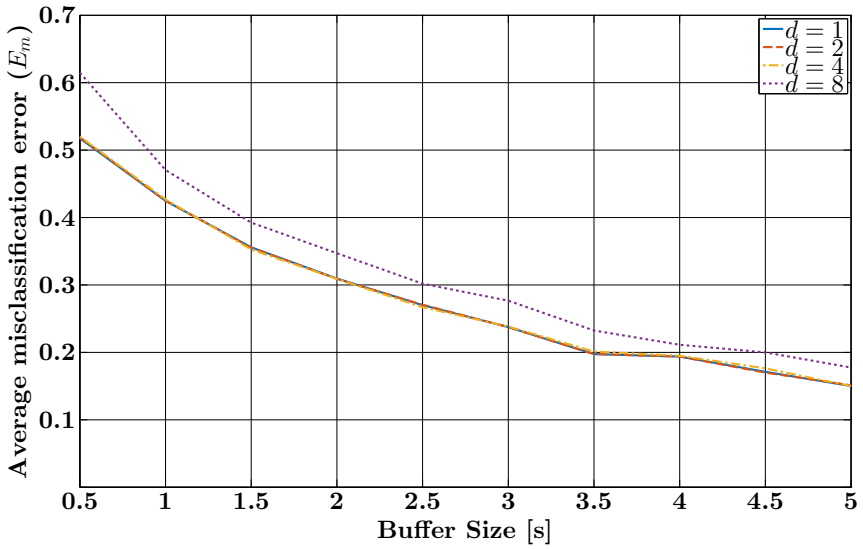


Figure 3.9: Average misclassification error for CCA averaged over all frequencies and subjects for different downsampling factors. Hardly any difference is noticed for $d \in \{1, 2, 4\}$. When the downsampling factor is set to 8 (the dotted purple line), significant reduction in accuracy is observed.

The estimates of the detection accuracies obtained so far using the non-overlapping data segments can be a bit misleading as classification is performed only on *homogeneous* EEG data segments, during which subjects attended to one single driving frequency. However, during online usage, EEG data collected in one segment can reflect two or more different states of user attention (e.g. user shifts his attention from one stimulus to another or from the idle state to one of the active states within the segment). This behavior becomes more probable with larger buffer sizes. To simulate this case, maximally overlapping data segments (i.e. $T = 1$ sample) were continuously extracted from the data and used to plot the CCA score evolution of the different stimulation frequencies over time. The resulting segments thus contain both homogeneous and heterogeneous data. In order to additionally provide more insights about the inter-subject variability in Fig. 3.6, the score evolution will be shown for the subjects S5 and S2, whose CCA results were among the best and the worst, respectively. CCA score evolution during the first full sequence (after viewing all stimulation frequencies once) for $(N, T) = (1024, 1)$ samples is shown in the upper row of Fig. 3.10. These plots show, to some extent, that during stimulation with f_l, the score $s(\mathbf{Y}, \mathbf{X}_{f_l})$ increases over time and surpasses the scores of the other frequencies. Additionally, among the used stimulation frequencies, one can observe for each subject, that a specific frequency is somewhat dominant throughout the whole sequence (12 Hz for S5 and 10 Hz for S2), and to this specific frequency most of the faulty detections and false alarms can be attributed. While high scores for the dominant frequency (most likely due to interference from the alpha brain band waves, within which a peak can be observed in most subjects' EEG[41]) starts to appear during the idle state, they get suppressed (though not

always) when subjects shift their visual attention to flickering light. This is in line with the results in [52] which indicates that visual stimulation reduces EEG background activity. Furthermore, one can categorize the SSVEPs of the two subjects into high and low SNR with respect to the obtained CCA score values, where S2 is the one with the low SNR.

3.5.2 Distribution of Canonical Correlations

In the following, we will refer to the canonical correlation values obtained when users attended to a specific stimulus with frequency f as the target canonical correlations (or target scores) for that frequency. Nontarget scores of a stimulation frequency f, on the other hand, refer to the values obtained when the user attended to other or no stimuli. By fixing $N_h = 3$ and $(N, T) = (1024, 1)$ samples, we computed the target and nontarget correlation coefficients for all stimulation frequencies from the two recording sessions recorded per subject in the monitor/binocular viewing condition. We assigned a data segment to a frequency f, if the most recent sample in that segment was obtained when the corresponding stimulus was then cued for viewing.

From the histogram of all these values, we estimated the distribution of all target and nontarget canonical correlations which are shown in Fig. 3.11. These plots show that the difference between the distributions of the target and nontarget canonical correlations ρ_i is most pronounced for ρ_1. Most importantly, we could see that the typical means (for all canonical correlations) per stimulation frequency differ significantly, in a way that reflects the general power density of EEG data which exhibits a characteristic $1/f$ profile [41, 117]. The mean and standard deviation of all target and nontarget canonical correlations are shown in Fig. 3.12 as a function of stimulation frequency f. Besides the $1/f$ profile, we can observe a peak at 10 Hz, which stems from interference of the dominant alpha brain waves at the same frequency. The distributions of the target and nontarget ρ_i's clearly justify why s_{fcca} defined as the sum of canonical correlations did deteriorate detection performance when compared to s_{cca}, as the bias towards the low frequencies and the subject-dependent peaks in the alpha band, increases by adding further correlations to the value of ρ_1. Therefore, the CCA scores need to be scaled differently for each frequency in order to correct for the observed bias. This scaling is done efficiently in the CVARS method.

3.5.3 Comparison of the Different Methods

The same non-overlapping EEG segments from section 3.5.1 were used to compare the CCA, MCC, MSI, MEC and CVARS methods. The mean misclassification error averaged over all subjects is shown in Fig. 3.13 for $N_h = 3$ and $F_s = 512$ Hz. MEC and CVARS were used with AR(7) model. By visual inspection, the results of the CCA, MCC and MSI methods don't differ significantly. The MEC and CVARS methods outperform CCA for all buffer lengths, except for 0.5 s buffers, in which case, all methods have comparable accuracies. The CVARS method performs slightly better than the MEC for almost all buffer lengths.

Furthermore, the upper three rows in Fig. 3.10 compare the CCA, MEC and CVARS with respect to their score evolution during the first full stimulation sequence with $(N, T) = (1024, 1)$ samples. CVARS and MEC methods show significant improvement over CCA for the subject with the low SNR (S2), especially for the stimulation frequency $f_1 = 15$ Hz and

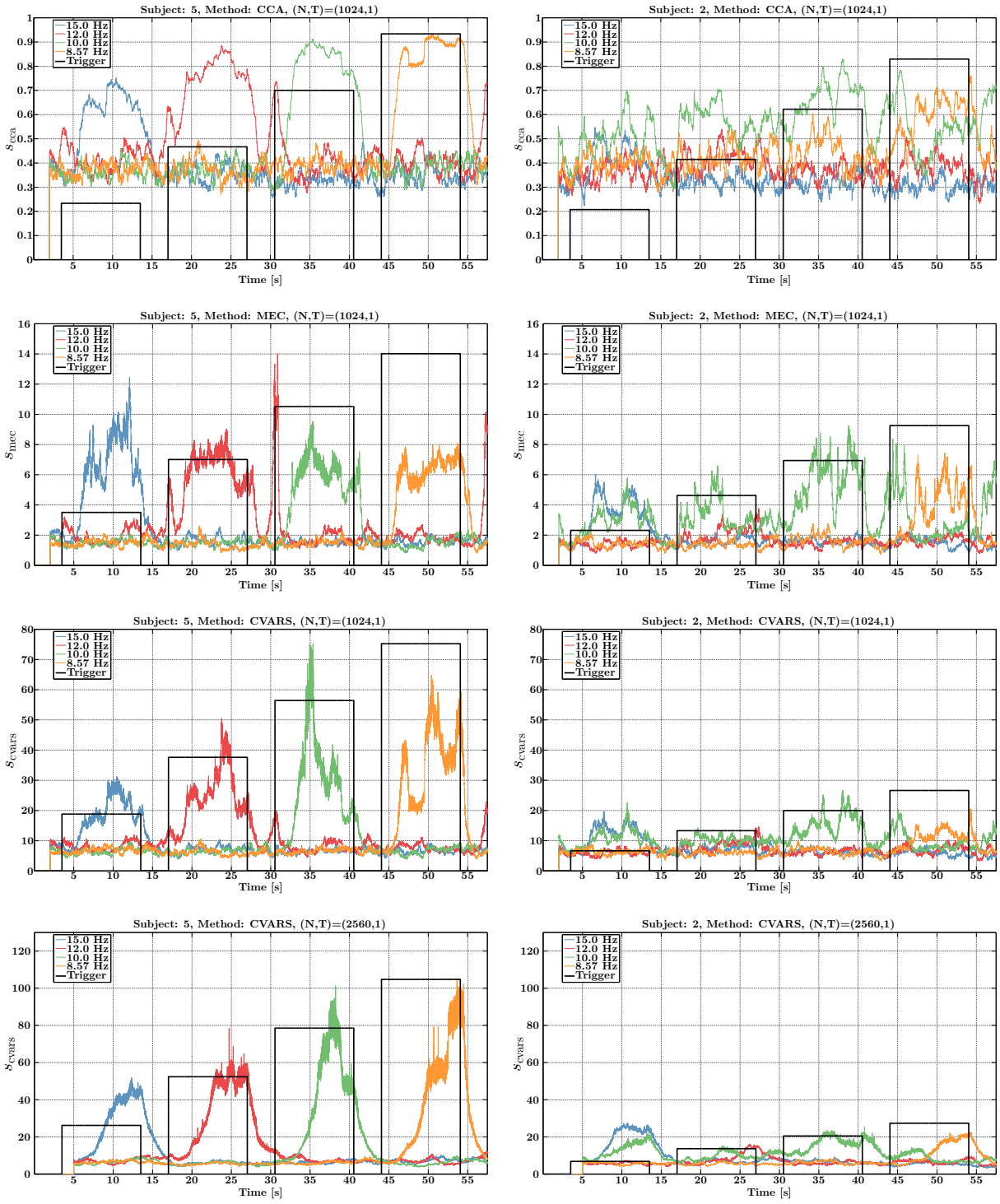


Figure 3.10: Score evolution for S5 (left) and S2 (right) obtained during the first stimulation sequence of the first recording session for buffers of 2 seconds (except for the last row this was 5 s), $T = 1$ sample and $N_h = 3$ with the CCA, MEC, and CVARS methods. AR(7) models were used when necessary. The trigger is shown as a staircase with 4 levels, each level corresponds to one stimulation frequency. For example, level 2 corresponds to $f_2 = 12$ Hz. The zero level corresponds to the idle state. The fourth level (i.e. for f_4) is scaled to the largest score obtained for better visibility.

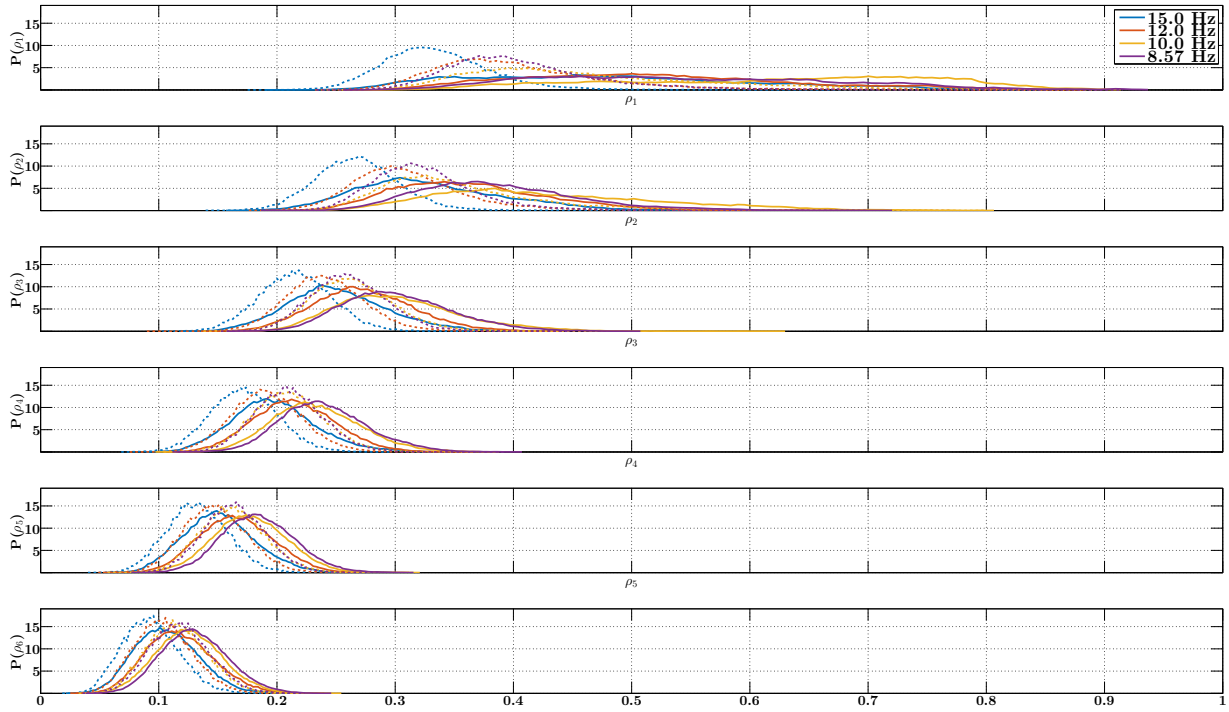


Figure 3.11: Distribution of target (solid) and nontarget (dotted) canonical correlations averaged over all subjects from overlapping EEG segments. For instance, the distributions of target ρ_i for $f_1 = 15$ Hz were obtained by computing the canonical correlations of EEG segments $\mathbf{Y} \in \mathbb{R}^{1024\times 16}$ where $g(\mathbf{Y}) = f_1$ with $s_{\text{cca}}(\mathbf{Y}, \mathbf{X}_{f_1})$. The nontarget distributions were obtained from segments with $g(\mathbf{Y}) \neq f_1$.

less for $f_2 = 12$ Hz. Less or no improvement can be observed for the subject with the high SNR (S5).

From the score evolution ($N = 1024, T = 1$) data of the two recording sessions, we have computed the average accuracy per subject over all stimulation frequencies (excluding the idle state) for the unsupervised CCA, standard MEC and CVARS. The results are summarized in the boxplot[1] of Fig. 3.14. A repeated measures ANOVA determined that mean classification accuracy differed statistically significantly between the three methods, $F(2, 18) = 19.902, p = 0.0012,$ partial $\eta_G^2 = 0.13$, where η_G^2 is the generalized eta squared [118]. Post hoc tests using the Bonferroni correction revealed statistically significant ($p = 0.005$) improvement of unsupervised CVARS (mean± std: 0.739 ± 0.099) over unsupervised CCA (mean± std: 0.64 ± 0.137). Furthermore, we found a trend for unsupervised CVARS being superior over unsupervised standard MEC, which given the current number of subjects though is not significant ($p = 0.069$). Unsupervised standard MEC (mean± std: 0.72 ± 0.107) performed significantly better than the unsupervised CCA ($p = 0.003$).

The effect of choosing longer buffers ($N = 2560$) with CVARS method is shown in the lower panel of Fig. 3.10. We can easily observe that fluctuations in the score function are reduced for both S2 and S5, with the side effect of introducing delays into the system

[1] Analysis was done using R software package and the report can be interactively viewed at http://rpubs.com/moh-marwan/detection-methods-evaluation

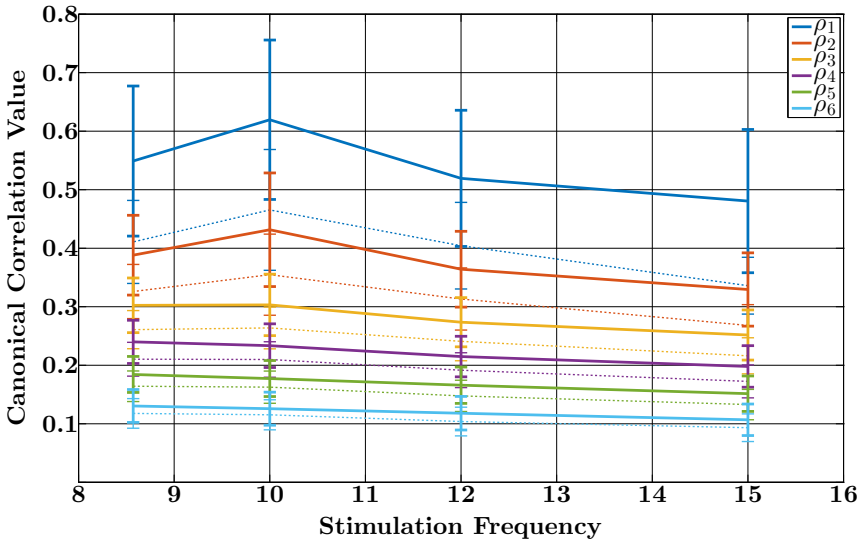


Figure 3.12: Mean and standard deviation of target (solid) and nontarget (dotted) ρ_i computed for the different stimulation frequencies.

(rise-up and decaying delays). This, however, did not guarantee reliable SSVEP detection for S2.

3.5.4 Supervised CVARS Method for SSVEP Detection

The unsupervised detection methods discussed so far provided estimates of the driving frequency in a WTA fashion. This way, the system is always in the control loop and the idle state is never reached. Furthermore, the target and nontarget scores distributions in Figs. 3.11 and 3.10 suggest that the simple argmax function in (3.4) will provide faulty detections due to the large visible difference in the mean scores for the different driving frequencies, and the resulting bias towards the lower frequencies. Learning a simple threshold from labeled data (obtained from a training session for each user), although can reduce the rate of false alarms and faulty detections, can lead also to a larger rate of misses as can easily be seen for CCA from the plots in Fig. 3.11 and for all methods in Fig. 3.10. For instance, setting a CCA threshold of 0.5 in Fig. 3.11 would result in a high probability of miss, as can be computed from the area under the curves of the target distributions for $\rho_1 < 0.5$. It can be additionally observed that it is more probable to miss the detection of higher frequencies than the lower ones. Linear Discriminant Analysis (LDA) lends itself naturally to such a problem, where labeled score vectors are obtained from one training session and an LDA classifier is thereby learned. In online sessions, the scores are computed as usual to produce the vector $\mathbf{s} \in \mathbb{R}^{N_f \times 1}$. The LDA classifier is then applied on $\mathbf{s}$ to provide the final estimate of the hidden attended stimulus. We summarize the comparison between the unsupervised and supervised CVARS with $(N, T) = (1024, 128)$ in the form of the confusion matrices shown in Fig. 3.15, for subjects S5 and S2, in addition to the average over all subjects. We chose CVARS here as it provided the best results in the unsupervised case. The confusion matrices were computed with each method applied on the two available recording sessions. In case of the supervised CVARS, a classifier was learned from each

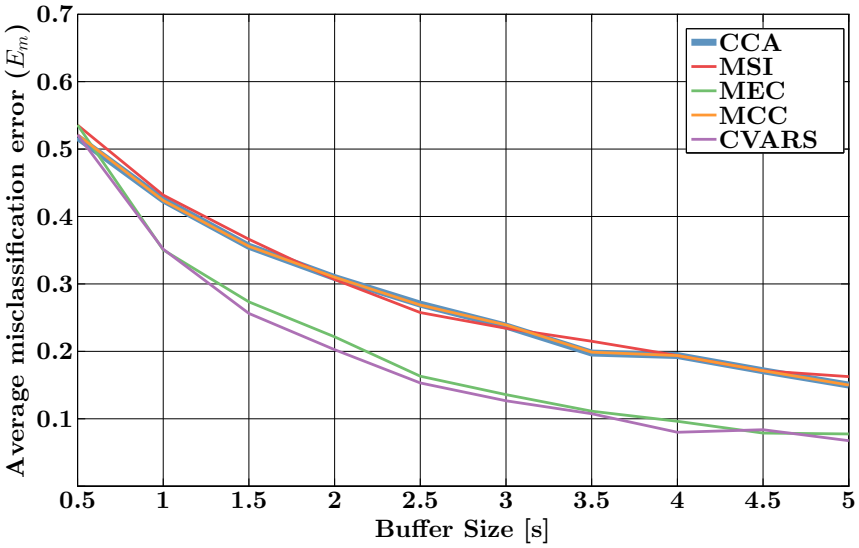


Figure 3.13: Average misclassification accuracy averaged over all frequencies and subjects. As expected, classification accuracies for CCA (plotted with thick line to highlight this fact) and MCC are the same.

session and applied on the other one, and results therefrom were averaged (i.e. a total of two classifiers were used per subject). These results clearly show a reduction in the rate of false alarms on the expense of higher probability of misses. The probability of miss and that of correct detections are quite uniform with respect to the stimulation frequencies, on the contrary to what is expected if a single threshold were used. Furthermore, a large portion of the probability of miss in the supervised CVARS replaces some of the wrong detections in the unsupervised case. Falsely detecting one control state as the idle state is generally favored over confusing it with another control state.

3.5.5 Discussion

From the results in the previous subsections, we can see that there are conflicting factors that affect the detection accuracy of SSVEP. On one hand, accuracy is a monotonically increasing function of the buffer size (i.e. N), should the buffer contain only homogeneous data, which is not the most probable case in online applications. On the other hand, larger buffer sizes introduce rise-up and decaying delays, as the old samples which contain no information about the currently attended stimulus still inhabit the buffer $\mathbf{Y}$ and contribute to the different values $s(\mathbf{Y}, \mathbf{X}_{f_l})$, leading to false alarms and faulty detections.

Comparing Figs. 3.6 and 3.16, which respectively show the misclassification results of non-overlapping segments as a function of buffer length for CCA ($N_h = 1$) and CVARS ($N_h = 3$), we can see that buffer lengths, at which acceptable accuracies are obtained, differ from one subject to another, regardless of the detection method used. Therefore, a trade-off between accuracy and speed should be optimized for each subject, based on a short training session. The same session is also used to learn the LDA classifier from the obtained scores.

Additionally, with our theoretical analysis and empirical results, we have shown that CCA and MCC give exactly the same estimates and thus the same misclassification error rates.

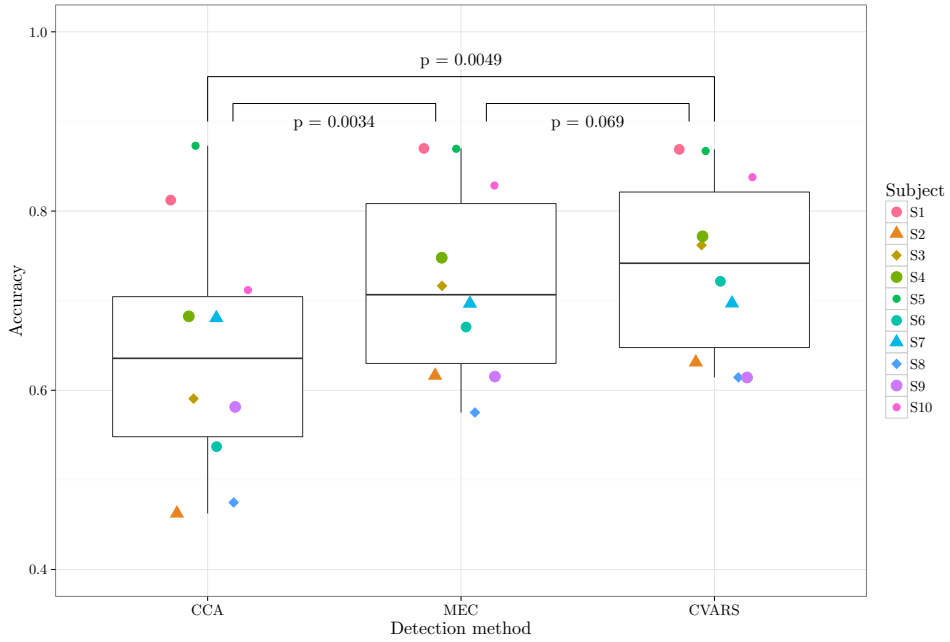


Figure 3.14: Pairwise comparisons between the unsupervised CCA, MEC and CVARS detection methods.

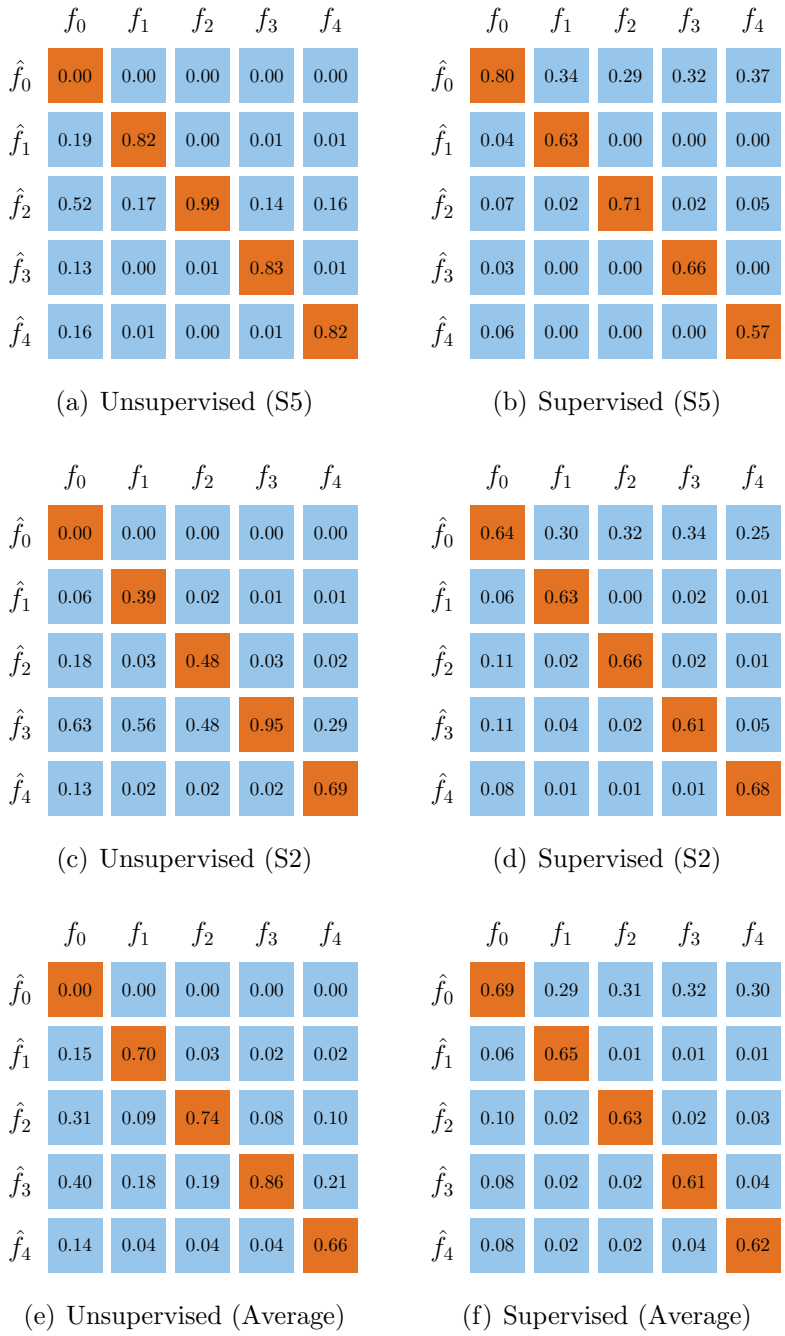
MSI and CCA have shown similar misclassification error rates, suggesting that ρ_1 plays a major role in calculating s_{msi}. This can be seen by rewriting (3.12) as $s_{\text{msi}} = \sum_{i=1}^{2N_h} f(\rho_i)$, where

$$f(\rho) = \frac{(1+\rho)\log(1+\rho) + (1-\rho)\log(1-\rho)}{P\log(P)}. \tag{3.21}$$

Fig. 3.17 plots $f(\rho)$ in the range of $[0, 1]$ in addition to the function $q(\rho) = \rho^2/C$, where C is a constant. The quadratic function $q(\rho)$ approximates $f(\rho)$ relatively well when $\rho < 0.5$. Therefore, for the case when all $\rho_i < 0.5$, the MSI score function (and its quadratic approximation) gives more weight to greater ρ_i's. On the other hand, when $\rho_1 > 0.5$, the contribution of ρ_1 becomes more emphasized as the difference $f(\rho_1) - q(\rho_1)$ grows very rapidly. The latter is the more probable case given the empirical values of target ρ_i's shown in Fig. 3.11.

Furthermore, we have shown that MEC outperforms CCA. This result differs from what was shown in [119]. However, the score function s_{mec} used there was defined with $\hat{\sigma}_{kl} = 1$, which ignores the noise power at the stimulation frequencies and consequently MEC scores will be biased towards the low frequencies. This reaffirms the need to scale the different correlations with regard to estimates about the noise power, which is done efficiently with the test statistics used in standard MEC and CVARS procedures. In [102], a test statistic similar to s_{mec} is used with the MCC spatial filter, where the spatial filter is obtained from a subset of the columns in $\mathbf{W}_y$. Again, this statistic should give similar results to the CVARS when the data is pre-whitened before applying the MCC filter.

Alternatively, Yin et al. [82] proposed the supervised CCA-RV method for the same purpose of reducing the variability in the final target scores of the different stimulation



(a) Unsupervised (S5) (b) Supervised (S5)

(c) Unsupervised (S2) (d) Supervised (S2)

(e) Unsupervised (Average) (f) Supervised (Average)

Figure 3.15: Confusion matrices for the unsupervised and supervised CVARS methods. The supervised CVARS method produces reliable results for all participants including S2, whose data has been shown to have low SNR. This proves the ability of supervised CVARS to deal with wide range of SNR levels.

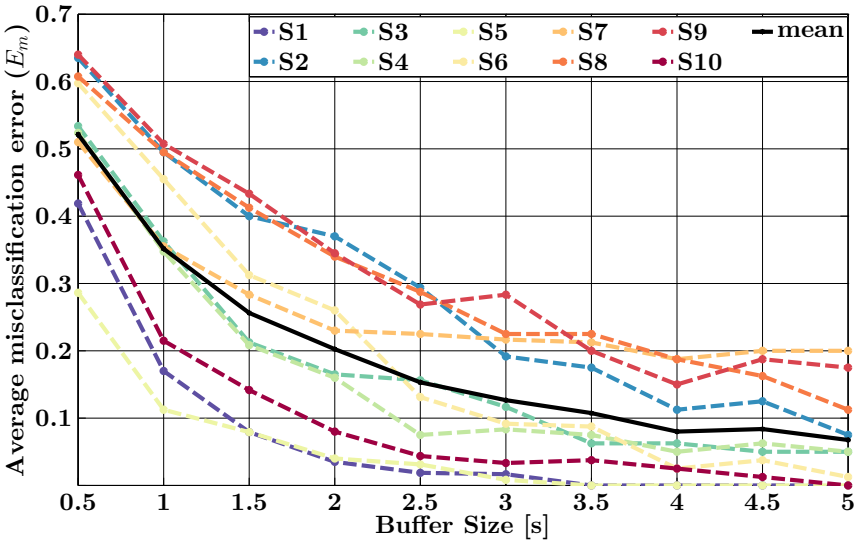


Figure 3.16: Average CVARS-method misclassification error computed as a function of buffer length for non-overlapping segments.

frequencies where $s_{\text{cca-rv}}$ was computed with

$$s_{\text{cca-rv}}(\mathbf{Y}, \mathbf{X}_f) = \frac{s_{\text{cca}}(\mathbf{Y}, \mathbf{X}_f) - s_{\text{cca-nt}}(\mathbf{Y}, \mathbf{X}_f)}{s_{\text{cca}}(\mathbf{Y}, \mathbf{X}_f) + s_{\text{cca-nt}}(\mathbf{Y}, \mathbf{X}_f)}, \tag{3.22}$$

where $s_{\text{cca-nt}}(\mathbf{Y}, \mathbf{X}_f)$ is the mean nontarget scores of a frequency f computed from a training session. Since the CCA-RV method was published very recently, we could not fully and fairly compare it to the CVARS method, especially also that real-time biofeedback mechanisms were employed in [82]. However, we claim that the supervised CVARS method described in this paper should outperform the supervised CCA-RV (when we ignore the effect of the biofeedback). Firstly, CVARS has been shown to outperform CCA when results for all subjects were averaged and therefore plugging the CVARS scores instead of CCA in (3.22) is expected to provide, on average, better results. Additionally, LDA learns from the available training session the optimal mapping from the CVARS scores to the stimulation frequencies as it takes into account not only the mean values as it is the case in (3.22) but also the variances and covariations of the individual scores.

Throughout this work, we had a number of channels $N_y = 16$, which was larger than the number of signals assumed in the source model $N_x = 2N_h = 6$. Reversing this relation does not affect the obtained results. For the CCA method, the number of canonical correlations is upper bounded by $\min(N_x, N_y)$, which will be N_y in this case. The eigenvalues in the MSI method, will be defined the same way as in section 3.3.2 except that there are $N_x - N_y$ eigenvalues which have the value 1. The CVARS method, in this case, will provide no dimension reduction, as the number of canonical variates will be the same as the original EEG channels. On the contrary, the MEC method usually produces less output channels than available in the original EEG signals. Friman et al. [102] reported typical values for $N_y = 6$ and $N_s \in \{4, 5\}$ which are comparable to $2N_h = 4$. For our dataset, we got typical values of $N_s \in \{14, 15, 16\}$ for $N_y = 16$ and $N_h = 3$. The result we proved here regarding the conditions when $\mathbf{W}_{\text{MEC}} = \mathbf{W}'_y$ makes the CVARS method more consistent with the

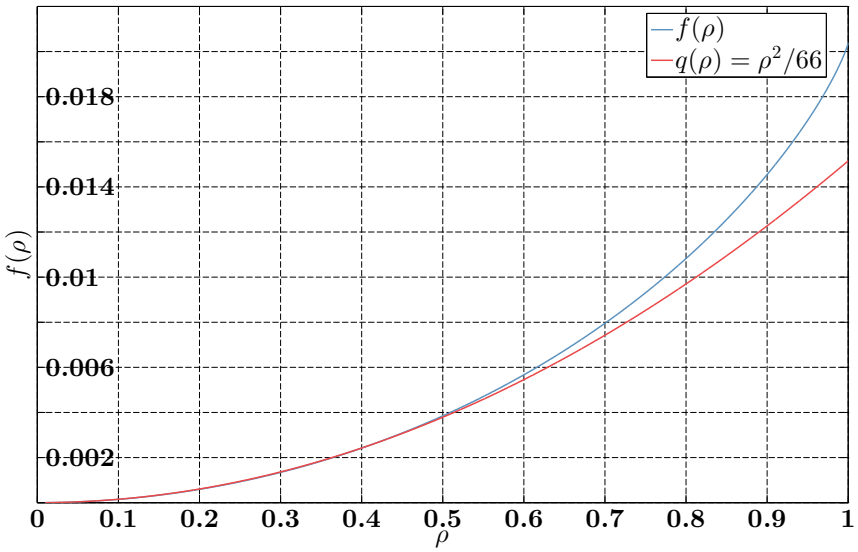


Figure 3.17: The function $f(\rho)$ qualitatively shows the relatively large contribution of the maximum canonical correlation in the final MSI score, especially when $\rho_1 > 0.5$. For $\rho_1 \leq 0.5, s_{\text{msi}} \approx \sum_{i=1}^{2N_h} \rho_i^2$.

assumptions about the number of source model signals.

3.6 Experimental Results - Viewing Methods

The experimental paradigm reported in Sec. 3.4.2 was designed to additionally investigate the effect of the different viewing and display methods on detection accuracy, and thus, to provide a first application of the different examined detection methods. The detection accuracy is considered here as an indirect measure that reflects to which extent SSVEPs are elicited by the different methods. The experiment as a whole is a repeated-measures design with two fully crossed within-subject factors. *Display* is the first factor and has two levels (monitor or HMD), whereas the second factor corresponds to the three *viewing conditions* (monocular left eye, monocular right eye and binocular viewing). To this end, the average unsupervised CVARS classification accuracy (i.e. the dependent variable) was computed for each subject and per viewing and display condition with maximally overlapping EEG segments of 2 s length (i.e. $N = 1024, T = 1$). The average was computed over all stimulation frequencies in the two recording sessions per condition. The segments corresponding to the idle state were not included in the average computation since they result in faulty detection in all display and viewing conditions. Obtained results are summarized in the boxplots [2] of Fig. 3.18 and the standard error barplots in Fig. 3.19.

Two-way repeated measures ANOVA was conducted to test the effects of the two main factors (i.e. Display type and Viewing Condition) on classification accuracy. A statistically significant interaction was found between the effects of the two factors, $F(2, 18) = 9.13, p < 0.005$, and therefore, a series of follow-up simple effect tests has been conducted. The results thereof are summarized as follows.

[2]The complete analysis can be found in http://rpubs.com/moh-marwan/viewing-methods-evaluation

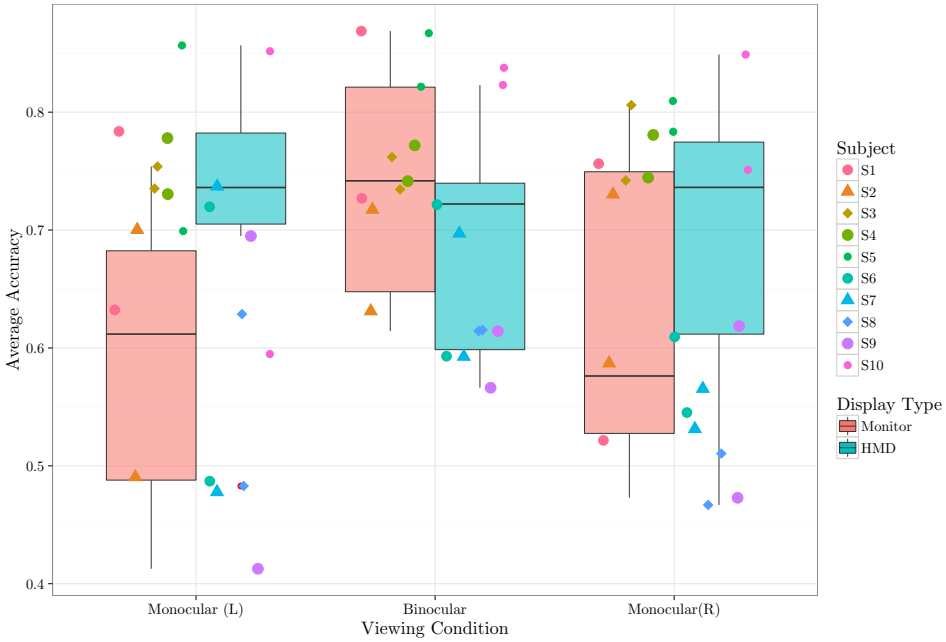


Figure 3.18: Classification accuracy for the different conditions summarized in boxplots.

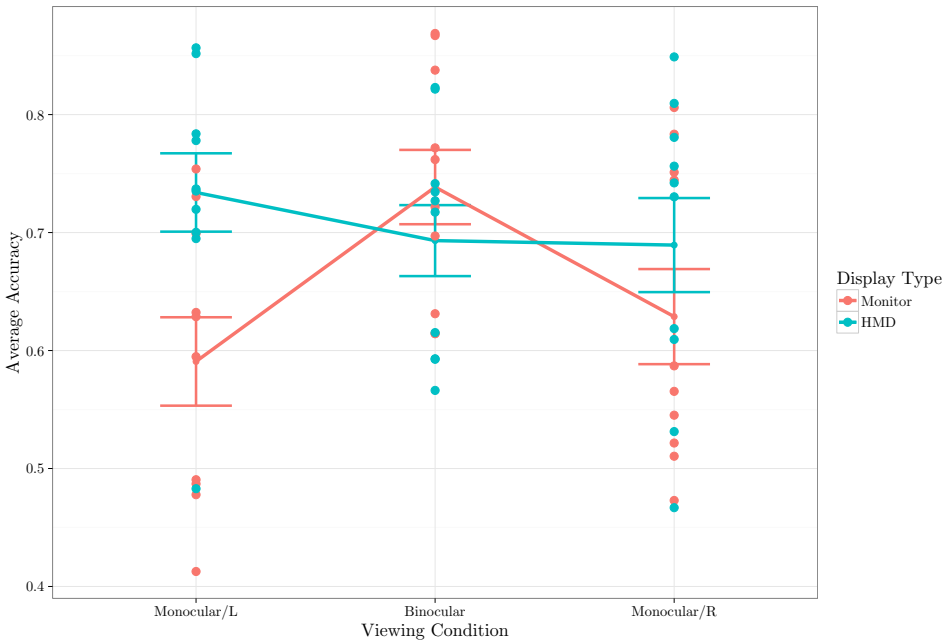


Figure 3.19: Classification accuracy for the different conditions summarized in standard error barplots.

