Pushing the boundaries of Brain-Computer Interface technology: challenges in novel applications and large-scale studies.

vorgelegt von
M.Sc.
Laura Acqualagna
geb in Jesi

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

Prüfungsausschuss:
Vorsitzender: Prof. Dr. Klaus Obermayer
1. Gutachter: Prof. Dr. Benjamin Blankertz
2. Gutachter: Prof. Dr. Klaus-Robert Müller
3. Gutachter: Prof. Dr. Peter Desain
Tag der wissenschaftlichen Aussprache: 21.06.2016

Berlin 2017

To my parents

Acknowledgements

These years in the BBCI group were a precious experience from many points of view. First of all, I had the opportunity to be in an international environment and to work and learn from very bright scientists coming from all over the world. This made me grow both as a scientist and as a person. Second, I had the chance to realize the dream I had since my second year of BSc, i.e. to work in the field of BCI, to deepen my knowledge of this fascinating field and to give my contribution. This experience and this thesis could not have been possible without the support of Prof. Benjamin Blankertz and Prof. Klaus Robert Müller, who believed in my capacities and gave me the chance to join the BBCI group, first as master student, then as PhD student. Not only their profound knowledge and expertise gave me the chance to learn much, but also their enthusiasm and kindness made the work environment a great place where to develop high quality science. Moreover, every time I had any doubt or questions, they were always available, even during busy times of approaching deadlines. To them it goes my biggest thanks. I would like to thank Prof. Andrea Kübler and Prof. Gabriel Curio, because i learned much from our joint work and conversations with them were always inspiring. Thank also to Prof. Peter Desain, who kindly agreed to review this thesis. My gratitude goes also to my former colleagues Matthias Treder and Martijn Schreuder, who supervised me at the beginning of my BBCI experience, and to Stefan Haufe, Sven Dähne, Anne Porbadnigk, Claudia Sannelli and Marija Uščumlić, who were always available for discussions if i had any doubts. I learned much from them and it was also fun to work together. I would like to thank all the other members of the BBCI group, former and current, because of the inspiring conversations at work, the good time at the BBCI evenings and at the conferences. These years would not have been the same without you guys. Thank Imke and Andrea for their organizational support, and also Dominik for his constant technical support and for being so patient in solving my endless problems with the cluster.

As all the PhD students know, during the years of the PhD there might be several moments of discouragement or lack of motivation, and in these cases is it very important to have the support of the people around you, especially if you are far away from home. I can consider myself lucky to have had so many great friends here in Berlin, who represented my second family and were always there for me in the 'down' moments and also in the happy events (and even to participate in the BCI experiments!). Thank Damien to be at my side during these important last months of the PhD and to give me a very good reason to speed

up the writing of the thesis. Finally, I would like also to thank my parents for being present in any moment, even though for them it was hard to see me leave Italy to pursue my objectives. They were ready to travel until Berlin when necessary and they always represented for me the safe harbor where to dock whatever it happens in life.

Abstract

Since the first studies in the 70s, Brain-Computer Interface (BCI) research has had an exponential development thanks to the high potential in improving life of severely impaired people. In the last years, the applicability of BCI extended also to non-medical applications as a complementary tool to provide interesting insights into the brain processes correlated with behavior. One of the most exploited measurement techniques used in BCI is Electroencephalography (EEG), because of its portability, contained costs and direct recording of neural activity. This thesis focuses on EEG-based BCI, with the aim of investigating cutting-edge applications, novel designs and the possibility of broadening this technology on large scale. According to the specific application, different brain features were analyzed, from visual to higher cognitive event-related potentials (ERPs), from the modulation of spontaneous oscillatory brain rhythms to that of the sensorimotor rhythm (SMR). Signal processing and machine learning algorithms were tailored to extract representative features and predict the user's state or the user's intention for the BCI operation. This thesis extends the boundaries of state-of-art BCI applications from several points of view. Firstly, the feasibility of a steady-state visual evoked potential (SSVEP)-based paradigm for video quality assessment was proven, which led to the collection of informative neural features in a much faster way than previous ERP-based paradigms. Secondly, two novel ERP-BCIs for mental typewriting independent of gaze-shifts were developed, which could be operated online with high performance by healthy users and could represent a valid alternative for patients with severely impaired oculomotor control. Thirdly, a revised version of the standard BCI design was investigated for the study of subjective relevance in the scope of information retrieval (IR) systems, bringing new insight into brain processes that can enhance the interaction between man and machine. Lastly, the widely employed SMR-BCI paradigm was applied on large scale and in a fully automatic way. The aim was to test how co-adaptive machine learning algorithms, successful when user-tailored, would perform in a realistic scenario without any manipulation from BCI-experts. The overall work shows the versatility of BCI as a successful supplement of the existing technology and as a tool to improve the lives of people, also pointing out the main limitations with critical view to trigger new research developments.

Zusammenfassung

Seit den ersten Studien in den 1970er Jahren, erlangte die Forschung in Gehirn-Computer Schnittstellen (engl. Brain-Computer Interface BCI) eine exponentielle Entwicklung dank der hohen Wahrscheinlichkeit der Lebensverbesserung von stark beeinträchtigten Personen. In den vergangenen Jahren hat sich die Verwendung von BCI auch auf nicht medizinische Anwendungen als ein ergänzendes Werkzeug ausgebreitet, welches interessante Einblicke in die Korrelation von Hirn-Prozessen und Verhalten gibt. Eine der in BCI meist verwendeten Messungstechniken ist die Elektroenzephalographie (EEG), aufgrund ihrer Mobilität, geringen Kosten und direkter Aufzeichnung von neuraler Aktivität. Diese Arbeit befasst sich mit dem EEG-basierten BCI. Ziel ist die Forschung an innovativen Anwendungen, neuen Designs und der Möglichkeit diese Technologie in großem Ausmaß zu verbreiten. Je nach spezifischer Anwendung wurden folgende verschiedene Hirnströme analysiert: Visuelle und kognitive Ereigniskorrelierte Potentiale (eng. event-related potentials (ERPs)), Modulation von spontan schwankenden Hirnsignalen und sensomotorischen Rhythmus (SMR). Algorithmen aus der Verarbeitung und dem Maschinellen Lernen wurden zugeschnitten um repräsentative Hirnströme zu extrahieren und den Zustand und die Absichten des Nutzers vorherzusagen. Diese Arbeit erweitert den aktuellen Stand der Technik der BCI-Anwendungen in vieler Hinsicht. Erstens, wurde die Durchführbarkeit eines BCI Paradigmas zur Bewertung von Video Qualität bewiesen. Dies ermöglicht eine wesentlich schnellere Sammlung von informativen Hirnströmen als bei bisherigen ERP-basierten Paradigmen. Zweitens, wurden zwei neue Blick-unabhängige ERP-BCIs zur mentalen Texteingabe entwickelt. Diese konnten online mit hoher Präzision von gesunden Nutzern durchgeführt werden. Diese Paradigmen stellen eine Alternative für Patienten mit einer hohen Beeinträchtigung der Augenbewegung an. Drittens, wurde eine veränderte Version des Standard BCI Designs zur Untersuchung von subjektiver Relevanz im Bereich der Systeme zur Informationsgewinnung erforscht. Dies brachte neue Einsicht in Hirnprozesse welche die Interaktion zwischen Mensch und Maschine fördern kann. Zuletzt wurde das sehr bekannte SMR-BCI Paradigma auf eine Großzahl von Nutzern voll automatisch angewandt. Das Ziel war zu testen, welche Ergebnisse co-adaptive maschinell lernende Algorithmen in einem realistischem Szenario ohne jegliche Manipulation von BCI-Experten erzielen würden. Die Arbeit zeigt im allgemein die Vielseitigkeit des BCI als eine erfolgreiche Ergänzung zu den bereits existierenden Technologien und Werkzeugen

welche der Verbesserung des Lebens der Menschen dienen. Die Arbeit deutet jedoch auch kritisch auf die Einschränkungen hin, um neue Forschungsentwicklungen auszulösen.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Brain-computer communication based on Electroencephalography (EEG).

In 1929 Hans Berger, Professor of Psychiatry at the University of Jena in Germany, published his first report of human electroencephalography (EEG) (Berger, 1929), after working five years on the results of his studies. This report included a description of the 10-Hz Alpha waves, which were subsequently called 'Berger waves' by Lord Adrian (Collura, 1993). Nearly a century after Berger's epochal discovery, EEG is a wide-spread neuroimaging technique used in hospitals and in research laboratories, which significantly helped the scientific community to understand human brain functions and brain diseases. Basic and clinical research shed light into the mechanisms of origin of the major EEG rhythms and evoked potentials and into their relationship with specific aspects of brain functions and mental tasks. Moreover, the beginning of the digital era together with the development of inexpensive and powerful computer hardware and software increased the clinical, scientific and commercial interest in this technology, also for alternative applications. In particular, EEG could infer a person's intention and the recorded EEG signals could be used for communication and control (Dornhege et al., 2007; Wolpaw and Wolpaw, 2012b). The practical implications of this achievement were striking. Patients affected by neuromuscular disorders like Amyotrophic lateral sclerosis (ALS), brainstem stroke, brain or spinal cord injury, cerebral palsy and other diseases that impair the neural pathways or the muscles, could highly benefit from a technology that helps them in restoring part of the motor functions or provides them with an additional channel to communicate and send commands to the external world. This is especially necessary in the late stages of diseases ending in a locked-in condition, in which other assistive technologies fail (Birbaumer et al., 2008b; Höhne et al., 2014a; Kübler et al., 2001a).

The concept of Brain-Computer Interface (BCI) was introduced by Vidal in 1973, who made the first attempt to

'evaluate the feasibility and practicality of utilizing the brain signals in a man-computer dialogue while at the same time developing a novel tool for the study of the neurophysiological phenomena that govern the production and the control of observable neuroelectric events' (Vidal, 1973).

A BCI records brain signals, extracts specific features and translates them into artificial outputs that act on the world or on the body itself. Brain signals can be recorded not only on the scalp surface with the EEG, but also with implanted electrodes (e.g. electrocorticography (ECoG), Leuthardt et al. (2004)), or with technologies that infer the neural activity by measuring other parameters, like magnetic fields (magnetoencephalography (MEG), Mellinger et al. (2007)) or hemodynamic responses associated with neuron behaviour (e.g. functional magnetic resonance imaging (fMRI), Sitaram et al. (2007); Weiskopf et al. (2004); functional near-infrared spectroscopy (fNIRS), Coyle et al. (2007)). All these alternative technologies have several drawbacks compared to EEG that make them more difficult to use for daily experiments or in-home situations with patients, for example high costs and occupancy area, necessity of invasive brain surgery, difficulty of using the signals for real-time BCI applications. Therefore, this thesis focuses on noninvasive BCI based on EEG technology.

As described by Wolpaw and Wolpaw (2012b), the BCI can serve for several purposes that can be summarized in *replacing, restoring, enhancing, supplement, improving.*

- The output of a BCI might replace natural output lost because of a disease or an injury. For example a person affected by ALS who is not able to speak or reliably move the eyes can use the BCI as a communication program. The output of a BCI might also be used to move a motorized wheelchair, in case of paralyzed people who lost the control of the limbs.

- The output of a BCI might restore lost natural outputs. For example, the BCI can stimulate electrodes placed on the muscles to contract them and move a paralyzed limb.

- The output of a BCI might enhance natural outputs. For example a person that is performing repetitive and monotonous tasks that require sustained attention, might be warned by the BCI in case of a lack of attention or excessive workload. Typical examples are the industrial production chain or the driving situation, in which the lack of attention can lead to decreased performance or dangerous car accidents.

- The output of a BCI might supplement natural output. For example a person might use the BCI in a game or an entertainment application as complement to the behavioral output.

- The output of a BCI might improve natural output. For example, in case of BCI as complement technology for rehabilitation engineering. The BCI can help in moving a robotic joystick used for rehabilitation of an arm whose natural movement was damaged

by a stroke. The feedback of such rehabilitation can boost the plasticity of the motor cortex and improve the natural motor output.

Recently, the Brain/Neural Computer Interaction (BNCI) Horizon 2020 roadmap (Brunner et al., 2015) described also a sixth category of BCI, i.e. as research tool for cognitive neuroscience, for example in the research area of decision making (Schultze-Kraft et al., 2015).

BCIs used mainly for communication and control are also referred as to *active BCI*, and are those that scientists mainly studied and developed so far, with the final goal to be used with target patients (Birbaumer and Cohen, 2007; del R. Millán et al., 2010; Kübler et al., 2005; Nijboer et al., 2008; Sellers and Donchin, 2006). However, the development of BCI systems for non-medical applications and for the use by healthy people has increased in the last years (Blankertz et al., 2010b; Müller et al., 2008; Zander and Kothe, 2011). Novel BCI designs have been explored to monitor the cognitive user's state and investigate mental features like attention, workload and task engagement (Berka et al., 2007; Borghini et al., 2014; Haufe et al., 2014b; Kohlmorgen et al., 2007; Müller et al., 2008; Venthur et al., 2010a). These kinds of BCI are also referred to as *passive BCI*. Another interesting and novel application uses BCI to infer the user's experience of music (Makeig et al., 2011; Miranda and Brouse, 2005; Treder et al., 2014).

Nowadays, BCI systems can be operated with high performance and many users can achieve the control of the BCI within one session, i.e. in one day. Especially in case of BCI based on the self-regulation of brain signals like Sensorimotor Rhythms (SMR) (Wolpaw et al., 2002) and Slow Cortical Potentials (SCPs) (Birbaumer et al., 1999), this was not common at the beginning of BCI research history. The first studies were based on an *operant conditioning* approach, in which users had to learn how to modulate their brain activity in order to control the BCI. This training could last several weeks or even months until reasonable performances could be established. The introduction of the machine learning approach moved the learning task from the user to the computer ('Let the machines learn!' motto of the Berlin Brain Computer Interface (BBCI) project) (Blankertz et al., 2002, 2008b; Krauledat et al., 2004; Müller et al., 2004; Müller et al., 2008). This approach used suitable signal processing and data analysis techniques to extract robust brain features specific for each user. In this way, the dimensionality of the EEG data was drastically reduced to allow a short training session of just 10-30 minutes and a fast online processing. Moreover, the variety of machine learning algorithms let researchers address complex challenges in BCI with successful results, e.g. single-trial classification and non-stationarity of EEG signals.

The typical BCI chain is summarized in several processing steps, that include: Signal acquisition, preprocessing and extraction of meaningful features, classification, feedback stimulation. It is important to specify that the output of the BCI is meant to change the environment around the user and therefore to provide sensory inputs that the user receives

as a feedback. In this sense, the human and the machine are seen as complementary systems that work in a closed loop (van Gerven et al., 2009). Therefore, systems aiming at monitoring the brain activity without providing any change in the interaction between the brain and the machine cannot be considered as a BCI (Wolpaw and Wolpaw, 2012b). In thesis, the two studies described in Chapters 4 and 6 were performed offline, i.e. no real-time feedback was provided by the BCI. The studies were meant to investigate novel designs, therefore being just the first step towards the realization of the novel BCI application. In fact, both studies comprised a machine learning analysis to predict specific features of the stimuli or user's intentions, in order to check the feasibility of the novel design as a potential BCI system.