- In the monitor case, the viewing condition was found statistically significant, $F(2, 18) = 12.3, p < 0.005$. Follow-up pairwise comparisons subjected to Bonferroni corrections revealed that binocular viewing resulted in accuracies that are significantly better than monocular left ($p < 0.005$) and right ($p < 0.05$) eye viewing. There was no significant difference between the monocular right and left eye viewing ($p = 0.62$).

- In the HMD case, no significant difference was found between the viewing conditions.

- In monocular left eye viewing, the HMD resulted in significantly higher accuracies than the monitor, $F(1, 9) = 10.42, p < 0.05$.

- In monocular right eye viewing, there was a trend for significance, $F(1, 9) = 4, p = 0.077$ between the HMD and the monitor conditions favoring the HMD.

- In binocular viewing, a trend for significance, $F(1, 9) = 4.56, p = 0.061$ was found between the HMD and the monitor conditions favoring the monitor.

3.6.1 Discussion

The previous results suggest that the main source of the significant interaction between the display type and the viewing condition was the data obtained from the binocular viewing with the HMD, which also can be seen from the barplots in Fig. 3.19. This can be attributed to the fact that in this case, the right and the left eyes were stimulated anti-phase with respect to each other, but with the same frequency. This leaves us with the following three observations. Firstly, based on the monitor data alone, it can be concluded that stimulation of both eyes simultaneously results in significantly higher detection accuracies when compared to single eye stimulation. This is an expected result as binocular stimulation recruits larger populations of neurons when compared to monocular stimulation. This is in complete agreement with [92], where it is argued that summation of the signals in case of binocular vision arise from or subsequent to binocular neurons in the visual cortex. Secondly, comparing results of the binocular viewing between the HMD and the monitor suggests that for binocular summation to take place, SSVEP stimulation of both eyes needs to be completely synchronized so that the eye-specific populations of neurons that respond to the flickering light can act in complete synchrony as well. Reciprocally, in case of anti-phase dichoptic stimulation as it was done in our experiment for the HMD/binocular viewing condition, results suggest that binocular inhibition (most likely due binocular rivalry) is brought about. Thirdly, comparing results of the monocular viewing between the HMD and the monitor, we observe that the HMD results in higher classification accuracies. Similar results were observed in [86] when comparing the two displays with binocular viewing. This observation is most likely due to the closeness of the stimulus to the eye in case of the HMD, which again results in recruitment of a larger population of neurons. Therefore, binocular stimulation is recommended for both eyes, yet monocular stimulation may provide satisfactory results as well. In-phase stimulation of the two eyes enhances the SSVEPs.

There is converging evidence in the literature about the first and the third observations. However, we believe the second observation is completely new and therefore deserves more clarification. Prior research in [120] comparing binocular and monocular sinusoidal

gratings stimulation has shown that binocular summation/inhibition may occur differently for different spatial frequencies. The work in [121] designed a visual stimulus composed of two halves, both flickering with the same frequency but were anti-phase. The setup herein allowed each half-field of the stimulus to separately excite the two halves of the retina, and consequently, out of phase SSVEPs are observed in the two sides of the visual cortex (i.e at the sites O1 and O2). Additionally, in case of dichoptic stimulation with flickering lights having frequencies f_r and f_l for the right and the left eye respectively, gave rise to second order inter-modulation components in the SSVEP signals [122], i.e. the sum and the difference of the two modulation frequencies. Our result hereby is complementary to these results, and shows that binocular interaction is sensitive to the phase difference in dichoptic stimulation when stimuli have the same frequency. Noteworthy here is that most of the cited works hereby relied on FFT measure for their comparisons, while we adopted the resulting detection accuracies as a basis for our evaluation, which only indirectly reflects the strength of the SSVEP signals.

3.7 Conclusions

Detection of SSVEP in continuous EEG signals lies under the general problem of detecting sinusoids in noisy measurements, a problem that has been thoroughly investigated in the array signal processing field in general and in EEG signal processing. We have theoretically shown the conditions in which state-of-the-art SSVEP detection methods share similar spatial filters, a step required to enhance the overall SNR. The equivalence of the discrimination power of the MCC and CCA methods has been proven and it was conjectured that MSI should have very similar results as well. Empirical evaluation was in line with results from the theoretical analysis.

The methods CCA, MCC and MSI rely on a single metric that is computed from the canonical correlations to provide an estimate about the stimulation frequency, to which a user is attending. Thereby they fail to provide reliable estimates when the signal is lost in the noise floor. On the other hand, the MEC method, bases its discrimination upon the estimated signal and noise powers in the spatially filtered signals at each considered frequency, i.e. the fundamental stimulation frequency and its higher harmonics. MEC thus outperforms other state-of-the-art methods, especially in low SNR regimes. However, MEC poses an artificial criterion for dimensionality reduction within the spatial filtering step. To overcome these issues combined, a novel method, namely CVARS, was proposed. The CVARS method was shown empirically to slightly outperform the standard MEC method. We have also shown that the CVARS and the MEC scores are the same, given that the EEG signal is spatially whitened before running the MEC algorithm and N_s is fixed to the number of canonical correlations rather than being artificially determined from the levels of the noise power.

It has been also shown that the supervised CVARS method based on a short training session can be used to learn a mapping function rather than the maximizer (argmax) that estimates the hidden driving frequency and the idle state from the obtained scores, reliably and accurately. The training session should also serve the purpose of finding the optimal buffer size for a specific subject to be used in online applications.

Due to its superior accuracy, the CVARS method can be applied to different fields other than BCIs, e.g. in physiological or vision studies of SSVEP signals. To our knowledge, theoretical analysis of the different SSVEP detection methods in this depth was not done before. As such, results provided here help to understand the sources of diverging results regarding the comparison of the different methods. For instance, MSI was found to outperform CCA and MEC in [105]. The result that MSI outperformed MEC can be attributed to the fact that the computation of the MEC scores ignored the noise power $\hat{\sigma}_{kl}$ as was computed in (3.16). On the other hand, the result that MSI was significantly better than CCA suggests that the number of channels used (at most 8 channels were used in [105]) might have an effect on the difference between the accuracy of the CCA and MSI methods, which we have argued here to be insignificant with a larger number of channels, i.e. $N_y = 16$.

Our application example targeting a comparative study of the visual stimuli presentation on the monitor has shown that binocular viewing of stimuli leads to higher classification accuracies. Additionally, conforming with other results in [86, 123], the HMD has shown superior results to the monitor. Results also suggest that anti-phase dichoptic stimulation with the HMD results in binocular inhibition.

Despite the reported advancement with the CVARS over other existing methods in detecting SSVEPs, we believe further research is needed to improve detection accuracies. Better understanding of the effects that repetitive visual stimulation has on ongoing EEG signals seems necessary in this regard. We have seen that the unsupervised CVARS scores of nontarget stimuli were suppressed when attention was directed to a target stimulus. This means that, in addition to the unsupervised CVARS scores, first order derivate of these scores for all frequencies, can be used in the feature vector for supervised detection. Additionally, throughout this work, we considered data from all recorded channels in our analysis under the assumption that spatial filtering, will be able to weight these channels accordingly. Though this assumption is valid to high extent, removing channels which contain no information from the analysis will make it easier for the spatial filters to arrive at better estimates. Therefore, it might be necessary to explore optimizing the channel selection process alongside the spatial filtering step. This can be for example performed on the same training session which is used to learn the LDA classifier. Unfortunately, simple approaches like exhaustive search are computationally expensive as the number of searches required grows exponentially with the number of electrodes, i.e. the number of searches equals $\sum_{i=1}^{i=N_y} \binom{N_y}{i} = 2^{N_y} - 1$. For $N_y = 16$, this is about 65K.

4 Interaction Error-related Potentials

The occurrence of errors in BCI-based systems seems inevitable at this stage of their development. In the previous chapter for instance, it was shown that even the best existing detection methods for SSVEP signals are prone to errors. It was also shown that while long data segments contribute to enhancing class separability, they lead to rise-up and decaying delays, which in turn can lead to false alarms, faulty detections and/or lower bandwidth, not to mention the effects all these have on the fluency of interaction. On the other hand, interaction error-related potentials (or interaction ErrPs) are typically observed in scalp EEG after feedback that indicates erroneous interface actions, and therefore potentially can be used to (in-)validate a previously selected interface element. However, classification of interaction ErrPs itself is a noisy process, and the fact that it is performed on single EEG trials brings along a new set of challenges. Often, quite long training sessions are required in order to arrive at adequate classification boundaries that can generalize to unseen EEG single trials time-locked to feedback presentations. Besides, classification algorithms which prove to be useful in one situation/interface, usually face generalization difficulties in new situations/interfaces.

Classification of interaction ErrPs is addressed in this chapter only in an indirect way. Our main focus was laid on extending the understanding for invariance and variability sources in interaction ErrPs, and the effects such sources have on classification and classifier transferability. The discussion hereby is based on results from three different experiments, specifically designed for this purpose and having a hybrid and adaptive P300-ErrP BCI in mind as an ideal application field. In particular, we examined the invariance and variability of interaction ErrPs with respect to (1) the mental processing required to assess interface actions (2) time (3) subjects and (4) the pre-processing of the raw EEG data. Scattered results from the ErrPs literature helped in formulating initial hypotheses about these four factors and to define our experimental design. We have found that interaction ErrPs are quite invariant with respect to (2) and (3) and exhibit high variability with respect to (1) and (4). Additionally, we found that the results from transferred classifiers agreed with the observed degrees of invariance and variability.

We have also introduced a new task that simulates adaptive and dynamic P300-based interfaces, whereby interface elements are updated with every new run to new ones. Additionally, the novel inplace feedback presentation feedback that is introduced in this chapter offers an application-agnostic form of feedback, that is expected to require the same mental assessment processes regardless of the nature of the P300 application. Our results suggest that inplace feedback is of great importance to adaptive and dynamic BCIs. This, however, requires further research.

This chapter proceeds as follows. Section 4.1 introduces the family of ErrPs and highlights some of the important aspects of interaction ErrPs. Section 4.2 provides a short review on related work and similar experiments in ErrPs research. Section 4.3 reports the materials

and the design of the different experiments which were conducted to answer the research questions. Experimental results are presented in section 4.4 followed by a discussion in section 4.5. This chapter concludes with section 4.6.

4.1 Background and Objectives

Error processing and awareness mechanisms in the brain lead to reproducible brain activity patterns, which can be observed time-locked to events of errors in scalp EEG. In general, these patterns are referred to as error-related potentials (ErrPs) and are typically taxonomized into four types: *response*, *observation*, *feedback* and *interaction* ErrPs [124]. This taxonomy basically reflects the variability in the error potentials with respect to the nature of tasks, in which they are observed. *Response* ErrPs were found to be elicited after incorrect responses in speeded choice reaction tasks [125, 126]. *Observation* ErrPs, on the other hand, have been shown to be elicited after observing errors committed by other humans [127] or virtual devices [128, 129]. *Feedback* ErrPs are elicited after negative feedback (e.g. feedback of unfavorable results in time estimation tasks) [130–132]. Finally, *interaction* ErrPs were observed after feedback that indicates erroneous interface actions [133, 134], and therefore they can be thought of as a special case of *observation* and *feedback* ErrPs.

The average difference waveform in the ERP structure between the error and correct trials (error-minus-correct) is usually used to highlight the ErrP components. For instance, the difference waveform in speeded choice reaction tasks has been characterized by Falkenstein et al. [125] with a negativity Ne (sometimes referred to as error-related negativity ERN) and a later, more extended positivity Pe. The sharp negative component, Ne, peaked at about 80 ms and Pe peaked in the interval 200-500 ms [135]. The negativity was also observed in correct trials, however with smaller amplitudes (referred to as correct-related negativity CRN). This observation made the authors believe that the negativity Ne reflects the comparison process itself (between the correct and performed response) and not its outcome, and that the independent component Pe reflects a later aspect of the error processing.

In addition to the temporal (phase-locked) signature of errors in scalp EEG, spectral (phase- and non-phase-locked) signatures were observed starting just before incorrect presses in speeded motor responses manifested as an increase in mid-frontal theta band activity accounting for 57% of ERN peak amplitude [136], and an increase in delta-power [126].

The respective temporal and/or spectral signatures vary across the different types of ErrPs. However, independent of the specific type of error potentials and independent of the task performed, EEG and fMRI studies [130, 135, 137, 138] have suggested the anterior cingulate cortex (ACC), the supplementary motor Area (SMA), and/or pre-SMA as candidates for a common neural generator.

Schalk et al. [133] were the first to report that EEG signals that follow erroneous and correct selections by a computer interface differ significantly. The term *interaction* ErrPs has been coined later by Ferrez et al. [134] to refer to this type of ErrPs. Thereafter, there has been a special interest in *interaction* ErrPs within the field of BCIs, mainly due to its potential use as a validation step to confirm a first-stage BCI selection.

First efforts to decrease error rates during BCI-mediated interaction considered adding a

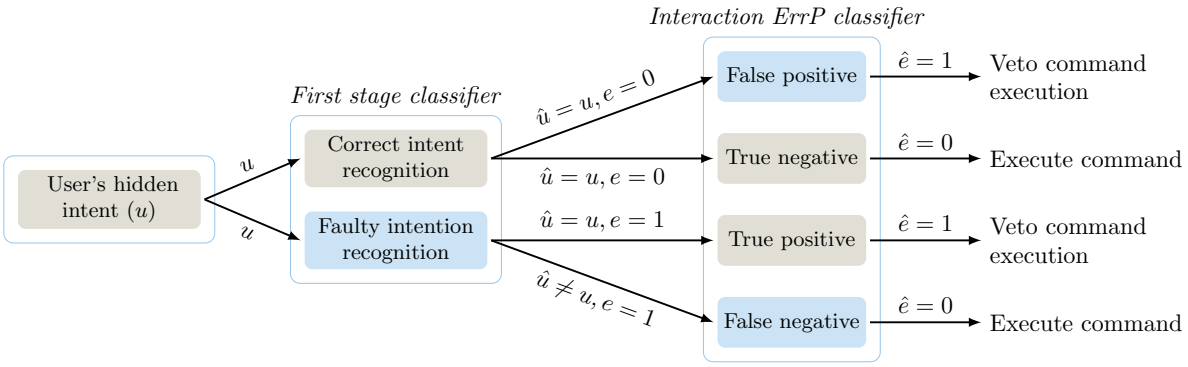


Figure 4.1: Potential use of interaction ErrPs as a validation step in BCIs (adopted and modified from [142]). User hidden input (u) is decoded by the first stage classifier into $\hat{u}$. The feedback presentation might evoke interaction ErrPs (e) which is then decoded by a specialized classifier into $\hat{e}$. In this work, we refer to EEG epochs collected for $e = 1$ as ErrP trials, and as noErrP otherwise.

response verification step[139], where users needed to confirm each selection by communicating an additional one to the system. This approach while improving the accuracy, degrades the achievable bit rates. On the other hand, integrating the detection of ErrPs into different BCI systems [133, 134, 140, 141], improved both the accuracy and the achieved bit rates. However, this gain does not come cost-free. Long training sessions are required to obtain considerable amount of ErrP/noErrP training examples that are used to learn classification boundaries between the correct and erroneous trials. Fig. 4.1 depicts a schematic view of how ErrPs might be integrated into BCIs. The first selection is typically mediated based on P300 or motor-imagery signals.

Therefore, it is desirable to find ErrPs classifiers which transfer across interfaces and over time. This was a concern of Schmidt et al. [143], who showed that a linear discriminant analysis (LDA) classifier for *interaction* ErrPs trained with data from a keyboard-based spelling task can transfer, however with a reduced performance, to a similar P300-based spelling task. Herein, visual inspection of the grand averages of the error and correct trials reveals a great similarity across the two tasks with respect to the general shape and timing of the different deflections. Furthermore, Kim and Kirchner [129] designed a task to compare *observation* ErrPs (users observed the movement of a cursor with no input whatsoever) and *interaction* ErrPs (users controlled the movement of a cursor with a noisy keyboard). Hereby, it can be observed that the grand average difference waveforms exhibit similar time courses for both ErrPs types, but with different amplitudes and latencies. The authors have shown that a linear support vector machine (SVM) classifier learned from the observation ErrPs successfully transfered to interaction ErrPs. In experiments of cursor movement task [134], when motor-imagery was used to mediate user input, a delay of 30 ms was observed in the difference waveform relative to the keyboard case [144]. This delay has been attributed by the authors to the extra time needed by subjects to stop focusing on the motor imagery as feedback was continuously shown to them. These results altogether suggest that if two tasks differed with respect to the type of the user input, and a sufficient time gap was introduced between the user input and the feedback about the

interface actions, then any observed variability in average waveforms consequent to the feedback onset across the two tasks is most likely caused by other factors than the different types of user input.

In this chapter, assuming invariance with respect to the type of user input, it is investigated whether there are other invariants of *interaction* ErrPs and the consequences these invariants, if any, might have for their classification. The main focus was laid on examining possible invariants with respect to: (1) human mental processes that are required after feedback onset (2) time (3) subjects and (4) the way the continuous data is pre-processed. In order to answer these questions, three different experiments and tasks were designed and conducted with different subjects. The first experiment is quite similar to the keyboard-based cursor movement in [134] and the other two experiments, similar to [141], were targeting P300-mediated interaction.

4.2 Related Work

There is a plethora of work in the literature that examined the presence of interaction ErrPs and their detection with many different interfaces and tasks. It can be observed across these studies that many aspects remained invariant whereas many others have shown great variability. Hereby, a brief summary of these studies is provided.

4.2.1 Invariance with respect to Human Mental Processes Consequent to the Feedback Onset

In one-dimensional cursor control using motor-imagery-based BCI (based on modulation of mu and beta rhythms) [133], it has been shown that the difference waveform is characterized by a positive potential centered at the vertex peaking around 180 ms. Despite that the cursor was required to be moved incrementally towards the goal, the error and correct trials were defined solely based on the correctness of the final destination. In quite similar experiments [134], the motor-imagery-based interface was simulated by keyboard presses and each intermediate step towards the goal was labeled either as a correct or erroneous trial. The time-locked difference waveform to cursor movements was shown to have a sharp negative peak after 250 ms (N2) followed by a positive peak after 320 ms (P3) and a second broader negative peak after 450 ms (N4). These peaks/components clearly differ from those in [133], and this difference can be attributed to the different mental processes required to evaluate whether or not a cursor arrives at the target goal and whether or not it just moves towards it [140]. Similarly, in experiments where subjects observed and evaluated the movements of a virtual device towards cued goals, Iturrate et al. [63] have shown that slight changes in the performed tasks lead to a statistical difference in the latencies of observation ErrPs components. The tasks differed only in the way the virtual device moved with respect to a cued goal (either with incremental steps in a horizontal and vertical grids or with a single jump). The observed signal variations were shown additionally, to make it difficult for a classifier trained with data from one task to straightforwardly transfer to other ones. Yet, recalibration and adaptation of the learned classifier (by adapting the means of correct and incorrect trials to the new task) provided fairly good results when a

few training examples were available for the new task [63].

On a different vein, some variability in the different grand average waveforms can be observed across the different hybrid P300-ErrP systems, where different feedback presentation methods have been deployed. For instance, the feedback about the estimated character of interest in [145] was presented in the area above the P300 matrix. The observed grand average difference waveform at the Cz site exhibited a negative peak around 300 ms and a later positivity around 400 ms. In an attempt to avoid ocular artifacts that accompany this style of feedback, authors in [140, 143, 146] adopted a central feedback presentation, whereby the selected character was shown overlaid at the center of the spelling matrix, 1 second or more after the row-column flashing is stopped. The central feedback strategy was also employed in a modified way in [141], where the character presentation is preceded by a presentation of an empty square at the center of the display aiming at attracting the user visual attention to that spot before the estimated character is presented at the same location on the display. This way, ocular artifacts can be further minimized. The observed grand average difference waveform at Cz for a group of healthy subjects was characterized by a negativity at around 348 ms and a later positivity at around 465 ms [141]. Alternatively, the feedback in [147] was done by replacing all the matrix elements with the estimated one.

It can be observed from this short review of the available hybrid P300-ErrP systems that the main concern of designing the feedback presentation was to avoid possible ocular artifacts that accompany the onset of the feedback. However, we argue that changing the way feedback was presented led also to different mental processing of the feedback stimulus, and therefore observed ErrPs signals varied as a byproduct. For instance, the central feedback, when used in language spelling tasks, requires that the users remember (though for a very short time) the last character, to which they attended, and compare it with the estimated one. Users may not need to perform this comparison (or even memorization) when the replacement feedback in [147] is used, since in this case a change of the letter at the attended place in the P300 matrix simply means that the interface made an error. Additionally, one can observe a great discrepancy in the two grand average difference waveforms in [141] and [147] (polarities of the different peaks in the two signals appear to be reversed).

4.2.2 Invariance with respect to Subjects, Time and Pre-processing of Continuous EEG Data

Iturrate et al. [148] have shown that despite existing inter-subject variability with respect to the non-stationarity of EEG, spatio-temporal filtering revealed stable and invariant features across subjects. Correct responses were found additionally to enjoy less variability across subjects when compared to error responses. In [141], no significant difference in the peak latency of interaction ErrPs has been observed across the groups of healthy subjects, subjects with motor impairment and subjects diagnosed with ALS.

With respect to time, Ferrez et al. [134] have shown interaction ErrPs potentials to be stable as the average waveforms and scalp topographies remained similar for two recordings spaced about three months.

Obviously as well, the way the continuous EEG data is pre-processed leads to different average ErrPs waveforms and scalp topographies, and the way features are extracted therefrom affects the obtained classification accuracies. For instance, it has been shown in [63] that most discriminability between error and correct trials comes from fronto-central electrodes when extracted trials were filtered with spatial ICA and otherwise from frontal electrodes. Furthermore, advanced spatial and spatio-temporal filtering methods were used to enhance SNR and proved to increase classification accuracy in [132, 149].

4.3 Material and Methods

In order to examine the different sources of invariance/variability a series of experiments were conducted. The experimental design for each of the experiments is described in the following subsections, and a summary that relates the respective tasks and explains the reasons behind their choice will be presented in Sec. 4.3.2

4.3.1 Subjects

A total of 12 healthy adults (4 female) aged 27.7 ± 5.6 (range $19 - 39$) served as paid volunteer subjects in this study. S10 was left-handed and all subjects except S2 had normal or corrected-to-normal vision. S2 had extreme hyperopia in the left eye. Subject S11 was excluded from the study for not being able to use the P300 speller. During these experiments, the participants were seated 0.7 m away from an LCD monitor on a comfortable armchair in a slightly dimmed room. All participants gave their written informed consent. Participants were additionally asked to fill in pre- and post-questionnaires, that were meant to collect data about the level of tiredness before and after the experiment in addition to some demographical data.

Scalp EEG signals were recorded from 28 electrodes positioned according to the international extended 10-20 electrode system at F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CPz, CP2, CP6, P7, P3, Pz, P4, P8, PO7, POz and PO8 as shown in Fig. 4.2. Similar to [150], the horizontal EOG (HEOG) was obtained from electrodes F9 and F10 whereas the vertical EOG (VEOG) was recorded from FP2 and an additional electrode placed directly below the right eye. EEG and EOG electrodes were referenced to the right earlobe and the ground electrode was positioned at FPz.

EEG and EOG data were measured with sampling rate of 256 Hz at full DC using g.USBamp acquisition system (g.tec medical engineering GmbH, Schiedlberg, Austria). All electrodes were filled with highly conductive gel in order to reduce impedance. Participants were free to move their eyes during the recordings, but were instructed to reduce all unnecessary muscular activity.

4.3.2 Experimental Paradigms

Participants visited the laboratory on three separate occasions spaced few days apart, and completed one experiment per visit. Each experiment consisted of multiple sessions, where each session lasted around 11 min. Since it is important to maintain a high attention

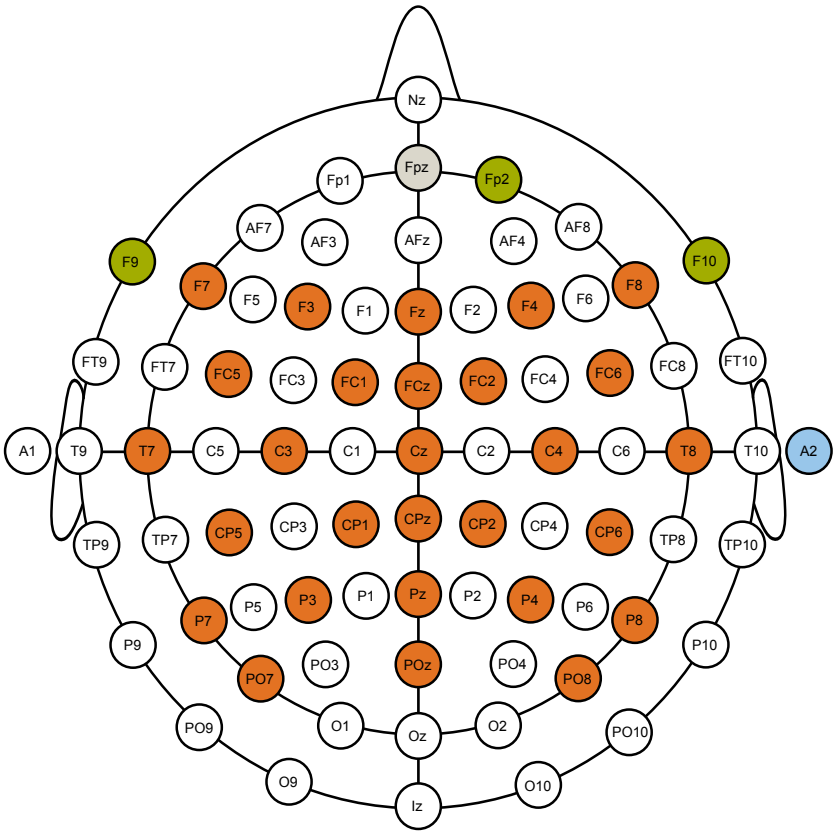


Figure 4.2: EEG/EOG electrode placement.

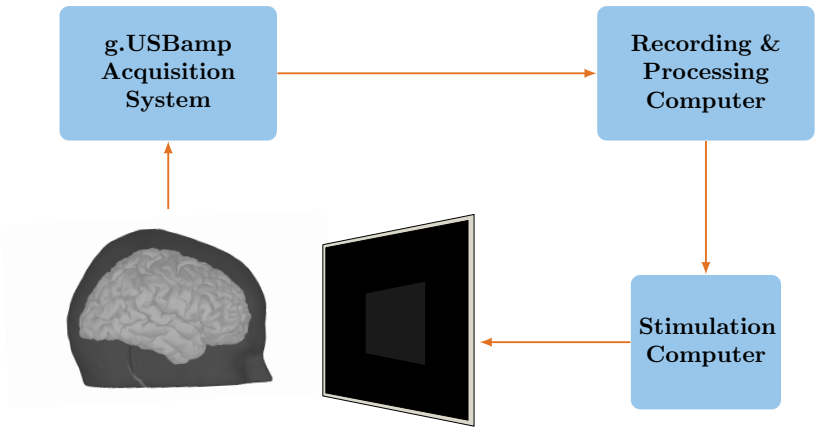


Figure 4.3: System overview showing the main hardware and software modules used in the experimental setup. Recording and stimulation were done on the same machine for experiment I.

level during task operation, breaks were given between sessions and subjects were free to stop a recording session at any time if they felt tired. Each experiment lasted for around 2 hours, including the preparation and break times. Break times varied between subjects and therefore number of recorded sessions and trials varied across subjects. Participants performed experiment I on the second visit, and the order of the remaining two experiments was counterbalanced across participants. Additionally, in order to examine the invariance vs. variability of ErrPs over time, some participants were invited to revisit the lab for additional times, where they performed one or more experiments for a second or third time. In all experiments, participants were instructed to mentally evaluate the interface actions as correct or erroneous. Fig. 4.3 shows the main modules used for the experimental setup in all three experiments, where the acquisition of the data and offline processing was facilitated with Simulink/MATLAB software (MathWorks, Massachusetts, United States). Visual stimulation in the different experiments was provided with a conventional LCD monitor.

Experiment I: Keyboard-mediated Ball Game Task

Very similar to [134], participants were instructed to use the left and the right arrow keyboard keys in order to move a ball towards a hole (respectively the sphere and the rectangle in Fig. 4.4), where both were aligned to the same horizontal line at the middle of the display. Each game run started with the ball randomly placed 5 steps away from the hole, either to the right or to the left with equal probabilities. Following each key press issued by the participant, the ball moved one step in the direction of the pressed key with a probability of 80% and in the other direction with the remaining probability. In order to isolate motor-related potentials due to key presses from potentials following the feedback presentation, the ball appeared in the new location τ s after each key press, where τ is uniformly drawn in the interval $[0.9, 1.1]$ s. Immediately after key presses, the color of the ball turned from green into red, indicating that further key presses will be ignored and the ball remained red for a period of 2 s. Subjects were instructed not to try to interact with

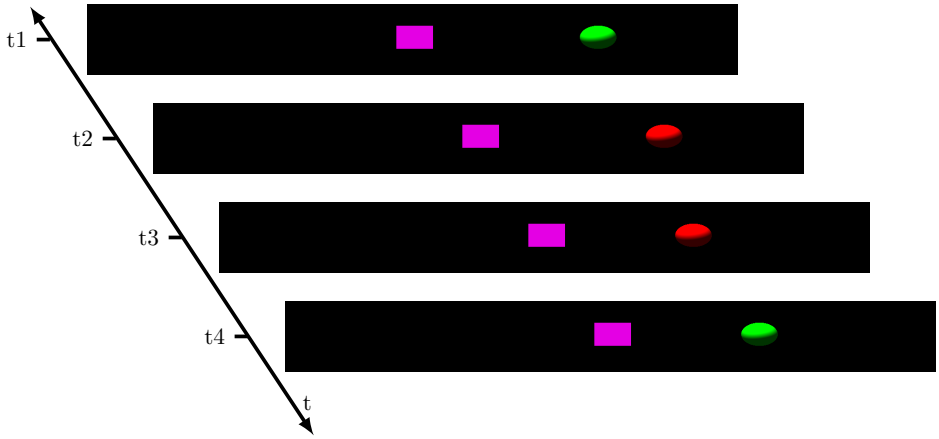


Figure 4.4: Key events in the keyboard-based interaction experiment. At time t_1, the user presses the key which brings the ball towards the hole (left key in the shown case). In the next display frame, the ball turns red and stays in place for a duration randomly drawn from the interval $[0.9, 1.1]$ s. Afterwards at t_3, the ball moves one step either to the right or the left according to user input and the error random generation. The ball remains inactive (red) after this movement till t_4. Shown is the correct case here and therefore EEG data time-locked to t_3 is considered a correct trial (noErrP). Note that $t_4 - t_1 = 2$ s.

the ball during this time. Once the ball reached the hole, a new game run was started after 2 s. Each subject finished multiple sessions and depending on the individual interaction pace, each session consisted of a varying number of runs and consequently a varying number of ErrP/noErrP trials.

Experiment II: P300-based Interaction with Central Feedback

A training session for P300 was first performed in copy spelling row-column flashing mode using the 6x6 spelling matrix, shown in Fig. 4.5 containing the alphanumerals. During this session, which lasted around 4 minutes (flashing of the first character started always after around one minute), users were instructed to copy spell 5 characters, with each new character cued for a short time before flashing starts. A full flashing sequence per character consisted of 16 repetitions, where in each repetition all rows and columns were flashed in a random order. In every flashing, a row or a column was highlighted on the screen for 100 ms, and the time between two consecutive flashes was set to 183.34 ms. An ML-LDA classifier was learned from this training session and was later used to estimate the hidden user intentions in the following online P300 sessions. In order to facilitate the acquisition of ErrP and noErrP training examples in the online P300 sessions, we have designed a simple mathematical task, wherein participants were instructed to attend to the maximum number in a 5x5 P300 matrix (example is shown in Fig. 4.6).

The P300 matrix in every new trial was filled with new random numbers, generated so all of them except one, were either 1 or 2-digit numbers. The remaining number, which was the maximum, consisted of 3 digits. Participants were informed about the fact that

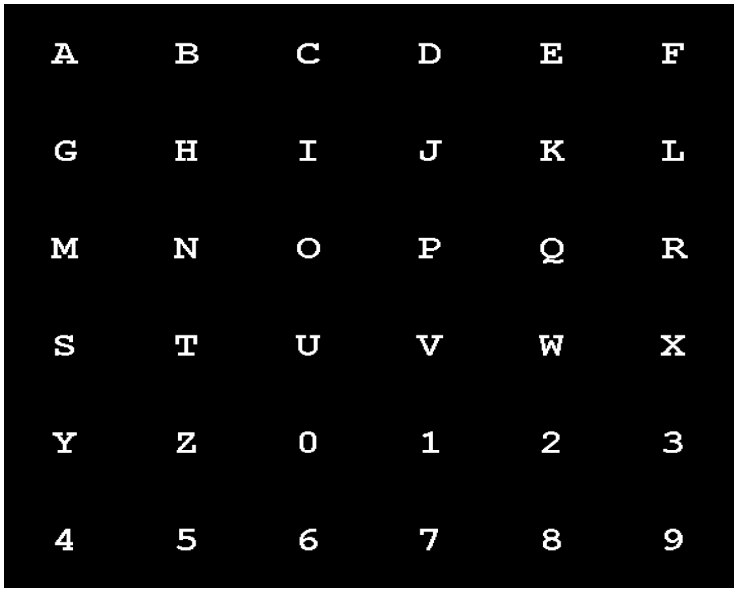
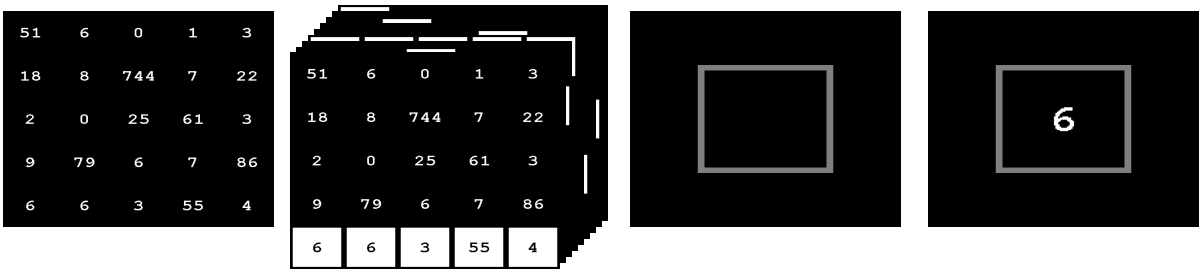


Figure 4.5: The alphanumerals P300 spelling matrix used in training sessions.

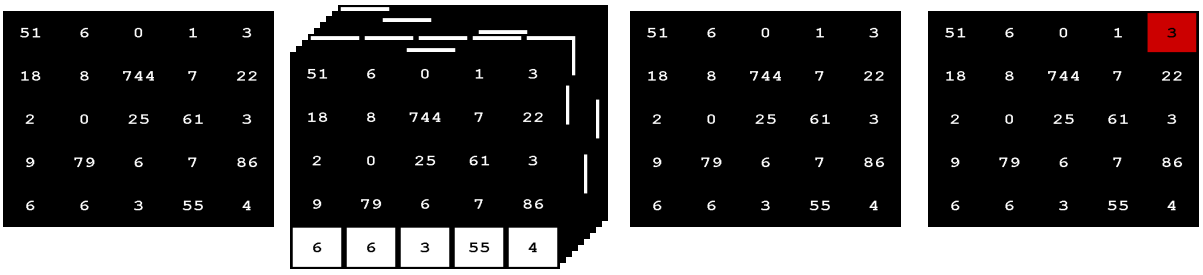
only one number consisted of 3 digits. This renders the mathematical task very simple or rather reduced to a simple visual search task, where the possibility a user makes a mistake by him/herself is minimized or rather nulled. Every ErrP/noErrP trial started with a new set of numbers randomly drawn and distributed within the P300 matrix, so that the location of the target maximum number was changed with every new trial. The update of the P300 matrix was facilitated by the XML interface described in [151]. In order to collect as many labeled ErrP and noErrP trials as possible, the flashing sequence in each trial was restricted in most cases to two repetitions. The accuracy of spelling, however, was monitored throughout the different sessions and the number of repetitions was sometimes modified to keep a relatively balanced number of ErrP and noErrP trials for each subject. After a decision was made about the user intent in the online P300 sessions, flashing was stopped and the mask was completely emptied for a duration of one second. Then, an empty square was shown at the center of the display for one second, aiming at directing the user's gaze to this location [141]. The estimated user intent (number) was shown afterwards inside the square for another second. The time between the end of the last flash and the presentation of the estimated number was therefore 2 seconds. Fig. 4.6(a) shows the key events in a single ErrP/noErrP trial. Each subject underwent a different number of sessions, each consisted of a different number of ErrP/noErrP trials.

Experiment III: P300-based Interaction with Inplace Feedback

This experiment shared all details of experiment II except the way the feedback was shown to users. Hereby, the mask remained displayed after flashing was stopped for 2 s and the estimated number was then highlighted for 1 s with a red square as shown in Fig. 4.6(b). After highlighting the estimated number, the P300 matrix was updated with a new set of random numbers.



(a) Central feedback presentation (Experiment II)



(b) Inplace feedback presentation (Experiment III)

Figure 4.6: Key events in experiments II and III.

Summary of the Experiments

Experiments II and III differ with respect to the feedback presentation and consequently lend themselves to examining invariance/variability of interaction ErrPs during P300-mediated interaction with respect to the different mental processes required to assess correct and incorrect interface actions. The reasons the maximum number task was introduced in these experiments are threefold. Firstly, this task requires no memorization of the last attended letter at the time of feedback onset as it is the case in spelling P300 applications with central feedback. Secondly, the task allows to collect ErrP/noErrP trials without relying on copy spelling and sham feedback modes. Thirdly, with the continuous updating of the P300 matrix, the task simulates the case of interacting with an adaptive P300 interface, which updates its contents based on the current context in a dynamic environment. This is of high importance to immersive robotic BCI applications (chapter 5).

Both experiments differ from experiment I with respect to the user input (keyboard vs. P300), where we also introduced about 1 s delay between the user input and the feedback onset in experiment I and 2 s in experiments II and III. The main difference, however, in our understanding is the mental processing required to assess the movement of the ball in experiment I compared to the processing of the central and inplace feedback in the P300-based interaction. As will become more clear in the following sections, both experiments I and II require a comparison step to assess the interface actions, which is absent in experiment III. Additionally, the presence of EOG artifacts that accompany feedback onset in experiment I will be useful to understand the effect of these artifacts in experiments I and III. These interrelations of the performed experiments can provide

more insights about the observed variability in ErrPs signals with respect to the mental processing that takes place after the feedback onset.