1.2 Motivation

This thesis describes several BCI applications, mainly focused on the previously introduced categories *BCI for restoring* and *BCI for supplement*. It is the result of about 250 experiments, among which about the half was run by myself. The thesis can be divided into two parts. In the first part, (Chapters 4, 5 and 6) cutting-edge BCI designs are investigated to push further the boundaries of the well-established paradigms towards novel applications. The feasibility of BCI technology is demonstrated also in contexts in which Neurotechnology is just starting to play a role, like in video quality assessment or in Information Retrieval systems. In all the described studies I made a particular effort in the design of the proper stimuli presentation that would elicit the desired brain response. Different data analysis techniques were tailored to gain the highest performances, both in the temporal and spectral domains. I also payed attention at single-subject level and tried, when possible, to find an answer for the high inter-subject variability typical of BCI data. The chapters develop describing increasing levels of engagement of the user in respect to the BCI. In the first study (Chapter 4), which describes how different levels of degradation of texture images modulate the brain signals, the user is completely passive in respect to the BCI and the brain responses are mainly elicited by the features of the presented stimuli. In the second study (Chapter 5), which describes two BCI systems for spelling application, the brain responses are still elicited in large degree by the intrinsic features of the stimuli, but the user becomes more active in the BCI operation since he/she is actively engaged in discriminating the target symbols. Here, 'top-down' processes involve knowledge-driven mechanisms that facilitate the discrimination between signal (i.e. the targets) and 'noise' or distractors (Sarter et al., 2001). These top-down processes become even more important in the third study (Chapter 6), in which they enhance the neural processing of relevant sensory input. The study addresses the topic of relevance and the user is the protagonist of the interaction with the BCI. The brain signal is modulated in a discriminable way by the stimuli that the user recognizes as relevant for that specific task, despite their intrinsic features. Therefore, top-down mechanisms prevail bottom-up processing in the stimuli evaluation. All the studies described in first part of

this thesis comprise a relative small amount of datasets (i.e. less than 20 for each study). A group size that is large enough to let meaningful statistical tests but still quite limited to facilitate single-subject inspection of the data is convenient in exploratory studies. In fact, especially in BCI experiments, inter-subject variability is still a major problem and an accurate investigation of the brain responses is necessary in order to tailor appropriate machine learning algorithm for that application. But what happens after a BCI paradigm is developed and tested in many studies, when pros and contras are known and it can be regarded as 'established' for the special application? Then, the following step is to test the BCI paradigm in bigger scale, with a new large sample of participants. This topic was addressed in the second part of this thesis. In this study, a fully-automatic co-adaptive motor-imagery BCI was evaluated for the first time with a pool of 168 participants. Even though this kind of studies is very demanding in terms of time and resources, I believe that it has to be encouraged in the BCI community. Particular effort should be made to test automatic systems that exploit the state-of-art machine learning algorithms. Indeed, we must not forget that not only BCI research investigates fascinating scientific hypotheses that give new insights into brain processing and novel algorithms development, but also the final purpose is to apply these systems outside of the lab (Allison et al., 2012; Blankertz et al., 2010b; Müller et al., 2008). This is worth for both clinical and non-clinical applications. Bringing the BCI system outside of the lab implies several consequences: the environment of the BCI operation is not anymore controlled; the use of the BCI itself has to be independent, or at least it should not rely completely, on the constant presence of a BCI expert to adjust the sensible parameters of the system. The latter topic is widely discussed in the second part of the thesis. The final users of the BCI system will be people naive in respect to the technology and, in many applications, also in an unfavorable physical condition. Therefore, it is essential to investigate whether a successful BCI paradigm is suitable to the majority of the people who aim at using it, and whether the machine learning techniques are robust enough or *adaptive* enough to cope with the unavoidable differences between the users' brain responses. These last crucial points can be properly addressed just bringing the BCI experimentation to a higher level, i.e. with large-scale studies and also with patients studies.

1.3 Outline of this thesis

While the previous section was meant to introduce the motivation of the present work and to give the reader an overview of the evolution of the chapters for a smoother reading, in the following a brief description of each chapter is given, excluding this short Introduction Chapter and the last Conclusions Chapter. The corresponding publications are referred and listed in the following section.

- Chapter 2 introduces the fundamentals about the generation of the EEG signal, the EEG standard notation and the neurophysiological responses mainly used in EEG-based BCI. Since the studies described throughout the thesis investigate different brain responses, this chapter is meant to provide a background knowledge of the main EEG rhythms, Event-Related Potentials (ERPs), Steady-State Visual Evoked Potentials (SSVEPs) and Sensorimotor Rhythm (SMR).

- Chapter 3 introduces the main methods used in this thesis. The chapter spans from methods for signal processing and extraction of significant EEG features, to techniques for dealing with artifacts in EEG recordings and machine learning algorithms for data analysis and classification.

- Chapter 4 describes a novel BCI application for video quality assessment based on a setting in which the presence of SSVEPs is a neural marker that objectively indicates the neural processing of the quality changes that are induced by the video coding. Two different machine learning methods to classify SSVEPs were tested, based on the modulation of the brain rhythm and on time-locked components, respectively. Furthermore, the correlation of the EEG-based measures with the outcome of the standard behavioral assessment was investigated. This work was published in [1].

- Chapter 5 describes two novel BCI paradigms for mental typewriting application, also referred to as *Speller*. The proposed BCI designs do not require gaze shifts towards different target locations and can be operated by non-spatial visual attention, therefore they can be considered as valid paradigms in applications for patients with impaired oculo-motor control. The first design uses rapid serial visual presentation (RSVP) for displaying the stimuli and it is here addressed as *RSVP Speller*. Three conditions were investigated, regarding the Stimulus Onset Asynchrony (SOA) and the use of color features. This work was published in [2,8-9]. As complementary analysis, a novel algorithm is presented that exploits the dependencies between EEG epochs overlapping in time to improve the selection accuracy and spelling speed. This work was not published, but it was presented as a poster at the BBCI workshop 2012 [12]. The second design investigates the feasibility of the first truly gaze-independent visual BCI, in which stimuli consist of colors displayed widescreen. This makes the focus of gaze irrelevant to the BCI-spelling. This design is called *Chroma Speller* and it won the Brain-Computer Interfacing HCI 2011 Challenge. The study was then published in [7].

- Chapter 6 describes a novel BCI application that aims at identifying EEG features that indicate subjective relevance of a perceived stimulus with respect to the current task. Such implicit information about the user state decoded online could be used in future Human-Computer Interaction (HCI) applications, like Information Retrieval (IR) systems, complementary to the explicit input in order to make the system as efficient as

possible. The transfer of BCI technology to HCI systems required the development of a novel experimental design and classification approaches that cope with a richer variety of stimuli than usual fundamental laboratory studies. A preliminary version of this work was presented as a poster at the EMBS conference 2015 [10] and the complete study was submitted as a journal paper [3].

- Chapter 7 describes a co-adaptive BCI system that for the first time was tested in a fully-automatic way and on a large scale. A pool of 168 participants naive to BCI operated the co-adaptive SMR-BCI in one single session. Different psychological interventions were performed prior the BCI session in order to investigate how motor coordination training and relaxation could influence BCI performance. A neurophysiological indicator based on the Power Spectral Density (PSD) (Blankertz et al., 2010a) was extracted by the recording of few minutes of resting state brain activity and tested as predictor of BCI performance. This work was published in [4].

1.4 Publications

1.4.1 Journal publications

[**1**] Acqualagna, L., Bosse, S., Porbadnigk, A.K., Curio, G., Müller, K.R., Wiegand, T. and Blankertz, B., 2015. EEG-based classification of video quality perception using steady state visual evoked potentials (SSVEPs). Journal of neural engineering, 12(2), p.026012.

[**2**] Acqualagna, L. and Blankertz, B., 2013. Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). Clinical Neurophysiology, 124(5), pp.901-908.

[**3**] Acqualagna, L. and Blankertz, B. 2016. Classification of subjective relevance through a Brain-Computer Interface, submitted.

[**4**] Acqualagna, L., Botrel, L., Vidaurre, C., Kübler, A. and Blankertz, B., 2016. Large-scale assessment of a fully automatic co-adaptive motor imagery-based Brain Computer Interface. PloS one 11(2), p.e0148886.

[**5**] Bosse, S., Acqualagna, L., Porbadnigk, A.K., Curio, G., Müller, K.R., Wiegand, T. and Blankertz, B., 2016. Assessing Perceived Image Quality Using Steady-State Visual Evoked Potentials and Spatio-Spectral Decomposition, submitted.

[**6**] Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., Uščumlić, M., Wenzel, M., Curio, G. and Müller, K.R., 2016. The Berlin Brain-Computer Interface: Progress Beyond Communication and Control, submitted.

1.4.2 Conference proceedings and other contributions

[7] Acqualagna, L., Treder, M.S. and Blankertz, B., 2013, November. Chroma Speller: Isotropic visual stimuli for truly gaze-independent spelling. In Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on (pp. 1041-1044). IEEE.

[8] Acqualagna, L. and Blankertz, B., 2011, August. A gaze independent spelling based on rapid serial visual presentation. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (pp. 4560-4563). IEEE.

[9] Acqualagna, L. and Blankertz, B. 2011, September. The Rapid Serial Visual Presentation Paradigm Tested in an Online Brain-Computer Interface Speller, in Proceedings of the 5th International Brain-Computer Interface Conference, Graz, Austria.

[10] Acqualagna, L and Blankertz, B., 2015, August. Neural correlates of relevant stimuli processing for Brain Computer Interfaces, In Engineering in Medicine and Biology Society, EMBC, 2015 Annual International Conference of the IEEE. IEEE.

[11] Acqualagna, L., Nicolae, IE, Wenzel, M. and Blankertz, B., 2015. Report on HCI-related cognitive responses that can be detected in the EEG. Internal Deliverable for the MindSee EU project.

[12] Murphy, R., Acqualagna, L. and Blankertz, B., 2012, September. Accounting for Dependencies in the Classification of ERPs Improves BCI Spellers. BBCI Workshop 2012, Advances in Neurotechnology, Berlin, Germany.

[13] Nicolae, IE, Acqualagna, L and Blankertz B. 2015a, August. Neural Indicators of depth of cognitive processing for user-adaptive neurotechnological applications, In Engineering in Medicine and Biology Society, EMBC, 2015 Annual International Conference of the IEEE, pp. 1484-1487, IEEE.

[14] Nicolae, IE, Acqualagna, L and Blankertz B, 2015b, October. Tapping Neural Correlates of the Depth of Cognitive Processing for Improving Human Computer Interaction, in Proceedings of the Fourth International Workshop on Symbiotic Interaction, Lecture Notes in Computer Science, Vol. 9359, pp. 126-131, Springer.

[15] Bosse, S., Acqualagna, L., Porbadnigk, A.K., Blankertz, B., Curio, G., Müller, K-R., Wiegand, T., 2014, October. Neurally informed assessment of perceived natural texture image quality. in IEEE International Conference on Image Processing, ICIP 2014, Paris, France, pp. 1987-1991, IEEE.

[16] Bosse, S., Acqualagna, L., Porbadnigk, A.K., Curio, G., Müller, K.R., Blankertz, B. and Wiegand, T., 2015, September. Neurophysiological assessment of perceived image

quality using steady-state visual evoked potentials. In SPIE Optical Engineering+ Applications (pp. 959914-959914). International Society for Optics and Photonics.

[**17**] Nicolae, IE., Ungureanu, GM, Acqualagna, L., Strungaru, R., Blankertz, B., 2015, November. Spectral Perturbations of the Depth of Cognitive Processing for Brain-Computer Interface Systems Health, Proceedings of The 5th IEEE International Conference on E-Health and Bioengineering - EHB 2015, Iasi, Romania. IEEE.

[**18**] Botrel, L., Acqualagna, L., Blankertz, B., Kaufmann, T. and Kübler, A., 2013, January. Relaxation Exercise improves SMR-BCI Performance. In Proceedings of TOBI Workshop IV. Practical Brain-Computer Interfaces for End-Users: Progress and Challenges, Sion, Switzerland.

[**19**] Gamberini, L., Spagnolli, A., Blankertz, B., Kaski, S., Freeman, J., Acqualagna, L., ... and Ferrari, E. (2015). Developing a Symbiotic System for Scientific Information Seeking: The MindSee Project. In Symbiotic Interaction (pp. 68-80). Springer International Publishing.

Chapter 2

Fundamentals: neurophysiology and EEG

2.1 Electroencephalography

Electroencephalography (EEG) is an electrophysiological monitoring method that records the electric signals produced by the brain activity through electrodes placed on the scalp. Each electrode reflects the space-averaged synaptic source activity occurring within large part of the underline cortical tissue, containing millions of neurons. This characteristic makes the EEG poor in spatial resolution, but excellent in temporal resolution, since the electrical signal travels in order of milliseconds from the sources to the recording sites. The EEG recordings are always bipolar, which means that what is measured is a difference of potentials between two electrodes. Usually, one electrode is designed as 'reference electrode' and all the others are recorded in respect to that one. There are many locations where the reference electrode can be placed, usually on a position less affected by the brain signals of interest (e.g. nose). Such choice can be arbitrary since the signal can be easily re-referenced to other sites in the data processing after the experiment. In order to make the recordings comparable across different laboratories and hospitals, the position of the electrodes, i.e. the electrodes montage, was standardized in different montages, for example the International 10-20 system (Klem et al., 1999)(Fig. 2.1). This standard electrode placement defines contours between skull landmarks and divide them in proportional distances, for example 20% of the total length along contours between landmarks for the 10-20 system (Wolpaw and Wolpaw, 2012b). The standard nomenclature uses odd numbers for electrodes on the left hemisphere side, even numbers for the electrodes on the right hemisphere side, and 'z' for electrodes along the midline. Each site has also one or more letters that indicate the lobe underneath the electrode, e.g. 'C' for central, 'O' for occipital and so on. Note that other electrodes systems have been developed in order to perform specific recordings, for example a system with higher electrodes density on the motor cortexes is often applied in motor imagery BCI experiments. In order to establish an electrical connection between the electrode and the skin, a conductive gel-paste is applied underneath the electrode, with light abrasive movements and being careful to not spread the gel between electrodes, in order to avoid electrical bridges. Recently,

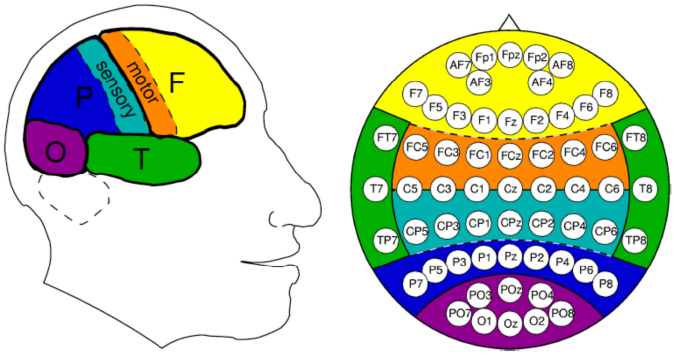


Fig. 2.1 Schematic of the cortical areas (left) and standard 64-channels EEG montage (right). F is frontal, P is parietal, O is occipital and T temporal. The sensory and motor areas are also highlighted with different color. In the EEG montage the same colors depict the corresponding cortical areas. Taken from the lecture *Brain-Computer Interfacing* at the *Bernstein Focus: Neurotechnology*, with permission.

water-based and dry electrodes have been introduced in the market, making the preparation of the EEG cap faster and the EEG itself a more portable technology (Guger et al., 2012; Popescu et al., 2007; Volosyak et al., 2010). However, this goes to the detriment of the signal quality, since these kinds of recording are more susceptible to external artifacts. In a typical EEG experiment, the EEG cap is connected to differential amplifiers, in which the signal is amplified 1,000-100,000 times, filtered and (in most research applications) digitalized, for example at 1 kHz sampling rate. The EEG signal is then visualized on a computer's monitor in a range between 10 microvolts and 100 microvolts. In all the studies described in this thesis the EEG was recorded at 1kHz using BrainAmp amplifiers and an actiCap active electrode system with 64 channels (both by Brain Products, Munich, Germany). The configuration of the EEG montage was slightly varied across experiments according to the specific application and will be specified in the respective sections. The electrodes were referenced to the left mastoid, using a forehead ground. For offline analyses, electrodes were re-referenced to linked mastoids. All the impedances were kept below 10 kΩ.

2.2 Neurophysiology of the EEG

The electrical brain activity measured by the EEG comes form ionic currents generated by biochemical processes at the cellular level. The principal neural generators are the pyramidal neurons in the cortical grey matter. The agents responsible for the scalp recorded EEG activity are the postsynaptic potentials (PSPs), which are local depolarizations of the post-synaptic neuronal membrane. For example, in a chemical synapsis of the central nervous system (CNS), the neurotransmitters released by a pre-synaptic neuron activate specific ionic channels in the membrane of the post-synaptic neuron. According to the type of synapsis, i.e.

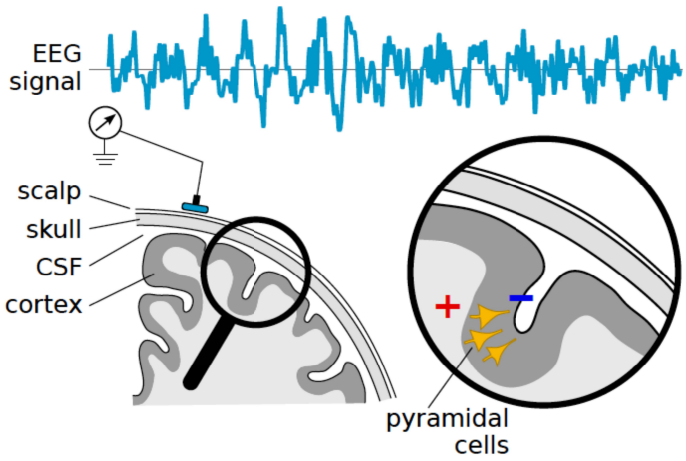


Fig. 2.2 Generation of the EEG signal. Schematic representation of the dipole generated by the pyramidal neurons in the cortex. After traveling through the cerebrospinal fluid (CSF), the skull and scalp, the current is measured by the EEG electrode. Taken from the lecture *Brain-Computer Interfacing* at the *Bernstein Focus: Neurotechnology*, with permission.

excitatory or inhibitory, the activation of such channels would produce a flux of ions through the membrane that would lead to a local change of polarity of the membrane itself. This phenomenon produces local currents along the body of the neuron that can integrate with the other PSPs from adjacent synapsis of the same neuron. This characteristic of spatial and temporal summation of PSPs makes the general phenomenon longer in duration and involves a larger membrane surface than the single action potential. Depending on the duration of the PSPs, the number of synchronous activated PSPs, the orientation of the layer of pyramid cells generating the currents and the distance between the sources and the recording electrode, the characteristic of the recorded EEG signals will be different (Schaul, 1998). In general, the electrical field around the pyramidal cells can be schematized as dipoles (Fig. 2.2), and the contribution of a dipole to the recorded EEG signal depends on its orientation in respect to the electrode. Radial dipoles would contribute more, tangential dipoles less. This refers mainly to the neurons of the folded cortex underneath the scalp. Deep brain sources such as the thalamus, basal ganglia, hippocampus and brainstem are very difficult to measure because of their large distance to the scalp and their geometrical arrangement.

In the generation of the EEG signal it is crucial to take into account the phenomenon of volume conduction. It refers to the manner in which the source currents spread through the brain, the cerebrospinal fluid (CSF), the skull and scalp tissue and it is determined by the geometry of the of the surface shape and by the resistivity of the different tissues. These characteristics make the amplitude of recorded EEG signal much more attenuated compared to e.g. intracranial recording (Ball et al., 2009). It also causes a current-spreading underneath the skull that makes the recorded signal at each EEG electrode a mixture of different sources,

originating the complex issue of estimating the original neural sources of the observed EEG activity (i.e. the so called 'inverse problem').

2.2.1 EEG rhythms

In the brain, the rhythmic fluctuations in the excitability of neurons or populations of neurons are referred as oscillations. The neurobiological mechanisms that produce the oscillations observed in the EEG involve interactions between inhibitory interneurons and excitatory pyramidal cells and the alternating balance between states of excitations and inhibitions (Cohen, 2014). Certainly the dynamics behind the generation of oscillations are very complex and influenced by many additional factors that modulate amplitudes, frequency and phase of the oscillations. The brain's natural rhythms govern the most general dynamic of the brain. The synchronous activity of oscillating neural networks at various frequencies in the same or different structures are associated with different brain states and can compete or interact with each other (Buzsáki and Draguhn, 2004). The first human EEG pattern was described by Berger as a large amplitude rhythm with frequency around 10 Hz, which was induced by eye closure in the awake subject at rest. It was called 'alpha' because it was the first that was observed (Buzsáki, 2006). Now we know that the oscillatory bands of the mammalian brain cover a wide range of frequencies, following the 1/f power low. They have been categorized as (in frequency ascendant order): 'delta' 0.5-3 Hz, 'theta' 3.5-7 Hz, 'alpha' 8-13 Hz, 'beta' 14-30 Hz, 'gamma' 30-70 Hz. As described in the review of Başar et al. (2001), the brain's natural oscillations provide links for the communication and associative functions of the brain, and they are functional correlates of sensory registration, perception, movement and cognitive processes related to attention, learning and memory (Engel et al., 2001; Varela et al., 2001). The oscillations can be also evoked or event-related to a sensory or cognitive event, and can be phase-locked or non-phase-locked to the related task. For example, together with the well-know 'spontaneous' alpha rhythm (i.e. that one observed by Berger) several forms of 'functional alpha' are observed during sensory, cognitive and motor processes (Başar et al., 1997). In the last decades, an extensive amount of studies has been conducted by neuroscientists emphasizing the importance of oscillatory brain activity and investigating the functional correlates of the different frequency bands and the interactions between them (Fries, 2005; Groß et al., 2001; Tallon-Baudry and Bertrand, 1999; Womelsdorf et al., 2007). The detailed review of the brain oscillations goes beyond the purpose of this thesis and the reader is redirected to the book 'Rhythms of the Brain' (Buzsáki, 2006) for a comprehensive description.

2.3 Neural signals that enable EEG-based BCI control

The following sections provide a summary of the main EEG signals used in BCI research. For a more detailed discussion, please refer to Luck (2005) and Wolpaw and Wolpaw (2012b).

2.3.1 Event-Related Potentials (ERPs)

Event-related potentials (ERPs) in the EEG are deflections of the neural activity triggered by a specific event. They can be elicited by different kinds of stimuli, for example visual cues, auditory tones, tactile stimuli. The amplitude of the ERP response to a single event is usually very little compared to the background EEG activity. Since the ERPs are time and phase-locked to the event and follow a constant time course, it is common practice to average the ERP responses over multiple trials in order to increase the signal to noise ratio (SNR). The result of the averaging is a series of negative and positive voltage deflections that are named ERP components. The components are usually named by their polarity, 'P' for positive and 'N' for negative, followed by a number that indicates the latency of the components in milliseconds, for example P300 or N200. Since the latencies can vary considerably between experimental paradigms, the components can be also named with the polarity following a number that indicate their position in the ERP waveform, for example P3 or N2. The latter nomenclature will be used throughout this thesis (except for the N400 component, which is always addressed as 'N400' in literature). The ERP components differ not only for polarities and latencies, but also for their scalp distribution and represent different steps of the neural processing of the event. The components elicited within the first 150 ms following the stimulus onset are called 'exogenous components', since they tend to reflect the early sensory responses of the eliciting stimulus. Their characteristics depend on the stimulus modality. ERP components with longer latencies are called 'endogenous', since they reflect the cognitive processing of the stimulus and are related to its significance and to the mental task that the user is performing. The following subsections introduce a general description of the major ERP components that can be elicited by BCI paradigms that use visual stimuli.

Early visual components.

The main ERP components that are linked to the sensory processing of visual stimuli usually arise within 300 ms after the stimulus onset, with both positive and negative polarities. The P1 wave is a positive component peaking between 100-130 ms after the stimulus onset and is larger at the lateral occipital cortex. It is sensitive to variation of the stimulus parameters like the stimulus contrast, but also to the direction of spatial attention (Hillyard et al., 1998). The P1 is followed by the N1, a negative ERP with several visual subcomponents peaking between 100-200 ms post-stimulus. One lateral occipital subcomponent of the N1

seems to reflect discrimination processes (Vogel and Luck, 2000). The anterior-central P2 subcomponent follows the N1 and it is enhanced by target stimuli of simple features that occur relatively infrequently (Luck and Hillyard, 1994). A well studied ERP subcomponent is the N2, which peaks between 200 and 300 ms post-stimulus and is characterized by several subcomponents (Luck and Hillyard, 1994; Näätänen and Picton, 1986). In the visual domain, the earlier anterior N2 subcomponent has been mostly investigated in conflict monitoring studies and is sensitive to the stimulus novelty, to the deviation from template stimulus, to stimulus complexity and percentage of non-target stimuli (Bocquillon et al., 2014; Folstein and Van Petten, 2008). The posterior N2 subcomponent is elicited by target stimuli, is sensitive to task difficulty and is thought to reflect classification of the stimuli (Folstein and Van Petten, 2008).

Cognitive components

The most investigated ERP cognitive component is the P3 wave, described originally by Sutton et al. (1965). Despite the large amount of psychophysiology studies that investigated the P3, it is not clear yet what this component exactly means, but we know what influences its amplitude and latency. Squires et al. (1975) made the first distinction about the frontal P3a and the parietal P3b subcomponents. The P3a has earlier latency and in Squires et al. (1975) it was elicited by infrequent and unpredictable shifts in tone peach or intensity, both when the subjects were attending to the stimuli and when they ignoring them. Other studies show that it is elicited by unusual or surprising task-irrelevant stimuli (Courchesne et al., 1975; Muller-Gass et al., 2007). In a recent study, Polich (2007) indicates the P3a originating from stimulus-driven frontal attention mechanisms during task processing. The classical P3 generally refers to the P3b sub-component, peaking between 300 and 500 ms in the parental cortex. It is elicited by target stimuli (i.e. task-relevant) that are infrequent, but not necessary surprising (Verleger et al., 1994). Its amplitude gets larger as the probability of task-defined stimuli gets smaller. Amplitude and latencies are also influenced by the attentional level of the subject, the task complexity, the target-to-target interval (Gonsalvez and Polich, 2002) and stimulus modality (Polich, 1996). Polich (2007) indicated that P3b appears related to memory processes succeeding the attentional operation. Actually, the positive cognitive ERP components cannot be reduced just to the P3a and P3b subcomponents, but there is a whole complex of positive components, richer in some experimental designs than in others (Sutton and Ruchkin, 1984). Vaughan and Ritter (1970) introduced the phrase 'late positive complex', since they found evidence of a late positive wave to unpredictable but irrelevant stimuli that differed in latency and scalp distribution from the P3 described by Sutton. Some investigators prefer to treat this ensemble as a unitary entity, while others prefer to examine in detail the distribution and response of the specific components (Dien et al., 2004; Spencer et al., 2001). Squires et al. (1975) referred to the late ERP subcomponents as 'Slow Waves',

that behave differently form the P3b in relation to experimental variables, but often overlap (Sutton and Ruchkin, 1984). The late positive complex (LPC) is elicited by stimulus of any modality, it has been associated with orienting, attention, stimulus evaluation and memory (Falkenstein et al., 1994; Gomarus et al., 2006). An increase of LPC amplitude was shown with the increase of stimulus intensity and significance, and increased reaction time (Roth et al., 1982; Squires et al., 1975).

2.3.2 Steady-State Visual Evoked Potentials

Vialatte et al. (2010) defined Steady State Visual Evoked Potentials (SSVEPs) as *'evoked responses induced by flickering stimuli. SSVEPs are periodic, with a stationary distance spectrum showing the characteristic SSVEPs peaks stable over time. SSVEPs are better observed in the frequency or time-frequency domains'*. They were investigated for the first time by Regan (1966), who ran experiments with long stimulus trains of sinusoidally modulated monochromatic light. The stimuli produced a stable Visual Evoked Potential (VEP), which could be extracted by averaging over trails and was called SSVEP. SSVEPs differ from transient VEPs, since their frequency components remain constant in amplitude and phase over a long time period (Regan, 1989). This means that SSVEPs contain stationary periodic oscillations whose spectral distribution remains stable over time. The frequency range of the flickering stimuli is usually between 3-50 Hz (Odom et al., 2004), but also frequencies below 3 Hz and up to 80 Hz can generate SSVEPs (Herrmann, 2001). SSVEPs can be generated by several types of stimuli, like flickering lights or images, alternately inverting bicolor checkerboards, etc., at a constant frequency. The elicited brain response will show oscillations at the same frequency and its even harmonics. According to most of the studies, SSVEPs are generated in the occipital cortex, but there are several theories that claim that broadly distributed sources are involved in the SSVEP propagation (Srinivasan et al., 2006) or that SSVEPs propagate from visual areas to non-visual areas in form of traveling waves (Silberstein et al., 1995).

2.3.3 Sensorimotor rhythms and Event-related Desynchronization/ Synchronization

Sensorimotor rhythms (SMRs) are oscillations in the electric or magnetic fields that are recorded over the posterior frontal and anterior parietal areas, i.e. over the sensorimotor cortices (Wolpaw and Wolpaw, 2012b). In EEG, SMRs occur in three main frequency bands: alpha (better know as 'mu' rhythm if localized in the motor cortexes), beta and lower gamma. Chatrian et al. (1959) and subsequent studies observed the decrease of SMRs of the mu rhythm during motor behavior. SMRs were deeply investigated especially in BCI applications since they decrease not only during motor execution (Jasper and Penfield, 1949), but also during motor imagination (Gastaut et al., 1965; Pfurtscheller and Neuper, 1997), motor

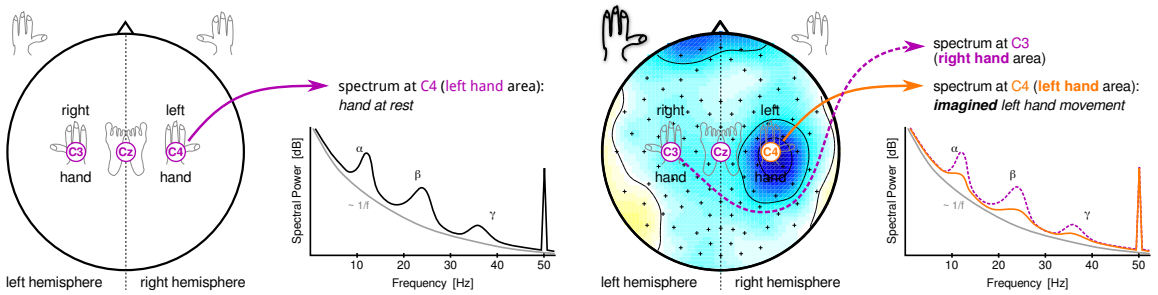


Fig. 2.3 Event-Related Desynchronization (ERD). Scalp plot and power spectrum at electrode C4 at rest (left) and during imagination of the left hand movement (right). Electrode C3 is placed over the motor area that desynchronizes during the imagination of the right hand movement, Cz over the area of the feet movement, C4 over the area of the left hand movement. During motor imagination of the left hand, the spectra peaks of alpha (mu), beta and gamma rhythms decrease their amplitudes at C4 (orange line), whereas the peaks of the rhythms at C3 do not change (magenta dashed line). The scalp distribution of the ERD is depicted as a blue deflection in the scalp plot at C4. Taken from the lecture *Brain-Computer Interfacing* at the *Bernstein Focus: Neurotechnology*, with permission.

observation (Pineda et al., 2000) and somatosensory stimulation (Nikouline et al., 2000). This characteristic makes them a suitable BCI input for paralyzed patients (Birbaumer et al., 2008a; Kübler et al., 2005; Neuper et al., 2003; Pfurtscheller and Neuper, 2001). The decrease of SMRs is known as Event Related-Desynchronization (ERD) (Pfurtscheller and Aranibar, 1977)(Fig. 2.3). ERD was initially described as a short lasting, phasic and localized amplitude attenuation or blocking of oscillations in the alpha and beta bands, in relation to an event (Pfurtscheller and Klimesch, 1992). The event can be external (e.g. stimuli) or internal (e.g. voluntary behavior). Focusing on motor behavior, during voluntary movements the mu rhythm displays a lower frequency ERD (8-10 Hz) widespread over the entire sensorimotor cortex, related to attentional processes, and a higher frequency (10-13 Hz) ERD that is task-specific and topographically restricted (Pfurtscheller et al., 2000). The localized ERD is typically accompanied by simultaneous increase of the SMRs in neighboring cortical areas, called Event-Related Synchronization (ERS) (Pfurtscheller, 1992; Suffczynski et al., 1999). ERS can be defined as a phasic and regional localized increase of alpha and beta bands that marks cortical areas at rest or in an idling state. Also the beta rhythm exhibits an ERD associated to motor behavior, followed by a brief ERS after the movement, called 'beta rebound' (Pfurtscheller et al., 2005). Many studies documented the similarities of the ERDs associated with motor imagery to those associated with actual movements (McFarland et al., 2000; Neuper et al., 2005). For example, imagery of the left or right hand movement would lead to the ERD of the mu and beta rhythm of the contralateral motor cortex, which is often associated to the ERS of the same rhythms in the ipsilateral motor cortex.