4.3.3 Analysis of Interaction ErrP Invariance and Variability

We base our analysis on the grand average of the correct and error potentials, to which we will refer as as the GAC, GAE respectively. The grand average difference waveform is simply the difference between GAE and GAC waveforms, and is referred to as GAD. The average waveforms are mainly computed from epochs time-locked to the feedback onset and plotted for a period of 1.5 s, i.e 0.5 s pre-stimulus and 1 s post-stimulus. In Appendix A, we reproduce the same waveforms for a period of 6 s centered around the time of feedback onset, i.e. 3 s pre-stimulus and 3 s post-stimulus. We chose this relatively long temporal epochs (when compared to other studies in the literature) to demonstrate the signal-to-noise ratio (SNR) of the obtained ErrP signals and other ERPs which can also be observed preceding and following the feedback onset (at $t = 0$). All waveforms will be shown for either the frontocentral electrode FCz or the vertex (i.e. Cz), as it is a common practice to do so and since the ERN is believed to have a frontocentral distribution [131]. Unless explicitly specified otherwise, the waveforms are computed for the epochs extracted from the pre-processed EEG data, where the continuous EEG data was first bandpass-filtered in the range $1 - 10$ Hz and then re-referenced to the common spatial average (i.e. the spatial mean was subtracted from each channel). The post-stimulus data in each epoch was corrected for baseline computed from the pre-stimulus data. The noisy trials, in which the EEG and EOG amplitudes surpassed certain thresholds, were excluded from the computation, aiming at preventing strong artifacts from appearing in the average signals. Additionally, the signed r^2 discriminability test is performed to highlight the spatial and temporal sources of variance between correct and error trials in the different experiments/tasks.

4.3.4 P300 Feature Extraction and Classification

The P300 feature extraction and classification pipeline is adopted from [152]. Hereby, only a subset of the EEG electrodes (Fz, Cz, P3, Pz, P4, PO7, POz and PO8) were used for P300 classification during experiments II and III. The continuous raw data from these 8 electrodes were first notch-filtered at 50 Hz, bandpass-filtered in the range $(0.5 - 20$ Hz) and downsampled to 64 Hz. Event-locked EEG epochs of 700 ms duration were extracted from the pre-processed continuous data after the onset of each target/nontarget flash. Features per electrode were obtained then by correcting these epochs for 100 ms pre-stimulus baseline followed by downsampling with a factor of 3. Features from the selected subset of electrodes were finally concatenated to form the labeled feature vectors (with a resulting dimensionality of 120). Each training session produced respectively around 160 and 800 target and nontarget training trials, which were used to train an ML-LDA classifier. In online sessions, feature vectors were obtained with the same pre-processing and feature extraction pipeline, where the unknown label of each feature vector was estimated with the learned classifier. Noteworthy here is that we did not aim at optimizing the accuracy of the P300-based interaction during the performed experiments, but rather at collecting as many ErrP/noErrP single trials as possible, with a reasonable number of flashing repetitions.

4.4 Experimental Results

4.4.1 Experiment I

The GAC, GAE and GAD for experiment I are shown in Fig. 4.7(a). The GAD exhibits an early negativity around 280 ms (N2) followed by a positivity around 340 ms (P3) and a later wider negativity around 460 ms (N4). By comparing the GAE and GAC waveforms, it becomes clear that these three deflections (especially P3 and N4) are specific to error trials.

The GAD waveform is also plotted in Fig. 4.8(a) alongside the GAD waveform obtained with the similar experiment in [134]. A considerable shift (around 20 ms) can be observed between the different peaks obtained in the two studies.

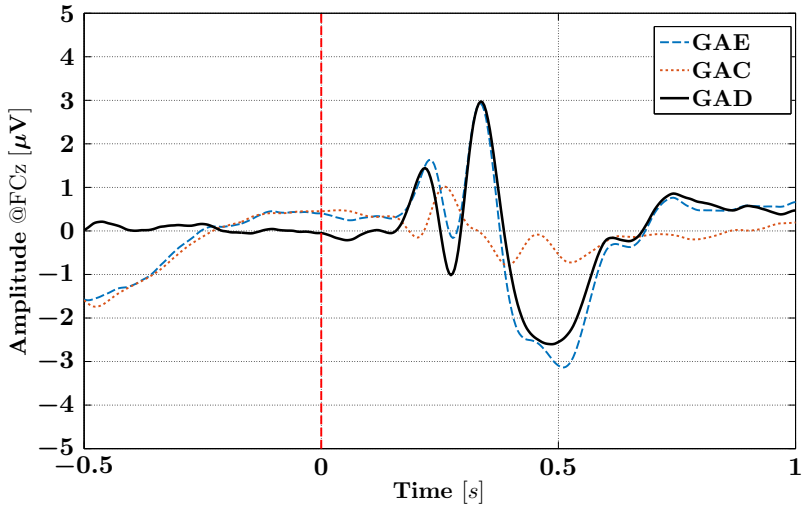
Furthermore, since the direction of the correct ball movement in experiment I was randomly alternating between the right and the left direction with each new run, correct and incorrect trials for these two conditions obviously resulted in different HEOG artifacts as can be seen in Fig. 4.9. This discrepancy in the HEOG signals, however, did not propagate to the electrode site FCz, as one can hardly observe any difference in the GAD waveforms computed separately for the two conditions. These results are in agreement with [134].

4.4.2 Experiment II

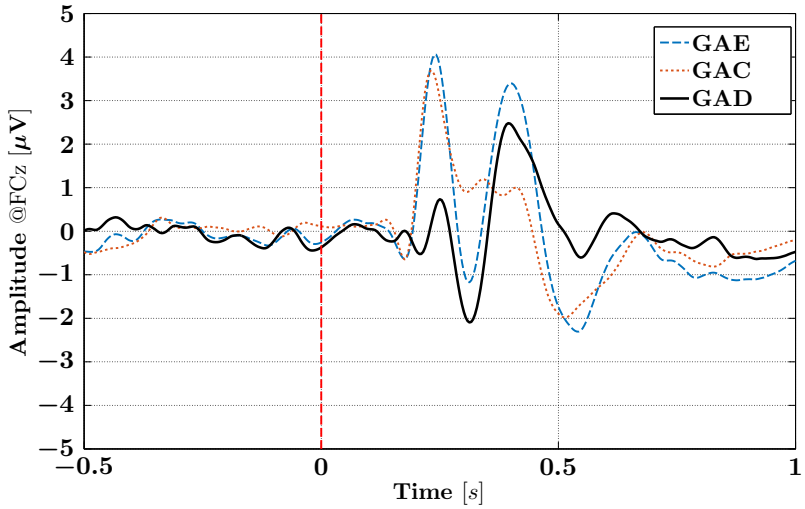
The GAD waveform at FCz, plotted in Fig. 4.7(b), is characterized with a negative peak at around 320 ms and a later positive peak at around 420 ms. Comparing the GAE and GAC waveforms reveals that these two deflections are specific to error trials. The GAD is also computed at Cz and compared to the results in [141, 146] in Fig. 4.8(b), where a difference in the latency and amplitude of the different peaks can be observed.

4.4.3 Experiment III

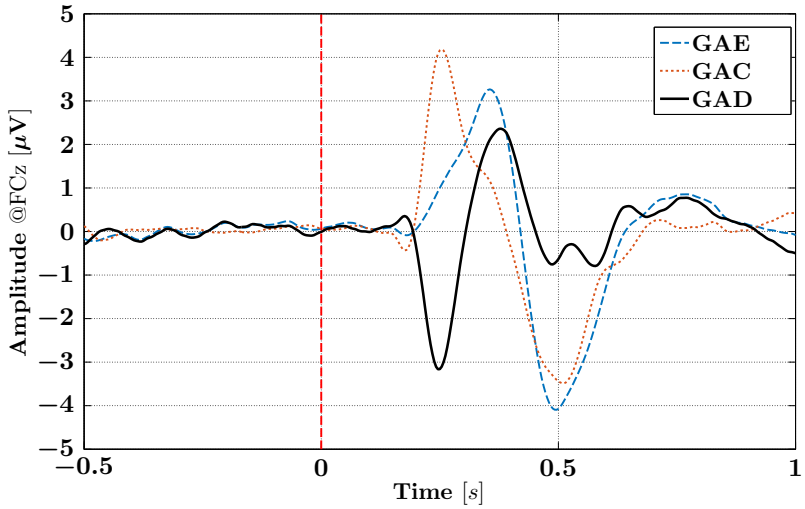
The GAD waveform at FCz, plotted in Fig. 4.7(c), is characterized by a negative peak at around 250 ms and a later positive peak at around 380 ms. These two deflections are a result of a small delay that can be noticed between the components of the GAE and GAC waveforms, which in turn can be attributed to the extra time required to notice the flash on the screen in case of error trials. The activity at the FCz site after $t = 0$ cannot be explained by the ocular artifacts that accompany errors, since similar activity is observed in correct trials where no eye movement is required. This can be easily seen by comparing the GAE and GAC signals. Additionally, it is expected that the ocular artifacts which accompany eye movement in case of errors to disappear as a result of averaging, since eye movements can be from any place to any other one on the P300 matrix. The ocular artifacts have been shown in the results of experiment I, even when these artifacts were more consistent within left and right trials, not to affect the respective GAD waveform at the FCz site.



(a) Experiment I

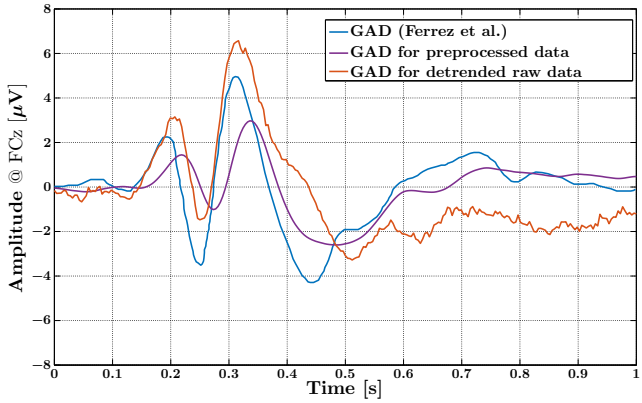


(b) Experiment II

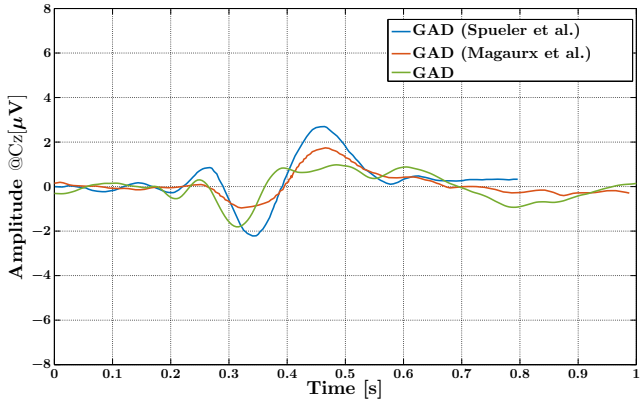


(c) Experiment III

Figure 4.7: The GAC, GAE and GAD waveforms were computed from the average of all subjects and recordings for all experiments.



(a) Experiment I



(b) Experiment II

Figure 4.8: Comparison of the obtained GAD waveforms with reported averages in the literature. (a) The GAD waveform at electrode FCz obtained with experiment I is compared to the GAD reported in [134]. There is a considerable shift between the two signals, but the shape remains very similar. The GAD is also shown when computed for detrended raw data, where it shows that the peaks observed for the raw data is closer to the ones in [134]. (b) The GAD waveform at Cz obtained from the data of experiment II, compared to the GAD in [141, 146].

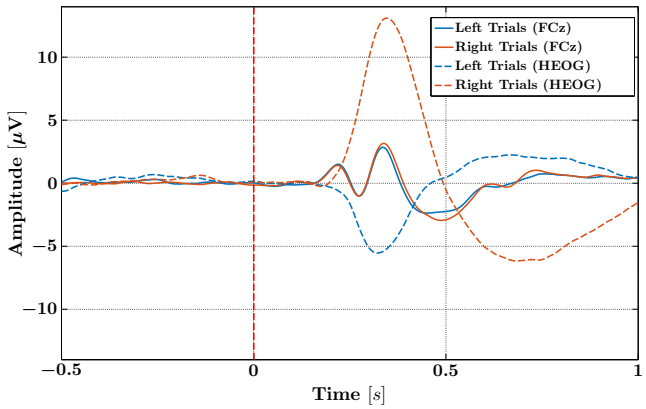


Figure 4.9: The GAD waveforms computed for from the left and right trials in experiment I at the electrode site FCz and for HEOG.

4.5 Discussion

4.5.1 Invariance with respect to Human Mental Processing of Interface Actions

In the previous section, we have compared the GAE and GAC waveforms to reason about the observed components in the GAD waveform for each experiment. In this section, we compare the GAC and GAE waveforms across experiments to reason about the variability/invariance in the computed GAD waveforms. To this end, figures 4.10(a), 4.10(b) and 4.10(c) rearranges the GAE, GAC and GAD waveforms from Fig. 4.7 and plots them over experiments.

Experiment II vs. III

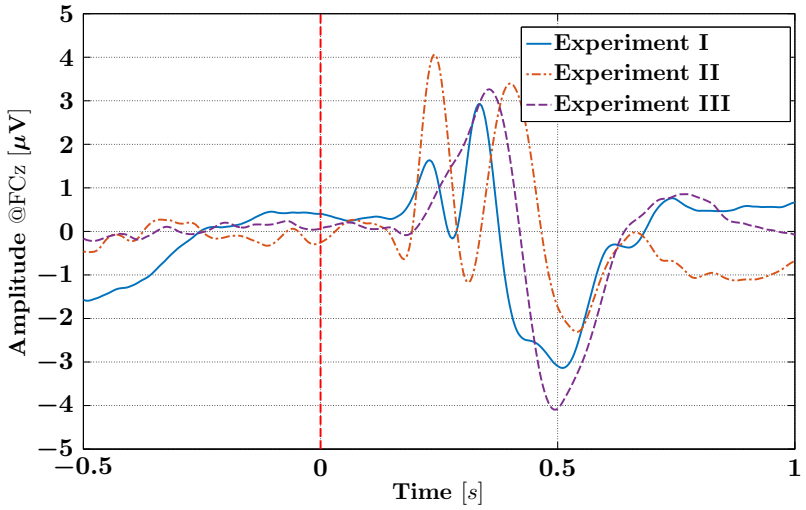
Both the GAD waveforms of experiment II and III are characterized by a negativity and a later positivity. So far, we have seen that not only the timing of the two deflections was different, but also their relation to the time course of the GAC and GAE waveforms. In experiment II, the two deflections were specific to error trials, whereas in experiment III, the deflections appeared as a result of the time delay of processing the error trials relative to the correct ones. With respect to mental processing of the interface actions (signaled to users by the feedback), no comparison whatsoever is necessary to assess whether these actions were correct or not in case of experiment III, as noticing that the visual feedback is not shown on the previously attended place on the P300 matrix is sufficient to realize that an error has occurred. Should the central method be used for feedback presentations as in experiment II, the subjects need to test whether the presented number was a 3-digit number or not. Obviously, the two feedback strategies required different mental processes to arrive at a decision whether the estimated number is correct or not, and this might explain the observed discrepancy in the GAD signals in Fig. 4.10(c). This discrepancy in the mental processing of interface actions manifested itself in a larger difference in the GAE waveforms across experiments when compared to the difference in the GAC waveforms, as can be seen from Figs. 4.10(a) and 4.10(b).

Experiment I vs. II

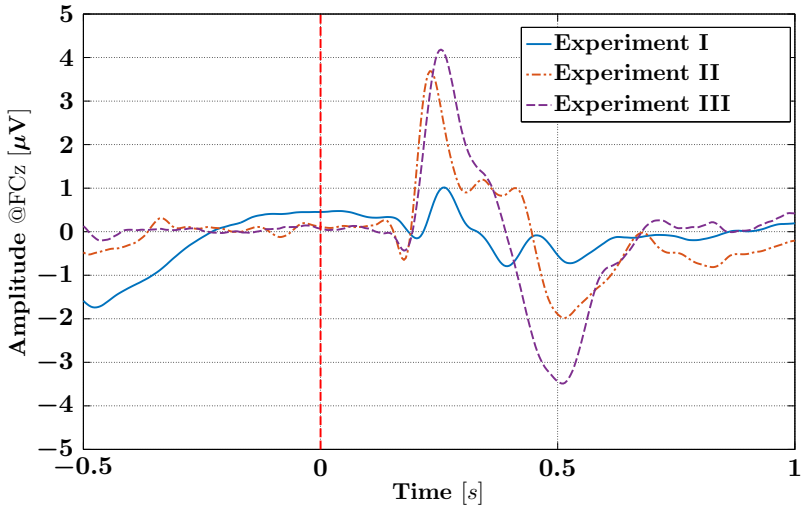
So far, we have seen that the deflections in the GAD waveforms of both experiments I and II are a result of similar deflections which were specific to incorrect trials. Additionally, one can observe in Fig. 4.10(a) a considerable similarity between the GAE of experiments I and II, which can be explained by the fact that both require a comparison step after the feedback onset, whereas this step is not required in experiment III.

Summary

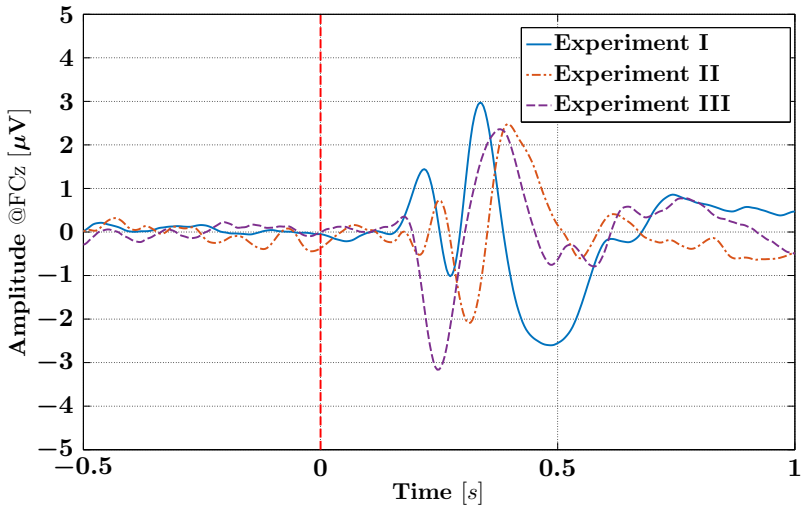
Altogether, the observed invariance/variability in the GAE, GAC and GAD waveforms can be attributed to differences in the mental processing required to assess interface actions.



(a) GAE



(b) GAC



(c) GAD

Figure 4.10: The GAE, GAC and GAD waveforms for the three experiments.

Therefore, different studies and the results obtained therefrom should be carefully compared, and a similarity in the GAD waveform should be always confirmed with respect to the separate averages of error and correct trials. It would have been certainly misleading to just show the similarity of GAD waveforms in experiments II and III, or to claim that the late N4 component is specific to experiment I.

4.5.2 Invariance and Variability in ErrPs over Time

Fig. 4.11 shows the GAD computed for some selected participants who completed one experiment or more for multiple times. Obviously, these plots show that the different components of ErrPs are empirically invariant over time for the same experiment/interface. This observation which is valid for all our three interfaces agrees with [134], who showed the same using an interface very similar to the one in experiment I. The SNR of the ErrPs per subject and per interface can also be seen quite stable over time.

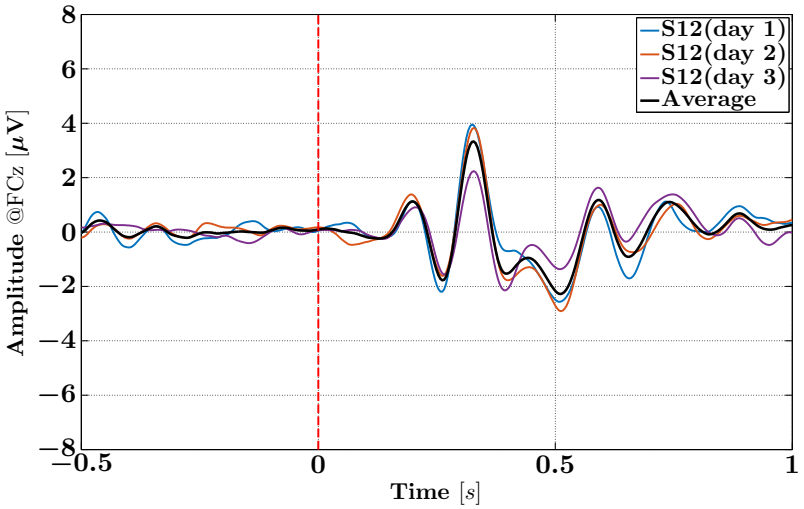
4.5.3 Invariance and Variability in ErrPs with respect to Subjects

Fig. 4.12 shows the GAD waveforms for all subjects in experiment I. One can observe inter-subject variability with respect to the amplitudes and SNR of the different deflections, whereas the timing of the different deflections seems to be consistent across subjects.

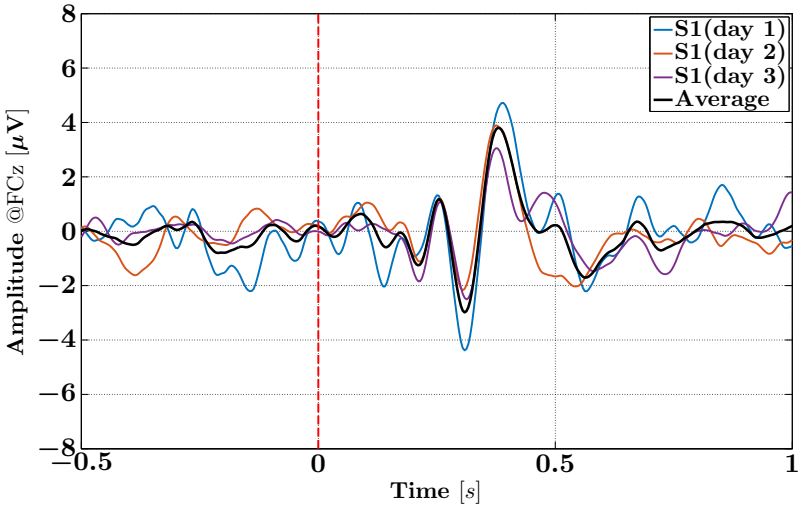
4.5.4 Invariance and Variability in ErrPs with respect to Pre-processing

In order to show how the pre-processing stage affects the obtained signals, we have introduced a simple modification to our pre-processing pipeline, whereby the grand average waveforms were also computed for all the EEG epochs extracted from the raw data of experiment I. These epochs were linearly detrended to remove the DC offset. The resulting GAD at FCz is shown in Fig. 4.8(a), suggesting that the relative shift and discrepancy of the amplitude values between our waveform and that in [134] can be attributed to a difference in the filtering stage or some other pre-processing details. In fact, the temporal filtering introduces time shifts in the data, and spatial mean subtraction reduces the amplitude of the observed deflections. Similarly, the discrepancies between our GAD waveform and those of [141, 146] for experiment II plotted in Fig. 4.8(b), can be attributed to differences in the pre-processing pipelines. However, one cannot rule out the contribution of the different mental processing to these discrepancies, as the maximum number and the language spelling tasks could have required different mental processing to assess interface actions.

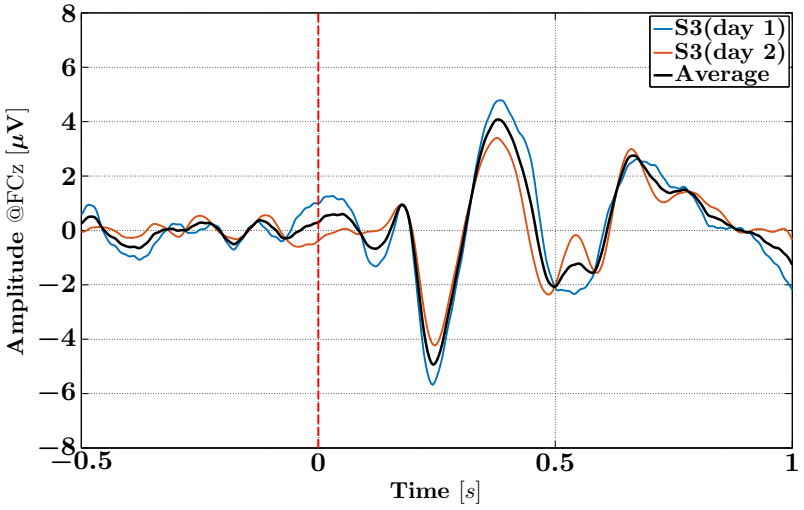
Furthermore, the results for the signed r^2 discriminability test performed for the two classes (ErrP vs. noErrP) problem is shown for the different experiments in Fig. 4.13. The test was performed twice. Firstly, with the trials extracted from the continuous EEG data, immediately after the step of the bandpass filtering. The second run was performed with the trials extracted from the continuous EEG after bandpass filtering and spatial mean subtraction (SMS). In both cases, noisy trials were removed by checking the EEG and EOG amplitudes, and therefore the only difference between the two runs was the step of the spatial mean subtraction. The plots show that, for all experiments, the discrimination



(a) Experiment I



(b) Experiment II



(c) Experiment III

Figure 4.11: Examples of the GAD waveforms computed for the different experiments for some subjects who performed one experiment or more on multiple occasions.

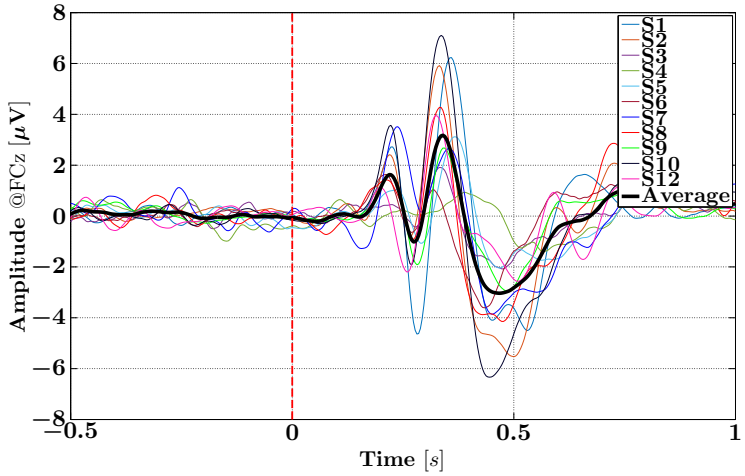


Figure 4.12: The average difference computed for all subjects in experiment I.

power between the ErrP and noErrP trials is temporally concentrated in the first second after the feedback onset.

4.5.5 Effects of Variabilities on the Classification of ErrPs

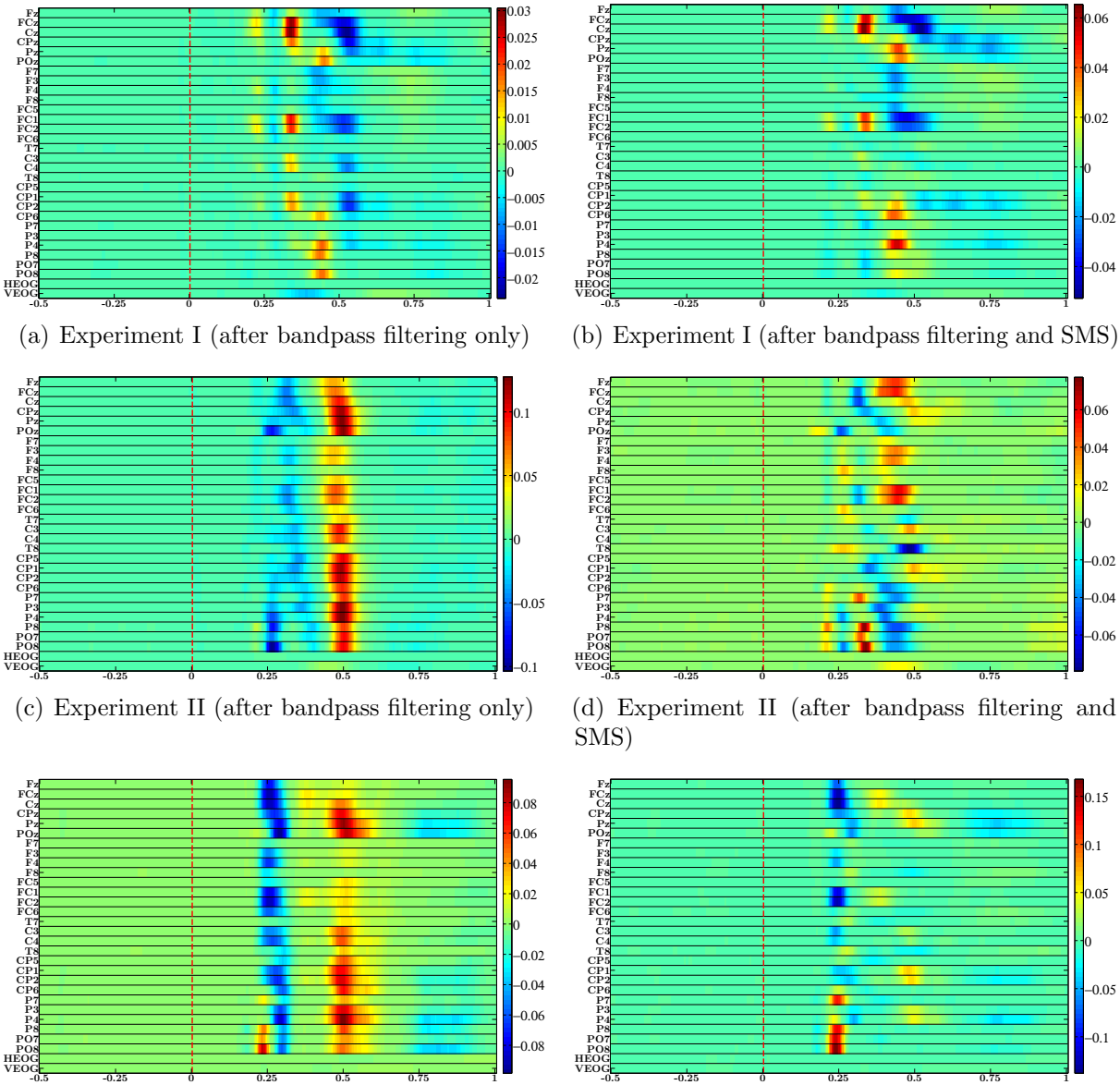
Feature Extraction and Classification

The nature of the observed ErrP signals guided the feature selection process. Hereby, EEG epochs were extracted as previously explained and features per electrode were obtained from downsampled post-stimulus data in the range $[\tau_1, \tau_2]$. Downsampling is achieved with averaging every 8 samples. Feature vectors were then obtained by concatenating features from 5 midline electrodes (Fz, FCz, Cz, CPz and Pz), leading to a dimensionality of 135, for $[\tau_1, \tau_2] = [0.15, 1]$ s. Shrinkage-LDA (see Sec. 2.4.2) was adopted to separate the correct (noErrP) and incorrect (ErrP) trials, assuming that the two classes differ only in their mean values.

Classification Results

A 10-fold cross-validation was used to evaluate the performance of the shrinkage-LDA classifier in predicting the noErrP and ErrP trials obtained during the different experiments. The results for each subject and per class are shown in table 4.1 for $\tau_1 = 0.15$ s and $\tau_2 = 1.0$ s. In order to better estimate the classification accuracy in online sessions, all extracted trials, including the ones with strong artifacts, were used for the classifier training and testing. The numbers of trials for each class per subject and per experiment are listed in table 4.2. The average accuracies are comparable with other results obtained with other interaction ErrPs studies. However, according to [153], most of these studies, including this one, involve a rather small number of subjects, and therefore performance differences might be a consequence of inter-subject variability.

Table 4.3 shows the classification accuracy of the shrinkage-LDA trained with data obtained from one experiment on one day, and applied to data from the same experiment on other days. The table only shows the results for the subjects who performed one or more



(a) Experiment I (after bandpass filtering only)

(b) Experiment I (after bandpass filtering and SMS)

(c) Experiment II (after bandpass filtering only)

(d) Experiment II (after bandpass filtering and SMS)

(e) Experiment III (after bandpass filtering only)

(f) Experiment III (after bandpass filtering and SMS)

Figure 4.13: Results for the signed r^2 test when (left) bandpass filtering was used for pre-processing and (right) when bandpass filtering and spatial mean subtraction were used in the pre-processing step.

Table 4.1: Mean classification accuracy

Subject	Class	Experiment I	Experiment II	Experiment III
S1	ErrP	0.76 ± 0.09	0.84 ± 0.13	0.78 ± 0.21
	noErrP	0.90 ± 0.03	0.80 ± 0.13	0.81 ± 0.15
S2	ErrP	0.80 ± 0.10	0.83 ± 0.17	0.87 ± 0.06
	noErrP	0.90 ± 0.04	0.86 ± 0.10	0.89 ± 0.10
S3	ErrP	0.58 ± 0.10	0.87 ± 0.17	0.95 ± 0.09
	noErrP	0.71 ± 0.05	0.84 ± 0.10	0.86 ± 0.11
S4	ErrP	0.57 ± 0.05	0.91 ± 0.04	0.88 ± 0.09
	noErrP	0.68 ± 0.04	0.85 ± 0.14	0.82 ± 0.17
S5	ErrP	0.66 ± 0.08	0.56 ± 0.14	0.62 ± 0.12
	noErrP	0.77 ± 0.05	0.45 ± 0.15	0.57 ± 0.18
S6	ErrP	0.70 ± 0.11	0.85 ± 0.13	0.71 ± 0.13
	noErrP	0.90 ± 0.02	0.65 ± 0.15	0.65 ± 0.19
S7	ErrP	0.73 ± 0.09	0.74 ± 0.13	0.85 ± 0.09
	noErrP	0.76 ± 0.05	0.82 ± 0.08	0.90 ± 0.07
S8	ErrP	0.78 ± 0.09	0.75 ± 0.14	0.88 ± 0.08
	noErrP	0.89 ± 0.03	0.68 ± 0.15	0.87 ± 0.06
S9	ErrP	0.66 ± 0.09	0.79 ± 0.09	0.87 ± 0.06
	noErrP	0.80 ± 0.03	0.65 ± 0.11	0.88 ± 0.12
S10	ErrP	0.81 ± 0.04	0.73 ± 0.07	0.78 ± 0.08
	noErrP	0.89 ± 0.03	0.63 ± 0.17	0.69 ± 0.12
S12	ErrP	0.82 ± 0.11	0.80 ± 0.17	0.82 ± 0.17
	noErrP	0.90 ± 0.02	0.77 ± 0.09	0.88 ± 0.10
S1(day 2)	ErrP	-	0.72 ± 0.10	0.85 ± 0.08
	noErrP	-	0.69 ± 0.12	0.83 ± 0.11
S1(day 3)	ErrP	-	0.83 ± 0.13	-
	noErrP	-	0.81 ± 0.10	-
S3(day 2)	ErrP	-	-	0.86 ± 0.08
	noErrP	-	-	0.88 ± 0.06
S4(day 2)	ErrP	0.57 ± 0.09	-	0.89 ± 0.05
	noErrP	0.67 ± 0.05	-	0.92 ± 0.05
S5(day 2)	ErrP	-	-	0.73 ± 0.13
	noErrP	-	-	0.67 ± 0.14
S7(day 2)	ErrP	-	0.74 ± 0.18	-
	noErrP	-	0.93 ± 0.04	-
S12(day 2)	ErrP	0.76 ± 0.07	0.79 ± 0.08	-
	noErrP	0.92 ± 0.02	0.76 ± 0.09	-
S12(day 3)	ErrP	0.76 ± 0.06	-	-
	noErrP	0.88 ± 0.03	-	-
mean	**ErrP**	**0.71± 0.09**	**0.78± 0.08**	**0.82± 0.08**
	noErrP	**0.83± 0.09**	**0.75± 0.12**	**0.81± 0.11**

Table 4.2: Number of extracted trials written in the form (Number of incorrect trials/Total number of trials). Percentage of incorrect to total trials is shown in parentheses.

Subject	Experiment I	Experiment II	Experiment III
S1	274/1315 (20.8%)	68/149 (45.6%)	62/125 (49.6%)
S2	228/1099 (20.7%)	71/149 (47.7%)	77/191 (40.3%)
S3	273/1382 (19.8%)	64/139 (46.0%)	111/186 (59.7%)
S4	337/1664 (20.3%)	164/245 (66.9%)	85/147 (57.8%)
S5	262/1381 (19.0%)	135/215 (62.8%)	152/248 (61.3%)
S6	340/1675 (20.3%)	124/191 (64.9%)	119/190 (62.6%)
S7	277/1432 (19.3%)	81/250 (32.4%)	88/278 (31.7%)
S8	285/1539 (18.5%)	174/295 (59.0%)	122/289 (42.2%)
S9	327/1532 (21.3%)	194/294 (66.0%)	190/313 (60.7%)
S10	307/1611 (19.1%)	212/331 (64.0%)	241/323 (74.6%)
S12	170/760 (22.4%)	111/211 (52.6%)	87/277 (31.4%)
S1(day 2)	-	174/341 (51.0%)	162/309 (52.4%)
S1(day 3)	-	174/338 (51.5%)	-
S3(day 2)	-	-	156/326 (47.9%)
S4(day 2)	313/1408 (22.2%)	-	178/333 (53.5%)
S5(day 2)	-	-	184/302 (60.9%)
S7(day 2)	-	66/286 (23.1%)	-
S12(day 2)	296/1594 (18.6%)	187/433 (43.2%)	-
S12(day 3)	290/1469 (19.7%)	-	-

experiments for multiple times. Despite the reduction in obtained accuracies, these results reemphasize our belief that ErrPs are empirically invariant to some extent over time for the same experiment/interface, which has been previously supported by the similar GAD waveforms in different days. On the other hand, when a classifier was learned from the data of one experiment and applied on data from another one, the classification accuracies were very close to random levels.

Table 4.4 shows the classification accuracy of the shrinkage-LDA when tested on the data of each subject and trained from the data of all other subjects for each experiment. The obtained accuracies are inferior to those in table 4.1, but they are clearly above chance-level on average and for most subjects, and therefore they suggest that stable features across subjects can be obtained, even with a simple pre-processing pipeline like the one we used in this work.

The relative robustness of the shrinkage-LDA classifier for all experiments/interfaces over time, rules out the possibility that the obtained chance-level classification accuracy of the shrinkage-LDA classifier, when trained with data from one experiment/interface and tested on a different one, was due the fact that the different interfaces were used on different days. Given that we used the same pre-processing and trial extraction pipeline for all experiments, this observation strengthens our belief that the discrepancies in the GAE, GAC and GAD waveforms for the different interfaces are a result of the different mental processes needed to evaluate the interface actions. Additionally, the GAD waveforms plotted throughout this work were mainly the difference between the means of ErrP/noErrP classes,

Table 4.3: Mean classification accuracy for classifiers trained on one day and tested on a different one.

Experiment	Subject	Training day	Testing day	ErrP	noErrP
I	S4	day 1	day 2	0.53	0.71
	S12	day 1	day 2	0.69	0.93
	S12	day 1	day 3	0.47	0.93
	S12	day 2	day 3	0.60	0.90
II	S1	day 1	day 2	0.47	0.83
	S1	day 1	day 3	0.57	0.86
	S1	day 2	day 3	0.74	0.75
	S7	day 1	day 2	0.77	0.86
	S12	day 1	day 2	0.79	0.68
III	S1	day 1	day 2	0.69	0.75
	S3	day 1	day 2	0.88	0.75
	S4	day 1	day 2	0.70	0.85
	S5	day 1	day 2	0.69	0.59

Table 4.4: Accuracies of LDA classifier tested on each subject when trained from data of all other subjects.

	Experiment I		Experiment II		Experiment III	
Subject	ErrP	noErrP	ErrP	noErrP	ErrP	noErrP
S1	0.83	0.76	0.84	0.41	0.81	0.30
S2	0.84	0.80	0.87	0.51	0.81	0.74
S3	0.55	0.73	0.77	0.49	0.82	0.85
S4	0.35	0.81	0.75	0.80	0.86	0.61
S5	0.43	0.81	0.53	0.47	0.57	0.53
S6	0.48	0.90	0.82	0.37	0.83	0.37
S7	0.75	0.54	0.65	0.64	0.76	0.86
S8	0.74	0.78	0.64	0.70	0.53	0.87
S9	0.57	0.83	0.78	0.44	0.95	0.61
S10	0.86	0.79	0.59	0.59	0.83	0.49
S12	0.79	0.82	0.50	0.88	0.63	0.82
mean	**0.66**	**0.78**	**0.70**	**0.57**	**0.76**	**0.64**

i.e. $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. Recall that for classification, we assumed that the two classes differ only with respect to their means. The differences observed in GAD, GAE and GAC waveforms across experiments, suggest that one needs to at least recalibrate trained classifiers for the new means, which requires to acquire new training examples. The effectiveness of recalibration with respect to the mean values has been already shown in [63].

4.6 Conclusions

Interaction ErrPs are special EEG manifestations of user's awareness of the incidents, at which the interface fails to recognize his/her intention in a first stage selection. As such, detecting signatures of interaction ErrPs in the brain activity after feedback presentations might be used to (in-)validate previous selection. By considering three different noisy interfaces, this chapter has analyzed the invariance vs. variability of interaction ErrPs with respect to: (1) the mental processes required to assess interface actions (2) time (3) subjects and (4) the pre-processing of the raw EEG data.

With a fixed pre-processing pipeline for all experiments, it has been shown that the interaction ErrPs for each experiment enjoyed invariance to some extent across subjects and over time. This invariance in turn resulted in a relative robustness of the shrinkage-LDA classifier across subjects and over time. On the other hand, the mental processes which are required to assess interface actions were shown to highly affect the observed interaction ErrPs. This has been supported by the differences in the observed GAC, GAE and GAD waveforms across interfaces/experiments. The observed variability with respect to the respective mental processes has also been shown to make it difficult for a classifier learned from the data of one experiment/interface to straightforwardly transfer to other experiments/interfaces. In particular, despite the similarity of the GAD waveforms in experiments II and III, where subjects performed exactly the same task but with different feedback presentation methods, shrinkage-LDA classifier which was learned from data of one experiment showed chance-level accuracies when tested on data from the other interface. Again, this proves the sensitivity of the ErrPs to the nature of the mental processing of correct and incorrect interface actions, which takes place immediately after the feedback onset.

Hereby, we didn't try to compare the many existing advanced spatial and spatio-temporal filtering methods that enhance the SNR of the ErrPs. But using a simple alteration of the pre-processing pipeline, we have shown that great variability can be introduced with respect to the timing and amplitude of the different components of interaction ErrPs. In summary, the main source of variability in the different interaction ErrP studies (including ours) was found to be the different mental processing required to assess interface actions. Conversely, given two interfaces which require the exact mental processing after feedback onset in correct and incorrect trials, our results suggest that any discrepancy with respect to the amplitude and latency of the different ErrP components are most likely caused by inter-subject variability, the non-stationarity of the EEG data or differences in the pre-processing pipeline. These discrepancies propagate as well to the classification step and affect the obtained accuracies.

Some relevant observations about the different sources of invariance can be found

scattered in different studies of related work. To our knowledge, this is the first time such invariance analysis is carried out for interaction ErrPs to this scale. This undoubtedly helps to extend the understanding of the different sources of invariance and variability in observed average ERPs across tasks and across different studies, which in turn might guide the design of transferable classifiers. In this regard, we have provided a sensible explanation to the necessary recalibration of class means for classifiers to transfer well across different interfaces, should these interfaces differ with respect to the mental processing of the feedback. The novel inplace presentation method, introduced in this work, has proven its adequateness as an alternative strategy for feedback in hybrid P300-ErrP systems. Assessing interface actions with this method boils down to assessing the correctness of the highlight location on the P300 mask, regardless of the nature of the application under consideration. This can be of great importance to adaptive interfaces, where the meaning of the interface elements change over time. On the contrary, the popular central feedback strategy might require application-specific assessment processes, since the feedback itself is application-specific. For instance, feedback of estimated letters in spelling tasks requires to assess if these letters match the previously attended ones and giving feedback in the max number task (introduced in this work), requires to assess if the shown numbers are composed of three digits. However, these claims need to be empirically tested in further research.

Despite the concrete results which we have found in this work, many aspects thereof still can be improved. As previously discussed in chapter 2, the peaks and troughs in observed ERPs do not have simple one-to-one mappings to the underlying source activity in the brain, and this is the main reason we did not try to give any interpretation to the observed peaks in the different experiments/interfaces. It might be beneficiary, therefore, to explore the different existing spatial and spatiotemporal filtering techniques to reveal possible source activity. Analyzing invariance and variability at the level of source activity is more fundamental to understand the discrepancies between the different mental assessment processes of interface actions.

Additionally, the problem of integrating interaction ErrPs in hybrid SSVEP-ErrPs systems has not been addressed so far. Since interaction with SSVEP happens usually in a faster rate than with P300, considering ErrPs within SSVEP-based interaction might bring some overhead, as it is required to block other stimulation (to reduce noise) around 1 s post-feedback in order to observe single trial ErrPs.

5 Adaptive BCIs for Robotic Embodiment

So far, the focus was laid upon advancing BCIs in a way that is relevant to many application fields. In particular, we aimed at increasing the rate of correct detections (chapter 3) and integrating interaction ErrPs into BCIs as a validation stage (chapter 4). In the present chapter, our focus is turned to the specifics of BCI applications in immersive robotic embodiment systems and the challenges inherent to them.

Fig. 5.1 depicts a schematic of a closed loop BCI-based immersive robotic embodiment system. Users continuously receive perceptual feedback from their robot avatar embedded in the remote environment and communicate their intentions by selectively attending (i.e. gazing) to one of the available interface elements. Other paradigms like motor imagery are possible [21], but we mainly consider BCIs based on selective attention. Ultimately, in order to fully immerse users in the remote environment, the perceptual channels from the local environment should be replaced by the ones that reflect the state of the avatar in the remote environment. In this regard, vision is the most important perceptual modality, but other modalities undoubtedly would enhance the user subjective experience of immersion and embodiment.

The robot, on the other side, continuously receives user commands and translates them using available low and high level controllers into robotic actions, that change the state of the environment or the robot itself. The objective of embodiment systems is to allow users to exercise physical and social interaction in the remote environment through their avatars, which are typically equipped with locomotion and manipulation capabilities in order to allow for enhanced interaction. Hereby, several successful demonstrations of BCI-based robotic embodiment systems are reported in the literature. In [21], a HOAP3 humanoid robot is teleoperated with MI-BCI on the basis of decoding fMRI scans, where the user moves the robot according to instructions. The work in [23] uses a teleoperated android

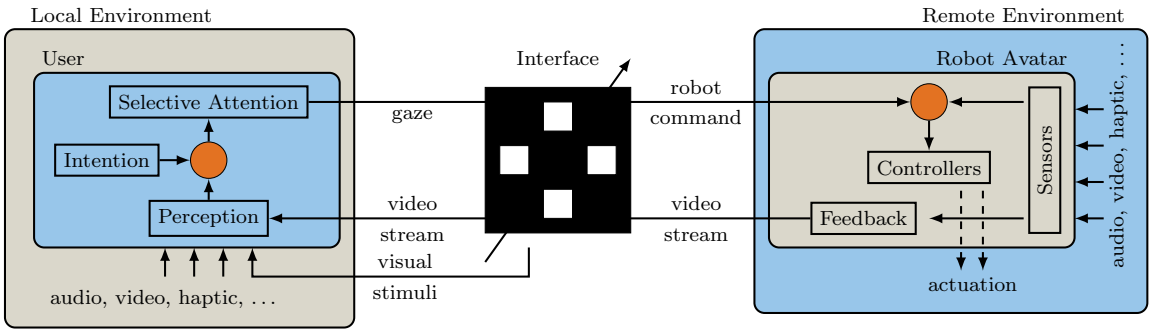


Figure 5.1: Closed loop BCI-based robotic embodiment system. Users communicate their intentions by selectively attending to one of the interface elements. An SSVEP-based interface is shown as an example here.

with MI-BCI to study embodiment and its effects on MI learning. In [22], the human-size humanoid robot HRP-2 is controlled to perform different tasks in the environment (e.g. navigation and object manipulation). Navigation here is performed either with incremental commands or in a goal-oriented fashion. Object manipulation includes touching/grasping objects, and handing objects to users.

Nonetheless, at each time during BCI-based interaction, only a limited number of interface elements can fit into the interface, whether that be based on P300, SSVEP or MI and consequently, only a few commands will be available to users. This gives rise to the primary challenge in robotic applications of BCIs, that is, the considerable mismatch between the low-bandwidth interface and the bandwidth-hungry application. Adaptive BCIs offer a way, and perhaps the only way, to overcome the bandwidth mismatch.

Adaptive user interfaces (AUIs), in general, have been thoroughly investigated within the field of human-computer interaction (HCI) aiming at reducing the cognitive load of users during interaction. In this regard, the reader might recall the new "New Tab" feature in modern web browsers, that shows the most recently and/or frequently visited websites. In graphical user interfaces (GUIs), adaptivity is often reserved for a small area in the interface, e.g. to present the user with the mostly used interface elements. Interface self-adaptation is expected to bring savings in time and in the user effort needed to arrive at the commands of interest. Savings can be computed for example as the difference between the cost of selecting a command through the adaptive interface (C_a) and the cost of selecting the same command with a deterministic hierarchical interface (C_h) [154]. C_a can be computed from the time needed to visually scan the predicted interface elements till the element of interest is found plus the required time for selecting that specific element. C_h, on the other hand, is computed from summation of the cognitive time needed to formulate the sequence of interface selections, and the time needed to make these selections. The design of AUIs is a highly challenging task due to the fact that interface self-adaptations need to (1) be unobtrusive to users (2) keep the users in control of the system (3) allow the user to maintain a coherent mental model of the interface. Adaptive BCIs inherit all these challenges, but given the limited bandwidth of the interface, self-adaptation is a necessity, rather than an extra feature.

In order for interface self-adaptations to be of any benefit, the system should be able to infer the hidden goals, which the user tries to achieve in the domain. This way, initiatives made by the interface can be received positively by the user. The problem of inferring user goals from observations is typically referred to as goal or intention recognition (IR), and is the main concern of this chapter. By considering navigation tasks in BCI-based robotic embodiment systems, a general user-agnostic Bayesian framework is proposed. Hereby, novel intuitive heuristics are used to model the user behavior, on the basis of general behavioral patterns that are observed in humans during task execution. The Bayesian framework is designed for the general population allowing to infer user intentions, without any prior training. The output of the Bayesian inference module is a belief vector that sorts all target goals in the environment according to how probable they are given the observed evidence. Moreover, a novel metric that measures the non-uniformness of the belief vector is proposed as to reflect upon the confidence of the inference system in its computed beliefs. To show the usefulness of the computed beliefs and confidence measures, a simple probabilistic shared control scheme is devised so that some adaptations are applied to the

robot movements, according to the belief/confidence information. In order to evaluate the performance of the Bayesian inference module, experiments were conducted with healthy subjects in an immersive robotic embodiment setup and in simulation. Results show that the intention recognition system is able to track the hidden user goals with relatively high accuracy. When the belief of the intention recognition module is used to modulate parameters of the robot movement, less user effort (measured by the number of interaction) is required to accomplish the assigned tasks.

This chapter proceeds as follows. Sec. 5.1 provides background information about adaptive interfaces in the context of robotic embodiment systems, to motivate the work on intention recognition. Sec. 5.2 provides a short review on state-of-the-art approaches to intention recognition for robotic navigation tasks. The problem statement for intention recognition and our Bayesian inference systems are detailed in Sec. 5.3. Detailed information about the experiments used to evaluate the proposed methods is described in Sec. 5.4. Experimental results are reported in Sec. 5.5 followed by a discussion in Sec. 5.6. This chapter concludes in Sec. 5.7.

5.1 Background and Motivation

User commands in BCI-based robotic navigation applications range from commanding the mobile base to move for some distance in one direction, to commanding the mobile base to move to a specific location or to autonomously perform complex maneuvers like “move to the kitchen” or “walk through the doorway”. Obviously, the varying degrees in the goal-directness of incoming user commands require adequate levels of autonomy at the robot side. And conversely, by deciding first on the level of autonomy, adaptive interfaces can suggest conforming sets of commands for interaction. The exact relation between the interface and the level of autonomy, therefore, can be decided upon either by the interface itself or by the user. Either way, the higher the robot autonomy, the less control is left to the user. In a general sense, robot autonomy is thought of here as a continuum where the fully autonomous and fully manual control modes are its extremes. Often the selection of the optimal point on the autonomy continuum is done in two steps. First, the continuum is discretized into different general levels (or modes) and the appropriate mode is selected. The optimal point of automation is then selected within that mode.

The autonomy continuum in teleoperation systems is discretized into several major levels by Sheridan et al.[155] with respect to the nature of the feedback given to the human user and the sharing of functions required to control the machine. Similarly, but for a wider set of applications, Endsley and Kaber [156] differentiates between 10 different levels. The functions which need to be shared are monitoring, generating and selecting plans and executing a selected plan. Clearly, the different levels of autonomy/assistance are dependent on the application at hand. In the context of robotic navigation applications, we borrow from the aforementioned taxonomies and that in [157] and adopt the 6 basic levels or robot operation modes shown in Fig. 5.2.

In *goal-oriented control* mode, users only communicate the end goal location they have in mind, and the robot autonomously navigates towards it. This requires global knowledge of the environmental map and path planning capabilities at the robot side. In the *help with*

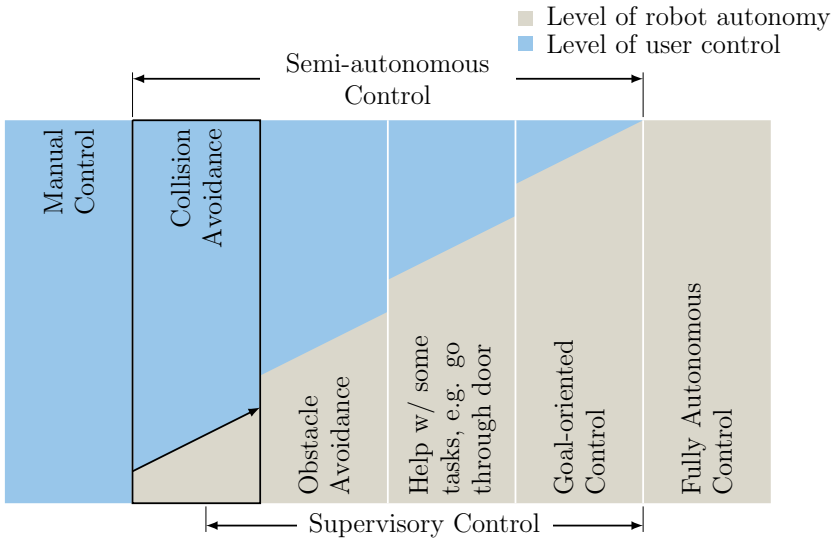


Figure 5.2: Robot operational modes for navigation tasks. The level of automation can be modulated within each mode of operation as well.

some tasks mode, like walking through a door, the robot needs to plan a safe local trajectory, by which it remains e.g. equidistant from the doorjamps. In the *obstacle avoidance* mode, it is the robot's job to execute maneuvers that avoid obstacles while the human user is issuing commands that move it through the environment. In the *collision avoidance* mode, a typical behavior of the robot is to halt movement in the face of obstacles. During the movement of the robot in these modes, the user might maintain supervisory control of the robot, e.g. to stop movement in case of emergency or a plan change. As hinted previously, the different automation levels define the nature of interface self-adaptations. For instance, in the *goal-oriented* mode, the interface might provide the user with the most probable next goals (or goal locations) in the environment. Such predictions can be performed on the basis of the transition probabilities between the different goals in the environment, which can be learned e.g. from the history of user interaction with the system. In absence of any knowledge about the user preferences in the domain, other interface adaptation strategies should be devised. One possible approach, is to allow users to interact with the robot in a lower autonomy mode, so that some evidence can be gathered throughout interaction and used to predict the target goals of the user.

In this work, we mainly focus on the problem of goal recognition during interaction within the collision avoidance mode, where user input is also restricted to incremental commands that translate or rotate the robot in the different directions with predefined steps. The reason for this is threefold. First, it has been noted in [158], that powered wheelchair users want to actively drive the wheelchair rather than being merely its passengers. The same can be said about the users of teleoperation/embodiment robots, and therefore, it is imperative to give users the sense of control they need [159], as much as can be allowed by other existing constraints like the safety of the human, the machine and the environment, the capabilities of the robot and the abilities/disabilities of the human users. The collision avoidance mode leaves most of the control in the hands (or rather the brains) of the users, and at the same time guarantees navigation safety. Second, and most importantly, it allows

us to develop the methodologies, by which gathered evidence from user commands in low autonomy levels can guide the interface self-adaptations in the higher autonomy modes. Third, with long term interaction with the robot, transition probabilities can be learned, and used later to guide interface adaptations in higher autonomy modes. The next section will provide a brief review on the problem of goal/intention recognition in general, with some emphasis on intention recognition in navigation tasks.

5.2 Related Work

The term *plan recognition* has been defined by Schmidt et al. [160] as the process of inferring an agent's goals from observing the actions the agent is performing in the domain and organizing these actions into a plan structure which explicitly describes the goal-subgoal relations among them. Conformingly, recent research on human action understanding within navigation contexts shows that a machine equipped with inverse planning is able to efficiently model this cognitive process [161]. We refer interchangeably to the agent whose actions are observed as the *actor* or the *user*, whereas the observing agent that makes the inference is referred to as the *recognizer*.

The *goals* refer to desired states of the world or states of knowledge about the world. In order to arrive at these states, an actor needs to perform a sequence of actions that defines a *plan*. *Actions* taken might change the state of the world, or the state of the actor's knowledge about the world. Actions are defined by their *preconditions*, i.e. possible states in which they can take place, and *effects* that describe the state transition when the action is performed. In some domains, it is possible to enumerate all possible goal states and plans an actor might have. The set of these plans is referred to as *plan library*. In the kitchen domain for example, goals might refer to meals the actor wishes to prepare and a cookbook defines the plan library, where the term plan in this case will refer to a recipe in the cookbook [162]. For the general domain of human plan recognition, the set of web pages in the website wikihow [163] might be considered as a linguistic encoding of a quite large set of possible plans pursued by humans.

The problems of plan and goal recognition have many details in common. Systems which try to infer the final target state only are referred to as *goal recognition* systems, and those which additionally predict the sequence of actions which moves the world from its current state to the desired final state are referred to as *plan recognition* systems. In navigation applications, these refer respectively to inferring the final goal location the user has in mind and the full trajectories, with which the user wants to arrive at these target locations. Depending on the level of abstraction used in recognition, both systems might be referred to as *intention recognition* systems [1].

Intention recognition systems can be classified with respect to the actor's attitude towards the recognizer into three categories. First, *the keyhole recognition* refers, as the term suggests, to the case when the actor is not aware of the existence of the recognizer [164, 165]. In the second category, which is referred to as *intended recognition*, the actor cooperatively chooses actions which reduce the ambiguity about his/her hidden intentions [166]. Finally,

[1] Other abstraction levels exist. For instance, the user might want to move to the kitchen to prepare a meal, and this very objective of the user can be considered the user intention.

adversarial recognition concerns itself with the case when the actors actively try to confuse the recognizer by choosing misleading actions [165, 167].

Charniak and Goldman [168] argue that the problem of intention recognition is largely a problem of inference under conditions of uncertainty, rendering the probabilistic approaches of best fit. Such approaches introduce a numerical measure of belief that reflects how likely or probable individual plans are and can thus, explicitly represent the uncertainty associated [169]. The vector that contains the probability of all plans is often referred to as the *belief vector*. Bayesian networks, and in particular, dynamic Bayesian networks (DBNs) are probably the mostly used probabilistic method for intention recognition in different domains, e.g. for predicting user plans in a game [164].

In robotic navigation contexts, Perrin et al. [170] use a DBN for goal recognition, in a system, with which the user interacts by either confirming or rejecting its propositions. The proposed Bayesian network additionally includes variables that integrate the time of the day, whether the phone is ringing and the previously visited goals. The conditional probability distributions (CPDs) used in the network are learned from training sequences. Alternatively, goal recognition is realized in [158] on the basis of simple metrics that are computed from the distances and relative orientations the wheelchair has to available goal locations (available doors in this case).

The works in [171, 172] use a recursive Bayesian update for plan recognition based on environmental data and history of interaction. The Bayesian update hereby is defined with

$$\begin{aligned} P_k(\mathbf{i}_{k-m:k} \mid u_{k-m:k}) = & && \text{(posterior)} \\ & P(u_k \mid \mathbf{i}_{k-m:k}, u_{k-m:k-1}) && \text{(user model)} \\ & P(\mathbf{i}_k \mid \mathbf{i}_{k-m:k-1}, u_{k-m:k-1}) && \text{(plan process model)} \\ & P_{k-1}(\mathbf{i}_{k-m:k-1} \mid u_{k-m:k-1}) && \text{(prior)} \\ & \eta, && \text{(normalization)} \end{aligned} \tag{5.1}$$

where $\mathbf{i}_k$ is the user mental plan to move from the current location to the target location in mind, u_k is the user input at time instant k, $u_{k-m:k}$ is the sequence of the user input from time instant $k-m$ to k, and m defines the past time instances that influence the plan and the user input at any time. Hereby, for instance, the user model defines the likelihood of the user input given that the user has the plan evolution $\mathbf{i}_{k-m:k}$ and issued previous commands $u_{k-m:k-1}$. The user and plan process models are defined differently for a variety of systems and inputs. For a BCI-based input and free space area [171], the user model is defined with a simple heuristic function that takes into consideration the relative orientations of the robot to the different goals and the distances to each of them. The plan process model is computed such that when the robot moves from the pose $\mathbf{x}_{k-1}$ to $\mathbf{x}_k$, the straight path between $\mathbf{x}_{k-1}$ and the j^{th} goal is transformed into the straight path between $\mathbf{x}_k$ and the j^{th} goal, and the j^{th} probability is transferred to the new path. More detailed user and plan process models are proposed in [172] for joysticks and deterministic discrete interfaces for different kinds of environments.

The different approaches to IR make use of available information about the environment and the history of interaction to arrive at good estimates about the user's hidden goals or

plans. The Bayesian approach allows to model this information in a sparse representation, and yet achieves reliable estimates. We conjecture that, in navigation tasks with discrete interfaces, the information that can be obtained with goal recognition (rather than the full plan recognition [171, 172]) is sufficient to guide interface adaptations. Moreover, the plans which were computed in [172] were obtained by path planners, and in [171] they were assumed to be straight lines from the robot to the different goals. Our approach to the goal recognition problem uses the previously computed (by a global planner) path plans to all goals as extra evidence regarding the probability of the target goals. In comparison to the approach in [170], we aim at intention recognition systems that require no training whatsoever, so that users can benefit from the recognition system in their first use of the system. On the other hand, straightforward methods like [158] may not be able to exploit all useful available information.

5.3 Intention Recognition in Obstacle Avoidance Mode

5.3.1 Problem Statement

The objective of the intention recognition module is to infer users' target goals throughout interaction, where evidence is gathered from observed users' commands and their directions of gaze. It assumed here that the users are unaware of the existence of the recognizer, i.e. keyhole intention recognition.

We denote the set of enumerable goals in the environment with $\mathcal{G} = \{{}^1\mathbf{g}^W, {}^2\mathbf{g}^W, \cdots, {}^n\mathbf{g}^W\}$, where each goal is defined by its pose with respect to the global coordinate system of the available map (W), such that ${}^m\mathbf{g}^W = [{}^mx^W, {}^my^W, {}^mz^W, {}^m\theta^W]^T$, where $m \in \{1, 2, \cdots, n\}$ and ${}^m\theta^W \in [-\pi, \pi]$ is the smallest angle between the x-coordinate of the m^{th} goal frame and that of the global frame (W), where the z-coordinate of all goals is parallel to that of the world frame. We assume that the number (n) and the poses of these goals, which might represent the location of salient objects the user frequently use, are accessible by the intention recognition system. For the sake of simplicity in the notation, the postscript W is dropped when we refer to goals and their poses in the global coordinate system. The unknown target goal, which the user has in mind, is defined here as a discrete random variable G with a sample space $\mathcal{G}$.

While having a goal location ${}^m\mathbf{g}$ in mind at time instant k, the user updates the mental path plan ${}^m\mathbf{i}_k$ of how to arrive there from the robot's current pose $\mathbf{x}_k = [x_k, y_k, \theta_k]^T$. We write this in the form, ${}^m\mathbf{i}_k = \mathbf{x}_k \rightarrow {}^m\mathbf{g}$. Additionally, we define the length of a path plan (in meters) with $l({}^m\mathbf{i}_k) = {}^ml_k$. In order to follow the plan in mind, the user issues a command $u_k \in \mathcal{U}$. In this work, we only consider incremental commands that can be issued with discrete interfaces (e.g. keyboard or SSVEP-BCI), and therefore, the set of all possible commands is enumerable, i.e. $\mathcal{U} = \{$move forward, move backward, move left, move right, turn left, turn right, stop$\}$. User commands are translated into robot actions a_k that change the state of the robot from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$. The change in the robot pose that is triggered by translational and rotational commands is set respectively to the default values δd_{def} (m/command) and $\delta\theta_{\text{def}}$ (rad/command).

In order to make the math easier to follow, we introduce the dummy user command u_0,

which implies $\mathbf{x}_0 = \mathbf{x}_1$ and ${}^m\mathbf{i}_0 = {}^m\mathbf{i}_1$ for all m. The user's gaze direction at time instant k is denoted by $\mathbf{h}_k$, which we assume here to be defined with the rotational angles of the gaze direction relative to an arbitrary reference frame.

We assume that the localization module provides a reliable estimate of $\mathbf{x}_k$, and that the mental path plans to all goals (i.e. $\mathbf{x}_k \rightarrow {}^m\mathbf{g}$) of the user can be estimated reliably with a global path planner on the basis of the 2D cost map of the environment, denoted by $\mathbf{M}_k$. Implicitly, with this assumption we hypothesize that users in the navigation domain act approximately optimally (rationality assumption [161]) and try to follow the path that minimizes some cost function. The global path planner is expected to find paths similar to the ones the user plans for the different goals.

Therefore, we can assume that at time instant k, the IR module has access to the following information

- The observed sequence of user commands up to time instant k, which is denoted by $u_{0:k} = (u_0, \cdots, u_k)$.
- The observed sequence of user gaze up to time k, which is denoted by $\mathbf{h}_{0:k} = (\mathbf{h}_0, \cdots, \mathbf{h}_k)$.
- The sequence of plans computed to all goals up to time instant k, denoted by ${}^{1:n}\mathbf{i}_{0:k} = ({}^1\mathbf{i}_{0:k}, \ldots {}^n\mathbf{i}_{0:k})$, where ${}^m\mathbf{i}_{0:k} = (\mathbf{x}_0 \rightarrow {}^m\mathbf{g}, \cdots, \mathbf{x}_k \rightarrow {}^m\mathbf{g})$.

The problem of goal recognition can be then formally defined as estimating the probability

$$ {}^mP_k = P_k(G = {}^m\mathbf{g} \mid u_{0:k}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}), \forall m, \tag{5.2} $$

where $\sum_{m=1}^{n} {}^mP_k = 1$ for all k.

The probabilities of all goals can be concatenated in the n-dimensional belief vector $\mathbf{P}_k$ which encodes the system belief about the user's hidden goal at time instant k, where $\mathbf{P}_k = [{}^1P_k, {}^2P_k, \cdots, {}^nP_k]^T$.

We assume that before the user starts navigating in the remote environment, the IR module has no prior knowledge about the next pursued goal, and therefore ${}^mP_0 = 1/n, \forall m$. Additionally, the subscript k is reset to 0 each time the robot arrives at one of the available goals, which automatically resets the belief vector to a near-uniform distribution, wherein the last visited goal is assigned a smaller probability relative to other available goals, i.e. ${}^mP_0 = 0.1/n$ if goal m was the last visited goal and ${}^iP_0 = 1/n, \forall i \neq m$. The belief vector $\mathbf{P}_0$ is then normalized to obey probability axioms.

5.3.2 Bayesian Framework to IR in Navigation Tasks

The objective of this section is to devise a recursive Bayesian update rule for the belief about user goals based on available information to the IR module. We begin by assuming that the recognizer has access to the true user input u_k, e.g. keyboard-based interface, but later, the update rule is extended to noisy interfaces, e.g. BCIs. At a later point in this section, it is discussed how to deal with the fact that the user commands and information about the user's gaze typically arrive asynchronously.

The conditional probability of a goal m given previous observations can be computed recursively according to

$$\begin{aligned} P_k({}^m\mathbf{g} \mid u_{0:k}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}) = & && \text{(posterior)} \\ & P(u_k \mid {}^m\mathbf{g}, u_{0:k-1}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}) && \text{(user input model)} \\ & P(\mathbf{h}_k \mid {}^m\mathbf{g}, u_{0:k-1}, \mathbf{h}_{0:k-1}, {}^{1:n}\mathbf{i}_{0:k}) && \text{(user gaze model)} \\ & P({}^{1:n}\mathbf{i}_k \mid {}^m\mathbf{g}, u_{0:k-1}, \mathbf{h}_{0:k-1}, {}^{1:n}\mathbf{i}_{0:k-1}) && \text{(plans evolution model)} \\ & P_{k-1}({}^m\mathbf{g} \mid u_{0:k-1}, \mathbf{h}_{0:k-1}, {}^{1:n}\mathbf{i}_{0:k-1}) && \text{(prior)} \\ & \eta, && \text{(normalization)} \end{aligned} \tag{5.3}$$

where η is a normalization factor which guarantees $\sum_{m=1}^{n} P_k({}^m\mathbf{g} \mid u_{0:k}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}) = 1$ for all k. The major difference between the two rules in (5.3) and (5.1), is that ${}^{1:n}\mathbf{i}_{0:k}$ is assumed known for all possible goals in 5.3, whereas in (5.1) its computation is the target of the update rule.

In the following, generic and user-agnostic models for the user input, the user gaze and the path plans evolution will be proposed.

User Input Model

The term $P(u_k \mid {}^m\mathbf{g}, u_{0:k-1}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k})$ models the likelihood a user issues a command u_k at time k, while having the goal ${}^m\mathbf{g}$ in mind and given the sequence of current and previous path plans to all goals ${}^{1:n}\mathbf{i}_{0:k}$, the sequence of previously issued commands $u_{0:k-1}$ and the sequence of user gaze $\mathbf{h}_{0:k}$ up to time instant k. We assume that the user input u_k is conditionally independent of $u_{0:k-1}, \mathbf{h}_{0:k}$ and ${}^{m'}\mathbf{i}_{0:k}$, for all $m' \neq m$, given the current target goal and the mental path plan at time instant k. This reduces the user input model to $P(u_k \mid {}^m\mathbf{i}_k)$. Intuitively, this means that the user command at time instant k is only influenced by the mental path plan that brings the robot to the target goal location at that time instant. We assume additionally that the issued command is mainly influenced by the local surroundings of the robot, and in particular by a sub-goal or a viapoint on the path ${}^m\mathbf{i}_k$, referred to as ${}^m\mathbf{g}_k$. Subgoals are determined on each path to each defined goal, i.e. $\mathbf{x}_k \rightarrow {}^m\mathbf{g}_k \rightarrow {}^m\mathbf{g}$ as the furthest point, to which a straight line can be drawn from the current robot position without touching any obstacle. This definition is similar to the one proposed in [172, 173] for discrete interfaces. Sub-goals are searched for within a predefined circle around the robot with a radius d_{subgoal}. The process of finding the subgoals is depicted in Fig. 5.3. These assumptions lead to the new approximation for the user model as $P(u_k \mid {}^m\mathbf{i}_k) \approx P(u_k \mid \mathbf{x}_k, {}^m\mathbf{g}_k)$.

Once the user command u_k arrives, the relative orientations of the robot with respect to the computed sub-goals are used to compute the approximated likelihood function $P(u_k \mid \mathbf{x}_k, {}^m\mathbf{g}_k)$ as shown in Fig. 5.4. For instance, when a turn left command is issued by the user, all computed sub-goals which lie in the left semi-circle with respect to the robot's heading are assigned a higher probability than those which lie in the right semi-circle. The user model ignores the z-component (the altitude) of the available goals.

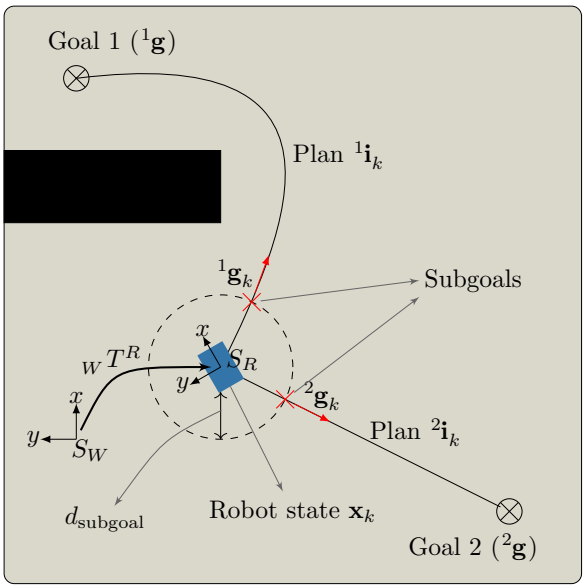


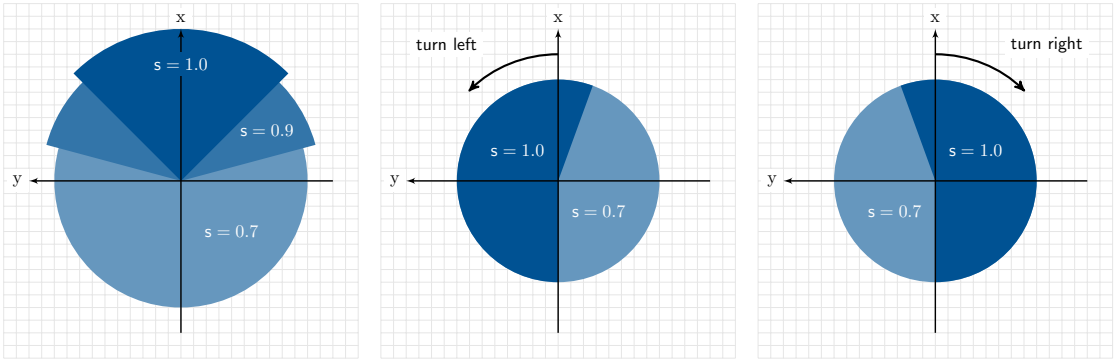
Figure 5.3: An example of estimating subgoals for goals 1 and 2. Subgoals in the figure are indicated by the red crosses and the black rectangle corresponds to an obstacle.

User Gaze Model

The term $P(\mathbf{h}_k \mid {}^m\mathbf{g}, u_{0:k-1}, \mathbf{h}_{0:k-1}, {}^{1:n}\mathbf{i}_{0:k})$ denotes the likelihood of a user's gaze $\mathbf{h}_k$ given the sequence of user commands $u_{0:k-1}$, the previous gaze sequence $\mathbf{h}_{0:k-1}$ and the sequence of all computed plans ${}^{1:n}\mathbf{i}_{0:k}$, where the user has the goal ${}^m\mathbf{g}$ as a target goal. We assume that the user always tries, if feasible, to bring the final goal ${}^m\mathbf{g}$ into sight from the current robot location $\mathbf{x}_k$. This assumption is supported by the work in [174], where it has been concluded that, in natural settings, look-ahead fixations represent a task-dependent strategy. In our application, look-ahead user fixations are assumed necessary to update the navigation plan every time the robot moves or objects in the environment move. This obviously ignores the possibility that the user's focus of attention can be shifted towards sub-goals, which he/she is trying to reach on the path towards the final goal. Nonetheless, such sub-goals can be accounted for by the user input model. The user gaze model can then be approximated as $P(\mathbf{h}_k \mid {}^m\mathbf{g}, \mathbf{x}_k)$.

In order to compute $P(\mathbf{h}_k \mid {}^m\mathbf{g}, \mathbf{x}_k)$, we note that the field of view of human vision is typically divided into an inner (i.e. foveal vision) and an outer (i.e. peripheral vision) part. The foveal vision corresponds to the sight area which maps on the central part of the retina with the highest receptor density and highest visual resolution. Thus, for sharp vision, one has to align the eyes/head to look directly at the point of interest. The peripheral vision, on the other hand corresponds to the remaining area of the visible field of view with lower resolution.

Inspired by these characteristics of the human vision system, we define two regions for the focus of attention. An inner region that is defined by the two angles ϵ_0 and δ_0, which respectively determine the horizontal and vertical openings around the gaze direction, as depicted in Fig. 5.5. Similarly, the outer region can be determined by the angles ϵ_1 and δ_1. With these assumptions, target goals are classified into three categories at any time instant



(a) Forward (b) Turn left (c) Turn right

Figure 5.4: The user input model shown for the translational and rotational commands. Subgoals are assigned a score depending on their location relative to the robot heading (corresponds to the x-axis in the plots). The lighter the area, the lower the score, e.g. subgoals, which are located in the negative x direction and have an angle greater than 180° (behind the robot) get the lowest scores assigned if a forward command were issued. The shown scores are exemplary.

k, according to their position with respect to the two regions of attentional focus, i.e. in the inner or outer region or out of sight. By conforming transformation, the goal locations (ignoring their relative orientation, i.e. ${}^{m}\theta$) can be defined with respect to the moving gaze frame as

$$ {}^{m}\mathbf{g}^{H} = {}_{W}T^{H} \cdot {}^{m}\mathbf{g}^{W}, \tag{5.4} $$

where ${}_{W}T^{H}$ denotes the transformation matrix from the fixed world frame (W) to the moving gaze frame (H).

Consequently, three conditions need to be fulfilled for a goal m to be assigned to the inner region. These are

$$ {}^{m}x^{H} > 0 \tag{5.5} $$

$$ \operatorname{atan}({}^{m}y^{H}, {}^{m}x^{H}) < \epsilon_0 \tag{5.6} $$

$$ \operatorname{atan}({}^{m}z^{H}, {}^{m}x^{H}) < \delta_0 \tag{5.7} $$

where ${}^{m}x^{h}$, ${}^{m}y^{h}$ and ${}^{m}z^{h}$ are the x, y and z components of the goal location in the moving gaze frame (H). Similar conditions can be derived for the outer region using the limiting angles ϵ_1 and δ_1.

Based on our assumption that users occasionally bring their goals into their field of view, we consider the gaze mode as

$$ P(\mathbf{h}_k \mid \mathbf{x}_k, {}^{m}\mathbf{g}) \propto \begin{cases} a & \text{if } {}^{m}\mathbf{g} \text{ lies within inner fov} \\ b & \text{if } {}^{m}\mathbf{g} \text{ lies within outer fov} \\ c & \text{otherwise,} \end{cases} \tag{5.8} $$

where $a > b > c$ are constants to be chosen at design time.

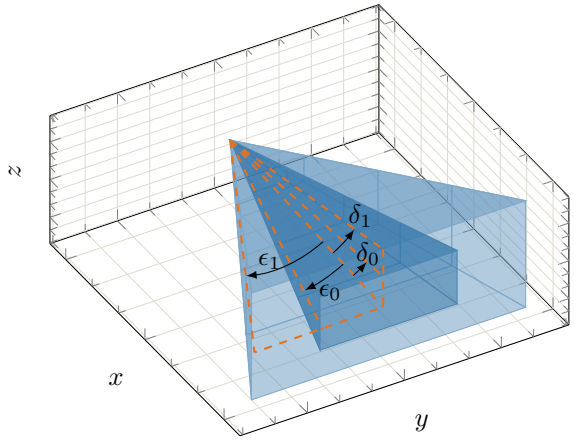


Figure 5.5: The inner and outer spatial regions of attentional focus are shown with respect to the coordinates system of the moving gaze frame.