Chapter 3

Methods for signal processing and machine learning in BCIs

The aim of a BCI is to translate brain signals into commands that can be used to drive the BCI according to the user's wish. Even in case of BCI for user-state monitoring, the final goal is to extract meaningful brain states associated to specific events or conditions, use them to make predictions and provide feedback. Two main steps are necessary in the analysis of the brain signals: feature extraction, that refers to the signal processing part, and classification, that refers to the application of machine learning techniques. With features we mean the characteristics of the brain signal. In literature a large variety of methods for the analysis and classification of the EEG signals is proposed. In this chapter, I will introduce the basics of the methods used throughout the thesis, in order to ease the reading of the following chapters. In the sections describing the specific studies, I will specify further details about the choice of a method and about the parameters tailored to the application. The signal processing and data analysis described in this thesis were performed with MATLAB (The MathWorks, Natick, MA, USA) using the BBCI toolbox (Blankertz et al., 2010b) .

3.1 Feature extraction

Feature extraction is the process of discriminating the components of the signal meaningful for the application from the background noise and artifacts, and organize them efficiently for the interpretation by the experimenter and the classification by the machine learning algorithms (Dornhege et al., 2007; Wolpaw and Wolpaw, 2012b). This step leads to a dimensionality reduction of the data set, in which the features resulting from temporal, spatial and/or spectral analysis contain the relevant information of the complex brain signal. Firstly, data are typically pre-processed with initial subsampling and basic temporal filtering for storage purposes. The temporal filtering can help remove high frequency artifacts, low-frequency drifts and attenuate the electrical line noise. In this thesis, data were low-pass filtered with a Chebyshev Type II filter whose order depended on the pass- and stop- bands and on the minimum attenuation allowed in the stop-band and maximum ripple allowed in the pass-band. In some ERP applications, data can also be high-pass filtered, usually at maximum 0.5

Hz to prevent that significant slow ERP components are filtered out. The choice of the cut-off frequencies, of the filter type and of the filter parameters are crucial since filters can sensibly distort or attenuate the signal and produce artifacts, like phase-shifts (Widmann and Schröger, 2012). Then, some methods for removing artifacts originating from noisy electrodes, eye blinks, ocular and muscular activity are applied to the continuous or epoched data. These methods can be grouped in two different categories: the rejection and the projection methods. The artifact rejection methods detect epochs contaminated by artifacts and exclude them from the following analysis. They are usually based on simple criteria easy to implement and apply, but the drawback is a considerable loss of information that may contain relevant neural activity. In the projection methods, instead, the amount of data is preserved. These methods decompose the EEG signal to find the neural and artifact components, with subsequent reconstruction of the EEG ideally artifacts-free. Remarkable examples are those based on Independent Component Analysis (ICA) (Delorme and Makeig, 2004; Jung et al., 1998; Makeig et al., 1996; Romero et al., 2009; Vigário, 1997; Viola et al., 2009; Winkler et al., 2014; Wübbeler et al., 2000). After that, further methods are applied to transform and re-arrange the data into the final feature vector that would feed the classifier.

The choice of the specific methods highly depends on the application. For example, in ERP-based BCIs usually a set of spatial and temporal features are selected that maximize the discrimination between the two classes. These features can be selected either manually by the experimenter looking at the time course and spatial distribution of the ERPs or by an heuristic, e.g. based on the signed pointwise biserial correlation coefficient ($sgn - r^2$) (Section 3.1.2, for a detailed description please refer to Blankertz et al. (2011)). In case of applications in which the frequency content of the EEG signal is informative, e.g. SMR or SSVEP based BCIs, further band-pass filtering can be applied around the frequency bands of interest (usually via Butterworth filter because of its sharp transition in the magnitude response). Whenever a filter that employs feedback from the past outputs is used, i.e. IIR filter, it can be applied in one direction for online application of in both direction in offline investigation. Applying the filter in both directions corrects the phase shifts introduced by the filter. This choice cannot be done in online applications since applying the filter backwards imply the knowledge of the samples from the future. Filters can also be applied in a more sophisticated fashion, in order to enhance the neural components of interest and improve their Signal-to-Noise ratio (SNR), like in case of the Spatio-Spectral Decomposition (SSD) (Section 3.1.5). Another important step in the feature extraction is the application of *spatial filters*, e.g Laplacian Filter, Common Spatial Pattern (CSP) (Section 3.1.4). This step is often required because the EEG signal has a poor spatial resolution because of volume conduction and some signals like the SMR rhythms are weak compared to other neural components like the visual alpha rhythm or muscular artifacts (Blankertz et al., 2008c). Therefore subject-specific spatial filters can help in enhancing the neural components of interest in a specific location, while attenuating the neighboring non-relevant components

of the EEG signal. Further details about these methods follow in Section 3.1.4. Finally, when interested in the time-course of the neural oscillations, feature extraction methods that preserve information of both time and frequency can be employed. A typical example in motor imagery BCI is the visualization of the ERDs (Section 2.3.3) that can be extracted by calculation of the envelope of the EEG power taking the absolute value of its Hilbert transform (Section 3.1.6). A similar approach for time-frequency decomposition is the wavelet convolution (Section 3.1.7) and the short-time FFT (Cohen, 2014).

3.1.1 Cleaning the EEG from artifacts

Artifact rejection

Artifact rejection methods aim at identifying EEG epochs contaminated by e.g. excessive eye movements, blinks and muscular artifacts and remove them from the following analysis. The same procedure can also be used to identify loose electrodes and remove them. While being computationally quite easy and fast to perform, these methods have the drawback to reduce the size of the dataset, since less epochs and a restricted montage will be available after the rejection. In this thesis, two methods were used on the raw data. The first method rejected epochs exceeding the maximal allowed absolute difference in EEG amplitudes of a selected group of electrodes. Frontal electrodes were selected, i.e. F9 and F10 for horizontal eye movements and a pair between AF3, AF4, Fp1, Fp2 and the EOG channel for vertical eye movements. The threshold was set according to how conservative the method should be, typically it ranged between $70\mu V$ and $100\mu V$. The second method aimed at estimating epochs containing muscolar artifacts and loose electrodes. The band-power in the broad-band between 5 Hz and 40 Hz was calculated and epochs containing too high variance on more than 20% of the channels were discarded. Channels containing too small variance, i.e. lower than 0.5 μV^2 in more than 10% of the trials could also be removed.

Artifact projection: ICA with automatic artifactual component classifier method (MARA)

The MARA classification method for artifact detection was proposed by Winkler et al. (2014) and it is a valid approach for the automatic selection of artifactual components in the EEG based on ICA decomposition (for more details about ICA, please refer to Hyvärinen et al. (2004) and for a general introduction on ICA applied to EEG to Makeig et al. (1996)). The idea is that artifactual components and neuronal activity are generated independently, therefore the independent components (ICs) of the artifacts can be hopefully separated from the brain signals. MARA classifies the ICs either as an artifact or as a neural component using a linear pre-trained classifier based on six features. Three spectral features describe the deviation of a component's spectrum from the prototypical 1/f and its shape, and whether

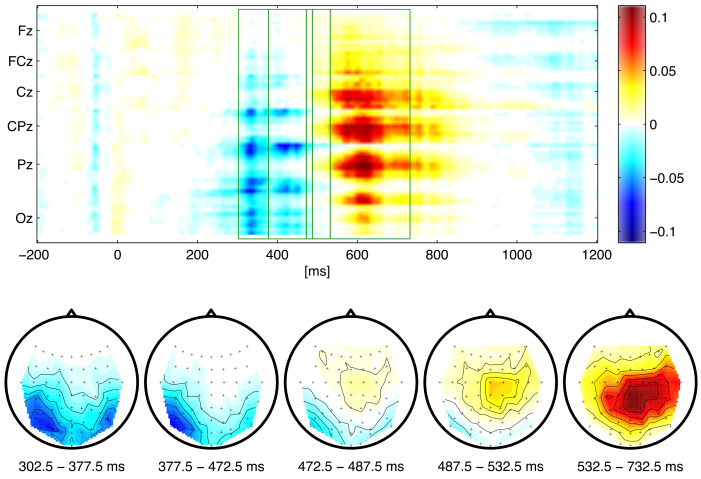


Fig. 3.1 $Sgn-r^2$ for the selection of spatio-temporal features. The matrix depicts the $sgn-r^2$ values of *target* vs *non-target* EEG epochs for each time point and electrode. The heuristic selects 5 time intervals with high values (the number of time intervals is a parameter and can be varied). The scalp plots underneath depict the distribution of the average $sgn-r^2$ over the time points of the respective intervals. Data taken from on participant of the offline RSVP Speller study (Acqualagna et al., 2010).

there is an alpha peak in components of neural origin. One feature represents outliers in the time series. The other two features extract information from the scalp map: one represents spatially localized scalp maps originating for example from muscle artifacts or loose electrodes, and the other seeks the component distribution with minimal l_2-norm, i.e. neural components.

3.1.2 Discriminability measurement: $sgn-r^2$

The point biserial correlation coefficient (r) is a correlation coefficient calculated between a continuous random variable (x) and a dichotomous variable (y).

$$r(x) := \frac{\sqrt{N_1 \cdot N_2}}{N_1+N_2} \frac{mean\{x_i|y_i=1\} - mean\{x_i|y_i=2\}}{std\{x_i\}}. \tag{3.1}$$

In this thesis the $sgn-r^2 := sign(x) \cdot r(x)^2$ is used as a measure of separability between distributions. A typical application of the $sgn-r^2$ is in the calculation of the spatio-temporal features in ERP analysis (Blankertz et al., 2011). In particular, the selection of the time intervals with high discriminability between the two classes is calculated with an heuristic that finds the intervals with high r^2 values ($Class_1$ minus $Class_2$ single trial ERPs) and fairly constant pattern (Fig. 3.1). This measure of separability can also be applied for the selection of the reactive frequency band and time intervals in a motor imagery experiment (Blankertz et al., 2006, 2007).

3.1.3 Features of ERP classification

As described in Blankertz et al. (2011), ERP components are characterized by their temporal evolution and scalp distribution and can be represented as the data matrix $\mathbf{X}^{(n)} \in \Re^{MxT}$ for each trial n, with M being the number of channels and T the number of sampled time points. Let be $C = \{c_1, \ldots, c_M\}$ a specific subset of channels. Whenever the term *spatio-temporal features* is used in the thesis, we refer to the formalism that follows. The scalp potential at channel c and time point t within trial n is denoted by $x_c^{(n)}(t)$. Given the subset of channels C, the vector of potential values for the subset of channels at time point t is defined as $\mathbf{x}_C(t) = [x_{c_1}(t), \ldots, x_{c_M}(t)]^\top$ (index of trial is omitted here). By concatenation of those vectors for all time points $t_1, \ldots, t_T$ of one trail, the following spatio-temporal features are obtained

$$\mathbf{X}(C) = [\mathbf{x}_C(t_1), \ldots, \mathbf{x}_C(t_T)]. \tag{3.2}$$

In order to reduce the dimensionality of the features for classification, it is beneficial to select specific time intervals in which the discrimination between two classes is the highest, like previously described in Section 3.1.2, and averaging across time in these time intervals. Let be $\mathcal{T}_1, \ldots, \mathcal{T}_K$ sets of time intervals. Given the selected channels C and time intervals $\mathcal{T} = \langle \mathcal{T}_k \rangle_{k=1,\ldots,K}$, the spatio-temporal features can be written as

$$\mathbf{x}(C, \mathcal{T}) = \begin{bmatrix} mean\langle \mathbf{x}_C(t) \rangle_{t \in \mathcal{T}_1} \\ \vdots \\ mean\langle \mathbf{x}_C(t) \rangle_{t \in \mathcal{T}_K} \end{bmatrix}. \tag{3.3}$$

3.1.4 Spatial filters

The scalp potential $\mathbf{x}(t) \in \Re^N$ measured by the EEG can be modeled as the sum of the linear contribution of K brain components $\mathbf{s}(t) = (s_1(t), s_2(t), \ldots)^T \in \Re^K$ multiplied by their propagation vectors that represent the strength and polarity of the coupling between the components and the sensors, aggregated in the matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots] \in \Re^{MxK}$. The coupling depends on several factors, like the orientation of the component with respect to the sensor, the distance between components and sensor, the impedance of the sensor and the conductivity of the tissue. The model takes into account also of the contributions not captured by $\mathbf{A}$ and referred as *noise* $\mathbf{n}(t)$, that can also comprise neural components not subject of the investigation:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t). \tag{3.4}$$

This model is called *linear forward model*. The propagation vector $\mathbf{a}_k \in \Re^M$ is called *activation pattern* and has a clear physiological interpretation. Haufe et al. (2014c) report that each entry of the activation pattern shows in which sensors the signal $s_k(t)$ is reflected and its value and sign are directly related to its strength and effect direction. The activation patterns can be

visualized as a scalp maps. Since the EEG data $\mathbf{x}(t)$ are known and the true components $\mathbf{s}(t)$ unknown, a *backward model* is needed to compute the linear estimate of the EEG components from the activity recorded at the sensors:

$$\hat{\mathbf{s}}(t) = \mathbf{W}^T \mathbf{x}(t). \tag{3.5}$$

The mapping from the scalp potential and the estimated components $\hat{\mathbf{s}}(t)$ is summarized in the transformation matrix $\mathbf{W} \in \Re^{MxK}$. Each column $\mathbf{w}_k \in \Re^M$ of $\mathbf{W}$ extracts one component $\hat{s}_k(t)$ and is referred to as *spatial filter* for that component. The interpretation of the spatial filter vector $\mathbf{w}_k$ is different from that of the activation pattern $\mathbf{a}_k$. The spatial filter should amplify the signal of interest and suppress at the same time the noise components, therefore they are complicated functions of signal and noise components in the data and require a complex spatial structure. For this reason, the weights of the filter cannot be used to draw conclusions about the features of the estimated component (Haufe et al., 2014c). A comprehensive review of spatial filters and details about how to reconstruct activation patterns from the filters can be found in Haufe et al. (2014c) and Parra et al. (2002), and a tutorial about spatial filters for extracting features from oscillatory activity can be found in Blankertz et al. (2008c).

Common Spatial Pattern (CSP)

The Common Spatial Pattern algorithm (Fukunaga and Koontz, 1970) is a feature extraction method that can learn spatial filters $\mathbf{W}^{CxC}$ (C being the number of channels) from band-pass filtered signals, maximizing the variance of the filtered data under one condition while minimizing it for the other condition. Since the variance of EEG signals filtered in a given frequency band corresponds to the signal power in this band, CSP analysis is applied to band-pass filtered signals in order to obtain an effective discrimination of mental states between two conditions based on band power features (Koles et al., 1990; Ramoser et al., 2000). CSP projects the signal $\mathbf{x}(t) \in \mathbb{R}^C$ from the original sensor space to the surrogate sensor space $\mathbf{x}_{CSP} \in \mathbb{R}^C$ as follows:

$$\mathbf{x}_{CSP}(t) = \mathbf{W}^\top \mathbf{x}(t). \tag{3.6}$$

Each column vector $\mathbf{w}_j \in \mathbb{R}^C$ (j =1, ... ,C) of $\mathbf{W}$ is a spatial filter; each column vector $\mathbf{a}_j \in \mathbb{R}^C$ (j=1, ... ,C) of a matrix $\mathbf{A} = (\mathbf{W}^{-1})^\top \in \mathbb{R}^{CxC}$ is a spatial pattern.

Let $\Sigma^{(+)} \in \mathbb{R}^{CxC}$ and $\Sigma^{(-)} \in \mathbb{R}^{CxC}$ be the estimates of the covariate matrices of the band-pass filtered signal in the two conditions, the CSP analysis is given by the simultaneous diagonalization of the two covariate matrices

$$\begin{aligned} \mathbf{W}^\top \Sigma^{(+)} \mathbf{W} &= \Lambda^{(+)}, \\ \mathbf{W}^\top \Sigma^{(-)} \mathbf{W} &= \Lambda^{(-)}, \qquad (\Lambda^{(C)} diagonal) \end{aligned} \tag{3.7}$$

where $\Lambda^{(+)} + \Lambda^{(-)} = \mathbb{I}$ (Fukunaga, 2013). This can be achieved by solving the generalized eigenvalue problem

$$\Sigma^{(+)} \mathbf{w} = \lambda \Sigma^{(-)} \mathbf{w}. \tag{3.8}$$

The spatial filters extremizing Eq. 3.7 are then the eigenvectors $\mathbf{w}_j$ (j=1, ... ,C) of Eq. 3.8, which correspond to the largest eigenvalues for one class and to the lowest eigenvalues for the other class. Eigenvectors with large eigenvalues close to one indicates that the corresponding spatial filter yields high variance for that condition and low variance for the other condition, respectively. Therefore, spatial filters with extreme eigenvalues maximize the difference in the variances for the two classes. Choosing D filters corresponding to extreme eigenvalues (either close to 1 or close to 0), the filtered data $\mathbf{x}_{CSP}(t) = \mathbf{W}_D^\top \mathbf{x}(t)$ will have smaller dimensionality D < C and the two classes will be maximally separated by their variance. The extracted CSP features are the logarithm of the EEG signal variance after projection onto the filters $\mathbf{W}^\top$. The classical measure for the selection of a subset of CSP filters is based on the eigenvalues. As suggested by Blankertz et al. (2008c), this measure is not robust to outliers because it is based on pooling the covariance matrices in each condition, therefore one single trail with high variance could have a strong impact on the CSP solution. In this thesis an heuristic for the selection of an optimal number of filters (up to six) was employed, described by Blankertz et al. (2008c) as more robust with respect to outliers than the classical eigenvalue score. In particular, the selection of the filters was done based on a score indicated as the ratio of medians, which measures the discriminability between the two classes:

$$score(\mathbf{w}_j) = \frac{med_j^{(+)}}{med_j^{(+)} + med_j^{(-)}} \tag{3.9}$$

where $med_j^{(c)}$ $(c \in \{+, -\})$ is the median of the variance in the $j-th$ CSP channel averaged across all trials of class c. As with the eigenvalues, a *ratio of medians* score near 1 or 0 indicates good discriminability of the corresponding spatial filter.