Plans Evolution Model

The path plans evolution model denoted by $P({}^{1:n}\mathbf{i}_k \mid {}^{m}\mathbf{g}, u_{0:k-1}, \mathbf{h}_{0:k-1}, {}^{1:n}\mathbf{i}_{0:k-1})$ defines the likelihood that the robot currently has the path plans ${}^{1:n}\mathbf{i}_k$ given that the user has ${}^{m}\mathbf{g}$ in mind, and the sequences of issued commands and gaze points and all path plans previously computed for all goals up to time instant $k-1$. Assuming that the plans ${}^{1:n}\mathbf{i}_k$ are conditionally independent of the previously issued commands and gaze points given previous path plans and the target goal, the path evolution model reduces to $P({}^{1:n}\mathbf{i}_k \mid {}^{m}\mathbf{g}, {}^{1:n}\mathbf{i}_{0:k-1})$. Despite this simplification, the computation of this model remains a bit tricky. In the following, a series of approximations allows to arrive at a reasonably simple model. Hereby, we consider first the path plan length, i.e. $({}^{m}l_k)$, as an approximate sufficient statistic of the plan ${}^{m}\mathbf{i}_k$. This yields $P({}^{1:n}l_k \mid {}^{m}\mathbf{g}, {}^{1:n}l_{0:k-1})$ as an approximation of the plan evolution model, where ${}^{1:n}l_k$ denotes the path plan lengths to all goals at time instant k. With this approximate statistic, only the lengths of the plans, rather than the complete path plans, need to be stored in memory in order to compute the score of the plans evolution model. With further simplification, the memory requirements are reduced to a single value. Fig. 5.6 shows a simple example with three different goals, and the evolution of their path plans. At each time instant k, the current and previous path plans are used to reason about the possible target goals.

As a heuristic estimate of the plan evolution model, we use the relative changes in path lengths returned by the path planner at each update to capture special trends towards (or away from) one (or more) end goals. Formally, we define $\Delta^{m}l_k = \frac{{}^{m}l_k - {}^{m}l_{k-1}}{{}^{m}l_k + C}$ as the relative difference in length at time k for goal m, where division by zero is mitigated by the constant $C \in \mathbb{R}^+$. Dividing by the current path length assures that closer goals are favored over farther ones. To account for path length differences further back in time, a weighted moving average over $\Delta^{m}l_k$ denoted by $\Delta^{m}\tilde{l}_k$ is adopted and computed with $\Delta^{m}\tilde{l}_k = \alpha\Delta^{m}l_k + (1-\alpha)\Delta^{m}\tilde{l}_{k-1}$, where $0 < \alpha < 1$ is a forgetting factor, and $\Delta^{m}l_1 = \Delta^{m}\tilde{l}_1 = 0$ for all m since $\mathbf{x}_0 = \mathbf{x}_1$.

The values of $\Delta^{m}\tilde{l}_k$ summarize approach (and depart) trends with respect to the different target goals in the environment, and thereby the plan evolution score can be approximated

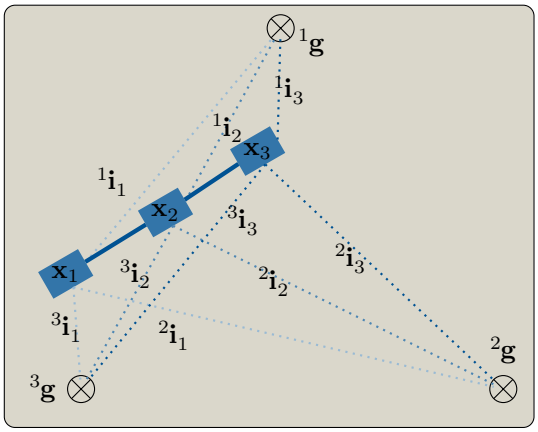


Figure 5.6: A simple example showing the evolution of trajectories towards three target goals. The plan evolution model assigns more scores to the goals whose record so far shows a trend of approach (e.g. $^1\mathbf{g}$ in the figure since $^1l_3 < {}^1l_2 < {}^1l_1$).

with

$$P(^{1:n}l_k \mid {}^m\mathbf{g}, {}^{1:n}l_{0:k-1}) \propto f(\Delta^m\tilde{l}_k), \tag{5.9}$$

where only the sequence of path plans towards goal m are used to compute $f(\Delta^m\tilde{l}_k)$. This is rather a simplification, but given that $P(^{1:n}l_k \mid {}^m\mathbf{g}, {}^{1:n}l_{0:k-1})$ needs to be computed for all goal in the environment, all available information from the computed path plans will be made use of.

$$f(\Delta^m\tilde{l}_k) = \beta_1 \cdot \left(1 - \frac{1}{\exp\left(-\beta_2 \cdot \Delta^m\tilde{l}_k\right) + 1}\right) + \beta_3, \tag{5.10}$$

where the parameters β_1, β_2 and β_3 to be chosen at design time.

Intuitively, the score function in (5.10) favors goals whose new plans are shorter than the previously computed and among those, favors the closer end goals to the farther ones.

IR in Noisy Interfaces

For noisy interfaces, the recognizer does not have access to the true user command u_k, but rather to a noisy version $\hat{u}_k$ thereof. Using the law of total probability, the update rule can be modified, similar to [171], to account for the partially observable user input in the

following manner:

$$
\begin{aligned}
P_k({}^m\mathbf{g} \mid \hat{u}_{0:k}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}) &= \sum_{u_{1:k}\in\mathcal{U}^k} P_k({}^m\mathbf{g} \mid u_{0:k}, \hat{u}_{0:k}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}) \cdot P(u_{0:k} \mid \hat{u}_{0:k}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}) \\
&\overset{(1)}{=} \sum_{u_{1:k}\in\mathcal{U}^k} P_k({}^m\mathbf{g} \mid u_{0:k}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}) \cdot P(u_{0:k} \mid \hat{u}_{0:k}) \\
&\overset{(2)}{=} \sum_{u_{1:k}\in\mathcal{U}^k} P_k({}^m\mathbf{g} \mid u_{0:k}, \mathbf{h}_{0:k}, {}^{1:n}\mathbf{i}_{0:k}) \cdot \prod_{i=1}^{i=k} P(u_i \mid \hat{u}_i) \qquad (5.11)
\end{aligned}
$$

The simplification of (1) is based on the assumption that $u_{0:k}$ is conditionally independent of $\mathbf{h}_{0:k}$ and ${}^{1:n}\mathbf{i}_{0:k}$ given $\hat{u}_{0:k}$. On the other hand, (2) is based on the assumption that consequent user commands are independent of each other and only depend on the current noisy measure thereof.

The formulation in (5.11) means that the recognizer should keep a record of all possible hypotheses about the user input which can be traced back to the first observation received, i.e. $\hat{u}_1$. The number of these hypotheses, however, grows exponentially with k, i.e. $|\mathcal{U}|^k$. Therefore, in this work, we limit tracing these hypotheses to the last issued command only, which implies that the uncertainty about each command is only accounted for once. In case of SSVEP-based interaction, the probabilities $P(u_i \mid \hat{u}_i)$ correspond to the individual elements in the interface confusion matrix, similar to the ones shown in Fig. 3.15. Hereby, (5.11) also includes the possibility that the BCI issues a command while the user wants to be in the idle state (i.e. NOOP command). However, this is taken into consideration in updating the posterior of the belief vector. The wrongly classified commands propagate to the execution phase.

Asynchronous Posterior Updates

In absence of any additional information, e.g. user navigation preferences, the system starts with $\mathbf{P}_0 = [1/n, \cdots 1/n]$. The update rule in (5.3) can be triggered either by the arrival of a user command or a new gaze point at the time instant k. Obviously, both triggers arrive in an asynchronous manner, and the probability that both will arrive exactly at the same time can be neglected. It is possible here for example to wait for the availability of the two signals to trigger the update synchronously. However, in this case many useful information, from which the recognizer can benefit will be lost. The other possibility is to trigger the update with every new information observed about the user. Consequently, the user input model contribution to the posterior in (5.3) will be ignored in case of gaze-based trigger, and the user gaze model contribution will be ignored when a new user command arrives.

In case of updates triggered by new user commands, a multiplication of the user input probability, the plan evolution probability and the prior followed by a normalization step yields the posterior. Since the path evolution model is independent of the user command which triggers the update, we compute its contribution every time the robot state $\mathbf{x}$ changes in order to react as quickly as possible to the arrival of user commands. In the case of SSVEP-based interaction, uncertainty in user commands is accounted for with the formula in (5.11). Gaze-triggered updates are computed in a similar manner.

Belief Confidence

The belief vector is of great importance on its own, but it is also necessary sometimes to have a metric that summarizes the confidence of the IR in such beliefs. To this end, we chose a novel metric (s) that reflects the non-uniformness of the belief vector. The new metric can be computed for any probability mass function (like the belief vector) characterizing a random variable X with n possible values, with the following steps:

1. The probability masses are ordered in ascending order, and the cumulative distribution function (CDF) is computed.

2. The zeroth moment M_0 and the first moment M_1 of the CDF are computed with $M_0 = \sum_m cdf[m]$ and $M_1 = \sum_m m \cdot cdf[m]$.

3. The x-coordinate of the centroid of the CDF is evaluated with $c = M_1/M_0$.

4. The non-uniformness metric is computed by normalizing c in the range $[0, 1]$ with $s = (c - c_{\min})/(c_{\max} - c_{\min})$, where $c_{\min} = (2 \cdot n + 1)/3$ and $c_{\max} = n$.

The metric s is a function of the belief vector, i.e. $s(\mathbf{P}_k)$ that ranges from 0 (for fully uniform beliefs) to 1 (for unit mass beliefs). The values of s can be considered as to reflect the recognizer's confidence about its belief, since the concentration of probability at one mass point typically reflects the accumulation of enough evidence about a specific goal location. One desired feature of the new metric is its linearity with respect to the number of goals that share the total probability. This is shown in Fig. 5.7 comparing the non-uniformness metric of $\mathcal{U}_i$ to Kullback-Leibler divergence $KL(\mathcal{U}_i \| \mathcal{U}_n)$, where $\mathcal{U}_n$ is the uniform assignment over the n possible values of X, i.e. $\mathcal{U}_n = [1/n, \cdots, 1/n]^T$ and $\mathcal{U}_i$ is the uniform assignment of total probability over a subset $i \leq n$ of the sample space, e.g. $\mathcal{U}_i = [1/2, 1/2, 0, \cdots 0]^T$ for $i = 2$. In Fig. 5.7, the KL-divergence measure is normalized with respect to the value $KL(\mathcal{U}_1 \| \mathcal{U}_n)$.

5.3.3 Performance Metrics for Intention Recognition

IR methods can be evaluated directly by calculating prediction accuracies from labeled and successful sequences (i.e. sequences that ended at the target goal). The predictions of the IR (i.e. the belief vectors) remain useless unless they are used in one way or another to improve interaction. Therefore, we adopt a shared control application of the belief vector, whereby when the IR module is confident about its belief, online modulation of the translational and rotational steps is performed so that the robot gets closer to high probable goals. A reliable IR system is expected to reduce the number of user commands required to accomplish navigation tasks. This way, shared control can provide indirect measures of the reliability of the IR method. It is worth repeating here, that SC should be performed in an unobtrusive way, so that users can benefit from its deployment, or otherwise, no reduction in the number of commands is expected. The following subsections formally defines the direct and indirect measures used to evaluate the IR method.

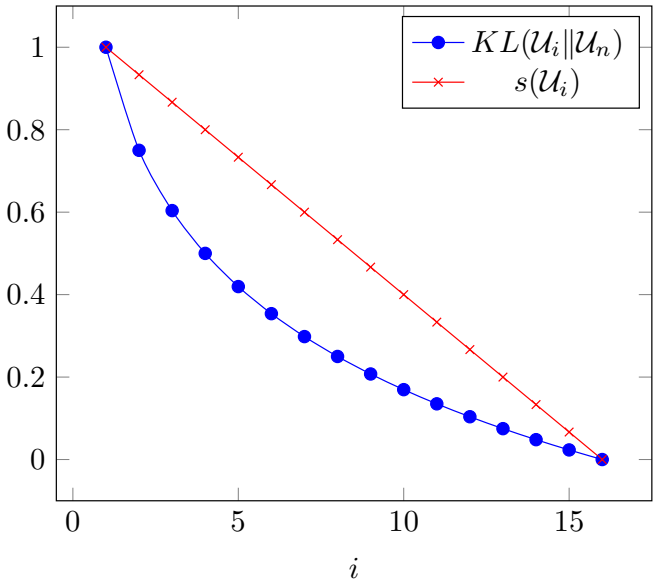


Figure 5.7: The KL-divergence measure compared to the non-uniformness metric.

Direct Measures

Recall and precision metrics are adopted in [175–177] to provide detailed evaluation of the performance of different IR methods. Recall is defined as the fraction of belief updates which has the true user's end goal in the set of N-best predictions. Precision is defined as the ratio of belief updates, in which the recognizer is confident and the true goal is among the N-best predictions, to the total number of belief updates, in which the recognizer is confident. The recognizer's confidence is defined here, as in Sec. 5.3.2, by the non-uniformness metric of its belief vectors , i.e. when $s(\mathbf{P}_k) > s_{\text{thresh}}$. Formally, let ${}^m\mathbf{P}_{1:k} = \{{}^m\mathbf{P}_1, {}^m\mathbf{P}_2, \dots {}^m\mathbf{P}_k\}$ be the sequence of belief updates during which the true user's hidden goal is ${}^m\mathbf{g}$, recall and precision can be computed as follows

$$\begin{aligned} \text{Recall} &= \frac{\sum_{i=1}^{k} I_N({}^m\mathbf{P}_i)}{k}, \\ \text{Precision}(s_t) &= \frac{\sum_{i=1}^{k} I_N({}^m\mathbf{P}_i) \cdot I_S({}^m\mathbf{P}_i)}{\sum_{i=1}^{k} I_S({}^m\mathbf{P}_i)}, \end{aligned} \tag{5.12}$$

where the indicator function $I_N({}^m\mathbf{P}_i) = 1$ if the goal m is among the N-best predictions and 0 otherwise and the indicator function $I_S({}^m\mathbf{P}_i) = 1$ if the non-uniformness metric of the belief vector $s({}^m\mathbf{P}_i) > s_t$ and 0 otherwise.

Indirect Measures with Shared Control

A simple shared control (SC) application of the belief vector is designed to help evaluate the performance of the intention recognition (IR) module. Hereby, we consider modulating the magnitude of the translational and rotational steps that define the robot movements in

the directions signaled by the user. The modulation is obtained with

$$x = \begin{cases} x_{\text{default}} & \text{if } s < s_{\text{thresh}} \\ s \cdot x_{\text{opt}} + (1-s) \cdot x_{\text{default}} & \text{if } s \geq s_{\text{thresh}}, \end{cases} \tag{5.13}$$

where x_{default} is the default translational (δd_{def}) or rotational steps ($\delta\theta_{\text{def}}$), $s \in [0,1]$ is the modulation factor and x_{opt} is chosen in a way so that the robot gets more attracted to the goals with the highest belief scores. Furthermore, we define s_{thresh} as a threshold, below which online parameters modulation is disabled. The modulation factor is chosen here to be the confidence of the IR module in its belief, i.e. $s = s(\mathbf{P}_k)$. The optimal translational and rotational magnitudes can be computed by considering the distance and relative orientation of the robot to all available subgoals (defined as described in Sec. 5.3.2), and the belief which the recognizer has about the corresponding goals.

(1) Optimal Translational Step

In order to compute the optimal translational step when the recognizer is quite confident about its belief, we model the different goals as different point attractors on a 2D plane. Hereby, the different goals (represented by their corresponding subgoals from the user input model in Sec. 5.3.2) exert different forces on the robot's translational movement in proportion to their squared probabilities, i.e. ${}^{m}P_k{}^2$.

We define δd_{opt} as the minimizer of the energy function

$$f(\delta d) = \frac{1}{2} \sum_m {}^{m}P_k{}^2 \cdot \|T(\mathbf{x}_k, \delta d) - {}^{m}\mathbf{g}_k\|_2^2, \tag{5.14}$$

where $\delta d \in [0, \delta d_{\max}]$ and $T(\mathbf{x}_k, \delta d)$ defines the new position of the robot after translating δd meters in the direction of travel. The term $\|T(\mathbf{x}_k, \delta d) - {}^{m}\mathbf{g}_k\|^2$ can be computed with

$$({}^{m}x_k - x_k - \delta d \cos(\theta_k))^2 + ({}^{m}y_k - y_k - \delta d \sin(\theta_k))^2, \tag{5.15}$$

where $[{}^{m}x_k, {}^{m}y_k]^T$ denotes the position of the subgoal m, and $[x_k, y_k]^T$ and θ_k, respectively denote the robot's position and heading at time instant k. Intuitively, minimizing the energy function in (5.14) minimizes the squared distance to all goals jointly, where the importance of each goal is weighted by its squared probability. For instance, when all the probability is concentrated on one goal (e.g. goal 2 in Fig. 5.8), δd_{opt} will be chosen so that it brings the robot very close to its corresponding subgoal. However, it is rare that IR converges to a point mass probability vector $\mathbf{P}_k$, and therefore the distance to all goals are taken into consideration in the minimization of (5.14). Noteworthy here is that the effectiveness of δd_{opt} is highly dependent on the accuracy of the $\mathbf{P}_k$.

Since the energy function is defined such that it has a single minimum, i.e.

$$\frac{\partial^2}{\partial x^2} f(\delta d) = \sum_{m=1}^{N} {}^{m}P_k{}^2 > 0 \qquad \forall \delta d. \tag{5.16}$$

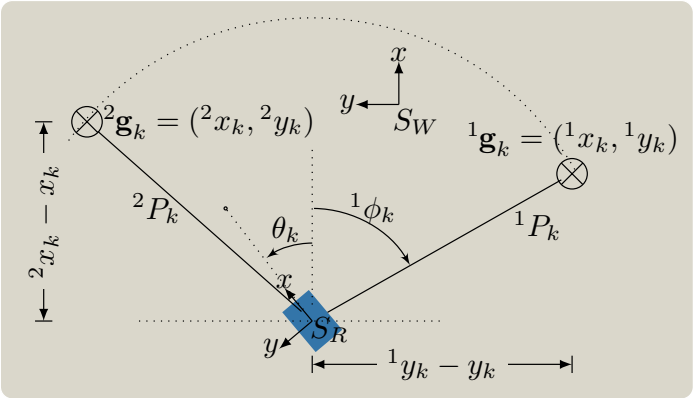


Figure 5.8: An example illustrating the quantities used in the computation of the optimal translational and rotational steps. Computed subgoals on the path to the different end goals often lie on a circle of radius d_{subgoal}. The optimal steps are computed so as to minimize an energy function determined by attraction forces of the different goals. The angle ${}^1\phi_k = \text{atan2}({}^1y_k - y_k, {}^1x_k - x_k)$

Solving $\frac{\partial}{\partial x} f(\delta d) = 0$, yields the following value of δd_{opt},

$$\delta d_{\text{opt}} = \frac{\sum_{m=1}^{n} {}^mP_k{}^2 \cdot [({}^mx_k - x_k)\cos(\theta_k) + ({}^my_k - y_k)\sin(\theta_k)]}{\sum_{m=1}^{n} {}^mP_k{}^2}. \tag{5.17}$$

(2) Optimal Rotational Step

Similarly, different goals apply different forces on the robot's rotational movements in proportion to their squared probabilities. The optimal rotational step $\delta\theta_{\text{opt}} \in [-\pi, \pi]$ is defined as the global minimizer of the energy function

$$f(\delta\theta) = \frac{1}{2} \sum_{m} {}^mP_k{}^2 \cdot (\text{d}(\mathbf{x}_k, {}^m\mathbf{g}_k) - \delta\theta))^2,$$

where $\text{d}(\mathbf{x}_k, {}^m\mathbf{g}_k)$ is defined as

$$\text{d}(\mathbf{x}_k, {}^m\mathbf{g}_k) = \begin{cases} \text{atan2}({}^my_k - y_k, {}^mx_k - x_k) - \theta_k. & \text{if } \|\mathbf{x}_k - {}^m\mathbf{g}_k\| \geq d_g \\ {}^m\theta - \theta_k & \text{if } \|\mathbf{x}_k - {}^m\mathbf{g}_k\| < d_g, \end{cases} \tag{5.18}$$

where d_g is a distance threshold, below which the robot is assumed close to an end goal. Intuitively, this means that when users get very close to specific goals, they are more likely willing to align the robot with the heading of that goal, or otherwise, they rather align the robot to face the corresponding subgoals. The value of d_g is set to 0.5 m. The optimal rotational step can be computed with

$$\delta\theta_{\text{opt}} = \frac{\sum_{m} {}^mP_k{}^2 \cdot \text{d}(\mathbf{x}_k, {}^m\mathbf{g}_k)}{\sum_{m} {}^mP_k{}^2}. \tag{5.19}$$

5.4 Experimental Evaluation

In order to empirically evaluate the Bayesian framework from Sec. 5.3, experiments were conducted with healthy subjects with a real and a simulated robot. Evaluation is based on the direct and the indirect measures from Sec. 5.3.3.

5.4.1 Hypotheses

Based on the characteristics of the Bayesian inference system, we have formulated the following hypotheses with respect to the direct and indirect measures from Sec. 5.3.3. First, the integration of shared control (SC) and intention recognition (IR) into the robotic system is expected to result in a fewer number of user commands when compared to the situation when intention recognition and shared control are disabled (H_1). Second, the incorporation of gaze information into the intention recognition system should produce more accurate estimates about the hidden goals and consequently should result in a fewer number of user commands when shared control is enabled (H_2). Third, we expect the intention recognition system, though mostly designed with 2D path planning for 2D navigation commands, to generalize to flat floor 3D environments, but perhaps with reduced performance (H_3). Fourth, BCIs are expected to require higher number of user commands when compared to deterministic interfaces like keyboards (H_4).

5.4.2 Conditions and Experimental Design

The experiments required that subjects drive a robot in a remote physical and a simulated environment and to visit a predefined subset of goal locations within these environments. We have varied the experimental setup within three factors:

1. The type of the robot/environment used (RBT): simulated 2D (S) or physical 3D (P).
2. The type of the interface ($INTFC$): keyboard (K) vs. SSVEP-based BCI (B).
3. The type of the IR-SC ($SCTRL$): IR is based on user commands only with no SC is applied ($L1$), SC is applied on the basis of command-triggered belief updates only ($L2$) or SC on the basis of command and gaze-triggered belief updates ($L3$).

A fully crossed design was not feasible, since the gaze (estimated with the head orientation) cannot be incorporated within the IR system for the 2D simulated robot/environment. Fig. 5.9 visualizes the resulting factorial design with the black cells referring to the infeasible conditions. The following subsections provide more details about the implementation of the different conditions.

5.4.3 Experimental Setup

Physical Robot

Fig. 5.10 shows the robot avatar used in the physical robot conditions (P). The robot has two 7 DoFs arms with a human-like reachable working space for manipulation and a

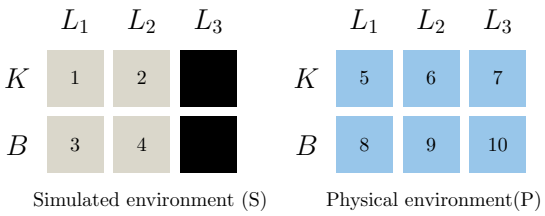


Figure 5.9: Experimental factorial design consisting of 10 different experimental conditions. Black cells correspond to infeasible conditions.

non-holonomic omnidirectional mobile base with rectangular footprint (dimensioned 68 cm × 82 cm) that consists of four wheels. For safety consideration, the mobile base is enclosed with a pressure sensitive bumper, which halts the robot immediately once it makes any contact with any rigid body. Additionally, a pan-tilt-roll unit is mounted onto the torso of the robot and serves as a 3 DoFs neck, where an emotional head is attached. Two RGB cameras (Point Grey, Richmond, Canada) serve as the eyes of the robot. Furthermore, two SICK S300 laser scanners (Waldkirch, Germany) are mounted on two opposite corners of the base to provide a 360° view and used for obstacle detection. However, since the scanners are static, they only can detect obstacles in the 2D plane parallel to the floor plane (which is assumed to be a flat surface) and having a distance $d = 10$ cm to it. A Kinect (Microsoft, USA) camera is fixed to the robot's chest, which is typically used in manipulation tasks to detect objects in front of the robot. Processing is done with two computers running Ubuntu 12.04 with real-time kernel patch. The Robot Operating System (ROS) [178] is used as the default interprocess communication infrastructure.

During the experiments, the robot received incremental commands from the user that define the direction of translation or rotation only. These commands were translated in turn into linear and angular velocity commands which the low-level controller of the mobile base can understand. The linear velocity is denoted by $\nu = [\nu_x, \nu_y]^T$ and the angular velocity by ω, which were assigned the default values: $\nu_x = \nu_y = \pm 0.25$ m/s and $\omega = \pm 0.25$ rad/s. The reason we chose such slow speeds is to keep a lower rate of optical flow in the visual feedback, which is known to be correlated with cybersickness [179]. A position controller, which continuously received the robot's location, made sure that the robot moved according to the received translation and rotations steps defined. The navigation stack from ROS was used for path planning and the `AMCL` ROS package [180] for robot's localization.

Simulated Robot

For simulated robot conditions (S), a 2D simulated version of the physical robot was constructed using stage simulator package in ROS [181]. Hereby, only the hardware components which are related to navigation were simulated, namely the robot body, the mobile base and the laser scanners.

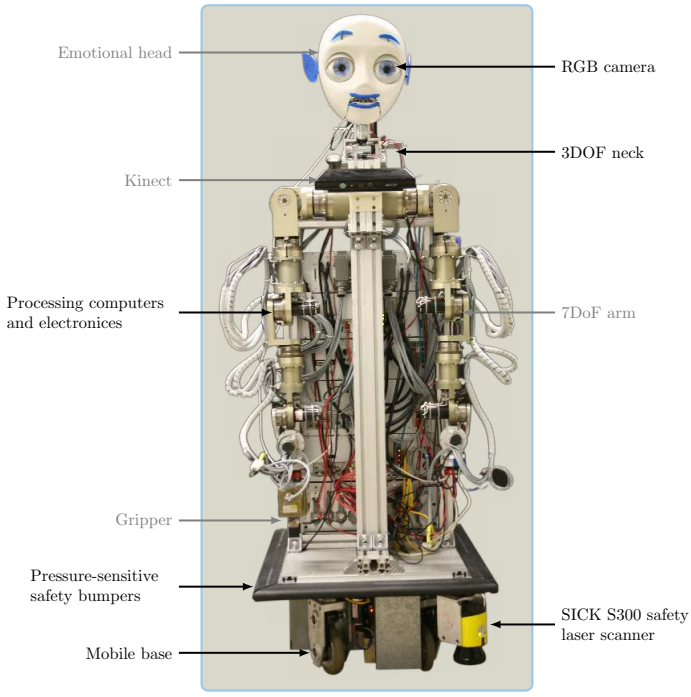


Figure 5.10: The robot avatar. The labels shown in gray indicate robot parts irrelevant to the scope of this work.

Collision Avoidance Mode

The collision avoidance behavior is realized with a velocity filter, whereby if the distance from the robot's body to the closest obstacle in the direction of travel (r) gets less than a predefined threshold r_{safe}, the incoming velocity commands undergo a reduction before they get delivered to the low-level controller (i.e. the robot slows down). The actual reduction is determined in proportion to the observed distance. If the distance to the closest obstacle in the direction of travel becomes equal to a second predefined threshold $0 < r_{\text{stop}} < r_{\text{safe}}$, the robot halts and any further velocity commands which might lead the robot closer to this obstacle will be filtered out and blocked. Consequently, this means that the robot always keeps at least r_{stop} distance to the closest obstacle. The default values are set to $r_{\text{stop}} = 0.3$ m and $r_{\text{safe}} = 0.6$ m during all experiments. For a wide range of applications, these values for r_{stop} and r_{safe} seem reasonable, but they can be tuned online as well, if it is required that the robot gets closer to a specific obstacle (or rather a possible target), e.g. to perform some manipulation tasks on objects located on top of a table.

User Interface

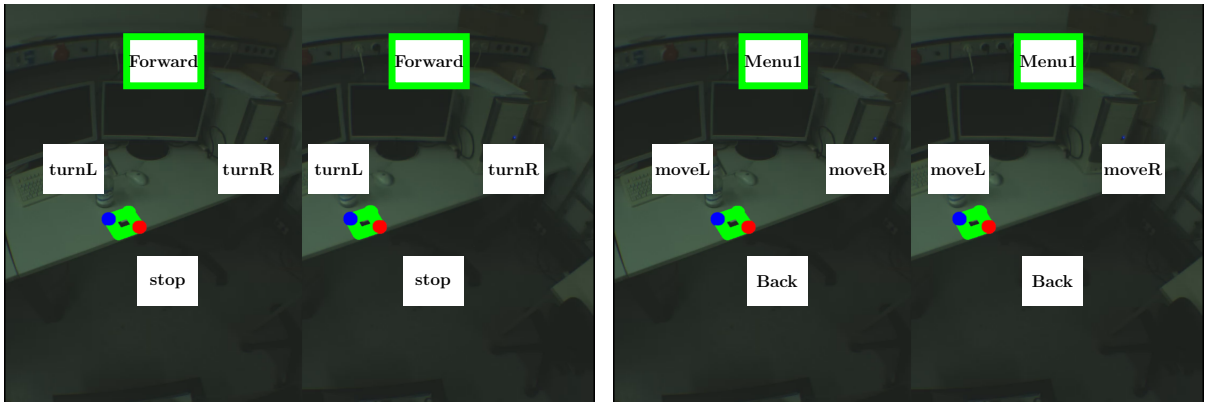
The robot avatar was embedded in a remote physical or a simulated environment and received teleoperation commands from the user who conveyed his/her commands to the system with a chosen interface, namely with an SSVEP-based BCI (B) or a keyboard (K). During physical robot operation, subjects received a continuous 3D stereoscopic

video stream from the ego-perspective of the robot avatar, which they viewed with the help of a head-mounted display (HMD) weighting 380 gram (Oculus VR, United States). Additionally, the user's head movement was continuously tracked, via the built-in head tracker available in the HMD, and transmitted to the robot side to be mapped into similar movements at the robot's neck. In the simulated environment conditions, the environment was shows on an LCD monitor. In the BCI conditions, visual stimuli were presented overlaid on the received video stream or the simulated environment, as can be seen in Figs. 5.11 and 5.12. The temporal resolution of the SSVEP-based interaction depends on the size of the EEG segment used for classifying user's commands and the level of overlap between two consecutive segments as has been pointed out in chapter 3. The segment size and the temporal resolution were respectively set by default to 2 and 0.25 s. The interaction rate of the BCI (4 Hz) was chosen higher than that of the control loop (e.g. a translational command takes on average 1 s), as it was also important to provide continuous visual feedback about the performance of the interface, so that users could better predict the responsiveness of the interface. This rate mismatch was more pronounced for the keyboard-based interaction, and therefore user commands which arrived while the robot was moving were completely ignored by the interface unless they were meant to stop the robot while it was moving. The SSVEP detection was based on the supervised CVARS algorithm.

Whereas keyboard interfaces potentially can provide a large set of commands using single buttons or combinations thereof, SSVEP-BCIs can only provide a limited set of commands. This is mainly due to the immersive nature of the application which necessitates that SSVEP stimuli to be shown overlaid on the video-stream received from the remote robot. The more stimuli are shown for display, the less will be the quality of the visual feedback. As a sensible trade-off, we fixed the number of possible commands to 4. By default, the commands move forward, turn left, turn right and stop are available to users. These commands, to which we will refer as *normal mode commands*, were chosen as they allow for human-like navigation around and in the direction of the symmetry plane. Naturally, backward and sideways movements are the exceptions, not the norm [182]. In situations where the robot receives a command which gets completely blocked as to avoid collisions, the set of user commands gets automatically replaced by a second set that includes move backward, move right, and move left. The set of the new commands is referred to as the *recovery mode commands*. Most if not all the time, the recovery commands will be sufficient to bring the robot to a free space, where the normal mode of operation can be resumed by the user through a dedicated interface element. Adaptations to the interface were accompanied by an auditory feedback that produced the speech of "normal mode commands" or "recovery mode commands", signaling the interface change to subjects.

Extraction of Gaze Direction

Typically, humans adjust their gaze by moving both their head (or more precisely the neck) and eyes in order to bring the focus of attention to the spatial regions of interest around them. Specific to immersive embodiment applications, the user's eye movement might not be spontaneous all the time, as it is the case in natural settings. In SSVEP-based interaction in particular, users overtly attend to one of the stimuli distributed at the sides



(a) Normal mode commands (b) Recovery mode commands

Figure 5.11: SSVEP stimuli overlaid on stereoscopic images coming from the remote physical environment. Recognized user commands are typically highlighted in green. The "Menu1" command in recovery mode resumes the normal mode commands. The scene shows an example target location, where a glass bottle can be observed behind the AR marker (which is highlighted in the figure with green). Part of the robot base is visible at the bottom side of the images. The visible lens barrel distortion (at the capturing side) is counteracted by the HMD which magnifies the images it receives (i.e. pincushion distortion).

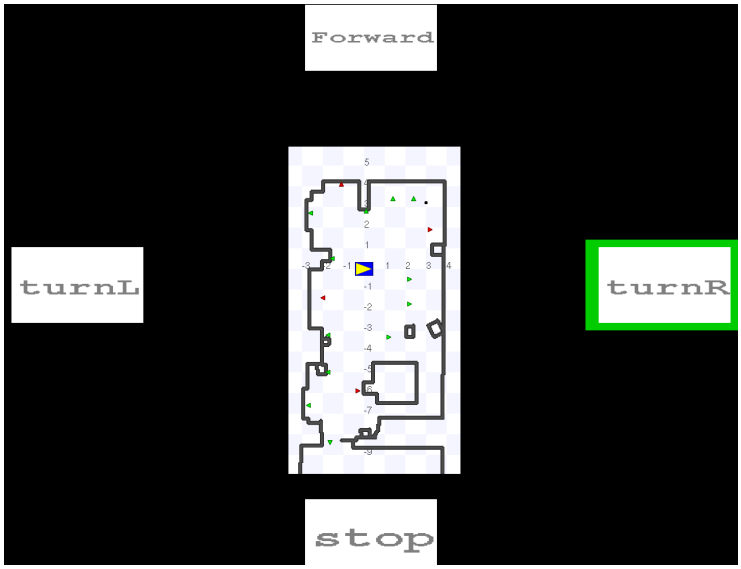


Figure 5.12: Simulated environment with SSVEP stimuli overlaid at the sides of the display. The stimuli constellation allows for intuitive interaction.

of the display when they decide to issue a control command. Therefore, the user's focus of attention in this work is estimated based on the head orientation only. This simplification is supported by the results in [183], which shows that the head orientation contributes 70% to the overall gaze direction on average and that head orientation data alone is sufficient to accurately estimate the focus of attention.

The gaze movements of a user observing a scene can be, for the sake of simplicity, broadly

separated into two classes. The first is characterized by sudden and rapid gaze movements known as *saccades*, and the second is characterized by a relative stability of the gaze for typically 200-600 ms [184] and is referred to as *fixations*. Ignoring the contribution of eye movements to gaze direction implies that fixations in this work are approximated by the steady head orientations which last for a certain amount of time, i.e. fixations of the head rather than the gaze.

At the technical level, we define the gaze frame (H) as a moving coordinate system, whose origin is positioned at the midpoint between the robot's eyes, x-axis is aligned with the gaze direction and z-axis is parallel to the robot's face plane and pointing upwards, i.e. from the neck to the forehead. In the following, we limit the tracking data to the yaw (Θ) and pitch (Φ) angles of head rotations, as the roll component, i.e. the head rotation around the axis of view, hardly affects the actual gaze direction.

In order to extract the gaze information from the raw head tracking data, we adopt a two-stage filter, to which we will refer as the *gaze filter*. In the first stage, fixation points are extracted from the head tracking data as the sample mean of consecutive yaw-pitch pairs, during which the head movements do not exceed a certain velocity threshold for a period of e.g. $t = 200$ ms. This stage is very similar to the I-VT filter algorithm used to classify eye movements [185]. In the second stage, the sample mean of the incoming classified fixations in the yaw-pitch plane is recursively computed to determine the sample gaze mean $[\bar{\Theta}, \bar{\Phi}]$. The newly incoming fixations continuously update the old sample mean as long as their Euclidean distance to the current mean is below a certain threshold denoted by $\mathrm{d_{thresh}}$. Otherwise, i.e. if the distance is larger than $\mathrm{d_{thresh}}$, the sample mean is set to the new fixation, as shown in Fig. 5.13. Every time the sample mean is set anew, it gets broadcasted as a new gaze point, i.e. $\mathbf{h}_k = [\bar{\Theta}_k, \bar{\Phi}_k]^T$. With the help of this information and the available transformation tree to the system, the transformation from the fixed world frame (W) to the moving gaze frame (H), i.e. ${}_W T^H$, can be easily computed.

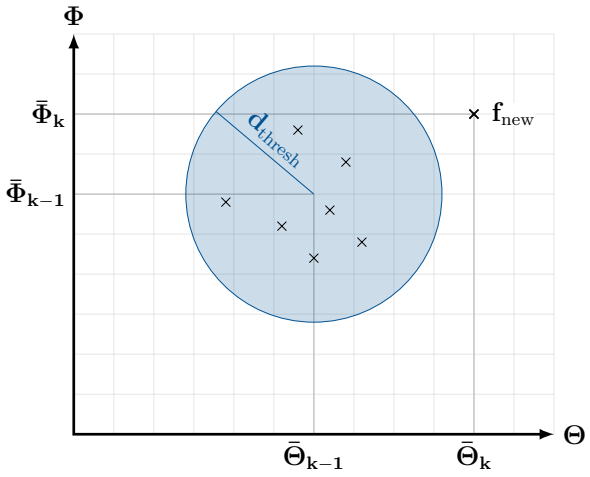


Figure 5.13: Mean gaze estimation from incoming fixation points (indicated as black crosses). When a new fixation is received, its distance to the old mean is checked against $\mathrm{d_{thresh}}$ and the new mean $[\bar{\Theta}_k, \bar{\Phi}_k]$ is computed accordingly either by updating the old mean $[\bar{\Theta}_{k-1}, \bar{\Phi}_{k-1}]$, or by setting the mean to the new value.

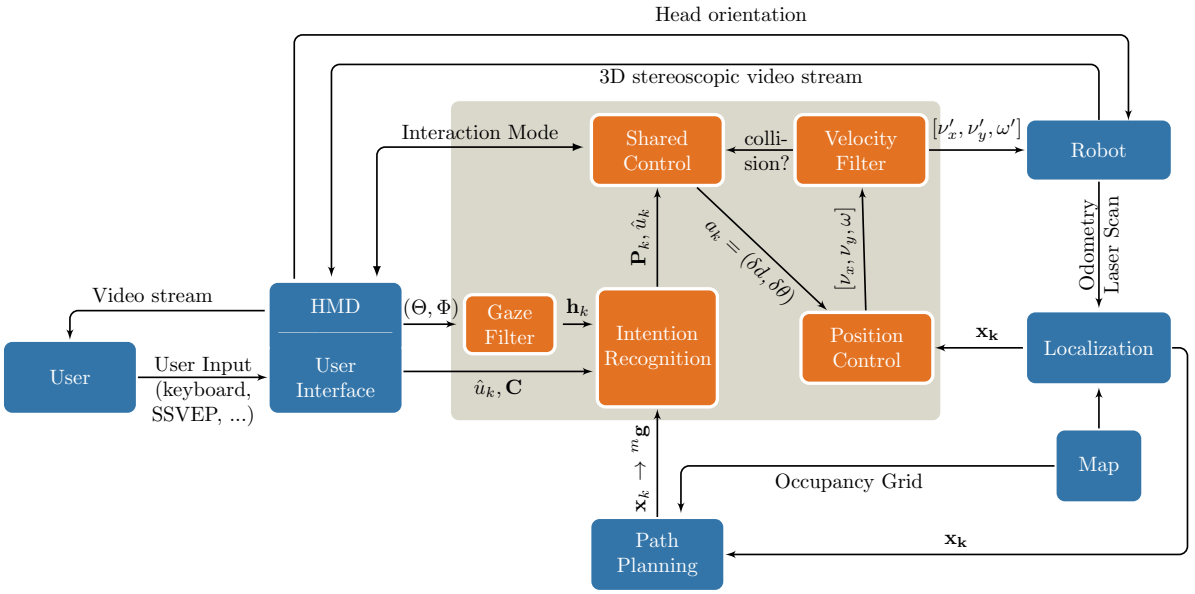


Figure 5.14: System block diagram. This work focuses on the blocks in the shaded area.

Intention Recognition and Shared Control

The IR module was active in all conditions, but its computed belief vectors were used to modulate the translational and rotational steps only in the conditions *L2* and *L3*. Additionally, the recovery mode and stop commands were excluded from the user input model as they are not typically oriented towards the final goal but rather dictated by the presence of obstacles in the robot's surroundings. In some cases, users might need to recede away from a goal in order to avoid the immediate obstacles. Additionally, it is expected that these commands are used for short periods of time only. The exact parameters which were used for IR and SC are reported in Appendix B. Fig. 5.14 shows a block diagram of the whole robotic embodiment system for navigation which highlights the interconnections and message passing between the IR and SC blocks with other components.

5.4.4 Subjects

A total of 22 healthy adults (5 females) aged 27.59 ± 5.66 (range $20 - 38$) with normal or corrected-to-normal vision served as paid volunteer subjects in this study.

5.4.5 Task and Procedure

A 2D occupancy grid of the physical 3D remote environment (a cluttered laboratory space, with many desks, tables and other robots) was built using a laser-based SLAM algorithm available in ROS, i.e. the `gmapping` package [186]. In total, 16 different goals were manually defined on the map, e.g. at the lab doors and other salient spots like desks, and were additionally marked in the physical environment with augmented reality (AR) markers. The occupancy grid of the environment and the set of all goals are shown in Fig. 5.15. The same occupancy grid was used to build a simulated environment with the stage simulator

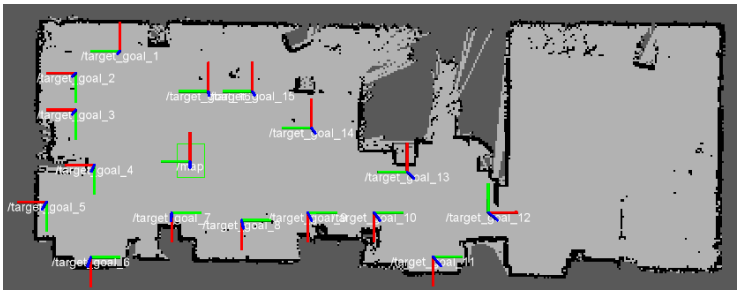


Figure 5.15: The environmental map was built using laser-based SLAM. The 2D poses of the 16 goals are shown, with respect to the world frame.

package in ROS [181].

Participants were instructed to accomplish the same navigational task in all conditions as fast as possible. The task was to drive the robot with the chosen interface and visit 4 target goals out of the available 16, where the different subjects were assigned different sets of goals. Participants were free to decide the order, in which they might visit the target goals, but they were instructed to stick to the same order in all yet to be performed sessions. Since the target goals were unfamiliar to the participants, exactly as the non-target ones, glass bottles were placed besides the AR markers of the target goals (on target desks or tables) to allow the participants to recognize them from afar, especially if neighboring goals were very close. A sample screenshot from the ego-perspective of the robot in a front of a target goal is shown in Fig. 5.11. Since doors have less ambiguity, they were only marked with the AR markers. Additionally, in the simulated environment, target goals were colored in red whereas the remaining ones were marked in green. The simulated environment for subject S1 is shown as an example in Fig. 5.12. A goal was considered reached if its distance to the robot became less than or equal to 0.8 m and its relative orientation with respect to the robot's heading was less than or equal to 0.5 rad. Additionally, the goal had to be positioned within the range$[-0.2, 0.2]$ m with respect to the y-coordinate of the robot's frame. Subjects were instructed simply to bring the robot to face each of the assigned goals. An auditory feedback signal was played back to participants when the robot arrived at any goal location. This way, we guaranteed that participants did not recede from a goal earlier (assuming they arrived) or later (assuming they did not arrive yet) than they should.

Upon their arrival at the laboratory, subjects were provided in written form, all information they needed about the course of the experiment and the different conditions they were going to perform. All participants gave their written informed consent. Participants were additionally asked to fill in a pre-questionnaire to collect some demographical data. They were additionally given a printed floor plan of the remote environment and were asked to decide upon the order in which they wanted to traverse the subset of goals consistently across conditions.

Each participant was assigned 1 session per condition and 10 sessions in total. The order of the experimental conditions was counterbalanced across subjects, where half of the participants were assigned to complete all simulated robot conditions first, and the other half were assigned the physical robot conditions first. If a subject was found to spend relatively long time trying to reach the first end goal in the BCI conditions, these sessions

were stopped since otherwise experiments might have extended over uncomfortable time lengths to participants, let alone the effect of long recordings on the quality of the EEG setup. Prior to the actual experiments, participants were familiarized with the different system components in the physical and simulated environments, e.g. the head tracking and mapping to the robot's head movement, 3D visual feedback and keyboard interface. In SSVEP-based interaction conditions, electrode placement and setup were done exactly as described in Sec. 3.4, and EEG signals were acquired with a sampling rate of 256 Hz at full DC. Two SSVEP training sessions were additionally collected during experiments with each subject. One training session preceded the simulated environment conditions with stimuli presented against a dark screen on an LCD monitor and viewed binocularly by the participants. The other session was collected immediately before the physical environment conditions. Stimuli in the latter case were presented on the HMD against a static view from the remote environment and were viewed binocularly as has been described in chapter 3. Training data were used to learn two linear discriminant analysis (LDA) classifiers for the CVARS scores as described in Sec. 3.5.4, which were later used in online SSVEP-based interaction conditions, i.e. conditions 3 and 4 for the first classifier and 8, 9 and 10 for the second one. The full training sessions were used to train the LDA classifier but a 5-fold cross validation classifier accuracies were used to estimate $P(u_i \mid \hat{u}_i)$ from Sec. 5.3.2. To elicit natural behavior, participants were not informed about the existence of the recognizer or the fact that online modulation of the system parameters was active in some conditions. On the other hand, participants were instructed to pay great attention to auditory feedback signaling task completion after each visited goal and signaling the automatic entry to the recovery mode of interaction. Again, if the familiarization phase of an individual subject took relatively long times, experiments were discontinued for him/her, as the actual experiments are predicted thereby to extend to longer times.

5.4.6 Performance Metrics

Direct Measures

The precision and recall are computed as described in Sec. 5.3.3. Evaluation based on these metrics takes into consideration all the 4 goals visited by each subject and per experimental condition.

Indirect Measures

As described in 5.3.3, the performance of the IR module can be indirectly measured by considering the effect of shared control on the number of commands issued to complete each task. The number of issued commands can also indirectly reflect the user workload or effort associated with each condition. Aiming at generalizing our results with respect to the task space, each subject was assigned a different set of goals, and hence the observed number of commands per condition and subject can be affected by the varying complexity and path length of each sequence of goals. In order to correct for this, a baseline was computed for each sequence, relative to which the observed number of commands can be computed. The baseline is computed as the expected number of discrete normal mode commands needed to visit these goals in the same order which was undertaken by the individual subjects,

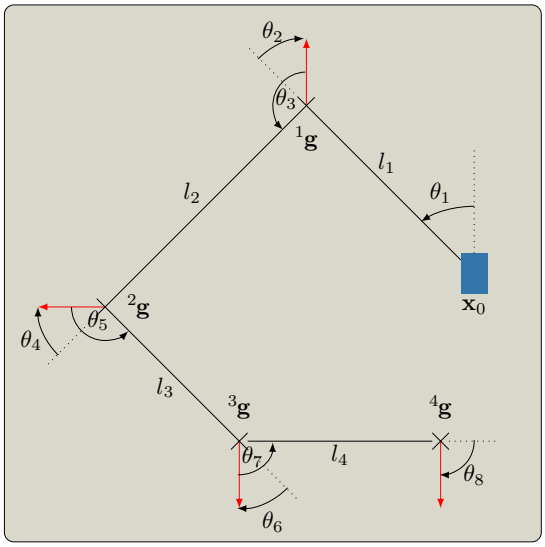


Figure 5.16: Baseline computation example showing 4 different goal poses visited in the order $^1\mathbf{g}, ^2\mathbf{g}, ^3\mathbf{g}$ and $^4\mathbf{g}$ starting from the initial pose $\mathbf{x}_0$. The baseline number of commands is computed as $\sum_{i=1}^{4} l_i + \sum_{j=1}^{j=8} \theta_j$

but assuming free space. Fig. 5.16 illustrates the computation of the baseline number of commands with a simple example.

5.5 Results

5.5.1 Task Completion

Experiments lasted for around 3 hours, including the time for reading the instructions, familiarization, training for SSVEPs, actual task performance and breaks between conditions. 12 subjects were able to complete all conditions. 2 subjects completed all conditions except the three *PB* conditions, i.e. the physical robot with BCI conditions. This was for a reason. These two subjects, namely S1 and S3 were the first to use the system, where at this stage of the experiment, we recorded only one training session for SSVEPs using the monitor stimulation. Since the *PB* conditions require visual stimulation through the HMD for online interaction, it has been observed with these two subjects that the classifier learned for the monitor-based stimulation could not generalize very well to the HMD stimulation, and therefore we decided hereafter to record another training session for the HMD stimulation as described earlier in Sec. 5.4.5. We decided to keep the incomplete data of these two subjects in our data analysis since other conditions still compare well to the procedure used for other subjects. Additionally, 3 other subjects were able to complete all the keyboard conditions, but none of the BCI's, as the accuracies for their SSVEP detection were relatively low. The familiarization step took longer than expected with 4 subjects, and therefore experiments were discontinued with them. 1 subject felt motion sickness during the familiarization stage and decided to drop out.

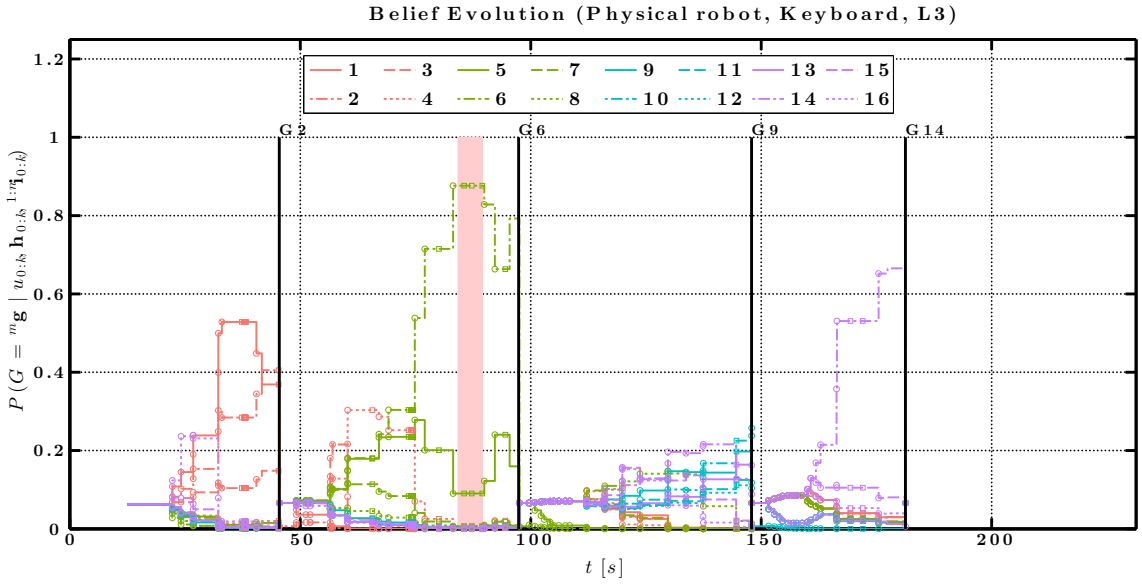


Figure 5.17: Belief evolution shown for subject S2 in condition $PKL3$. The shaded area shows the time interval(s), during which the recovery mode was active and belief updates were not. The vertical lines correspond to the time instants when the robot arrived at a goal and are annotated accordingly.

5.5.2 IR Evaluation based on Direct Measures

Fig. 5.17 shows an example of the belief vector $\mathbf{P}_k$ (in condition $PKL3$ for subject S2) as it unfolds over time during the complete session in the form of 16 staircase plots. The rising and falling edges correspond to time instances, at which the belief vector underwent gaze or command-triggered updates. Upon arrival at any goal, the belief vector is reset. Correspondingly, Fig. 5.18 shows the robot's pose during the complete session. These two figures combined can give a first insight about the performance of the Bayesian IR system.

Fig. 5.19 summarizes the results from all subjects and all finished sessions with respect to the precision and recall metrics introduced in Sec. 5.4.6, where precision was evaluated for different values of s_{thresh}. The recall plots indicate that the recognizer was able to correctly estimate the user's hidden end goals around 40% of the time in all experimental conditions. This rate increases with increasing N, e.g. it becomes around 70% for $N = 4$. On the other hand, the increased recall observed with higher s_{thresh} in Fig. 5.19(b) indicates the suitability of our proposed non-uniformness metric as a confidence measure for the recognizer.

5.5.3 IR Evaluation based on Indirect Measures

As previously mentioned, the relative number of commands issued to complete the task is adopted as the comparison criterion (i.e. the dependent variable) using the indirect measures. Since this measure reflects in a way the user's effort needed to accomplish the assigned tasks, we will hereafter refer to it as the user effort.

Our experiment in this study is a partially crossed $2 \times 2 \times 3$ repeated-measures within-subjects factorial design (with missing data). The fact that the incorporation of gaze into

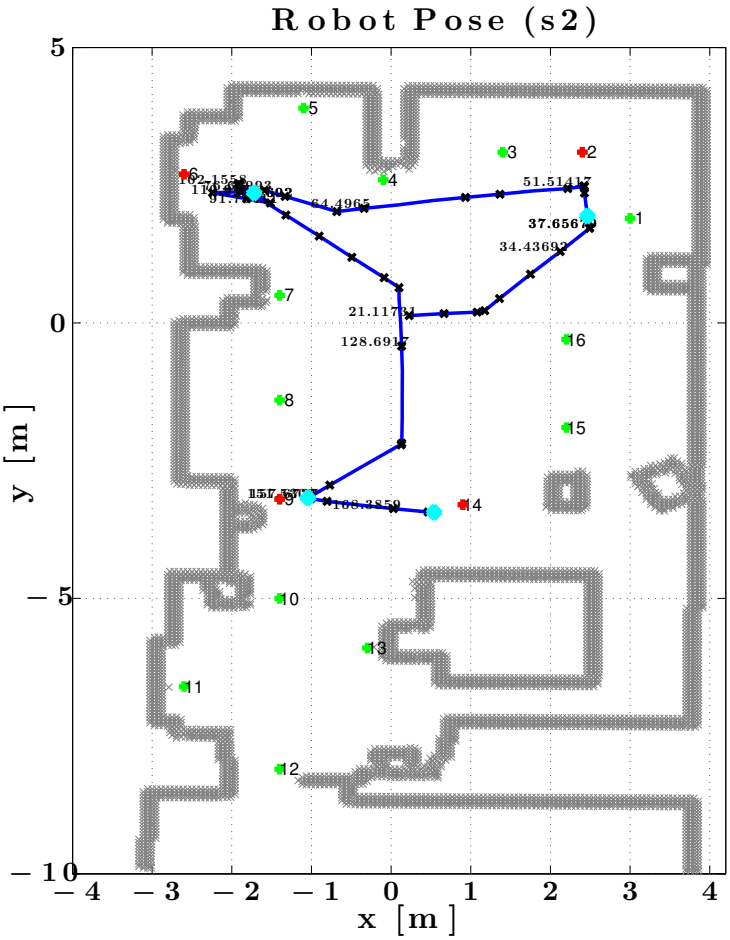
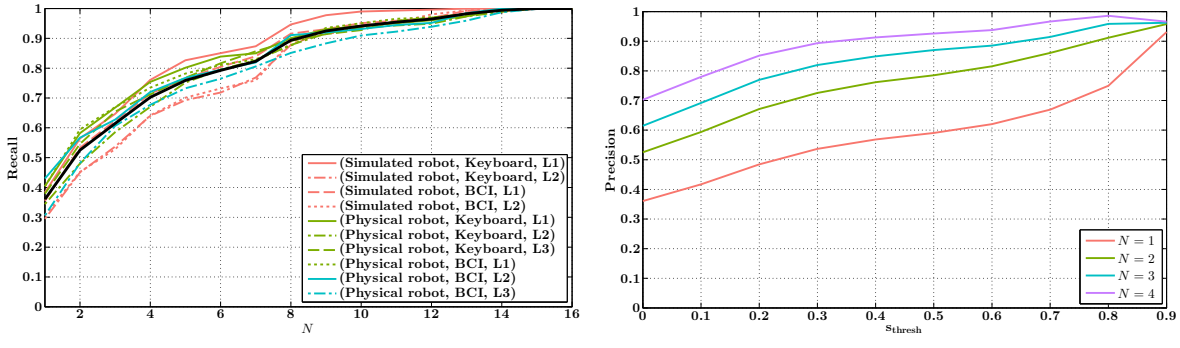


Figure 5.18: The complete robot's path $\mathbf{x}_k$ for subject S2 in condition $PKL3$. Target goals are marked in red. Additionally, points marked with cyan represent the points (in time and space) when the user received the auditory signal signaling goal-reaching. All goals in the environment are shown with their identification numbers. The time in seconds is shown for some points along path, where $t = 0$ represents the time when the session started.



(a) Recall as a function of N

(b) Precision as a function of s_{thresh} and N

Figure 5.19: Recall and precision of the IR module. Around 40% of the time, IR gave a correct estimate of the hidden user intention. From the precision plot, it can be seen that s_{thresh} can be used as a confidence metric.

the IR system is only feasible in the physical robot conditions renders it legitimate to run the analysis in two steps. Firstly, we consider the fully crossed $2 \times 2 \times 2$ design that excludes the level ($L3$) in the factor $SCTRL$ which leaves us with the levels S/P, K/B and $L1/L2$ of the main factors in the experiment. Secondly, we consider the $L3$ level within the fully crossed 2×3 design corresponding to the K/B and $L1/L2/L3$ levels of the $INTFC$ and $SCTRL$ main factors, respectively.

Statistical Analysis of the $2 \times 2 \times 2$ Design ($RBT \times INTFC \times SCTRL$)

The data obtained from all subjects and all sessions (excluding $L3$ sessions) was analyzed using linear mixed-effects model, where the subject factor was treated as a random factor, and the RBT, $INTFC$ and $SCTRL$ were treated as fixed factors. To this end, the function `lmer` from R `lmerTest` package [2] was used since it is able to handle unbalanced data, as it is the case here, by approximating the denominator degrees of freedom using either Satterthwaite's or Kenward-Roger's approximations [187]. The three-way ANOVA of type III has shown no significance in the three way interaction, $F(1, 97.12) = 0.045, p = 0.83$. However, the two-way interaction terms of $RBT \times INTFC$ and $INTFC \times SCTRL$ were significant, respectively with $F(1, 97.97) = 4.68, p < 0.05$ and $F(1, 97.12) = 4.33, p < 0.05$. The main factors $INTFC$ and $SCTRL$ were found to be significant as well. Follow-up simple effects tests were performed with the help of `multcomb` [188] and `lsmeans` [189] packages. The results are summarized in figures 5.20. Hereby, the RBT was found non-significant across all levels $INTFC$, whereas the levels $INTFC$ were found to be significantly different across the levels of the RBT, as can be seen in Fig. 5.20(a). Additionally, one can see in Fig. 5.20(b) that all simple effects of $INTFC \times SCTRL$ were significant, but with different significance levels and therefore we can attribute the significant interaction to these different levels of significance.

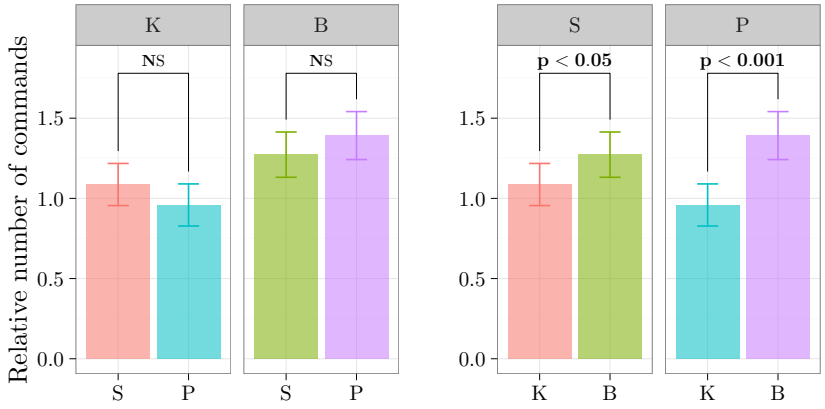
Statistical Analysis of the 2×3 Design ($INTFC \times SCTRL$)

The two-way ANOVA of type III tests revealed significance for the two main factors, i.e. $INTFC$ and $SCTRL$, with no significant interaction. Post hoc pairwise comparisons with Bonferroni corrections revealed that the $L2$ required significantly less effort than $L1$. There was also a trend that in case of $L3$ less user effort is required, but given our sample size this did not reach significance. Conforming with the three-way ANOVA from previous section, the BCI required higher effort.

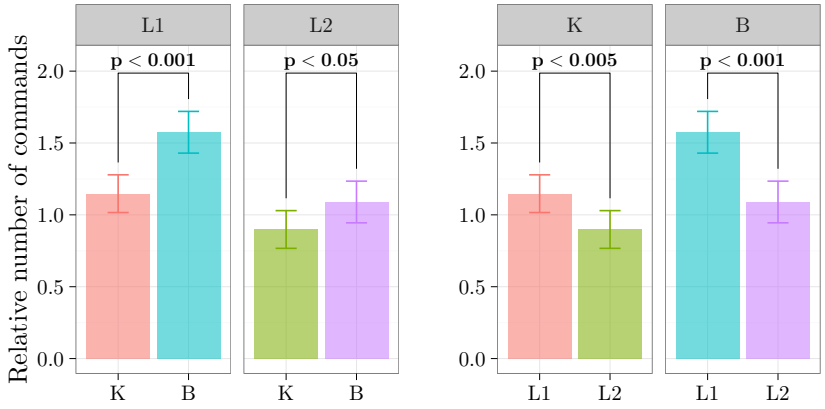
5.6 Discussion

The precision and recall results have shown that the IR system was able to infer the correct target goals, as the best prediction, 40% of the time. Comparing the $L1$ and $L2$ conditions shows that a small reduction in recall is observed, despite that IR was exactly the same in

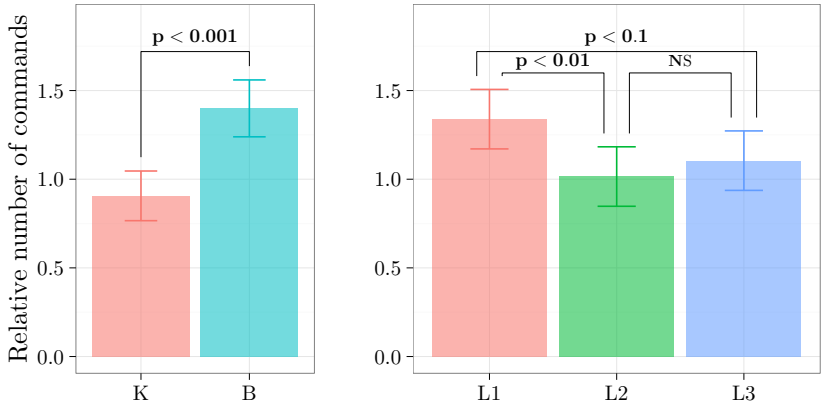
[2] The complete analysis can be viewed in this link http://rpubs.com/moh-marwan/rembodiment



(a) Simple effects for the interaction between the robot type and the interface type



(b) Simple effects for the interaction between the interface type and shared control type



(c) Main factors in the 2×3 model

Figure 5.20: Barplots of the least squares means and 95% confidence interval plotted for (a,b) the $2 \times 2 \times 3$ design and (c) the 2×3 design.

the two conditions. This can be explained by the fact that SC is applied in $L2$, which yields a smaller number of user commands and therefore less evidence about the possible target goals is available to the goal recognition module. The non-uniformness metric, i.e. $s(\mathbf{P}_k)$, was shown additionally to be an adequate measure for the confidence of the IR module, as higher values of $s(\mathbf{P}_k)$ resulted in higher precision of the IR module as can be seen in Fig. 5.19(b). Our statistical analysis, on the other hand, has demonstrated that SC based on IR beliefs resulted in less user effort (H_1). The incorporation of gaze information into the IR module did not help to enhance the performance of our IR-SC system as was expected (H_2). Several reasons might have contributed to this result. First, we have assumed a simple model for the user gaze based on the hypothesis that users frequently fixate their vision at the goal of interest to help update their global plans. While this is true in general, the constant time which we assumed in this work (i.e. 200 ms) for head fixations seems not suitable to capture fixation times realistically. Second, users tend to look more often at the immediate vicinity of the robot for better local trajectory planning. Therefore, we predict that considering these details into the user gaze model in future developments will be beneficial.

The results from the three-way ANOVA have shown no significant difference between the simulated and the physical robots/environments (H_3). This is indeed a desirable feature of simulation systems as further developments in IR can be, if feasible, tested first in simulation. Additionally, it comes as no surprise that the BCI resulted in higher user effort, with average increase of around 30% compared to the keyboard (H_4). This is mainly due to the imperfect detection of the SSVEP signals. However, most of the participants were able to use the system and accomplished relatively complex navigation tasks during the different conditions. Subjects, for which the SSVEP detection did not get over 70% accuracy, encountered difficulties to accomplish the task in the BCI conditions, and therefore, BCI sessions were discontinued with them as has been previously reported. With respect to the BCI inclusion in the IR algorithm, Fig. 5.19(a) clearly shows that this did harm the precision of the IR module. Importantly, the effect of $L2$ shared control has been shown to be more significant in the BCI sessions compared to the keyboard case as can be seen in Fig. 5.20(a). This also should come as no surprise since shared control is certainly more useful for noisy interfaces.

Even when the IR module was highly confident about its prediction, the correct target goal was sometimes confused with other goals in the environment. The belief evolution plots for all subjects and conditions show that such confusion happens most of the time with adjacent goals. Therefore, the performance of the IR module is expected to improve for environments with less goal density. Reciprocally, with a higher goal density, the performance of the system might be negatively affected. The number of goals in the environment per se is not expected to have a high influence on the precision or recall, but will definitely affect the processing speed. This is due to the fact that the computation of the path plans to all goals, which takes place at the end of each movement, constitutes the processing bottleneck of the IR system. In order to efficiently generalize to environments with higher number of goals, and environments with higher goal density, a hierarchical structure of target goals in the environment might be of high benefit. Hereby, target goals are clustered, with respect to their distance to each other, within different levels. IR is applied to the goal hierarchy from the coarse to the refined representation. In other words,

recognition is applied first between clusters, and then between the goals within each cluster. Further investigation of this approach is required.

Furthermore, online modulation of the system parameters on the basis of the IR belief resulted in an average reduction by a factor of 30% (for level $L2$) in the total number of commands required to accomplish the tasks. Given the observed precision levels of the IR system shown in Fig. 5.19, more proactive assistance can also be provided in further developments to the system, e.g. to autonomously maneuver the robot towards a goal, when the system is highly confident about its predictions. This brings us back to the autonomy continuum we discussed in Sec. 5.1, where the belief vector lends itself as a plausible criterion to automatically change the operational mode of the robot. Additionally, the belief vector can be used to automatically trigger interface adaptations, whereby e.g. the most probable goals are shown for selection.

During the experiments reported in this chapter, most subjects had difficulties, which were also reported verbally by some of them, to issue stop commands as quick as required in the BCI conditions. This is due to the limiting factors of the buffering step in the SSVEP detection method. Therefore, devising a reliable way to stop the robot, e.g. with an EMG channel [190], if feasible, might be necessary to ensure that users can maintain supervisory control over the actions of the robot.

We assumed throughout this work that the map of the remote environment and the possible goal locations are known a priori. If not available, this information can be also learned throughout interaction with users. For instance, states of interest (goals or sub-goals) were extracted from successful task episodes based on the averaged occurrence frequency in [191]. The map itself is easily obtained with the SLAM algorithm, e.g. with `gmapping` ROS package [186] which was used for the purposes of this work.

We assumed additionally that users, at each time, have a specific end goal in mind. The proposed IR module is expected to perform also well in situations when users change the targeted goal during movement. The reason for adopting a *keyhole* approach in this work was primarily to elicit natural behavior of users so that neglecting user cooperation while evaluating the performance of the IR-SC algorithms becomes legitimate. In this regard, *intended* IR approaches might be appropriate for robotic embodiment systems as well. For instance, should the user become aware of gaze-triggered belief updates about their end goals, users might choose to fixate purposely at these goals for extended periods of time, so that some ambiguity can be resolved at the recognizer side.

Due to the length of our experiments, we were not able to include another factor to test the effects of the adaptive BCI. However, we argue that adaptations to the BCI played a major role in making the task easy to complete and in keeping the number of interactions required to complete the assigned tasks comparable with the baseline. For instance, consider the scenario where the robot is facing a desk where the target goal is only some tens of centimeters away from the robot. With the normal mode commands only, the user needs to convey several commands to arrive at the goal, e.g. perform a half-turn, move forward, perform a 90° turn, move forward, move for another 90° and move forward. This sequence is also not guaranteed to arrive at the goal location in case of BCIs. However, with the adaptive interface, the target goal is only one command away from the user. This very advantage of the adaptive BCI has been observed often during the experiments.

Throughout this work, we also assumed that the localization module provided perfect

estimates about the robot's location. Given the underlying probabilistic nature of the `AMCL` ROS package that implements the adaptive particle filter for localization [180], the estimated robot's pose is a noisy version of the true pose. We expect that this noise in localization affected all experimental conditions uniformly, and therefore its effect on the obtained results can be ignored. However, future developments might benefit from taking the uncertainty in the robot's pose into consideration.

There are some technical details that we skipped so far for the sake of not cluttering the board with too many details that obscure the most relevant issues to IR, but these details remain part of the bigger picture of robotic embodiment. For instance, the stereoscopic visual feedback provided to users did not perfectly match direct line-of-sight. Though the robot was equipped with cameras that provide large field of view, the complete FOV of humans was not reached. As a result, the perceived distances through the visual channel did not map one-to-one to real distances. Obviously, users were able to learn the actual mapping as they could navigate freely in the remote environment accordingly. Furthermore, the requirements for real-time video streaming necessitate compression and decompression of captured images at the robot and the user side, respectively. This leads to some compression artifacts that reduce the quality of viewed video stream, which is further reduced by the resolution of the used display, i.e the HMD. Additionally, due to physical constraints, the 3 DoFs neck of the robot did not exactly match the range of human neck movement. As a result, it has been observed during experiments, that such mismatch often affected the user ability to explore the environment. Improvement to such technical details will undoubtedly contribute to improved quality of user experience in robotic embodiment systems.

5.7 Conclusions

In this chapter it is argued that adaptive BCIs offer a way out of the bottleneck of bandwidth limitation in BCI-based robotic applications. In order for interface-self adaptations to be effective, reasoning about the user hidden intention is a crucial element in adaptive BCIs. On this account, we have focused on intention recognition within a specific robotic application, namely navigation tasks in collision avoidance mode. The proposed intention recognition module is based on a recursive Bayesian update rule, for which some simplification is done by considering some intuitive heuristics that can model the behavior of the general population, and therefore can be used in a plug-and-play fashion. Based on the obtained belief vector from the intention recognition module and the confidence about such beliefs, the shared control module adapts the translational and rotational movement steps, so that the robot can be brought closer to the goals of highest scores. Confidence about the belief estimates was computed with a new metric, that reflects the non-uniformness of the belief vectors.

In order to evaluate the proposed intention recognition system, experiments were conducted with healthy subjects within robotic embodiment settings. These experiments varied along three factors: type of the robot/environment (simulated and physical), type of the interface (keyboard or BCI), the way intention recognition is used by shared control (IR is based on user commands only with no shared control, intention recognition is based on user commands only with shared control applied, and intention recognition is based on user gaze

and commands with shared control applied). Our results have shown that the intention recognition algorithm was able to track and infer the hidden user goals with relatively high precision and recall. Results have shown also that a simple shared control scheme on the basis of the intention recognition belief/confidence significantly reduced the user effort needed to accomplish specific tasks, where effort was thought to be reflected in the number of issued commands. Additionally, we found that there was no significant difference between the simulated and the real environments, suggesting that the simulated environment can be used in the test phases of further developments to intention recognition algorithms. Despite the fact that BCI required higher effort compared to the keyboard conditions, most subjects were able to complete the assigned tasks. The proposed intention recognition system was additionally shown able to handle the uncertainty in user input. Indirect measures have shown that the effect of the intention recognition was more pronounced for BCIs than in the case of keyboard interfaces in reducing the user effort. However, subjects, who had relatively low SNR for SSVEPs encountered considerable difficulties in the task. As a result, corresponding BCI experiments were discontinued with them since otherwise, experiments would last for extended periods of time, given the fact that frequent erroneous interactions give rise to oscillatory behavior of the robot, where e.g. random in-place rotations are very frequent.

Moreover, the collision avoidance mode was adopted since it allows to leave most of control in the hands of the user. Setting the collision avoidance as the default operational mode brings other advantages. This mode allows the intention recognition module to benefit from high rate of user commands, on which inference can be based. In higher autonomy modes of robot operation, where less user input is available, other intention recognition methods are required. Additionally, the obtained belief vector from the IR module can be used for other purposes than the simple shared control application. For instance, it can be used as a basis for interface self-adaptations. That is, when the belief reaches high confidence regarding target goals, the most probable goals can be shown to the user for selection.

In summary, we have devised a recursive Bayesian rule to infer and track the target goal locations pursued by users while navigating in robotic embodiment systems. Being based on different intuitive heuristics and assuming that users typically exhibit similar navigation behavior in structured environments, the proposed method can be used without prior training of system parameters to individual users. The output of the goal recognition module lends itself as a plausible criteria to guide more advanced shared control driving schemes and strategies for interface self-adaptation.

6 Conclusions and Outlook

This thesis tackled the problem of advancing brain-computer interfaces (BCIs) for robotic embodiment applications, where EEG-based selective attention BCIs were the main focus. Sec. 6.1 summarizes the main results of the thesis and Sec. 6.2 reflects on potential future directions for research in BCIs and their deployment for robotic embodiment systems.

6.1 Summary

There are two major limitations that hinder fluent BCI-based interaction in immersive robotic applications. First, decoding of brain signals is yet error-prone and leads to erroneous interactions, which often need to be corrected or reversed. In some cases, however, reversal of wrongly executed actions is not feasible, e.g. entering a narrow dead-end corridor. Second, the bit rate, bandwidth, temporal resolution and DoFs available with BCIs are far beyond the requirements of embodiment applications. We have addressed these issues from different perspectives aiming at minimizing their implications.

On this account, chapter 3 was devoted to the problem of detecting SSVEPs from continuous multi-channel EEG. With genuine theoretical analysis of the state-of-the-art methods, it was observed that these methods have many foundations in common, and particularly, they all build upon the auto-covariance of the EEG signals and the cross-covariance between the EEG signals and a commonly assumed source model for the SSVEP signals. Additionally, we found that the maximum contrast combination (MCC) and multivariate synchronization index (MSI) methods are variations of the canonical correlation analysis (CCA) method, and it was argued, therefore, that their detection results should not differ significantly. It was also shown that these three methods ignore the fact that background activity might overlap with the used stimulation frequencies. The minimum energy combination (MEC) method was shown to effectively capture the noise level in the spatially filtered signals. At the same time, it was argued that MEC poses an artificial criterion for dimensionality reduction. Based on the findings of this analysis, we proposed the novel CVARS detection method as an alternative. The new method combines the power of the canonical variates and that of autoregressive spectral analysis in estimating the signal and noise powers, respectively. The new method was shown to significantly outperform CCA (and thus MSI and MCC) and slightly outperforms the MEC method. The supervised version of CVARS proved its accuracy and reliability in estimating the hidden user intention including the idle state. Furthermore, as an application of the new method and using a fully crossed factorial experimental design, we compared the effect of the display (LCD monitor and HMD) and viewing condition (with monocular right eye, monocular left eye and binocular viewing) on obtained accuracies for the unsupervised CVARS. Hereby, the HMD has proved its adequateness in delivering the visual stimuli and was shown to provide higher accuracies than the LCD monitor condition. Moreover,

it was shown that binocular viewing leads to higher detection accuracies than monocular viewing, should visual stimulation of the two eyes be done in complete phase synchrony. Results have suggested that binocular inhibition is brought about by anti-phase dichoptic stimulation of the two eyes.

Despite the progress made in detecting different EEG patterns, here and elsewhere, detection errors remain part of the decoding process. In order to further reduce the effects of these errors, it was demonstrated in chapter 4 that interaction ErrPs can be used as a validation step. This has been also shown in the literature to improve the bit rate of typical BCI systems, e.g. P300-based BCIs. Our main focus in this chapter was laid upon understanding potential invariant features of interaction ErrPs across tasks, subjects and over time, and the effects of such invariance, if any, on the classification of ErrPs. In order to answer these research questions, we have designed and conducted three experiments, by which ErrPs were observed and examined. Our analysis was primarily based on observed grand average waveforms and classification results. Results have shown that interaction ErrPs are relatively invariant over time, and to a less extent across subjects, but not across tasks. The variability of ErrPs across tasks, manifested by different grand average waveforms and by chance-level classification accuracy for transferred classifiers, was attributed to the different mental processes engaged in assessing whether interface actions are correct or not. Altogether, our results suggest that a classifier learned from one task should be learned anew or recalibrated for new tasks that require different mental processes to assess interface actions. This new approach to the classifier transferability problem adds a fundamental layer to the design of transferable classifiers. We have also proposed a new simple task that simulates adaptive BCI systems, whereby interface elements are updated continuously throughout interaction. The novel inplace feedback presentation strategy introduced with one of our experiments lends itself as an alternative to other feedback strategies in hybrid P300-ErrP BCIs. The inplace feedback is expected to result in better classifier transferability as the mental process of assessing interface actions should be the same, regardless of the nature of the P300 application in use.

Chapter 5 discussed application-specific advances to robotic applications, where interface self-adaptations were proposed as a means to circumvent the limited bandwidth of BCIs. In particular, the problem of robot navigation was discussed and a taxonomy for the levels of robot autonomy was suggested. In this regard, we have drawn a link between interface self-adaptations and the different levels of automation. We have also shown the central role, which user intention recognition methods play in making interface self-adaptations effective. The problem of intention recognition was thus the central point in our treatment of interface self-adaptation. Hereby, we have particularly considered the application of robot navigation in collision avoidance mode, where it was additionally assumed that the robot is embedded in a remote environment with a known map, and a well defined set of goal locations. Moreover, it was assumed that interaction is mediated with discrete interfaces, and user incremental commands defined only the direction of robot rotation or translation. To this end, a Bayesian intention recognition framework was devised such that the probability of each goal was updated as a response to arriving evidence from the user commands and user gaze and on the basis of interaction history and other available contextual information. The Bayesian inference system was realized with different novel and intuitive heuristics. These heuristics, inspired by general behavioral patterns observed

in navigation tasks, were adopted for the general population and therefore allowed the system to be used without prior training. Additionally, a novel metric was proposed to reflect upon the confidence in the obtained beliefs. Direct and indirect measures were adopted as performance metrics of the resulting Bayesian model. Direct measures reflect the accuracy of the inferred belief vectors, whereas indirect measures are obtained by making use of the belief vectors, in one way or another. To this end, we have developed a simple probabilistic shared control scheme that uses the belief and the confidence of the intention recognition module, and silently modulates the magnitude of the translational and rotational robot steps. The number of commands needed to finish specific navigation tasks therefore was used to indirectly measure the efficiency of the intention recognition module. Moreover, this chapter provided an empirical evaluation of the proposed approach with one experiment having 10 different conditions. Hereby, participants were instructed to visit different goal locations in the remote environment. Experimental conditions varied with respect to three main factors: type of the environment (simulated or physical), type of the interface (BCI or keyboard) and the way shared control and intention recognition were realized (IR based on user commands only with shared control disabled, IR based on user commands only with shared control enabled, and IR incorporating user commands and user gaze with shared control enabled). Most of the subjects were able to finish all tasks, and those who had SSVEP detection accuracies below 70% encountered difficulties in the BCI conditions. Results have shown that the intention recognition system was able to estimate the target goals after a number of interactions within all experimental conditions. Additionally, intention recognition and shared control were shown to reduce the user effort, measured by the number of interactions required to accomplish the well specified navigation tasks. The decrease in required effort, and thus the effect of intention recognition and shared control, was found more significant in case of BCIs compared to the keyboard conditions. Simulated and physical environments have shown comparable results, and therefore future development can be firstly tested on simulated environment, reducing the cost of these tests. The incorporation of gaze information into the intention recognition module did not show any improvement. Therefore, more accurate models/parameters for modeling user gaze are necessary to better exploit this available information in the recognition process. Our experiments are quite unique in the context of BCI robotic embodiment applications. We believe that the fact that we aimed at a user agnostic adaptive BCIs helped to recruit many subjects with no prior experience in these systems.