For a more detailed review on CSP analysis and its application to EEG signal processing, please refer to Blankertz et al. (2008c), Blankertz et al. (2008a), Lotte et al. (2007), Lemm et al. (2005), Samek et al. (2012), Sannelli et al. (2011).

CSP algorithm can produce suboptimal results in certain situations in which the class covariance matrices cannot be properly estimated. One example is when the covariance matrices have high dimensionality and the number of samples is limited. The estimation of the covariance matrix can be negatively affected also by the contamination of the EEG by artifacts, or by the variation of the extracted feature distribution, i.e. *non-stationarity problem*. Samek et al. (2014) reviewed the state-of-art variants of the basic CSP algorithm, which included algorithms for increasing the robustness of the estimation of the covariance matrix, for regularizing the CSP solution towards stationarity, for incorporating data from other subjects, end so on. The authors also introduced a unifying optimization framework in which the CSP algorithm can be formulated as a divergence maximization problem. In particular, they investigated the β-divergence, but other kinds of divergence could be applied. The fact that all quantities used in the optimization process were measured as divergences let them combine and compare different regularization schemes. The proposed framework was tested on EEG data of 80 participants of a previous motor imagery BCI study (Blankertz et al., 2010a), and it was compared to several other CSP variants considered as baseline.

3.1.5 Spatio-Spectral Decomposition

Spatio-Spectral Decomposition (SSD) is a method for the extraction of neural oscillations from multi-channel EEG recordings (it can also be applied to other neuroimaging data, for details please see Nikulin et al. (2011)), based on linear decomposition. It maximizes the signal power at a peak frequency while simultaneously minimizing it at the neighboring, surrounding frequency bins. This procedure leads to the optimization of the signal-to-noise (SNR) ratio and allows the extraction of components with a characteristic peaky profile typical for oscillatory processes. An important assumption of this approach is that noise components produce signals with a relatively broad frequency range. Algorithmically, the variance of the noise around the spectral peak of interest is minimized while simultaneously the variance of the signal at the peak frequency is maximized. In this thesis, SSD was employed as pre-processing step before CSP for dimensionality reduction instead of the commonly used Principal Component Analysis (PCA) (Haufe et al., 2014a). The advantage of SSD lies in selecting components explaining oscillations-related variance instead of just any variance as in case of PCA. Thus, the objective function optimized by the SSD filters is given by

$$argmax_{\mathbf{w}} \frac{\mathbf{w}^\top \Sigma_{sig} \mathbf{w}}{\mathbf{w}^\top \Sigma_{noise} \mathbf{w}}, \tag{3.10}$$

where Σ_{sig} is the covariance of the data filtered in the frequency band of interest, while Σ_{noise} is the covariance of the data filtered in the sidebands. Filtering in the left and right side bands can be implemented by applying a bandpass filter covering both side bands followed

by a bandstop filter cutting out the band of interest. The entire SSD filter matrix can be obtained by means of the generalized eigenvalue decomposition of Σ_{sig} and Σ_{noise}

$$\Sigma_{sig}\mathbf{w} = \lambda\Sigma_{noise}\mathbf{w}. \tag{3.11}$$

where λ denotes the generalized eigenvalue associated with the eigenvector $\mathbf{w}$.

3.1.6 Hilbert transform

The Hilbert transform allows the extraction of a complex signal from a signal that contains only a real part. In order to extract time-varying estimates of power and phase form EEG time series, the signal has to be represented as $Me^{i2\pi ft}$, which can be expressed as $Mcos(2\pi ft) + iMsin(2\pi ft)$ (*analytic signal*). Without any processing, EEG data has the form $Mcos(2\pi ft)$, i.e. an oscillatory signal that has only the real (cosine) component and no imaginary ($i$sine) component. The Hilbert transform extracts the imaginary part, $iMsin(2\pi ft)$, of a real-valued signal, $Mcos(2\pi ft)$. This is done by creating and adding the phase quadrature component to the $Mcos(2\pi ft)$ part, by rotating parts of the complex Fourier spectrum of the real-valued signal (Smith, 2007). It can be demonstrated (Cohen, 2014) that in practice it is not necessary to explicitly rotate the positive- and negative-frequency Fourier coefficients to compute the Hilbert transform, but simply the positive-frequency coefficients can be doubled and the negative-frequency coefficients set to zero.

3.1.7 Wavelet convolution

Wavelet convolution is a method for time-frequency representation that let explore the changes in the frequency structure of a signal over time (Daubechies, 1990). Many different kinds of wavelets exist, also called 'mother wavelets'. The common characteristic is that they must have values at or very close to zero at both ends, and mean value of zero. Given the mother wavelet $\psi(t)$, the corresponding family of wavelets consists of a series of son wavelets, which are generated by dilation and translation from $\psi(t)$ as follows:

$$\psi_{a,b}(t) = |a|^{1/2}\psi\left(\frac{t-b}{a}\right), \tag{3.12}$$

where a is a scale factor and b is time location. The wavelet transform of signal $x(t)$ is defined as the inner product in the Hilbert space of the L^2 norm as follows:

$$W(a,b) = \langle\psi_{a,b}(t), x(t)\rangle = |a|^{1/2}\int x(t)\psi^*_{a,b}(t)dt. \tag{3.13}$$

The asterisk stands for complex conjugate. $W(a,b)$ gives the information of $x(t)$ at different levels of resolution and also measures the similarity between the signal $x(t)$ and each son wavelet $\psi_{a,b}(t)$. $W(a,b)$ is the convolution between $x(t)$ and the wavelet function. Not

all wavelets are well suited for time-frequency decomposition of EEG data (Cohen, 2014), therefore only the Morlet wavelet will be introduced. The Morlet wavelet is a sine wave windowed with a Gaussian. Gaussian windows are good options because they have no sharp edges that produce artifact and allow the control of the trade-off between temporal and frequency precision. In the wavelet convolution, the frequency information at each time point is a weighted sum of the frequency information of surrounding time points, with the weight decreasing with increasing distance form the center of the wavelet. An important parameter to set is the number of cycles of the wavelet, that defines the trade-off between temporal and frequency precision. More number of cycles leads to higher frequency resolution, less number of cycles to higher temporal resolution. Therefore, this is a non-trivial parameter to set. Another important parameter is the number of frequencies to use, like with the Fourier transform. In fact, wavelet convolution involves using many wavelets of different frequencies keeping unvaried the other proprieties, i.e. the family of wavelets. In general, for most EEG experiments a number of frequencies between 15 and 30 (e.g. from 3 Hz to 60 Hz) is sufficient (Cohen, 2014). Importantly, in order to extract power and phase information from EEG data, complex Morlet wavelets are necessary. From the convolution of the complex Morlet wavelet with EEG data, the amplitude (and its square, the power) and the phase between the kernel and the signal can be estimated.

3.2 Classification

3.2.1 Linear Discriminant Analysis (LDA)

The simplest approach to the classification problem involves constructing a *discriminant function* that directly takes an input vector $\mathbf{x}$ and assignes it to one of the K classes C_k (Bishop, 2006). *Linear discriminants* are those for which the decision surfaces are hyperplanes. In this section, just the case with two classes will be introduced. A linear discriminant function can be obtained by taking a linear function of the input vector so that

$$y(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_0 \tag{3.14}$$

where $\mathbf{w}$ is called a *weight vector* and w_0 is a *bias*. An input vector $\mathbf{x}$ is assigned to class C_1 if $y(\mathbf{x}) \geq 0$ and to class C_2 otherwise. The corresponding decision boundary is therefore defined by the relation $y(\mathbf{x}) = 0$, which corresponds to a (D-1)-dimensional hyperplane within the D-dimensional input space. The vector $\mathbf{w}$ determines the orientation of the decision surface, while the bias parameter w_0 determines the location of the decision surface.

In LDA, which is equivalent to Fisher Discriminant Analysis, the components of the vector $\mathbf{w}$ are adjusted such that the projection of the input vector $\mathbf{x}$ into one dimension

$$y = \mathbf{w}^\top \mathbf{x} \tag{3.15}$$

maximizes the class separation as follows. Consider N_1 points of class C_1 and N_2 points of class C_2, and define the mean vectors of the two classes as $\mathbf{m}_1$ and $\mathbf{m}_2$. The simplest measure of separation of the classes when projected onto $\mathbf{w}$ is the separation of the projected class means. Therefore, $\mathbf{w}$ is chosen to maximize

$$m_1 - m_2 = \mathbf{w}^\top(\mathbf{m}_1 - \mathbf{m}_2) \tag{3.16}$$

where

$$m_k = \mathbf{w}^\top(\mathbf{m}_k) \tag{3.17}$$

is the mean of the projected data from class C_k. Moreover, in order to minimize the classes overlap when projected onto the line joining the means, the function to maximize should give a large separation between the projected class means while also giving a small variance within each class. Therefore, the objective function to maximize in LDA is

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_W \mathbf{w}} \tag{3.18}$$

where $\mathbf{S}_b$ is the between-class covariance matrix and $\mathbf{S}_w$ is the total within-class covariance matrix. It can be shown that the optimal weight vector is given by

$$\mathbf{w} \propto \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \tag{3.19}$$

This result represents a specific choice of direction for the projection of the data to one dimension, but the projected data then can be used to construct a discriminant by choosing a threshold y_0 so that a new point will be classified as belonging to C_1 if $y(\mathbf{x}) \geq y_0$ and classified as belonging to C_2 otherwise. For Gaussian distributions with the same covariance matrix of all classes, it can be shown that LDA is the optimal classifier in the sense that it minimizes the risk of misclassification for new samples drawn from the same distributions (Duda et al., 2012). Though more sophisticated discriminant functions exist (Müller et al., 2003), the assumptions of LDA appear to be satisfied for EEG data, e.g. ERP features (Blankertz et al., 2011).

Shrinkage for improved classification

Let $\mathbf{x}_1, ..., \mathbf{x}_n \in \Re^d$ be n feature vectors, the standard estimator of the covariance matrix is the empirical covariance matrix

$$\hat{\mathbf{\Sigma}} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^\top \tag{3.20}$$

This estimator is unbiased and has good proprieties under usual conditions. However, for high-dimensional data and only a limited amount of data points available, the sample estimator might become imprecise, since the number of unknown parameters that have to be estimate is quadratic in the number of dimensions. This leads to a systematic distortion that degrades classification performances: Large eigenvalues of the original covariance matrix are estimated too large and small eigenvalues are estimated too small. In BCI, this issue mainly affects LDA classifiers trained on ERP data (see Blankertz et al. (2011); Tomioka and Müller (2010); Vidaurre et al. (2009) for detailed discussion). Therefore, a regularization term can be introduced in order to compensate for these distortions

$$\tilde{\mathbf{\Sigma}} := (1-\gamma)\hat{\mathbf{\Sigma}} + \gamma\nu\mathbf{I} \tag{3.21}$$

where γ being the regularization parameter $\gamma \in [0,1]$ and ν being the scaling parameter, defined as the average eigenvalue of $\hat{\mathbf{\Sigma}}$. LDA with such modified covariance matrix is called LDA with shrinkage (also known as covariance-regularized LDA or regularized LDA). With the use of such regularization, it can be demonstrated that extreme eigenvalues are modified towards the average ν. A regularization parameter $\gamma = 0$ yields to unregularized LDA, while $\gamma = 1$ assumes spherical covariance matrices. The key question in this procedure is how to select an optimal value for the shrinkage parameter (Schäfer and Strimmer, 2005). Often the appropriate way to choose γ is in a data-driven fashion, minimizing a risk function like the Mean Square Error (MSE) between the shrunk and the true covariance matrices

$$R(\gamma) = E(\sum_{i,j}(\tilde{\Sigma}(\gamma)_{i,j} - \Sigma_{i,j})^2) \tag{3.22}$$

One common but also computationally very intensive approach to estimate the minimizing γ is by using cross-validation (Friedman, 1989). Alternatively, the optimal regularization parameter may be determined analytically (Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005), which is computationally less intensive:

$$\gamma^\star = \frac{\sum_{i,j}\{Var(\hat{\Sigma}_{i,j}) - Cov(\hat{\Sigma}_{i,j}, \nu I_{i,j})\}}{\sum_{i,j} E[(\hat{\Sigma}_{i,j} - \nu I_{i,j})^2]}. \tag{3.23}$$

The method of regularized LDA with shrinkage parameter selected by the analytical solution of Schäfer and Strimmer (2005) is called *shrinkage LDA*.

3.2.2 Classification accuracy

Two different accuracy measures are used throughout this thesis. One is the *single-trial accuracy*, which assesses the performance of the classifier on individual epochs. In case of two classes, the single-trial accuracy is a binary measure that refers to the ability of separating epochs belonging to class C_1 to epochs belonging to class C_2. In the following chapters the

quality of a binary classifier will be assessed by the Area Under the Curve (AUC) of the Receiver Operator Characteristic, also known as ROC curve (Fawcett, 2006). The second accuracy measure is the *selection accuracy* and is evaluated on run-level after accumulating the classifier outputs from the individual epochs for that run. It reflects the final prediction of the BCI and indicates the percentage of correctly classified runs, i.e. when the final chosen class matches the target class that the user wanted to select.

Area Under the ROC curve (AUC)

The ROC curve is a way of graphically displaying the performance of a binary classifier when its discrimination threshold is varied (Hanley and McNeil, 1982). The curve is created by plotting the true positive rate (TPR), also known as *sensitivity*, versus the false positive rate (FPR), also known as *1-specificity*. The TPR defines how many correct positive results occur among all positive sample available during a test (e.g. a patient with a disease who is correctly diagnosed as having the disease), while the FPR defines how many incorrect positive results occur among all negative samples (e.g. a healthy patient wrongly diagnosed as having the disease). The ROC space is defined by FPR and TPR as x and y axes respectively and the ROC curve plots parametrically the TPR versus the FPR varying the discrimination threshold parameter. The AUC can be regarded as a way to reduce the ROC curve to a single scalar value, in order to compare classifiers. The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than as randomly chosen negative instance. Since the AUC is a portion of the area of the unit square (of the ROC space), its value will always be between 0 and 1. However, since random guessing is represented in the ROC space as a diagonal line between (0,0) and (1,1) with area of 0.5., a realistic classifier should have AUC>0.5.

Selection accuracy

In many BCI systems (like the BCI Spellers described in Chapter 5), each run usually includes several repetitions of the target and non-target classes in order to have a more robust prediction of the final target selected by the BCI. For example, in case of thirty stimuli, e.g. thirty different letters, for a specific run there is one stimulus that belongs to the target class and 29 stimuli that belong to the non-target class. Therefore, the single-trial classification outputs are accumulated and sorted according to their stimuli membership and repetition. The ten classifier outputs corresponding to the same stimuli are averaged together to get one single value for each letter. Among the thirty values, the minimum will be selected as classified target latter, since for convention negative classifier outputs refer to the target class. In this case the accuracies are expressed in percentage and the level referring to random classification (chance level) is calculated as 100%/N, with N being in this example the number of different letters (i.e. chance level=3.33%).