In summary, this thesis has proposed different methods and algorithms that aim at overcoming some of the limitations in current BCIs and/or at reducing their implications. Further steps in these directions are yet necessary to allow for easy-to-use intuitive BCIs in robotic applications.

6.2 Outlook

It has been recognized in [192], that the lack of standardization and the little agreement on the most promising future directions are currently among the major challenges in the field of BCI. Efforts in this direction are currently carried out by researchers within the EU project “BNCI Horizon 2020 project: the future of brain/neural computer interaction”.

Besides these efforts, BCIs are expected to grow in the following directions.

6.2.1 Applications for Patients

Throughout this work, we have mainly reported results from experiments with healthy subjects. Generalization of these results to patients remains to be empirically validated. As a first step in this direction, in [219], we have compared a small group of people with spinal cord injury (SCI) to a control group of healthy participants during P300-based virtual and robotic social interactions. Results have shown that both groups used the BCI within the immersive scenarios with good levels of accuracy and perceived control of the virtual and robotic avatars. Several other studies with Amyotrophic lateral sclerosis (ALS) and locked-in syndrome (LIS) patients, as in [193–195], reported that patients were able to use P300 and SSVEP-based BCIs. The results of all these studies, however, were based on a limited number of patients. Therefore, there is a need for more studies to examine applicability of BCIs with larger groups of patients and with patients with different medical conditions.

6.2.2 Adaptive BCIs

In this work, we have touched the surface of adaptive BCIs. Application-specific adaptive BCIs, as argued hereby, can be of great importance to enhance the communication and control channels between the user and the artificial devices under consideration. Probably the best interface one could think of, is a single-button device by which the user can, infrequently, object what the controlled device is currently doing completely autonomously. An impressive step towards this goal has been realized by [196] for navigation tasks, whereby the interface suggests the most probable actions based on the state of the robot in the environment. The user either accepts or rejects the propositions using a single-button interface. Several sequential proposition might be needed to arrive at the user's command of interest. Error-related potentials were suggested as a means to communicate the approval/refusal of the user [196]. The single-button interface based on ErrPs was successfully used to move a cursor inside a simple grid [197].

In robotic embodiment applications, the single-button interface requires machine actions to match, within a margin of tolerance, the expectations of the user within different contexts and for extended periods of time. In fact, there is some evidence that human users tolerate some deviations of automated actions, and sometimes even attribute completely autonomous movements to themselves. Since modeling users to the level which enables the realization of such interfaces is currently not feasible, incremental advancements in the fields of shared control and user modeling can eventually bring us closer to this hypothetical single-button interface.

In addition to adaptations in the interface elements, BCIs can benefit from adaptations to their processing pipeline, including the preprocessing, feature extraction and classification steps. Shenoy et al. [198] have shown that simple adaptive classification schemes based on labeled data from online sessions can improve MI-BCI performance significantly. Llera et al. [142] have proposed to use interaction ErrPs as neural feedback, by which adaptations to a linear classifier of a first stage selection (i.e. the task classifier) can be performed online.

These approaches make good use of labeled data collected during online sessions, and therefore, potentially can improve the generalization power of the used classifiers. Efforts in this direction are expected to result in more robust and reliable BCIs.

6.2.3 Novel EEG Patterns and Novel Processing Methods

Current BCIs build upon different ERPs that were originally discovered in the field of neuroscience and related disciplines. It is expected that this trend will continue in the future and novel ERP patterns will be available for BCI-based interaction. Additionally, spontaneous EEG signals carry information about ongoing brain activity other than the time-locked ERPs (which is treated as noise in ERP studies). Motor-imagery-related spectral perturbations and spectral signatures of errors are good examples here. Extracting other information from spontaneous EEG has the potential to advance BCIs to a great extent. Moreover, as we have seen throughout this work, BCIs, especially those based on selective attention, have a relatively poor interaction temporal resolution (i.e. in the range of seconds), despite the fact that EEG can be acquired with relatively high sampling rates (i.e. in the range of milliseconds). Modeling EEG data (e.g. by advanced source localization techniques) in a way that allows to interpret EEG changes as they unfold millisecond per millisecond, can further improve the overall temporal resolution of BCIs.

6.2.4 Hybrid BCIs

Throughout this work, the only hybrid BCI we encountered was of the P300-ErrP type. There are other systems which were successfully used in the literature. For instance, the works in [199, 200] have developed different hybrid SSVEP-P300 applications. Relying on more than one source of evidence helps to enhance the discrimination power between target and nontarget stimuli. Other innovative hybrid BCIs are expected to replace currently used BCIs which rely on a single type of ERPs. In the same vein, the BCI which was used in our experiments to control the robot avatar on the basis of SSVEP signals, might benefit from additionally incorporating motor-imagery-related perturbations into the decoding process.

6.2.5 Portable BCIs

Novel wearable technologies that allow BCI systems to be more convenient and more portable are highly desirable, as this contributes to bringing the BCI technology to the market of commercial applications. In this regard, some examples already exist. A prototype for a wearable BCI is proposed in [201], where a front-end wearable EEG device with retractable comb-shaped active dry electrodes is used to measure the EEG activity. A back-end host system (a tablet) receives the EEG signals wirelessly and performs EEG monitoring and detects MI-related signals.

6.2.6 BCI-based Robotic Embodiment

In this work, we have primarily discussed the specific problem of robot navigation. However, the objective of robotic embodiment systems is to allow interaction in a general purpose

manner, where in addition to navigation, applications include object manipulation and social interaction. To realize such general purpose embodiment systems, further developments are necessary. The conceptual framework shown in Fig. 6.1 can be used to guide future research and development. This framework has been developed and envisioned based on discussions within the EU VERE (Virtual Embodiment and Robotic Re-embodiment) project.

The *recognition* block comprises of modules that extract information from EEG and physiological signals. Outputs routed to other modules include high and low-level intentions, error-related signals, affective and cognitive state of users (fatigue, cognitive load, etc.). Hereby, the high-level intentions refer to *what* the user wants to do and to *which* actions they aim, and the low-level intentions refer to *how* they want these actions to be done. This block additionally provides a quantitative measure for the quality of experience (QoE). The *embodiment optimization agent* (EOA) continuously receives subjective measures of embodiment and QoE from the recognition module and decides upon appropriate operating points of other modules aiming at maximizing the user's feeling of embodiment. As can be seen in Fig. 6.2, this includes deciding upon the appropriate:

- levels of robot autonomy: which operational mode on the autonomy continuum should be used, and whether shared/collaborative control should be applied.
- feedback: how rich the feedback and which modalities should be used.
- allowed latency: what should the maximum allowed time for a single command be.
- levels of action personalization: to which extent actions should be user-specific, e.g. by modulating action parameters based on the affective state of the user.
- nature of intentions: whether actions are triggered by user sensorimotor intentions (e.g. willingness to move) or driven by the user's willingness to use the visual interface elements.

The previous list is not exhaustive. Factors that prove to have effect on the feeling of embodiment should be added to the design of the EOA as well.

The *interface manager* coordinates transparent interface adaptations (i.e. adaptations in the interface elements) and decides upon the BCI paradigm that is appropriate for interaction. Similarly, the *action manager* takes into account user intentions, error-related signals and parameters for personalization to decide upon robotic commands and their parameters, which are forwarded to the robot for execution. Hereby, for instance, incremental movement commands are directly executed by the low-level control of the robot, while goal-oriented commands are interpreted and carried out by the action models and motion planning modules. Furthermore, the interface manager signals the action manager to execute user inputs provided by the recognition module. Information about the current state of the environment is continuously gathered from sensory data, and shared between the different modules, which also borrow knowledge from the task, skill, and preference databases. Such databases can be gathered throughout long-term interaction.

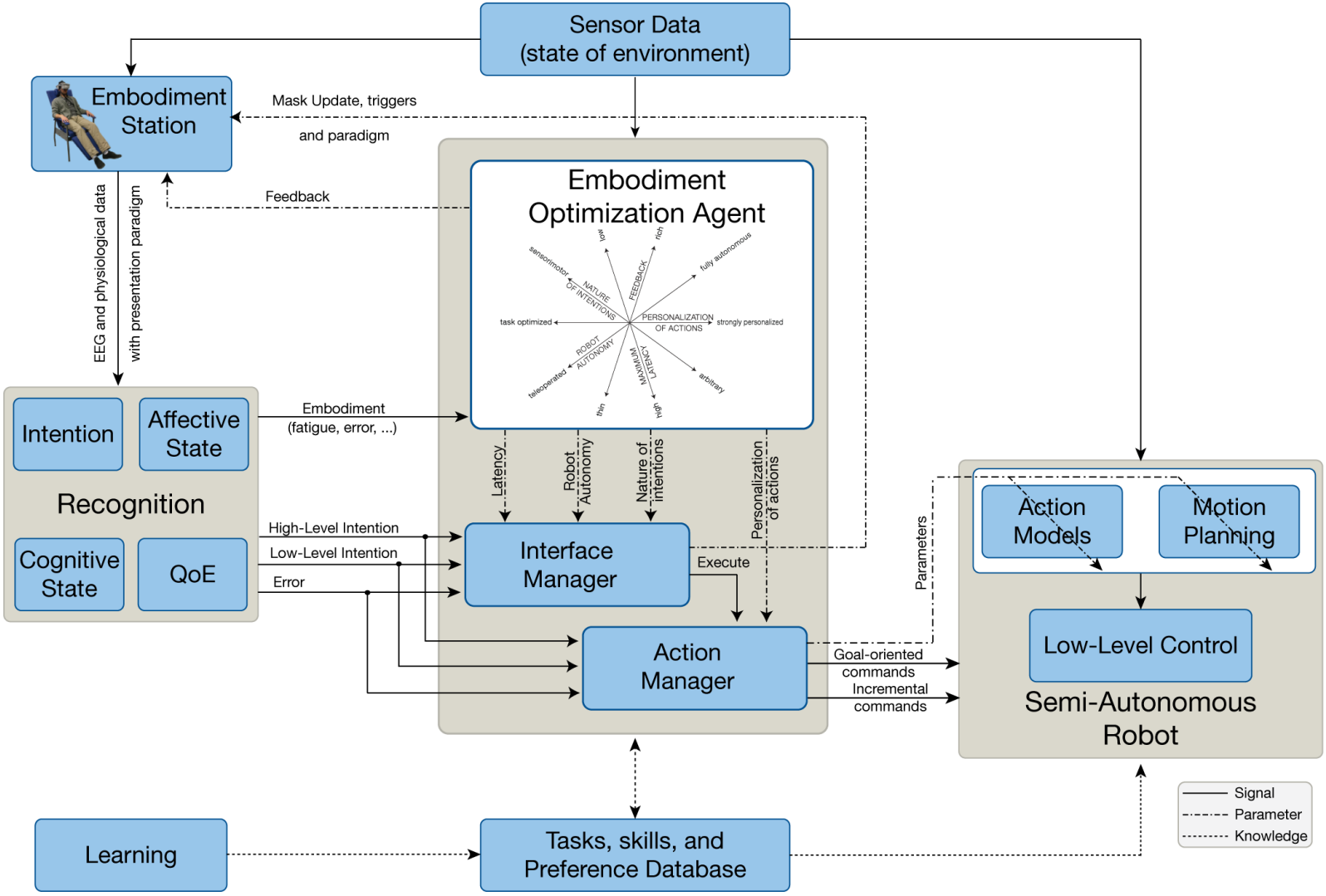


Figure 6.1: Conceptual framework for BCI-based immersive robotic embodiment.

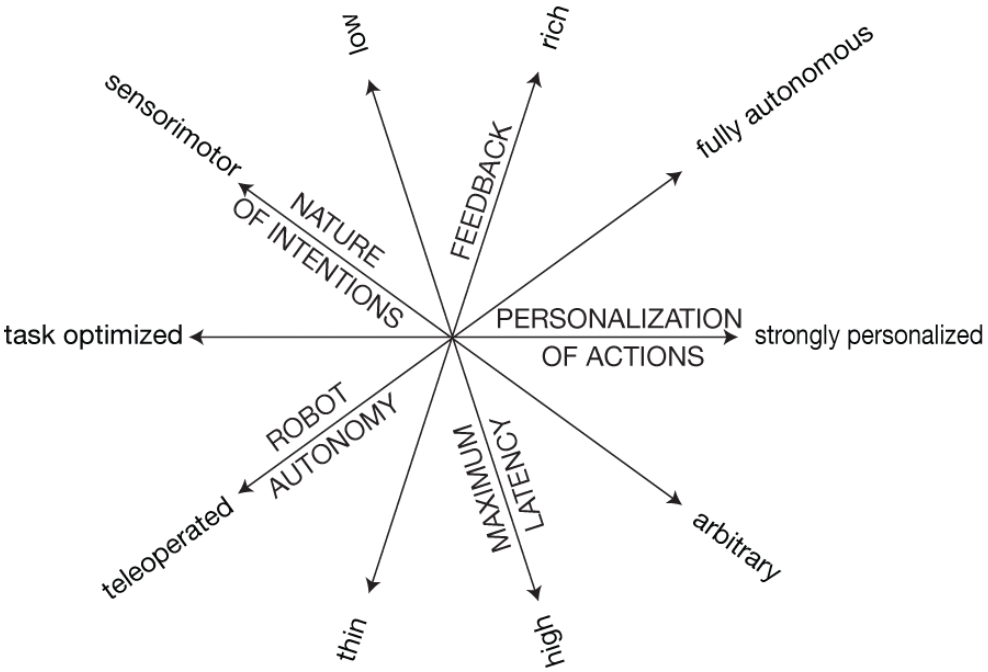


Figure 6.2: Embodiment optimization agent.

A Further Results on Interaction ErrPs

In Sec. A.1, we first show the average event-related spectral perturbation (ERSP) and inter-trial coherence (ITC) for the three different experiments reported in chapter 4. These plots are obtained with the function `newtimef` from the EEGLAB toolbox [59].

Additionally, in Sec. A.2, we reproduce the GAC, GAE and GAD plots encountered in chapter 4 for 6 s duration centered around the feedback onset at $t = 0$. The new figures give more insights about the SNR level of the interaction ErrPs, and additionally allow to observe event-related potentials (other than ErrPs) which are associated with the different experiments prior to the feedback onset.

Experiment I The activity which can be observed similarly in GAC and GAE in Fig. A.4 around $t = -1$, i.e. 1 s before the feedback onset, represents the EEG potentials related to the event of pressing the keyboard keys. Recall that there is a jitter between the feedback onset and the time of the key presses (i.e. the time difference is in the range $[0.9 - 1.1]$ s).

Experiment II The strong alpha activity which appears after $t = -2$ s, i.e. the time when the flashing of the P300 matrix stops, in GAC and GAE waveforms in Fig. A.5, might be explained by the fact that subjects tend to defer eyeblinks until task performance ends [133, 202]. The activity at the FCz site in GAC and GAE waveforms following $t = -1$ corresponds to the brain response to the central square when it appears on the screen..

Experiment III Exactly as in experiment II, one can observe the strong alpha activity in the GAC and GAE waveforms in Fig. A.6, time-locked to the event when P300 flashing stops at $t = -2$ s. The alpha activity in experiments II and III completely overlaps after $t = -2$ s.

A.1 ERSP and ITC

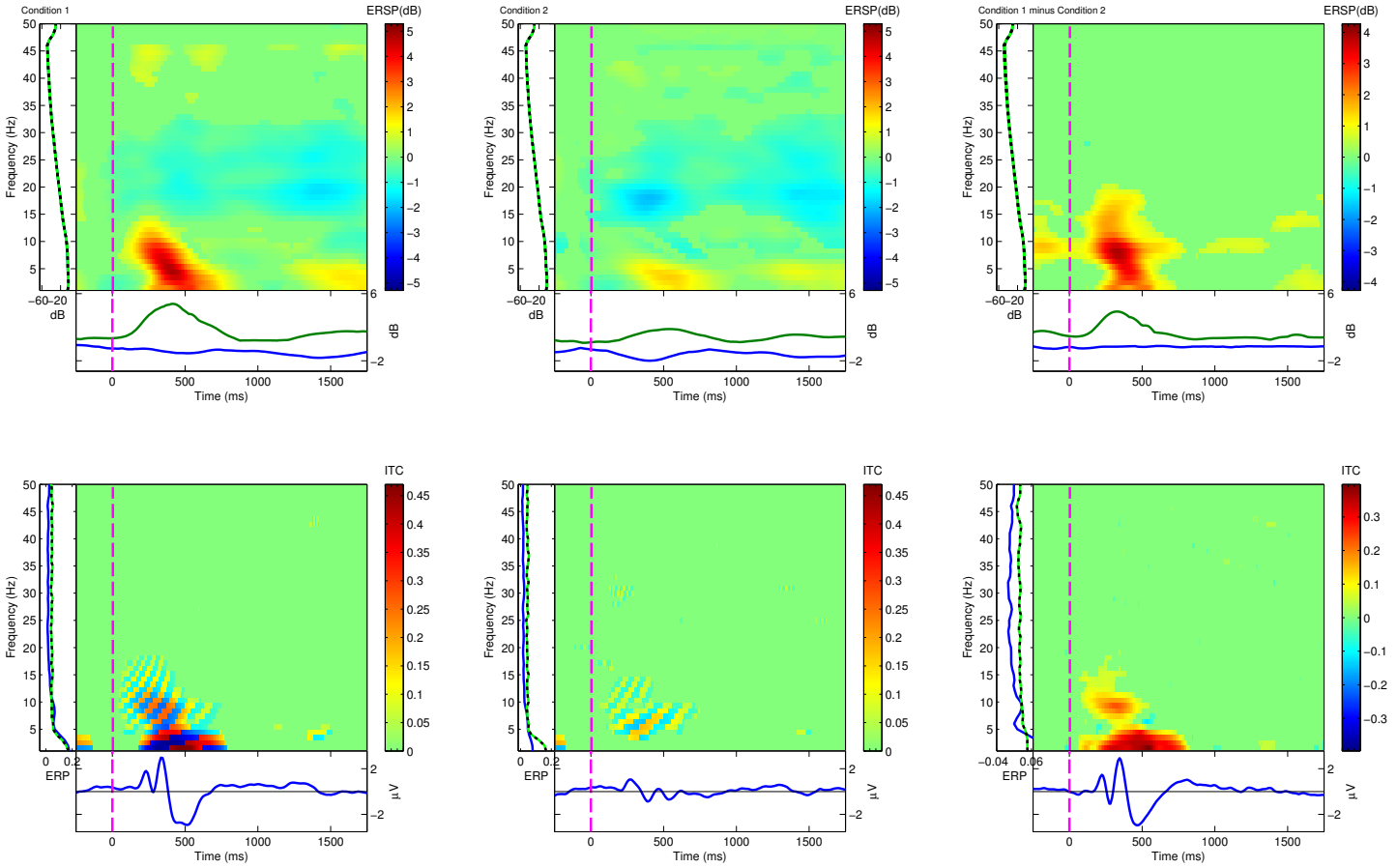


Figure A.1: ERSP and ITC for experiment I. The left, middle and right panels respectively show resluts for error, correct and error-minus-correct trials.

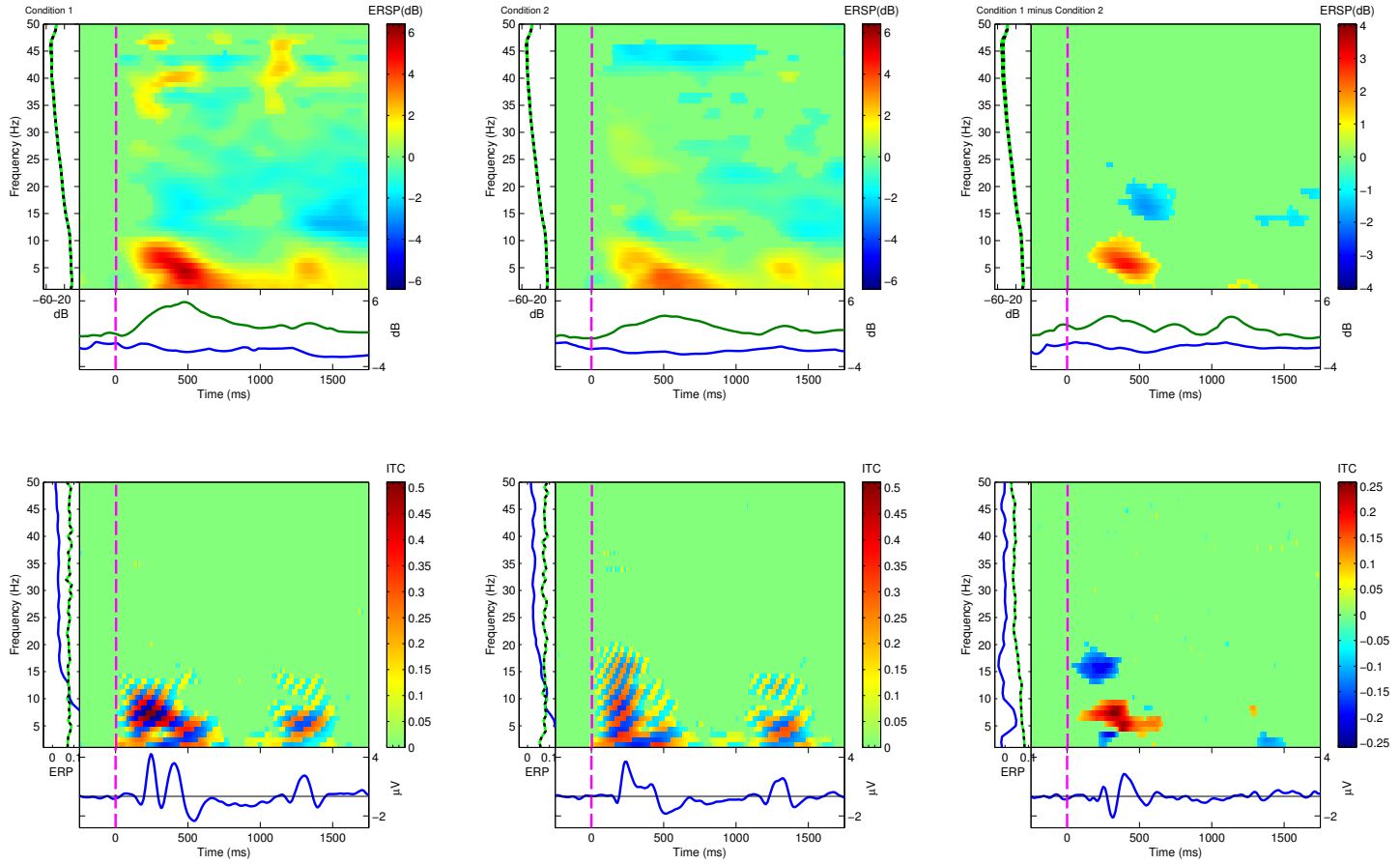


Figure A.2: ERSP and ITC for experiment II. The left, middle and right panels respectively show resluts for error, correct and error-minus-correct trials.

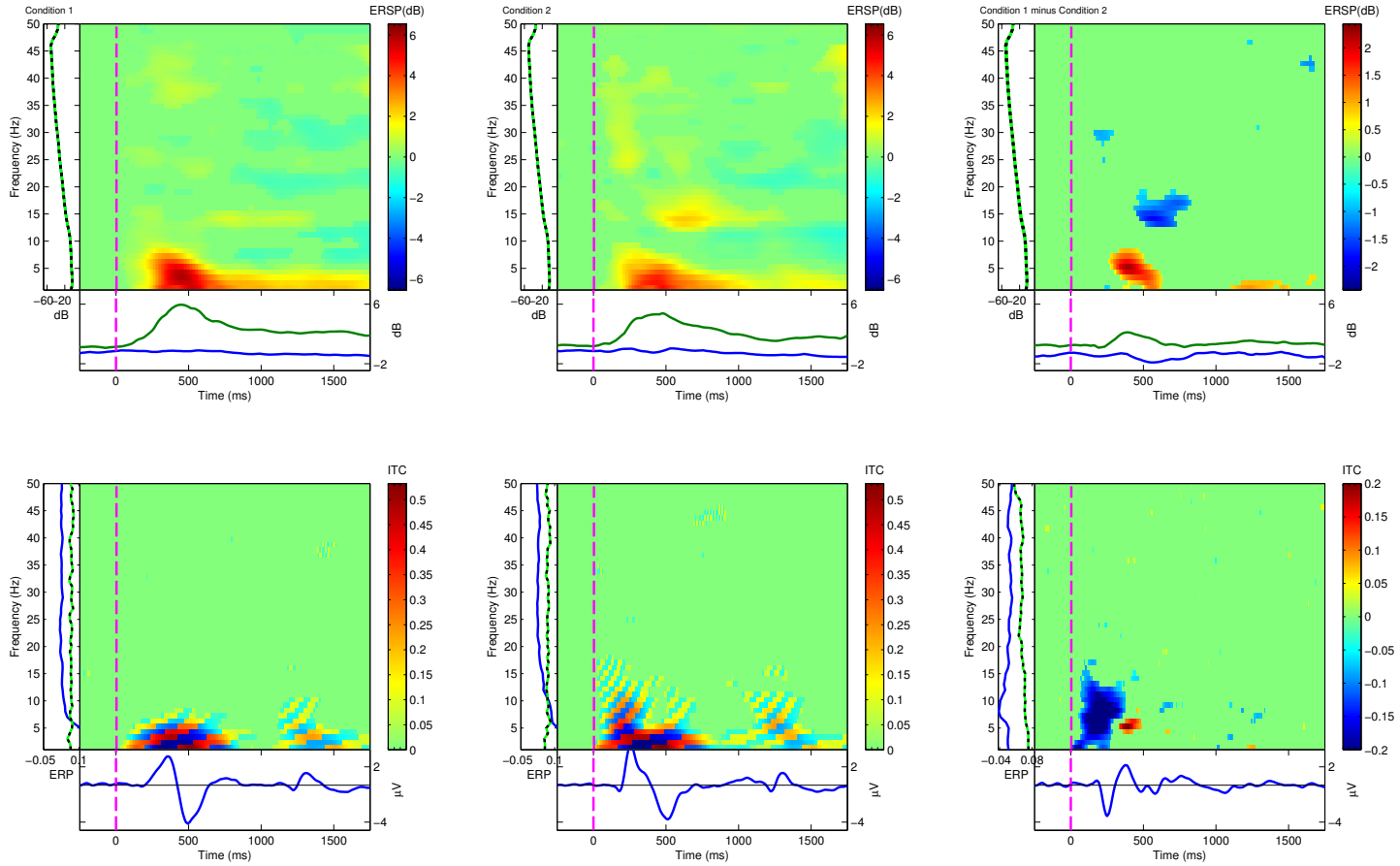


Figure A.3: ERSP and ITC for experiment III. The left, middle and right panels respectively show resluts for error, correct and error-minus-correct trials.

A.2 Grand Averge Figures - The Full Picture

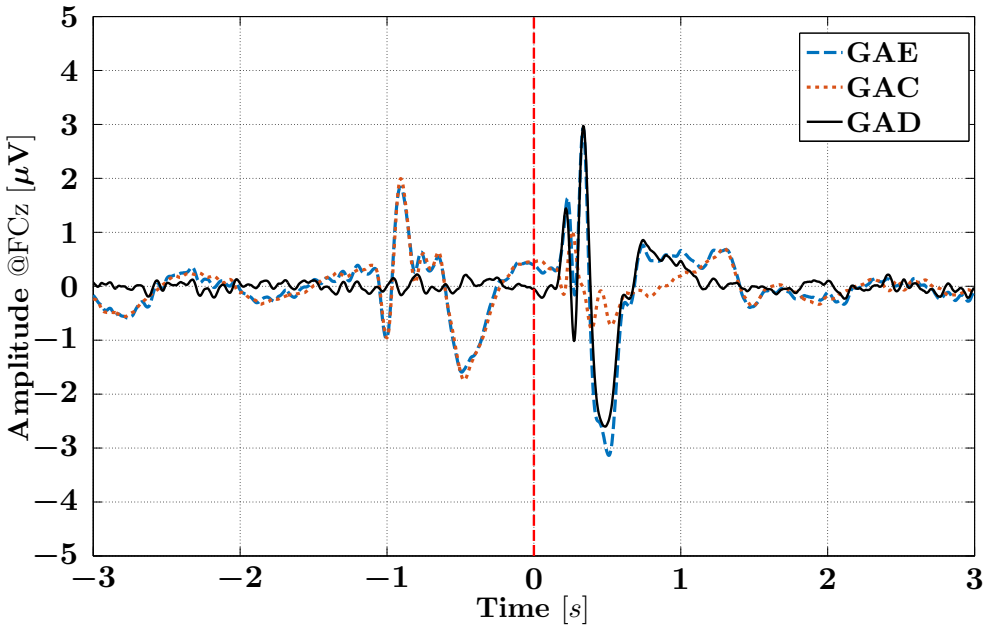


Figure A.4: The GAC, GAE and GAD waveforms were computed from the average of all subjects and recordings for experiment I

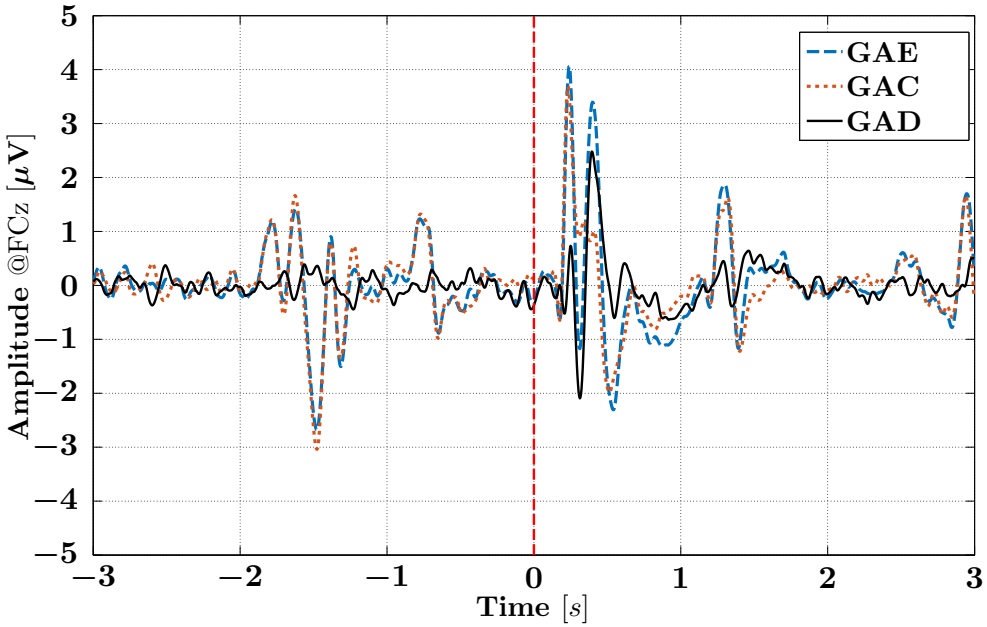


Figure A.5: The GAC, GAE and GAD waveforms were computed from the average of all subjects and recordings for experiment II

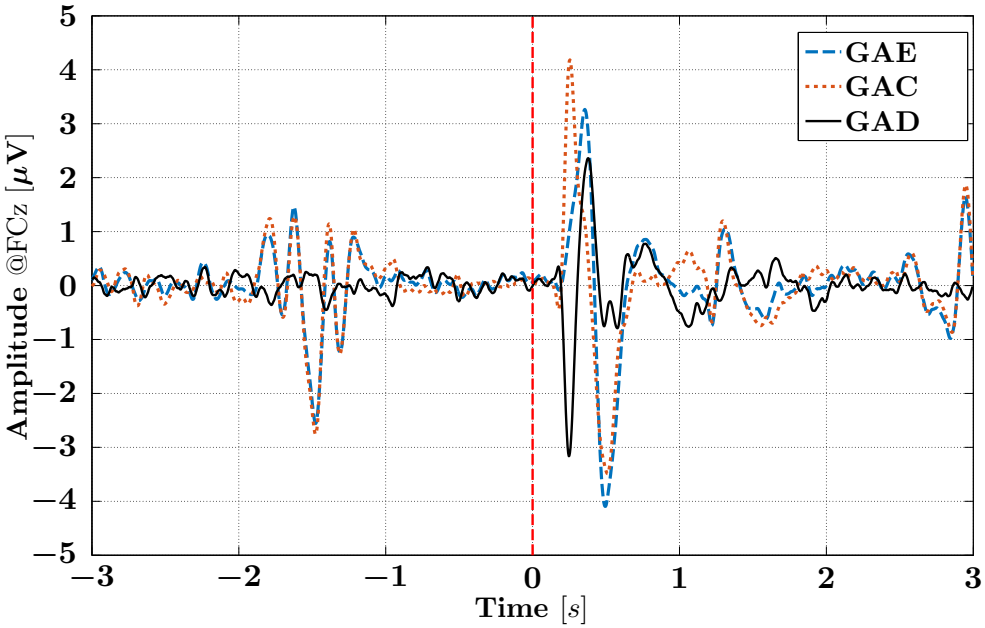


Figure A.6: The GAC, GAE and GAD waveforms were computed from the average of all subjects and recordings for experiment III

B Parameters for the Robotic Embodiment Experiments

User Input Model

(1) Forward command model:
inner opening: $\pi/4$ rad
mid opening: $5\pi/12$ rad
inner opening score=1
mid opening score=0.9
elsewhere score=0.7
(2) Turn right model:
left area score=1
right area score=0.7
(3) Turn left model
left area score=0.7
right area score=1
(4) $d_{\text{subgoal}} = 1.5$ m

User Gaze Model

(1) $\epsilon_0 = \delta_0 = 0.1$ rad
(2) $\epsilon_1 = \delta_1 = 0.2$ rad
(3) $a = 1$, $b = 0.9$, $c = 0.8$

Plans Evolution Model

(1) $\alpha = 0.8$
(2) $\beta_1 = 0.8$, $\beta_2 = 10$ and $\beta_3 = 0.2$. This renders ${}^m s_k \in [0.2, 1] \forall m, k$

Indirect Measures with Shared Control

(1) $\delta d_{\text{def}} = \pm 0.25$ m/command
(2) $\delta \theta_{\text{def}} = \pm \pi/12$ rad/command
(3) For modulating translational commands, $s_{\text{thresh}} = 0.3$
(4) For modulating rotational commands, $s_{\text{thresh}} = 0.8$

Bibliography

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

[2] Jean-Jacques Vidal. Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1):157–80, 1973.

[3] Effie Teo, Wei-Peng and Chew. Is motor-imagery brain-computer interface feasible in stroke rehabilitation? *PM&R*, 6(8):723–728, 2014.

[4] José del R Millán, Rüdiger Rupp, Gernot R Müller-Putz, Roderick Murray-Smith, Claudio Giugliemma, Michael Tangermann, Carmen Vidaurre, Febo Cincotti, Andrea Kübler, Robert Leeb, Christa Neuper, Klaus Robert Müller, and Donatella Mattia. Combining Brain-Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges. *Frontiers in neuroscience*, 4:1–33, jan 2010.

[5] United Nations. Magnitude and speed of population ageing, 2002.

[6] GertWillem R. B. E. Römer, Harry J. A. Stuyt, and Albér Peters. Cost-savings and economic benefits due to the Assistive Robotic Manipulator (ARM). In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005*, pages 201–204, 2005.

[7] Surjo R Soekadar, Niels Birbaumer, and Leonardo G Cohen. Brain-computer-interfaces in the rehabilitation of stroke and neurotrauma. In *Systems Neuroscience and Rehabilitation*, pages 3–18. Springer, 2011.

[8] Jan B.F. Van Erp, Fabien Lotte, and Michael Tangermann. Brain-computer interfaces: beyond medical applications. *Computer*, 45(4):26–34, 2012.

[9] Alexander Y. Kaplan, Sergei L. Shishkin, Ilya P. Ganin, Ivan a. Basyul, and Alexander Y. Zhigalov. Adapting the P300-based brain-computer interface for gaming: a review. *Computational Intelligence and AI in Games, IEEE Transactions on*, 5(2):141–149, jun 2013.

[10] Daniel Perez-Marcos, Mel Slater, and Maria V Sanchez-Vives. Inducing a virtual hand ownership illusion through a brain-computer interface. *Neuroreport*, 20(6):589–594, 2009.

[11] Joseph Lee. Brain-computer interfaces and dualism: a problem of brain, mind, and body. *AI & SOCIETY*, 31(1):29–40, 2016.

[12] Anne Marie Brouwer, Jan Van Erp, Dirk Heylen, Ole Jensen, and Mannes Poel. Effortless passive BCIs for healthy users. In Constantine Stephanidis, , and Margherita Antona, editors, *Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion: 7th International Conference, UAHCI 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July*

21-26, 2013, Proceedings, Part I, pages 615–622. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[13] Christoph S Herrmann. Human EEG responses to 1-100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, 137(3-4):346–353, apr 2001.

[14] G Pfurtscheller and F H Lopes. Event-related EEG / MEG synchronization and desynchronization : basic principles. *Clinical Neurophysiology*, 110(11):1842–1857, 1999.

[15] Tom Carlson and José del R Millán. Brain-controlled wheelchairs: A robotic architecture. *IEEE Robotics and Automation Magazine*, 20(1):65–73, mar 2013.

[16] Luca Tonin, Tom Carlson, Robert Leeb, and José del R Millán. Brain-controlled telepresence robot by motor-disabled people. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 4227–4230, jan 2011.

[17] Kapil D. Katyal, Matthew S. Johannes, Spencer Kellis, Tyson Aflalo, Christian Klaes, Timothy G. McGee, Matthew P. Para, Ying Shi, Brian Lee, Kelsie Pejsa, Charles Liu, Brock A. Wester, Francesco Tenore, James D. Beaty, Alan D. Ravitz, Richard A. Andersen, and Michael P. McLoughlin. A collaborative BCI approach to autonomous control of a prosthetic limb system. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 1479–1482, 2014.

[18] Emily M Mugler, Carolin A Ruf, Sebastian Halder, Michael Bensch, and Andrea Kubler. Design and implementation of a P300-based brain-computer interface for controlling an internet browser. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(6):599–609, 2010.

[19] Peter Brunner, Anthony L. Ritaccio, Joseph F. Emrich, Horst Bischof, and Gerwin Schalk. Rapid communication with a "P300" matrix speller using electrocorticographic signals (ECoG). *Frontiers in Neuroscience*, 5:5, 2011.

[20] Bertrand Rivet, Antoine Souloumiac, Guillaume Gibert, and Virginie Attina. "P300 speller" Brain-Computer Interface: Enhancement of P300 evoked potential by spatial filters. *16th European Signal Processing Conference*, pages 1–5, 2008.

[21] Ori Cohen, Sébastien Druon, Sébastien Lengagne, Avi Mendelsohn, Rafael Malach, Abderrahmane Kheddar, and Doron Friedman. fMRI based robotic embodiment: a pilot study. In *IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 314–319, 2012.

[22] Damien Petit, Pierre Gergondet, and Abderrahmane Kheddar Andrea Cherubini. An integrated framework for humanoid embodiment with a BCI. In *Robotics and Automation (ICRA), 2015 IEEE International Conference*, pages 2882–2887, 2015.

[23] M. Alimardani, S. Nishio, and H. Ishiguro. BCI-teleoperated androids; a study of embodiment and its effect on motor imagery learning. In *2015 IEEE 19th International Conference on Intelligent Engineering Systems (INES)*, pages 347–352, 2015.

[24] Jürgen Mellinger, Gerwin Schalk, Christoph Braun, Hubert Preissl, Wolfgang Rosenstiel, Niels Birbaumer, and Andrea Kübler. An MEG-based Brain-Computer Interface (BCI). *Neuroimage*, 36(3):581–593, 2007.

[25] Ranganatha Sitaram, Andrea Caria, Ralf Veit, Tilman Gaber, Giuseppina Rota, Andrea Kuebler, and Niels Birbaumer. FMRI brain-computer interface: A tool for neuroscientific research and treatment. *Computational intelligence and neuroscience*, 2007, 2007.

[26] Olaf Blanke. Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, 13(8):556–571, 2012.

[27] Saeid Sanei, Saideh Ferdowsi, Kianoush Nazarpour, and Andrzej Cichocki. Advances in electroencephalography signal processing. *Signal Processing Magazine, IEEE*, 30(1):170–176, 2013.

[28] Robert W Williams and Karl Herrup. The Control Of Neuron Number. *Annual review of neuroscience*, 11(1):423–453, 1998.

[29] Gerard J. Tortora and Bryan H. Derrickson. *Principles Of Anatomy And Physiology*. John Wiley & Sons, 2008.

[30] Dale Purves, George J Augustine, David Fitzpatrick, Lawrence C Katz, Anthony-Samuel LaMantia, James O McNamara, and S Mark Williams. Synaptic Transmission. *Neurosceince*, 2001.

[31] Jonathan Wolpaw and Elizabeth Winter Wolpaw. *Brain-computer interfaces: principles and practice*. OUP USA, 2012.

[32] Michal Teplan. Fundamentals of EEG measurement. *Measurement science review*, 2(2):1–11, 2002.

[33] Matti Hämäläinen, Riitta Hari, Risto J. Ilmoniemi, Jukka Knuutila, and Olli V. Lounasmaa. Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 1993.

[34] György Buzsáki. *Rhythms of the Brain*. 2006.

[35] Gina R. Kuperberg. *Electroencephalography, Event-Related Potentials, and Magnetoencephalography*. American Psychiatric Pub, 1 edition, 2008.

[36] Paul H. Schimpf, Ceon Ramon, and Jens Haueisen. Dipole models for the EEG and MEG. *IEEE Transactions on Biomedical Engineering*, 49(5):409–418, 2002.

[37] Peyman Adjamian. The application of electro- and magneto-encephalography in tinnitus research-methods and interpretations. *Frontiers in Neurology*, 5:1–24, 2014.

[38] S. A. Bunge and I. Kahn. Cognition: An Overview of Neuroimaging Techniques. *Encyclopedia of Neuroscience*, 2:1063–1067, 2009.

[39] Antti Lahtinen, Päivi Laarne, Hannu Eskola, and Jaakko Malmivuo. Approximation of cortical EEG from scalp recordings. In *10th Nordic-Baltic Conference on Biomedical Engineering, 1st International Conference on Bioelectromagnetism*, pages 259–260, 1996.

[40] A. Ebner, G. Sciarretta, C. M. Epstein, and M. Nuwer. EEG instrumentation. *Electroencephalography and clinical neurophysiology*, 52:7–10, 1999.

[41] Stefan Haufe. *Towards EEG source connectivity analysis*. PhD thesis, Technische Universität Berlin, 2012.

[42] Steven J. Luck and Emily S. Kappenman. *The Oxford handbook of event-related potential components*. Oxford university press, 2011.

[43] Francesco Di Russo, Antígona Martínez, Martin I. Sereno, Sabrina Pitzalis, and Steven A. Hillyard. Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, 15(2):95–111, 2002.

[44] Steven J. Luck. *An introduction to the event-related potential technique*. 2005.

[45] Pouya Ahmadian, Stefano Cagnoni, and Luca Ascari. How capable is non-invasive EEG data of predicting the next movement ? A mini review. *Frontiers in Human Neuroscience*, 7:1–7, 2013.

[46] Garipelli Gangadhar, Ricardo Chavarriaga, and José Del R Millán. Fast recognition of anticipation-related potentials. *IEEE transactions on bio-medical engineering*, 56(4):1257–60, apr 2009.

[47] Saeid Sanei and Jonathon. A. Chambers. *EEG Signal Processing*. John Wiley & Sons, 2013.

[48] David A. Paskewitz and Martin T. Orne. Visual Effects on Alpha Feedback Training. *Science*, 181(4097):360–363, 1973.

[49] Xiaoxuan Jia and Adam Kohn. Gamma rhythms in the brain. *PLoS Biology*, 9(4):2–4, 2011.

[50] Toshimitsu Musha and Mitsuaki Yamamoto. 1/f Fluctuations in Biological Systems. In *The 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 6, pages 2692–2697, 1997.

[51] Richard F. Voss and John Clarke. Music from 1/f noise1/F Noise in Music and Speech:. *The Journal of the Acoustical Society of America*, 63(1):258—-263, 1978.

[52] Peter C.M. Vijn, Bob W. van Dijk, and Henk Spekreijse. Visual stimulation reduces EEG activity in man. *Brain research*, 550(1):49–53, 1991.

[53] Shin'Ichiro Kanoh, Ko Ichiro Miyamoto, and Tatsuo Yoshinobu. Generation of spatial filters by ICA for detecting motor-related oscillatory EEG. In *The 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 1703–1706, 2012.

[54] Scott Makeig and Julie Onton. ERP Features and EEG Dynamics : An ICA Perspective. *Oxford handbook of event-related potential components*, pages 1–58, 2011.

[55] John Makhoul. Linear Prediction: A Tutorial Review. *Proceedings of the IEEE*, 63(4):561–580, 1975.

[56] Sanford Weisberg. *Applied Linear regression*. John Wiley & Sons, 3 edition, 2005.

[57] Hans-Werner Jansen, Theo Gaaser, and Franke Juergen. Fitting Autoregressive Models to EEG Time Series: An Empirical Comparison of Estimates of the Order. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(1):143–150, 1985.

[58] Scott Makeig. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, 86(4):283–293, 1993.

[59] Arnaud Delorme and Scott Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[60] Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1):1–30, 2005.

[61] Christian Andreas Kothe and Scott Makeig. BCILAB: a platform for brain-computer interface development. *Journal of neural engineering*, 10(5), 2013.

[62] Isabelle Guyon and Andre Elisseeff. An Introduction to Variable and Feature Selection. *The Journal of Machine Learning Research*, 3(3):1157–1182, 2003.

[63] Iñaki Iturrate, Luis Montesano, and Javier Minguez. Task-dependent signal variations in EEG error-related potentials for brain-computer interfaces. *Journal of neural engineering*, 10(2):1–13, 2013.

[64] Yasuhiro X. Kato, Tomoko Yonemura, Kazuyuki Samejima, Taro Maeda, and Hideyuki Ando. Development of a BCI master switch based on single-trial detection of contingent negative variation related potentials. In *The 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 4629–4632, 2011.

[65] Fei Guo, Bo Hong, Xiaorong Gao, and Shangkai Gao. A brain-computer interface using motion-onset visual evoked potential. *Journal of Neural Engineeringeural*, 5(4):477–85, 2008.

[66] Rui Zhang, Peng Xu, Rui Chen, Teng Ma, Xulin Lv, Fali Li, Peiyang Li, Tiejun Liu, and Dezhong Yao. An Adaptive Motion-Onset VEP-Based Brain-Computer Interface. *IEEE Transactions on Autonomous Mental Development*, 7(4):349–356, 2015.

[67] Yijun Wang, Bo Hong, Xiaorong Gao, and Shangkai Gao. Implementation of a brain-computer interface based on three states of motor imagery. In *The 29th Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pages 5059–5062, 2007.

[68] Sheng Ge, Ruimin Wang, and Dongchuan Yu. Classification of four-class motor imagery employing single-channel electroencephalography. *PloS one*, 9(6):1–7, 2014.

[69] G. Reshmi and A. Amal. Design of a BCI system for piloting a wheelchair using five class MI Based EEG. In *Advances in Computing and Communications (ICACC), 2013 Third International Conference on*, pages 25–28, 2013.

[70] Weibo Yi, Shuang Qiu, Hongzhi Qi, Lixin Zhang, Baikun Wan, and Dong Ming. EEG feature comparison and classification of simple and compound limb motor imagery. *Journal of neuroengineering and rehabilitation*, 10(1):1–12, 2013.

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

[72] Johannes Höhne, Konrad Krenzlin, Sven Dähne, and Michael Tangermann. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *Journal of Neural Engineering*, 9(4):1–12, 2012.

[73] Anne Marie Brouwer and Jan B F van Erp. A tactile P300 brain-computer interface. *Frontiers in Neuroscience*, 4:1–11, 2010.

[74] Sangtae Ahn, Minkyu Ahn, Hohyun Cho, and Sung Chan Jun. Achieving a hybrid brain-computer interface with tactile selective attention and motor imagery. *Journal of neural engineering*, 11(6):1–14, 2014.

[75] Ken Nakayama and Manfred Mackeben. Steady state visual evoked potentials in the alert primate. *Vision Research*, 22(10):1261–1271, 1982.

[76] Athanasios P Liavas, George V Moustakides, Günter Henning, Emmanuil Z. Psarakis, and Peter Husar. A periodogram-based method for the detection of steady-state visually evoked potentials. *IEEE Transactions on Biomedical Engineering*, 45(2):242–248, 1998.

[77] Almudena Capilla, Paula Pazo-Alvarez, Alvaro Darriba, Pablo Campo, and Joachim Gross. Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PloS one*, 6(1):1–15, 2011.

[78] S. T. Morgan, J. C. Hansen, and S. A. Hillyard. Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences*, 93(10):4770–4774, may 1996.

[79] Xiaorong Gao, Dingfeng Xu, Ming Cheng, and Shangkai Gao. A BCI-based environmental controller for the motion-disabled. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):137–140, 2003.

[80] H. Dyball. Tapping the brain. *Electronics Letters*, 46(15):1042, 2010.

[81] Ivan Volosyak, Hubert Cecotti, and Axel Gräser. Optimal visual stimuli on LCD screens for SSVEP based brain-computer interfaces. In *2009 4th International IEEE/EMBS Conference on Neural Engineering*, pages 447–450, 2009.

[82] Erwei Yin, Zongtan Zhou, Jun Jiang, Yang Yu, and Dewen Hu. A Dynamically Optimized SSVEP Brain-Computer Interface (BCI) Speller. *IEEE Transactions on Biomedical Engineering*, 62(6):1447–1456, 2015.

[83] Ricardo Parafita, Gabriel Pires, Urbano Nunes, and Miguel Castelo-Branco. A spacecraft game controlled with a brain-computer interface using SSVEP with phase tagging. In *Serious Games and Applications for Health (SeGAH), 2013 IEEE 2nd International Conference on*, pages 1–6, 2013.

[84] Robert Prueckl and Christoph Guger. A brain-computer interface based on steady state visual evoked potentials for controlling a robot. *International Work-Conference on Artificial Neural Networks*, pages 690–697, 2009.

[85] Xavier Navarro, Sebastian Campion, Fabrizio De Vico Fallani, Pierre Pouget, Thomas Similowski, Mathieu Raux, and Mario Chavez. Steady state visual evoked potentials-based patient interface under breathing constraints. In *International IEEE/EMBS Conference on Neural Engineering, NER*, pages 138–141.

[86] Bonkon Koo, Hwan Gon Lee, Yunjun Nam, and Seungjin Choi. Immersive BCI with SSVEP in VR head-mounted display. In *The 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 1103–1106, 2015.

[87] D Regan. Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and clinical neurophysiology*, 20(3):238–248, 1966.

[88] Fahrettin F. Gonen and Gleb V. Tcheslavski. Techniques to assess stationarity and gaussianity of EEG: An overview. *International Journal of Bioautomation*, 16(2):135–142, 2012.

[89] Danhua Zhu, Jordi Bieger, Gary Garcia Molina, and Ronald M. Aarts. A survey of stimulation methods used in SSVEP-based BCIs. *Computational Intelligence and Neuroscience*, 2010:1–12, 2010.

[90] Francesco Di Russo, Sabrina Pitzalis, Teresa Aprile, Grazia Spitoni, Fabiana Patria, Alessandra Stella, Donatella Spinelli, and Steven A. Hillyard. Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential. *Human Brain Mapping*, 28(4):323–334, 2007.

[91] Fei Teng, Yixin Chen, Aik Min Choong, Scott Gustafson, Christopher Reichley, Pamela Lawhead, and Dwight Waddell. Square or sine: Finding a waveform with high success rate of eliciting SSVEP. *Computational Intelligence and Neuroscience*, 2011:1–5, 2011.

[92] F. W. Campbell and L. Maffei. Electrophysiological evidence for the existence of orientation and size detectors in the human visual system. *The Journal of Physiology*, 207(3):635–652, 1970.

[93] Yu-Te Wang, Yijun Wang, Chung Kuan Cheng, and Tzyy Ping Jung. Developing stimulus presentation on mobile devices for a truly portable SSVEP-based BCI. *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5271–5274, 2013.

[94] Chuan Jia, Honglai Xu, Bo Hong, Xiaorong Gao, Zhiguang Zhang, and Shangkai Gao. A Human Computer Interface Using SSVEP-Based BCI Technology. *Foundations of Augmented Cognition*, pages 113–119, 2007.

[95] Gido Hakvoort, Boris Reuderink, and Michel Obbink. Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system. Technical Report Technical Report TR-CTIT-11-03, Centre for Telematics and Information Technology University of Twente, Enschede, 2011.

[96] Ming Cheng, Xiaorong Gao, Shangkai Gao, and Dingfeng Xu. Design and implementation of a brain-computer interface with high transfer rates. *IEEE Transactions on Biomedical Engineering*, 49(10):1181–1186, 2002.

[97] Kian B Ng, Ross Cunnington, and Andrew P Bradley. Enhancing the classification accuracy of steady-state visual evoked potential-based Brain-Computer interface using component synchrony measure. *IEEE world congress on computational intelligence*, pages 1–6, 2012.

[98] Jonathan D. Victor and Joelle Mast. A new statistic for steady-state evoked potentials. *Electroencephalography and clinical neurophysiology*, 78(5):378–388, 1991.

[99] Matthias M. Müller, Terence W. Picton, Pedro Valdes-Sosa, Jorge Riera, Wolfgang A. Teder-Sälejärvi, and Steven A. Hillyard. Effects of spatial selective attention on the steady-state visual evoked potential in the 20-28 Hz range. *Cognitive brain research*, 6(4):249–261, apr 1998.

[100] Simon P. Kelly, Edmund C. Lalor, Ciarán Finucane, Gary McDarby, and Richard B Reilly. Visual spatial attention control in an independent brain-computer interface. *IEEE transactions on biomedical engineering*, 52(9):1588–1596, sep 2005.

[101] Thomas Meigen and Michael Bach. On the statistical significance of electrophysiological steady-state responses. *Documenta Ophthalmologica*, 98(3):207–232, 2000.

[102] Ola Friman, Ivan Volosyak, and Axel Gräser. Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE transactions on biomedical engineering*, 54(4):742–750, apr 2007.

[103] Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency recognition based on canonical correlation analysis for SSVEP-Based BCIs. *IEEE Transactions on Biomedical Engineering*, 53(12):2610–2614, jun 2006.

[104] Niya Wang, Tianyi Qian, Qing Zhuo, and Xiaorong Gao. Discrimination between idle and work states in BCI based on SSVEP. In *The 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010*, volume 4, pages 355–358, 2010.

[105] Yangsong Zhang, Peng Xu, Kaiwen Cheng, and Dezhong Yao. Multivariate synchronization index for frequency recognition of SSVEP-based brain-computer interface. *Journal of neuroscience methods*, 221:32–40, jan 2014.

[106] Carlos E. Davila, Richard Srebro, and Ibrahim A. Ghaleb. Optimal detection of visual evoked potentials. *IEEE transactions on bio-medical engineering*, 45(6):800–803, 1998.

[107] Guangyu Bin, Xiaorong Gao, Zheng Yan, Bo Hong, and Shangkai Gao. An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. *Journal of neural engineering*, 6(4):1–6, aug 2009.

[108] Hongya Ge, Ivars P. Kirsteins, and Xiaoli Wang. Does canonical correlation analysis provide reliable information on data correlation in array processing? In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2113–2116, 2009.

[109] Harold Hotelling. Relations Between Two Sets of Variates. *Biometrika*, 28(3/4):321–377, 1936.

[110] David R. Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: an overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, dec 2004.

[111] Louis L. Scharf and John K. Thomas. Wiener filters in canonical coordinates for transform coding, filtering, and quantizing. *IEEE Transactions on Signal Processing*, 46(3):647–654, mar 1998.

[112] Ake Björck and Gene H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973.

[113] Cristian Carmeli. *Assessing cooperative behavior in dynamical networks with applications to brain data*. Phd thesis, École Polytechnique Fédérale de Lausanne, 2006.

[114] Yangsong Zhang, Peng Xu, Tiejun Liu, Jun Hu, Rui Zhang, and Dezhong Yao. Multiple frequencies sequential coding for SSVEP-based brain-computer interface. *PloS one*, 7(3):1–9, 2012.

[115] Christoph Kapeller, Kyousuke Kamada, Hiroshi Ogawa, Robert Prueckl, Josef Scharinger, and Christoph Guger. An electrocorticographic BCI using code-based VEP for control in video applications: a single-subject study. *Frontiers in systems neuroscience*, 8, 2014.

[116] Christoph Kapeller, Christoph Hintermüller, Mohammad Abu-Alqumsan, Robert Pruckl, Angelika Peer, and Christoph Guger. A BCI using VEP for continuous control of a mobile robot. In *The 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 5254–5257, 2013.

[117] Charmaine Demanuele, Christopher J James, and Edmund Js Sonuga-Barke. Distinguishing low frequency oscillations within the 1/f spectral behaviour of electromagnetic brain signals. *Behavioral and brain functions*, 3(1):1–14, 2007.

[118] Roger Bakeman. Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3):379–384, 2005.

[119] Wenya Nan, Chi Man Wong, Boyu Wang, Feng Wan, Peng Un Mak, Pui In Mak, and Mang I Vai. A comparison of minimum energy combination and canonical correlation analysis for SSVEP detection. In *2011 5th International IEEE/EMBS Conference on Neural Engineering*, pages 469–472, Cancun, apr 2011. IEEE.

[120] Shozo Tobimatsu and Motohiro Kato. The effect of binocular stimulation on each component of transient and steady-state VEPs. *Electroencephalography and Clinical Neurophysiology - Evoked Potentials*, 100(3):177–183, 1996.

[121] A. Materka and M. Byczuk. Alternate half-field stimulation technique for SSVEP-based brain-computer interfaces. *Electronics Letters*, 42(6):321–322, 2006.

[122] Steve Suter, Penelope S Suter, Denise T Perrier, Kerrie L Parker, James A Fox, and Jacqueline S Roessler. Differentiation of VEP intermodulation and second harmonic components by dichoptic, monocular, and binocular stimulation. *Visual neuroscience*, 13(6):1157–1166, 1996.

[123] Vitontonio Bevilacqua, Giacomo Tattoli, Domenico Buongiorno, Claudio Loconsole, Daniele Leonardis, Antonio Frisoli, and Massimo Bergamasco. A novel BCI-SSVEP based approach for control of walking in Virtual Environment using a Convolutional Neural Network. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 4121–4128, 2014.

[124] Alex Kreilinger, Christa Neuper, Gert Pfurtscheller, and Gernot R. Müller-Putz. Implementation of error detection into the Graz-Brain-Computer Interface, the interaction error potential. In *9th European Conf. for the Advancement of Assistive Technology*, pages 1–5, Florence, Italy, 2009.

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

[126] Gert-Jan Munneke, Tanja S. Nap, Eveline E. Schippers, and Michael X Cohen. A statistical comparison of EEG time- and time-frequency-domain representations of error processing. *Brain Research*, 1618:222–230, 2015.

[127] Hein T van Schie, Rogier B Mars, Michael G H Coles, and Harold Bekkering. Modulation of activity in medial frontal and motor cortices during error observation. *Nature neuroscience*, 7(5):549–54, may 2004.

[128] Wolfgang HR Miltner, Jens Brauer, Holger Hecht, Ralf Trippe, and Michael G H Coles. Parallel brain activity for self-generated and observed errors. In *Errors, conflicts, and the brain: current opinions on performance monitoring*, pages 124–129. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences., 2004.

[129] Su Kyoung Kim and Elsa Andrea Kirchner. Classifier Transferability in the Detection of Error Related Potentials from Observation to Interaction. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3360–3365, 2013.

[130] Wolfgang H. R. Miltner, Christoph H. Braun, and Michael G. H. Coles. Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience*, 9(6):788–798, 1997.

[131] Nick Yeung, Clay B. Holroyd, and Jonathan D. Cohen. ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex*, 15(5):535–544, 2005.

[132] Sandra Rousseau, Jutten Christian, and Congedo Marco. Time window selection for improving error-related potential detection. In *5th International Conference on Neural Computation Theory and Applications*, 2012.

[133] Gerwin Schalk, Jonathn R. Wolpaw, Dennis J. McFarland, and Gert Pfurtscheller. EEG-based communication: presence of an error potential. *Clinical neurophysiology*, 111(12):2138–2144, dec 2000.

[134] Pierre W. Ferrez and José Del R. Millán. Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE transactions on biomedical engineering*, 55(3):923–929, mar 2008.

[135] Michael Falkenstein, Jörg Hoormann, Stefan Christ, and Joachim Hohnsbein. ERP components on reaction errors and their functional significance: a tutorial. *Biological psychology*, 51(2):87–107, jan 2000.

[136] Phan Luu, Don M. Tucker, and Scott Makeig. Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, 115(8):1821–1835, 2004.

[137] Cameron S. Carter, Todd S. Braver, Deanna M. Barch, Mathhew M. Botvinick, Douglas Noll, and Johathan D. Cohen. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280(5364):747–749, may 1998.

[138] Iñaki Iturrate, Luis Montesano, and Javier Minguez. Single trial recognition of error-related potentials during observation of robot operation. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 4181–4, jan 2010.

[139] Jonathan R. Wolpaw, Herbert Ramoser, Dennis J. McFarland, and Gert Pfurtscheller. EEG-based communication: improved accuracy by response verification. *IEEE transactions on rehabilitation engineering*, 6(3):326–333, sep 1998.

[140] Adrien Combaz, Nikolay Chumerin, Nikolay V. Manyakov, Arne Robben, Johan A. K. Suykens, and Marc M. Van Hulle. Towards the detection of error-related potentials and its integration in the context of a P300 speller brain-computer interface. *Neurocomputing*, 80:73–82, 2012.

[141] Martin Spüler, Michael Bensch, Sonja Kleih, Wolfgang Rosenstiel, Martin Bogdan, and Andrea Kübler. 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, jul 2012.

[142] Alberto Llera, Marcel A. J. Gerven, Ole G{\'o}mez, Vicen{\c{c}} Jensen, and Hilbert J. Kappen. On the use of interaction error potentials for adaptive brain computer interfaces. *Neural Networks*, 24(10):1120–1127, 2011.

[143] Nico M. Schmidt, Benjamin Blankertz, and Matthias S. Treder. Online detection of error-related potentials boosts the performance of mental typewriters. *BMC Neuroscience*, 13(1):1, 2012.

[144] Pierre W. Ferrez. *Error-related EEG potentials in brain-computer interfaces*. Phd thesis, École Polytechnique Fédérale De Lausanne, 2007.

[145] G. Visconti, B. Dal Seno, M. Matteucci, and L. Mainardi. Automatic Recognition of Error Potentials in a P300-Based Brain-Computer Interface, 2008.

[146] Perrin Magaurx, Maby Emmanuel, Daligault Sébastien, Bertrand Olivier, and Mattout Jérémie. Objective and subjective evaluation of online error correction during P300-based spelling. *Advances in Human-Computer Interaction*, 2012, 2012.

[147] Hiromu Takahashi, Tomohiro Yoshikawa, and Takeshi Furuhashi. A study on combination of reliability-based automatic repeat request with error potential-based error correction for improving P300 speller performance. In *International Conference on Neural Information Processing*, pages 50–57, 2010.

[148] Iñaki Iturrate, Luis Montesano, Ricardo Chavarriaga, José Del R. Millán, and Javier Minguez. Minimizing calibration time using inter-subject information of single-trial recognition of error potentials in brain-computer interfaces. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6369–6372, 2011.

[149] Iñaki Iturrate, Luis Montesano, Ricardo Chavarriaga, José Del R. Millán, and Javier Minguez. Spatio-temporal filtering for EEG error related potentials. In *5th Int Brain-Computer Interface Conference*, pages 1–4, 2011.

[150] Johannes Höhne, Konrad Krenzlin, Sven Dähne, and Michael Tangermann. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *Journal of Neural Engineering*, 9(4):045003, 2012.

[151] Christoph Hintermüller, Christoph Kapeller, Günter Edlinger, and Christoph Guger. BCI Integration : Application Interfaces. In Dr. Reza Fazel-Rezai, editor, *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, pages 21–41. InTech, 2013.

[152] Christoph Guger, Shahab Daban, Eric Sellers, Clemens Holzner, Gunther Krausz, Roberta Carabalona, Furio Gramatica, and Guenter Edlinger. How many people are able to control a P300-based brain-computer interface (BCI)? *Neuroscience Letters*, 462(1):94–98, 2009.

[153] Ricardo Chavarriaga, Aleksander Sobolewski, and José del R Millán. Errare machinale est: The use of error-related potentials in brain-machine interfaces. *Frontiers in Neuroscience*, 8(208), 2014.

[154] Saul Greenberg, John J. Darragh, David Maulsby, and Ian H. Witten. Predictive interfaces: What will they think of next? *Extra-ordinary human-computer interaction: interfaces for users with disabilities*, pages 103–140, 1995.

[155] Thomas B. Sheridan and William L. Verplank. Human and Computer Control of Undersea Teleoperators. Technical report, DTIC Document, 1978.

[156] Mica R. Endsley and David B. Kaber. Level of automation eŒ ects on performance , situation. *Ergonomics*, 42(3):462–492, 1999.

[157] Eric Demeester. *User-adapted plan recognition and shared control for wheelchair driver assistance under uncertainty.* Phd thesis, Katholieke Universiteit Lueven, Belgium, 2007.

[158] Tom Carlson and Yiannis Demiris. Human-wheelchair collaboration through prediction of intention and adaptive assistance. In *IEEE International Conference on Robotics and Automation*, pages 3926–3931, 2008.

[159] Aditya Goil, Matthew Derry, and Brenna D. Argall. Using machine learning to blend human and robot controls for assisted wheelchair navigation. In *IEEE International Conference on Rehabilitation Robotics*, pages 1–6, 2013.

[160] Charles F. Schmidt, N. S. Sridharan, and John L. Goodson. The plan recognition problem: an intersection of psychology and artificial intelligence. *Artificial Intelligence*, 11(1-2):45–83, 1978.

[161] Chris L. Baker, Rebecca Saxe, and Joshua B. Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–49, dec 2009.

[162] Henry A. Kautz. A formal theory of plan recognition and its implementation. In Josh D. Brachman, Ronald J. and Allen, James F. and Kautz, Henry A. and Pelavin, Richard N. and Tenenberg, editor, *Reasoning about plans*, pages 69–124. Morgan Kaufmann Publishers Inc., 1991.

[163] How to do anything, www.wikihow.com.

[164] David W Albrecht, Ingrid Zukerman, and Ann E. Nicholson. Bayesian models for keyhole plan recognition in an adventure game. *User modeling and user-adapted interaction*, 8(1-2):5–47, 1998.

[165] Taro Kanno, Keiichi Nakata, and Kazuo Furuta. A method for team intention inference. *International Journal of Human-Computer Studies*, 58(4):393–413, 2003.

[166] Karen E. Lochbaum. An algorithm for plan recognition in collaborative discourse. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, pages 33–38, 1991.

[167] Sviatoslav Braynov. Adversarial planning and plan recognition: Two sides of the same coin. In *Secure Knowledge Management Workshop*, 2006.

[168] Eugene Charniak and Robert P Goldman. A Bayesian model of plan recognition. *Artificial Intelligence*, 64(1):53–79, 1993.

[169] Marcelo Gabriel Armentano and Analía Amandi. Plan recognition for interface agents. *Artificial Intelligence Review*, 28(2):131–162, feb 2009.

[170] Xavier Perrin, Francis Colas, Cédric Pradalier, Roland Siegwart, Ricardo Chavarriaga, and José Del R Millán. Learning user habits for semi-autonomous navigation using low throughput interfaces. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 1–6, 2011.

[171] Eric Demeester, Hüntemann Alexander, José del R Millán, and Hendrik Van Brussel. Bayesian Plan Recognition for Brain-Computer Interfaces. In *IEEE International Conference on Robotics and Automation*, pages 653–658, 2009.

[172] Eric Demeester, Alexander Hüntemann, Dirk Vanhooydonck, Gerolf Vanacker, Hendrik Van Brussel, and Marnix Nuttin. User-adapted plan recognition and user-adapted shared control: A Bayesian approach to semi-autonomous wheelchair driving. *Autonomous Robots*, 24:193–211, 2008.

[173] Eric Demeester, Alexander Hüntemann, Dirk Vanhooydonck, Gerolf Vanacker, Alexandra Degeest, Hendrik Van Brussel, and Marnix Nuttin. Bayesian estimation of wheelchair driver intents: Modeling intents as geometric paths tracked by the driver. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5775–5780, 2006.

[174] J B Pelz and R Canosa. Oculomotor behavior and perceptual strategies in complex tasks. *Vision research*, 41(25):3587–96, 2001.

[175] Han The Anh and Luis Moniz Pereira. Corpus-Based Incremental Intention Recognition via Bayesian Network Model Construction. In *Proceedings of the First Workshop on Goal, Activity and Plan Recognition*, pages 1–8, 2011.

[176] Marcelo Gabriel Armentano and Analía Amandi. Goal Recognition with Variable-Order Markov Models. In *International Joint Conference on Artificial Intelligence*, pages 1635–1640, 2009.

[177] Nate Blaylock and James Allen. Recognizing instantiated goals using statistical methods. In *IJCAI Workshop on Modeling Others from Observations*, 2005.

[178] Morgan Quigley, Brian Gerkey, Ken Conley, Josh Faust, Tully Foote, Jeremy Leibs, Eric Berger, Rob Wheeler, and Andrew Ng. ROS: an open-source Robot Operating System. In *ICRA Workshop on Open Source Software*, 2009.

[179] Joseph J. LaViola Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, jan 2000.

[180] Dieter Fox. KLD-sampling: Adaptive particle filters. In *Advances in Neural Information Processing Systems 14*, 2001.

[181] Brian P. Gerkey, Richard T. Vaughan, and Andrew Howard. The Player / Stage Project : Tools for Multi-Robot and Distributed Sensor Systems. In *Proceedings of the International Conference on Advanced Robotics (ICAR 2003)*, pages 317–323, 2003.

[182] Zygmunt Pizlo, Yunfeng Li, Tadamasa Sawada, and Robert M Steinman. *Making a machine that sees like us*. Oxford University Press (UK), 2014.

[183] Rainer Stiefelhagen and Jie Zhu. Head orientation and gaze direction in meetings. In *Extended Abstracts on Human Factors in Computing Systems*, pages 858–859, New York, New York, USA, 2002. ACM Press.

[184] Robert J.K. Jacob. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction*, 4:151–190, 1993.

[185] Dario D. Salvucci and Joseph H. Goldberg. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium*, pages 71–78, 2000.

[186] Giorgio Grisetti, Cyrill Stachniss, and Wolfram Burgard. Improved Techniques for Grid Mapping. *IEEE Transactions on Robotics*, 23(1):34–46, 2007.

[187] Alexandra Kuznetsova, Rune H. B. Christensen, Cecile Bavay, and Per Bruun Brockhoff. Automated mixed ANOVA modeling of sensory and consumer data. *Food Quality and Preference*, 40:31–38, 2014.

[188] Torsten Hothorn, Frank Bretz, and Peter Westfall. Simultaneous Inference in General Parametric Models. *Biometrical Journal*, 50(3):346–363, 2008.

[189] Russell V Lenth. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69(1):1–33, 2016.

[190] Tom Carlson, Luca Tonin, Serafeim Perdikis, Robert Leeb, and José del R Millán. A hybrid BCI for enhanced control of a telepresence robot. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3097–100, jan 2013.

[191] Sang Hyoung Lee, Sanghoon Lee, Il Hong Suh, and Wan Kyun Chung. Learning of subgoals for goal-oriented behavior control of mobile robots. In *International Conference on Neural Information Processin*, pages 64–71, 2009.

[192] Clemens Brunner, Benjamin Blankertz, Febo Cincotti, Andrea Kübler, Donatella Mattia, Felip Miralles, Anton Nijholt, and Begonya Otal. BNCI Horizon 2020: Towards a Roadmap for Brain / Neural Computer Interaction. In *International Conference on Universal Access in Human-Computer Interaction*, pages 475–486, 2014.

[193] Guillermo Gallegos-Ayala, Adrian Furdea, Kouji Takano, Carolin A Ruf, Herta Flor, and Niels Birbaumer. Brain communication in a completely locked-in patient using bedside near-infrared spectroscopy. *Neurology*, 82(21):1930–1932, 2014.

[194] Eric W. Sellers, David B. Ryan, and Christopher K. Hauser. Noninvasive brain-computer interface enables communication after brainstem stroke. *Science translational medicine*, 6(257), 2014.

[195] D. Lesenfants, D. Habbal, Z. Lugo, M. Lebeau, P. Horki, E. Amico, C. Pokorny, F. Gómez, A. Soddu, G. Müller-Putz, S. Laureys, and Q. Noirhomme. An independent SSVEP-based brain-computer interface in locked-in syndrome. *Journal of neural engineering*, 11(3):035002, 2014.

[196] Xavier Perrin, Ricardo Chavarriaga, Francis Colas, Roland Siegwart, and José Del R. Millán. Brain-coupled interaction for semi-autonomous navigation of an assistive robot. *Robotics and Autonomous Systems*, 58(12):1246–1255, dec 2010.

[197] Inaki Iturrate, Luis Montesano, and Javier Minguez. Shared-control brain-computer interface for a two dimensional reaching task using EEG error-related potentials. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, volume 2013, pages 5258–5262, jan 2013.

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

[199] Minpeng Xu, Hongzhi Qi, Baikun Wan, Tao Yin, Zhipeng Liu, and Dong Ming. A Hybrid BCI Speller Paradigm combining P300 Potential and the SSVEP Blocking Feature. *Journal of Neural Engineering*, 10(2):026001, 2013.

[200] Brendan Z. Allison, Jing Jin, Yu Zhang, and Xingyu Wang. A four-choice hybrid P300/SSVEP BCI for improved accuracy. *Brain-Computer Interfaces*, 1(1):17–26, jan 2014.

[201] Chi Chun Lo, Tsung Yi Chien, Yu Chun Chen, Shang Ho Tsai, Wai Chi Fang, and Bor Shyh Lin. A wearable channel selection-based brain-computer interface for motor imagery detection. *Sensors*, 16(2):1–14, 2016.

[202] Hideki Ohira. Eyeblink activity in a word-naming task as a function of semantic priming and cognitive load. *Perceptual and motor skills*, 82(3):835–842, 1996.

Supervised Students' Theses

[203] Dullal Ghosh. *An adaptive brain-computer interface for navigational tasks*. Master's thesis, KTH Royal Institute of Technology, 2012.

[204] Luis Vergara. *Online recognition of error-related potentials during P300-based brain-robot interaction*. Bachelor's thesis, Technical University of Munich (TUM), 2014.

[205] Felix Ebert. *Bayesian intention and plan recognition in BCI-mediated robotic re-embodiment systems*. Master's thesis, Technical University of Munich, 2014.

[206] Vincent Bobinski. *Offline analysis of simultaneous EEG and EOG data after P300 in-place visual feedback*. Bachelor's thesis, Technical University of Munich, 2014.

[207] Sharat Embrandiri. *Online detection of steady-state visually evoked potentials using deep canonical correlation analysis*. Master's thesis, Indian Institute of Technology, Madras, 2014.

Author's Publications

[208] Georg Schroth, Robert Huitl, David Chen, Mohammad Abu-Alqumsan, Anas Al-Nuaimi, and Eckehard Steinbach. Mobile visual location recognition. *IEEE Signal Processing Magazine*, 28(4):77–89, 2011.

[209] Christoph Kapeller, Christoph Hintermueller, Mohammad Abu-Alqumsan, Robert Prueckl, Angelika Peer, and Christoph Guger. SSVEP based Brain-Computer Interface combined with video for robotic control. In *Proceedings of the Fifth International Brain-Computer Interface Meeting*, 2013.

[210] Nikolas Martens, Robert Jenke, Mohammad Abu-alqumsan, Christoph Kapeller, Christoph Hintermüller, Christoph Guger, Angelika Peer, and Martin Buss. Towards robotic re-embodiment using a Brain-and-Body-Computer Interface. In *IEEE International Conference on Intelligent Robots and Systems*, pages 5131–5132. IEEE, oct 2012.

[211] Mohammad Abu-Alqumsan and Angelika Peer. Advancing the detection of steady-state visual evoked potentials in brain-computer interfaces. *Journal of neural engineering*, 13(3):36005, 2016.

[212] Christoph Kapeller, Christoph Hintermüller, Mohammad Abu-Alqumsan, Robert Pruckl, Angelika Peer, and Christoph Guger. A BCI using VEP for continuous control of a mobile robot. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 5254–5257, 2013.

[213] Georg Schroth, Robert Huitl, Mohammad Abu-Alqumsan, Florian Schweiger, and Eckehard Steinbach. Exploiting prior knowledge in mobile visual location recognition. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 2357–2360, 2012.

[214] Mohammad Abu-alqumsan, Robert Jenke, Angelika Peer, and Martin Buss. Robotic Re-Embodiment using a Brain-Computer-Interface. In *Neuroscience*, 2012.

[215] Emmanuele Tidoni, Mohammad Abu-Alqumsan, Daniele Leonardis, Christoph Kapeller, and Gabriele Fusco. Interacting with my virtual and robotic surrogate: a study on healthy and spinal cord injured people (Abstract). In *London Virtual Social Interaction Workshop, Institute of Cognitive Neuroscience UCL*, 2014.

[216] Mohammad Abu-Alqumsan, Nikolas Martens, Robert Jenke, Christoph Kapeller, Christoph Hintermüeller, Angelika Peer, and Martin Buss. Adaptive brain computer interface - Application in robotics. In *Workshop on Human-Friendly Robotics, Twente, The Netherlands*, 2011.

[217] Mohammad Abu-Alqumsan, Felix Ebert, and Angelika Peer. Bayesian intention recognition for shared control in robotic embodiment systems. (under preparation), 2016.

[218] Mohammad Abu-Alqumsan, Christoph Kapeller, Christoph Hintermüller, Christoph Guger, and Angelika Peer. Invariance and variability in interaction error-related potentials and their consequences for classication. *Journal of Neural Engineering*, (submitted), 2016.

[219] Emmanuele Tidoni, Mohammad Abu-Alqumsan, Daniele Leonardis, Christoph Kapeller, Gabriele Fusco, Christoph Guger, Christoph Hintermüller, Angelika Peer, Antonio Frisoli, Franco Tecchia, Massimo Bergamasco, and Salvatore M Aglioti. Local and remote cooperation with virtual and robotic agents: a P300 BCI study in healthy and people living with spinal cord injury. (submitted), 2016.