3.2.3 Validation

Throughout this thesis both offline and online experiments are described. In order to validate the classification models, different approaches are used according to the type of study. In offline analyses, typically a *cross-validation* approach is employed. Cross-validation is a technique that estimates the predictive value of a model in a way that generalizes to unseen data (Lemm et al., 2011; Müller et al., 2001). The dataset is divided into N disjoint folds (N-fold cross-validation) and the classification is run N times. For each time, N-1 folds are used to train the model parameters and the remaining fold is used to test the performance of the model's prediction. The method is run until each fold has been used exactly once as test-set. The final prediction accuracy of the model is obtained by averaging the accuracies in the test-sets. Besides, the cross-validation can be repeated K times shuffling the whole dataset and dividing it into N folds like already described. The final accuracy will be derived by averaging the performance across the K shuffles. Cross-validation is a very useful method to estimate the classifier performances especially if the size of the dataset is limited, since it exploits the available data in an optimal way. If the number of samples is large compared to the number of features, an alternative could be dividing the dataset just in one training set and one test set, like in an online operation. In an online experiment, the parameters of the classification model are learned using the data collected during the calibration phase. The classifier is trained on the dataset in cross-validation mode and gives a first estimation of the model performances (typically binary classification results are reported). At this point, typically the experimenter decides whether to change the selection of the features used for the training of the classifier. In case that the experimenter changes the features[1], then the classifier is re-trained on the new features and a new performance estimation is given. It is important to notice that the accuracy obtained by cross-validation of the calibration data is just a rough indication of the performances that we can expect from the user, since many other factors play a role in the online validation, e.g. motivation, frustration, tiredness, etc., which can change the feature space of the new test data. For a more detailed discussion about the problematics of the classifier transfer between offline calibration and online validation, the importance of the appropriate feature selection strategy, please see Chapter 7.

[1]For example, in some experiments described in this thesis the parameters were first chosen automatically using the heuristic described in 3.1.2, then they could have been manually modified in order to check whether better performances could be achieved.

Chapter 4

BCIs for video quality assessment

4.1 Introduction and state of the art

Perceived quality of a set of images or videos is usually assessed using some opinion tests, in which participants are asked to rate the quality of a given visual stimulus on a rating scale ITU (2002, 2008). This type of behavioral tests has some well known limitations. First, they require a large number of participants to obtain statistical significance and the results often suffer from large inter-subject variance. Moreover, the judgment can be biased by several factors not depending on the quality of the stimuli, such as the nature of the task, the rating scale used, the internal state of the participant, expectations, emotions. A possible way to quantify visual distortions more objectively would be modeling some key features of the visual system (Jayant et al., 1993; Seshadrinathan and Bovic, 2010). However, a precise model of subjective perception is not yet available and the process underlying the judgment itself is not fully understood. Therefore, in recent years, there has been an increasing interest to assess perceived image quality through the direct measurement of brain signals via EEG. EEG has been widely exploited in visual and audio perception research (Babiloni et al., 2006; Busch et al., 2009; Demiralp et al., 2007; Porbadnigk et al., 2013a) and recently also for assessing user's perceived multimedia quality (Antons et al., 2012; Arndt et al., 2014; Bosse et al., 2014; Hayashi et al., 2000; Kroupi et al., 2014; Lindemann and Magnor, 2011; Lindemann et al., 2011; Moldovan et al., 2013; Mustafa et al., 2012; Norcia et al., 2014; Perez and Delechelle, 2013; Porbadnigk et al., 2010, 2011, 2013b; Scholler et al., 2012). Most of these studies focus on the measurement of specific ERP components like the P3, which is largely independent on the sensory modality. They demonstrate that an EEG-based approach in classifying quality changes or artifacts in videos and images is a feasible measurement that, together with behavioral tests, can lead to a more objective rating of perceived quality. Notwithstanding, all studies that capitalize on the P3 component have several limitations:

- The P3 component is not directly linked to sensory processing, but reflects higher cognitive processing of the stimulus.

- P3-based experiments usually require a high number of trials in order to have a sufficient number of target events and a good signal-to-noise ratio.

- The design of such experiments needs to be carefully done. For example, the target events have to be sufficiently distant in time between each other, in order to significantly modulate the brain signal (Duncan-Johnson and Donchin, 1977; Sellers et al., 2006).

4.2 Video quality assessment based on SSVEPs

This chapter presents a novel SSVEP-based paradigm in which the flickering effect was caused by changes in video quality (owing to compression). The goal was to elicit a direct visual response to the quality changes and investigate whether a meaningful quantification by single-trial classification using machine learning techniques could be obtained. A characteristic that makes SSVEPs preferable to P3 as stimuli in video quality assessment is that SSVEPs are directly linked to the sensory processing and can give a more straightforward indication of the level of perception of the visual stimuli. Moreover, for the reasons described above, a P3 approach requires long inter (target-) stimulus intervals in order to achieve a suitable signal-to-noise ratio, while SSVEP stimulation allows to collect a large number of trials within a short amount of time. For example, in the study of Scholler et al. (2012), a total of 600 ERPs was collected in 120 minutes, while more than 10.000 ERPs (single VEPs of the SSVEPs) can be collected using the proposed SSVEP approach within the same amount of time. Of course, such a comparison can only give an indication of the potential to speed up the assessment by using SSVEPs, since comparing just the number of trials ignores that both studies had different stimulus material and presumably a different signal-to-noise ratio in single-trials.

The standard behavioral test for video quality assessment was performed, either before or after the EEG measurement and the Mean Opinion Scores (MOS) for the same group of textures was collected. Therefore, the classified SSVEP features could be correlated not only with the quality changes intrinsic to the video signals, but also with the judgments of the participants. Textures in condition of natural luminance were used as stimuli, instead of artificially generated stimuli. This makes the stimuli more realistic and the experimental condition closer to real world video sequences. Yet, the textures were simple enough to avoid influences of semantic content, and chosen to be homogeneous in order to minimize the effects of eye movements and to avoid that attention was captured by salient objects not related to the purpose of the study.

Data and results of this chapter were previously published in Acqualagna et al. (2015).

4.3 Methods

4.3.1 Stimuli

Six gray-level texture images (Kylberg, 2011; Ojala et al., 2002) were chosen as the basis for stimulus generation (Figure 4.1). The size was 512x512 pixels and they all had the same mean luminance. In order to make the measurement independent of the image statistics and of the actual gaze position during the experiment, the texture images were spatially roughly stationary. The quality of each texture image was then degraded in six different levels. The distortions were introduced by coding the textures using the HM10.0 test model (JCT-VC, 2014) of the emerging High Efficiency Video Coding (H.265/MPEG-HEVC) standard (Sullivan et al., 2012). In this standard, statistical redundancies were exploited by block-wise temporal and spatial linear prediction. The residual signal was transformed block-wise, and coefficients were quantized in the transform domain. The quantization was controlled by the Quantization Parameter (QP). Coding artifacts, which are perceived by the human observer as a loss of visual quality, were introduced by the quantization of the transform coefficients. In order to investigate how the visual cortex responds to distortions at the threshold of perception, the first three distortion levels were chosen to be perceived as high quality. The QP-values used in the experiment were estimated in a pilot study in order to meet consistent MOS values. All the texture images in all the different levels of degradation were displayed as videos, 117 s long. Details about the structure of the videos are described in Section 4.3.2.

4.3.2 Experimental design

Participants

Sixteen participants (seven females and nine males, in the age group 21-46) took part in the experiment. All had normal or corrected-to-normal vision and none of them had a history of neurological diseases. They were all native German speakers or at least with a level of German comprehension of five, on the six level scale of competence laid down by the Common European Framework of reference for Languages (Little, 2007). All of them were naïve in respect of video quality assessment studies and were paid for their participation. The study was performed in accordance with the declaration of Helsinki and all participants gave written informed consent.

Apparatus

The electrodes configuration used in this study was: Fp1,2, AF3,4,7,8, Fz, F1-10, FCz, FC1-6, FT7,8 Cz, C1-6, T7, CPz, CP1-6, TP7,8, Pz, P1-10, POz, PO3,4,7,8, Oz, O1,2. The electrode that in the standard EEG montage is placed at T8 was placed under the right eye and used to measure eye movements. The stimuli were shown on a 23" screen (Dell U2311H) with a native

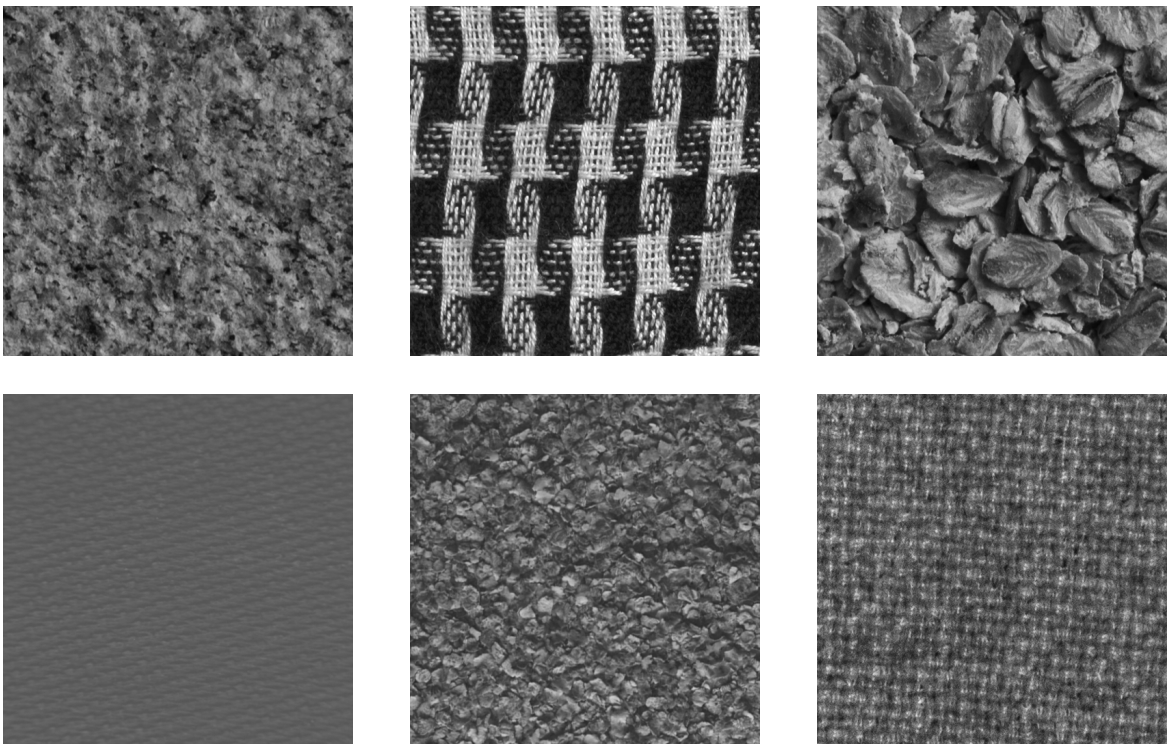


Fig. 4.1 Stimuli. Six natural images were chosen as basis for stimuli generation. The textures in the upper row represent stone, scarf and oatmeal, in the lower row grey rubber, grey flakes and blanket, all in their undistorted form. They have been degraded in six level of quality coded with the HM10.0 test model HEVC standard and grouped together to form videos with a frame rate of 3 Hz.

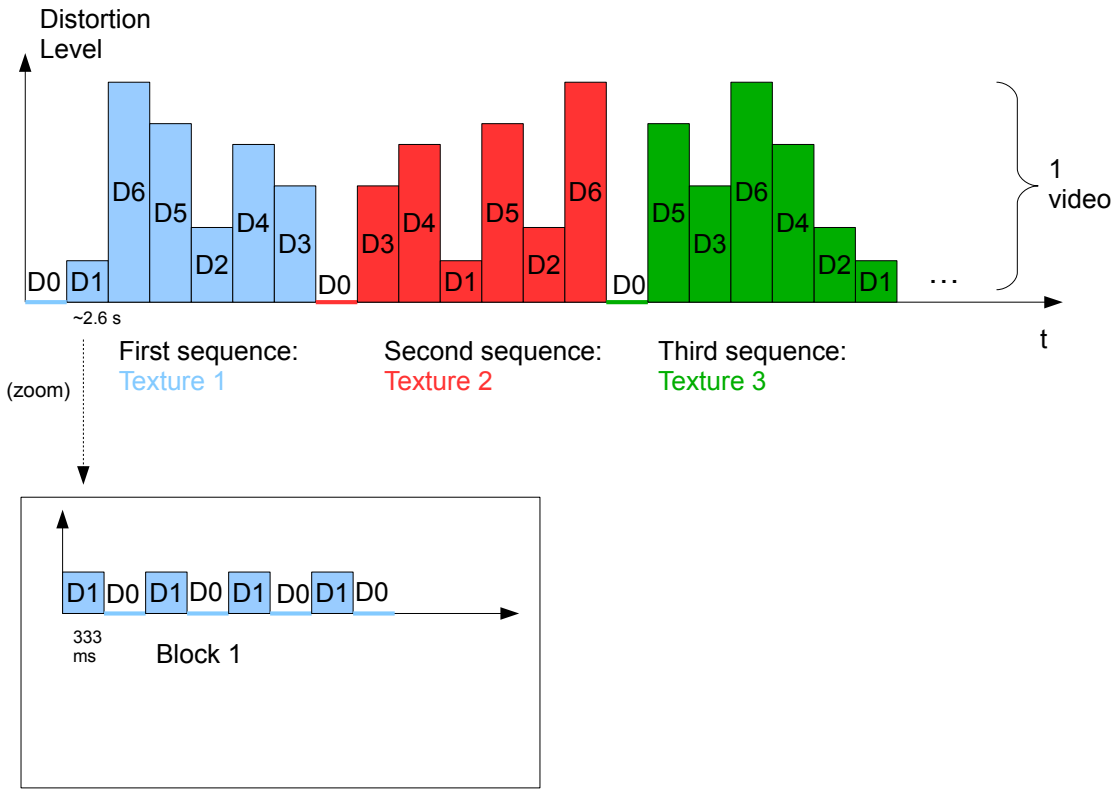


Fig. 4.2 Video structure. Each video (trial) comprised the six textures presented in all the levels of distortion (D1, ... , D6) in random order. Each texture was displayed distorted for 333 ms, followed by the undistorted form for 333 ms (D0) and the same succession was repeated four times for each level. Such alternation of quality changes produced a flickering effect eliciting SSVEPs when perceived by the participant.

resolution of 1920x1080 pixels at a refresh rate of 60Hz. The screen was normalized according to the specifications in ITU (2002). The stimuli resolution was 512x512 pixels (128x128 mm), which corresponds to 7.15° visual angle. The size of the images in the behavioral part of the experiment was the same as in the videos. The viewing distance was 110 cm, in compliance with specifications in the ITU-T Recommendation P.910 ITU (2008). Participants sat in front of the display in a dimly light room.

Procedure

The experiment consisted of an EEG measurement and a behavioral part. After a general introduction to the experiment and the preparation of the EEG cap, half of the participants started with the EEG recording and half with the behavioral assessment. In the EEG part, they had to watch a series of 51 videos, divided into three runs (20+15+16). Between each

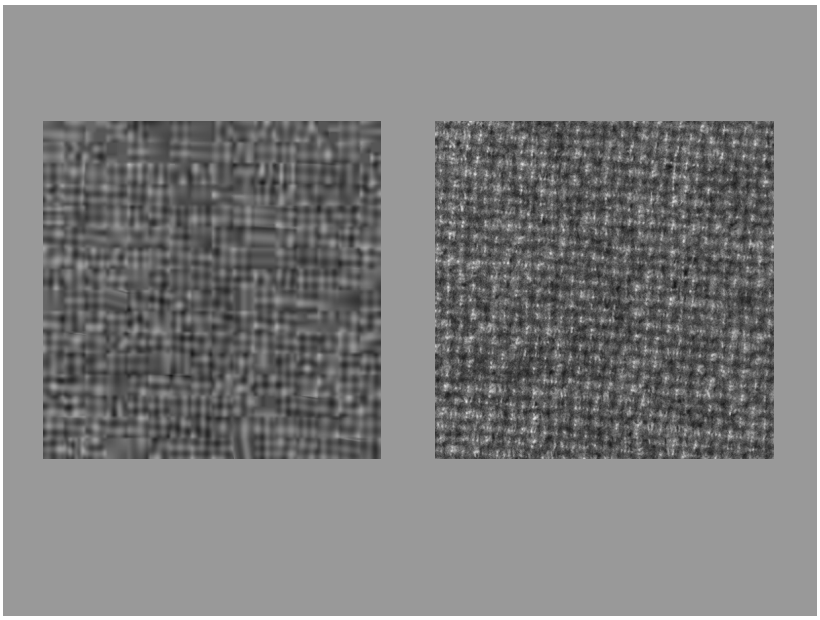


Fig. 4.3 Behavioral assessment. Textures were presented in random order, the distorted form on the left side of the display and the undistorted on the right (here an example of blanket with degradation level D6). Participants had 10 s to look at them in free viewing condition before the evaluation of the quality level. Each level of distortion was presented for three times, for each texture.

run, there was a break of about 10 minutes, to give the participants the chance to relax and stretch. Each video had a length of 117 s , followed by a pause of 5 seconds. Each video began with a fixation cross that was displayed for 3 s at the center. In order to minimize artifacts, participants were instructed to not move their eyes during the presentation of the video and to blink as little as possible. Figure 4.2 depicts the structure of one video. Each video comprised all six textures and degrees of quality, presented in random order. The stimulus onset asynchrony (SOA) was 333 ms, that is, each texture image was displayed for 333 ms. At the beginning, the texture was presented in its undistorted form (D0) for 2664 ms (333 ms x 8). Then, the first quality change occurred and the distorted texture was displyed for 333 ms, followed by 333 ms of the same texture in its undistorted version. This cycle *distorted-undistorted* was repeated four times for a total of 2664 ms. Then, the same texture was displayed with another level of quality change, for a cycle *distorted-undistorted* of the same length. This procedure was performed until all the distortion levels were displayed for that texture (randomized order). After that, the texture was switched and the new one was displayed at the beginning in its undistorted version for 2664 ms, before starting the cycle of quality changes. This presentation elicited SSVEPs if the changes due to altered quality were processed in the visual cortex. A SOA of 333 ms resulted in a flickering frequency of 3 Hz. In the behavioral part of the experiment, participants had to evaluate the perceived quality of the textures, following the standardized Degradation Category Rating quality assessment

(ITU, 2002) in a presentation mode. Each texture was presented in the display in pairs for 10 sec: on the right hand side in its original undistorted version (reference) and simultaneously on the left hand side with changed quality (Figure 4.3). After that, a new window was displayed, with the nine-grade degradation (distortion) scale, according to the ITU-T P-910 recommendation (ITU, 2008). The scale was displayed in German language, and in English can be translated in the following: 1- Very annoying; 3- Annoying; 5- Slightly annoying; 7- Perceptible, but not annoying; 9- Imperceptible. Grade 8 is commonly interpreted as the perceptibility threshold, that is the distortion level where the observer is not completely sure to perceive the distortion. Participants had up to 10 seconds to decide about the level of distortion of the previously displayed left image compared to the reference one on the right, scrolling a bar until the selected grade and confirming by button press. After that, the presentation switched to the next pair of textures. If the person did not make any choice within the 10 s, the presentation automatically went ahead with the next comparison. In the behavioral assessment, each texture image was presented in all the levels of distortions (comprising the comparison reference-reference). For each level, there were three evaluations. The order of the evaluations was randomly shuffled. At the beginning of the assessment, a calibration block was displayed in order to make the participants confident with the test: each texture was displayed for just 2 evaluations, worst quality level vs reference and reference vs reference, for a total of 12 calibration evaluations. Like in the actual behavioral assessment, in this short test participants were not aware of the quality level of the displayed textures. The data of this calibration block were not considered in the analysis.

4.3.3 Preprocessing and data analysis

The EEG signal was lowpass filtered from 0 to 40 Hz with a Chebyshev filter Type II of order ten (3 dB of ripple in the passband and 40 dB of attenuation in the stopband) and down-sampled to 100 Hz. For the SSVEP visualization, the continuous signal was divided into epochs ranging from 0 ms to 2664 ms, relative to the onset of the first distorted texture for each quality change. Each epoch comprises all four repetitions of the same quality level. Epochs referring to the same distortion level were averaged over all the videos and all the textures. In the frequency-domain analysis, EEG spectra were calculated between 1 Hz and 18 Hz for all epochs and then averaged for each distortion level, over all videos and textures. For single-trial offline classification, two different methods were tested. The first one exploited the oscillatory nature of the SSVEPs and used Common Spatial Pattern (CSP) analysis (Section 3.1.4). The second one is based on methods used in ERP analysis and exploits spatio-temporal features (Section 3.1.2).

CSP method

CSP was used to extract spatial filters in order to enhance the signal of interest in the occipital cortex. The CSP filters were calculated between the epochs of maximum distortion level, which will be named class D6, and epochs of the undistorted level, which will be named class D0. Since the performance of this spatial filter depends on the operational frequency band of interest, manually selecting a specific frequency range is commonly used with the CSP algorithm (Ang et al., 2008; Dornhege et al., 2006). In this case, CSP was performed after filtering the data with a 5-th order Butterworth filter bank at 3 Hz (pass-band 2-4 Hz) and at 6 Hz (pass-band 5-7 Hz), applied forward and backward. The filter bank concatenated data filtered at 3 Hz and 6 Hz, which were subsequently spatially filtered with the respective CSPs. In this way, features referring to both frequencies could be exploited simultaneously. Continuous EEG signal was divided into epochs of 667 ms length, time-locked to the onset of the distorted image. For each texture and level of distortion, the four repetitions of the cycle *distorted-undistorted* were averaged. For D0, the first block of epochs at the beginning of each video was discarded, because it is often affected by artifacts due to subject's movements between the videos. One up to three CSP filters per class were automatically selected for each subject. The CSP filters that maximize the variance for class D6 while minimizing the variance for D0 were used. The selected filters were then applied to the training epochs comprising D6 and D0 and the log-variance in three equally spaced intervals of the filtered data was used as feature matrix for training a classifier based on linear discriminant analysis (LDA) (Section 3.2.1). The CSP filters were then applied to the testing epochs, from D1 to D6, and classification was made between each level of distortion and D0. Classification was performed using five-fold cross validation with five shuffles.

Spatio-temporal features method

In this method, single-trial visual evoked potentials time-locked to the onset of the distorted texture were classified, using spatio-temporal features. The EEG data was divided into epochs as follows: for the 'Class 1' events, epochs were time-locked to the onset of the distorted textures. For the 'Class 2' events, the starting point of epochs was shifted 160 ms after the onset. Being the period of the oscillatory visual response of about 333 ms, a shift of 160 ms would lead to a high discrimination between the two classes. In both classes, the length of the epochs was 667 ms. For feature extraction, five temporal windows were selected individually for each participant by the heuristic based on the $sign - r^2$. Features were calculated from 36 channels (FC1,3,z,2,4, C1,3,5,z,2,4,6, CP1,3,5,z,2,4,6, P1,3,2,4 Pz,7,9,8,10, PO3,7,4,8, O1,z,2) by averaging voltages within each of the five chosen time windows resulting in 36x5 = 180 dimensional feature vectors. Classification was performed using ten-fold cross-validation with ten shuffle and LDA with shrinkage of the covariance matrix (Section 3.2.1).

1	Very annoying
2	
3	Annoying
4	
5	Slightly annoying
6	
7	Perceptible but not annoying
8	
9	Imperceptible

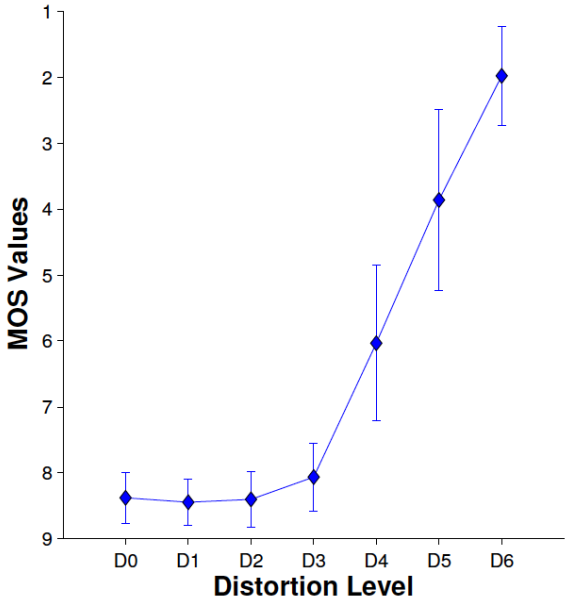


Fig. 4.4 Behavioral assessment. Left: table representing the nine level scale of scores that the participants could give to the quality of the distorted texture compared to the undistorted. Right: mean opinion scores given by participants for each quality level. Error bars refer to the standard deviation from the mean.

Classification results were correlated with the average magnitude of the alpha rhythm during the experiment. The alpha peak was searched within a range of frequencies. In order to estimate alpha power, the difference between the value of the alpha power and the linear interpolation between the flanking frequencies (as baseline) was calculated. For this estimate, flanking frequencies have been searched within the intervals 6-10Hz and 11-13Hz, and the alpha peak was determined as maximum of the spectra between the flanking frequencies. For all participants, alpha peak was detected between 9-12 Hz.

4.4 Results

4.4.1 Behavioral data

Figure 4.4 displays on the left the nine-degradation scale of the MOS, in which grade 8 corresponds to the perception threshold of the degradation, that is the level where the

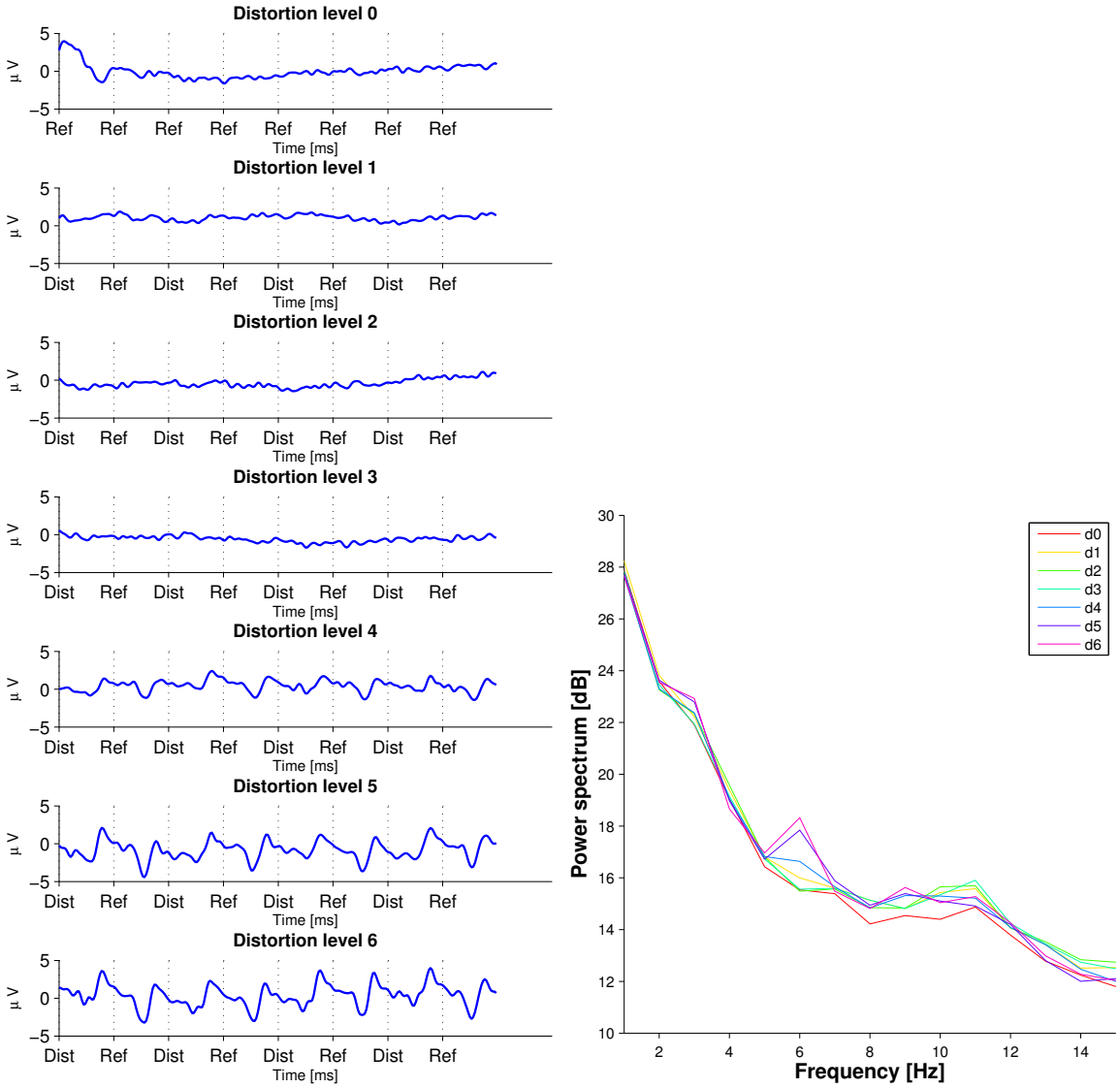


Fig. 4.5 SSVEPs. Left: Brain activity of participant VPib averaged over videos and textures at channel Oz is displayed from D0 to D6. *Dist* refers to the onset of the distorted texture, *Ref* to the undistorted. Right: Spectra of the brain activity of participant VPib at channel Oz averaged over videos and textures.

observer is not completely sure anymore to perceive any degradation. On the right, the mean MOS values over all the participants, textures and repetitions is plotted as a function of the distortion level. The plot shows that on the behavioral level the participants were not able to discriminate the two most subtle distortion levels (D1 and D2) from the reference image (D0). For level D3, the mean MOS is lower but still remains above the value of the perception threshold. From distortion level D4 to D6, the mean MOS values decrease linearly with the level of distortion, and the error bars also display an increase in variability among participants.

4.4.2 Neurophysiological data

Figure 4.5 (left) displays representative SSVEP waveforms of participant VPib recorded at an occipital scalp site (electrode position Oz). The plots are ordered according to increasing distortion levels, that is, the top row refers to the reference texture of highest quality (D0) and the bottom row to the maximum distortion level (D6). Each plot represents the average EEG activity over all the videos and textures at that specific quality level. The time zero is locked to the onset of the first repetition of the texture of each block. As described in the Section 4.3.2, the time interval between the onset of each frame is 333 ms. For the first three levels of distortion (D1-D3), which are around the perception threshold, the ongoing EEG activity is not visibly modulated by the quality change. From D4 onwards, the SSVEPs become clearer and their amplitude increases with increasing distortion. In all the plots where the SSVEPs are evident, it can be noticed that the onset of the reference images (Ref) elicits a more pronounced negative peak than the onset of the distorted ones, at the same latency. This result suggests that the transition Dist-Ref has a stronger impact on the modulation of the visual evoked potentials than the transition Ref-Dist. The results of the analysis in the frequency domain analysis of the SSVEPs (electrode position Oz) of the same participant are displayed in Figure 4.5 (right). Power spectra were calculated on single trials before averaging over the videos and textures. The plots represent the average power for each level of distortion, coded with different colors. For the first three levels of distortion (yellow, light green, emerald green) there is no clear increase of the power in any of the frequencies of interest (3 Hz and higher harmonics). From D4 to D6 (blue, violet, magenta), the spectra display two clear peaks at 3 Hz and 6 Hz, whose amplitudes increase with increasing distortion level. At D5 and D6 a small peak at 9 Hz becomes visible, but this modulation is much smaller compared to the first two harmonics (therefore not taken into account in further analysis).

4.4.3 Classification

CSP filtering is a gold standard of the processing of EEG oscillatory signals in BCIs. In the same way, spatio-temporal features have been successfully used in the analysis of ERPs.

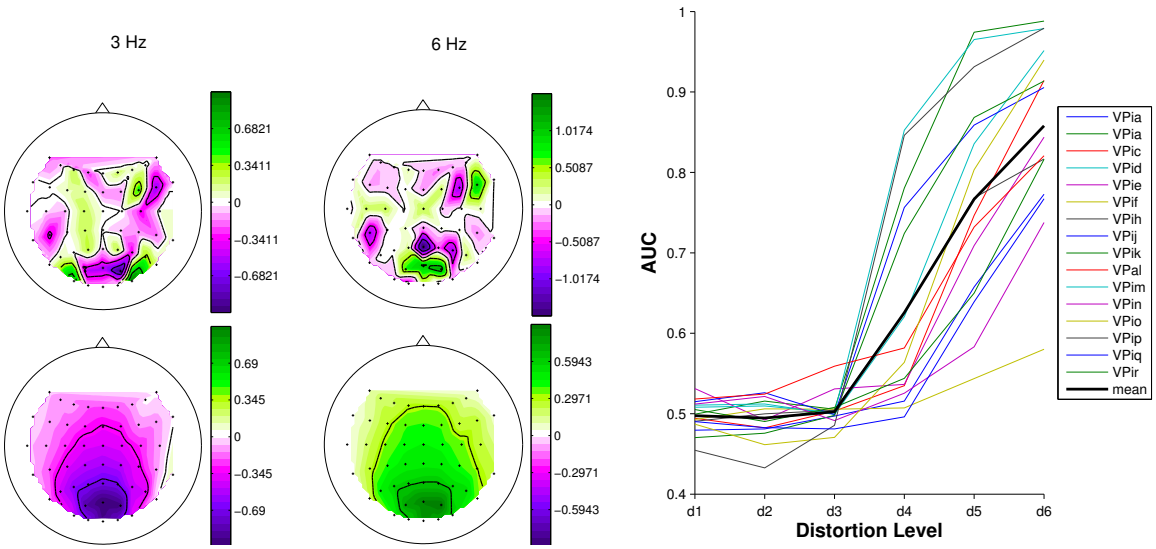


Fig. 4.6 CSP analysis. Left and middle: Scalp plots display the CSP filters (upper row) and patterns (lower row) of participant VPib, for brain activity filtered around 3 Hz (left) and 6 Hz (middle). Note that the colormap has no direct association to signs because the signs of the vectors are irrelevant in this analysis. Right: Classification performances using LDA after CSP filtering, for all participants (colored lines) and average (black thick line). The x axis represents the class that was classified versus D0, and the y axis the classification performances.

Both processing methods in combination with LDA (with shrinkage when necessary) were considered in this study in order to exploit both natures of the SSVEPs.

CSP method

Figure 4.6 displays the results of the CSP analysis as color coded scalp topographies (for one representative participant, VPib). For each participant, only those CSP filters were selected which maximized the variance of D6 while minimizing the variance of D0, for each frequency band (i.e. centered in 3 Hz and 6 Hz). The patterns clearly display that the highest variance of the neural signal takes place in the occipital cortex, where the visual information is processed. After the training of the LDA classifier and the application of the CSP filters to the testing data, the classifier was evaluated for all the levels of distortions. The results are displayed in the plot in Fig. 4.6 (right). For all participants, the mean classification performance for the first three levels of distortion (D1 to D3) is around chance level and not affected by the quality change. Between D4 and D6, the mean classification performance increases linearly and significantly with increasing distortion ($p<0.01$), reaching at D6 0.86 (SD=0.11). Average performances at D5 and D4 are 0.77 (SD=0.13) and 0.63 (SD=0.12). Note that some participants have performances at chance level also at D4.

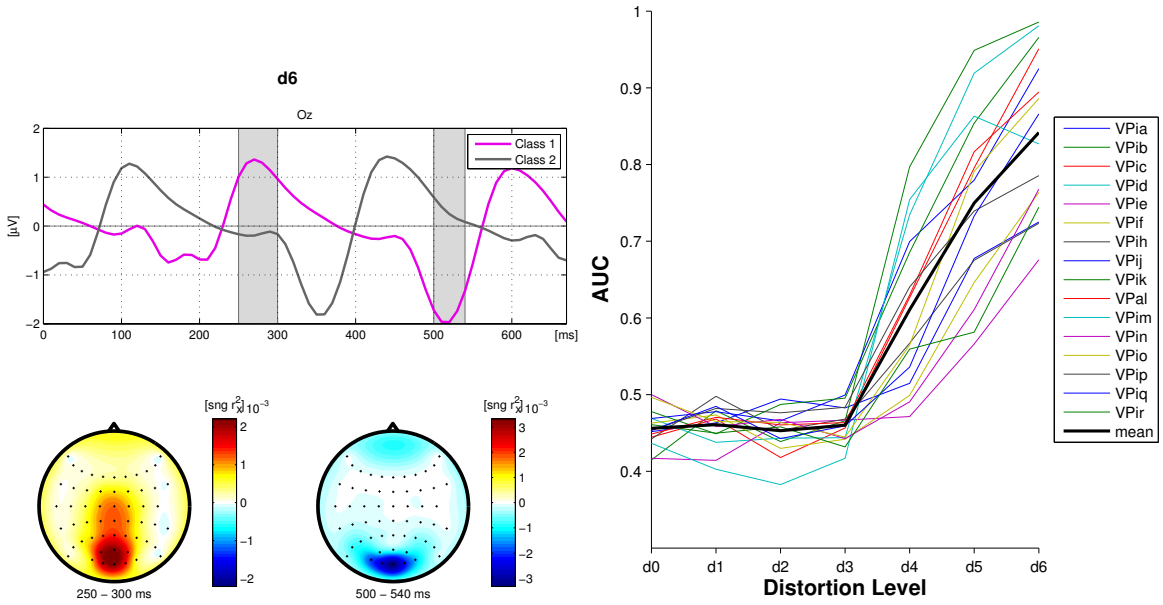


Fig. 4.7 Classification based on spatio-temporal features. Left: Grand average brain activity over all participants at channel Oz, at maximum distortion level D6. The magenta line represents Class 1 and the gray line Class 2. Scalp plots underneath refer to the shaded areas in the time plot and display the magnitude of the $sign - r^2$ for each channel. Right: Classification performances using shrinkage LDA for all participants (colored lines) and mean (black thick line).

Spatio-temporal features method

In the second method of classification based on spatio-temporal features, epochs of 667 ms length were considered time-locked to the onset of the distorted images for Class 1, and with a lag of 160 ms for Class 2. Figure 4.7 (left) displays the grand average of evoked potentials over all participants and videos for distortion level D6. Considering just Class 1, a first negative deflection is visible between 150-210 ms after the onset of the distorted image, followed by a pronounced positive peak around 270 ms. As already displayed in Figure 4.5, the onset of the reference texture at 333 ms elicits a more pronounced negative visual evoked potential than that elicited by the onset of the distorted texture. The scalp plots underneath display the topographies in two time intervals where the class difference is large. They visualize the distribution of the sgn-r^2 values as a measure of discriminability between Class 1 and Class 2, which is highest in the occipital cortex with focus around the central channels. In other words, the visual processing of the stimuli leads to maximal discrimination between good and degraded quality. A mean AUC of 0.84 (SD=0.1) is achieved at the maximal distortion level D6. At D5 and D4 the mean AUC drops to 0.75 (SD=0.12) and 0.61 (SD=0.1) respectively. The mean AUC at D3 down to D0 are around chance level. Figure 4.7 (right) displays the trend of the classification performances for all participants (colored lines) and the mean (black thick line) as a function of the distortion level. From distortion D3 upwards, the mean performance increases significantly with the level of distortion ($p<0.01$). Since classification is based on the spatio-temporal features derived from the evoked potentials, this trend tightly follows the trend of the modulation of the visual evoked potentials in the occipital cortex. The SSVEP plots in Figure 4.5 refer to participant VPib, who reaches the highest classification performance in the spatio-temporal classification method, that is 0.99, and also in the CSP method, that is 0.99. Participant VPib shows clear visual evoked potentials and spectra peaks at both 3 Hz and 6 Hz.

MOS values significantly linearly correlated with classification performance obtained using the spatio-temporal features for all the participants ($p<0.01$). MOS values were also correlated with the performance obtained by the CSP method with filter bank, which takes into account contemporary both the frequencies of interest. Also in this case, a significant linear correlation ($p<0.01$) was found for all participants.

4.4.4 Investigating individual differences

While for the first three levels of distortion there are no substantial inter-participant differences in classification results, the variability of the classification performance at higher levels of distortion becomes statistically significant ($p<0.05$). For example, in the classification based on spatio-temporal features four participants reach a mean classification performance of more than 0.95 for the maximal level of distortion D6, while five participants never exceed 0.75. The

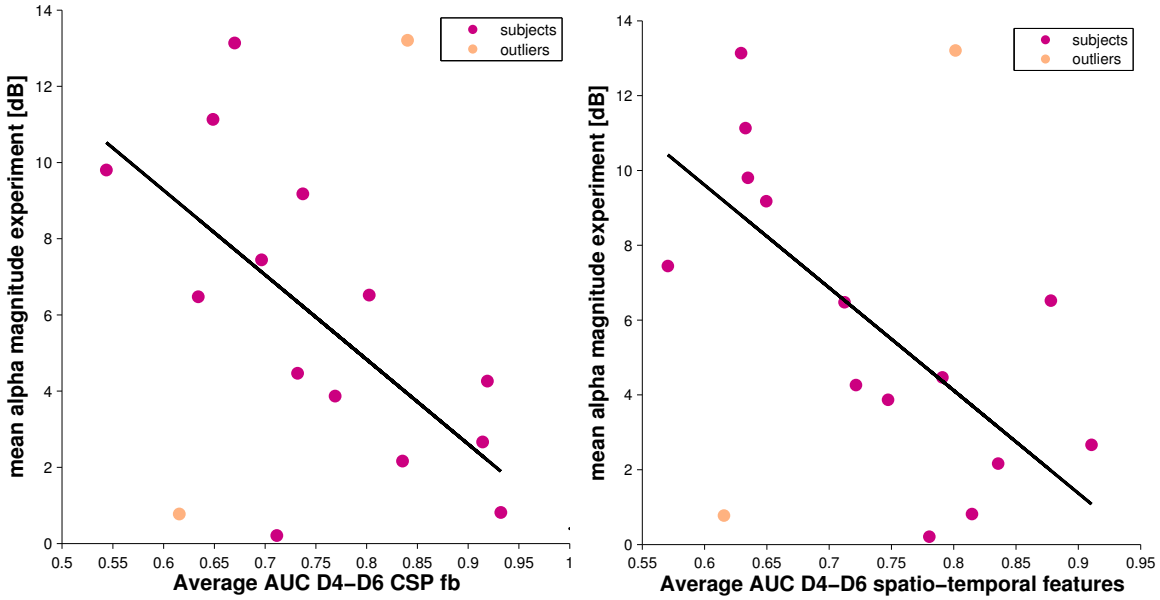


Fig. 4.8 Correlation amplitude of alpha rhythm-classification with average performance D4-D6. Plots refer to CSP with filter bank (left) and to classification based on spatio-temporal features (right). Each dot represents one participant, the yellow ones refer to participants considered as outliers. Regression lines are also displayed.

latter even do not pass chance level in the classification of distortion level D4, thus seeming less sensitive in general to changes of visual quality. In order to find neurophysiological correlates of classification performance, the average peak magnitudes of the alpha rhythm during the experiment for each participant was calculated and correlated with the average classification result that the participant reached at distortions over the threshold of perception, i.e. average AUC D4-D6. The results are displayed in Figure 4.8 as scatter plots, for the two classification methods described in the previous section. Each dot represents a participant. The black line is the result of the linear regression between the average magnitudes of the alpha peak of each participant and the respective classification performances. The yellow dots represent participants having the 10% largest Mahalanobis distances to the data center, and therefore considered as outliers and removed from the analysis (Huber and Ronchetti, 1975). In both classification methods used, a significant linear negative correlation between the average alpha activity during the experiment and the performance of classification of the quality change was found. The Pearson correlation coefficients were r=-0.73 for classification based on spatio-temporal features ($p<0.01$), r=-0.66 for CSP classification with the filter bank ($p=0.01$). Accordingly, brain signals of participants with a high level of alpha activity were less modulated by quality changes. One reason for an increased level of alpha activity is decreased visual attention.

4.5 Discussion

EEG data can be analyzed from two main points of view, the one that refers to the modulation of brain rhythms and the one that takes into account the time-locked components of the brain activity. The two feature extraction methods used in this study took into account both oscillatory and temporal nature of the SSVEPs, and led to results with comparable trends. Classification results showed mean offline performances up to 0.86 and 0.84 for the maximum distortion level D6, respectively for the CSP method and the spatio-temporal features method. It has to be pointed out that performances referring to the spatio-temporal features method were based on classification of single-epochs (i.e. what is generally referred to as *single-trial* in literature), while the performances of the CSP method were based on 4 epochs of the same distortion level (i.e. one block, Fig. 4.2). Both methods showed the same trend in the classification results, that is, the performances did not exceed chance level significantly for the first three levels of distortion, while from D4 to D6 they increased significantly and linearly with the quality change. These results suggested that the quality changes introduced in the textures by the compression algorithm modulated the ongoing EEG activity eliciting SSVEPs which could be classified over the perception threshold.

4.5.1 Comparison with previous studies

The performances achieved in the proposed design were in the same range of those exploiting the P3 component as neural feature for video quality assessment via EEG (Lindemann et al., 2011; Mustafa et al., 2012; Scholler et al., 2012). Note that, performances between these two types of studies are very difficult to compare because of the different design of the stimuli used in P3-based and SSVEP-based experiments. The classification accuracies reported by Scholler and Lindemann referred only to trials correctly identified by the participants at the behavioral level. Scholler et al. (2012) also reported for three participants an average 65% accuracy in classifying the trials in which the quality change was present but not detected by the subjects, advancing the hypothesis of higher sensitivity of the EEG compared to the behavioral response. Since this study capitalized on SSVEPs elicited in condition of passive viewing, trials potentially 'detected' by the participants and trials in which participants might have had a lower level of attention could not be differentiated. So the performances referred to the overall number trials, without pre-screening or rejections. About the number of trials, it is important to notice that an SSVEP-based paradigm allowed the recording of an amount much higher compared to P3-based paradigms. In fact, in the latter the number of epochs containing the actual target event is just a fraction of the total amount of epochs used in the experiment. In this design, each epoch was a 'target', meaning that all the trials contained useful information for the quality assessment. As already mentioned in Section 4.1, the number of events (as event referring to the occurrence of evoked potentials useful for the

evaluation) was increased up to eleven times if compared to the P3-based study of Scholler et al. (2012). That is to say that a SSVEP-based assessment could be much faster than a P3 one.

4.5.2 Correlation between behavioral and neural assessment

In this study, a significant correlation between the neural and the behavioral assessment using the MOS values was demonstrated. This result was valid for all participants for both the spatial-temporal features and CSP based classification. The comparison between the neural assessment and the behavioral one is a fundamental aspect to consider in video quality assessment. Studies that foresee participants pressing a button at detection of artifacts or distortions, allow a more straightforward comparison between the neural and the behavioral response (Porbadnigk et al., 2013a). However, results might be affected by 'label noise' (Porbadnigk et al., 2015). Besides, in these cases there is no indication of the actual level of quality perceived by the participants, that is, how well they could detect the distortions and how annoying the detected artifact could be. And this is actually an important detail to take into account in the implementation of video codecs. Therefore, unlike most of previous works, the behavioral assessment was performed in a separate test in which the MOS values were collected. They represented the actual subjective rate normally used in image quality assessment. These results were then linked to the performances of the classification of the SSVEPs modulated by the quality changes. This represented a key factor in the design of the paradigm, since previous studies usually referred the classification performances to the presence of an artifact or correlated the results with the absolute value of the quality change.

4.5.3 Understanding individual differences

In both spatio-temporal features and CSP based the methods, mean classification remained at chance level for distortions D1, D2, D3 and increased linearly and significantly from D4 to D6. In general, the same trend was visible in most of the participants. Clearly, not all participants were expected to be sensitive in the same manner to the flicker of the quality change. For example, considering the classification method based on spatio-temporal features it can be noticed that five participants had a classification performance which stayed at chance level at D4, a distortion level which was supposed to be above perception threshold. Even though also for these participants performance increased significantly with increasing distortion, at D6 it does not go beyond 75%. For these participants the SSVEPs were less pronounced in the time domain, as well as the spectra peaks at the discriminative frequencies. They seemed to be 'less sensitive' to the quality changes in general. There might be different reasons for such phenomenon, which lie in the nature of perception itself. One further reason can be found in the results of the analysis of the occipital alpha rhythm during the experiment.

Correlation between alpha power and classification performance

The magnitude of the alpha-power during the experiment (evaluated at electrode Oz) significantly correlated negatively with the average classification performance of distortions levels over the perception threshold, which directly reflected the modulation of the SSVEPs. This could justify the attenuated modulation and poor classification performances of some participants. In literature, the relationship between the spontaneous oscillatory activity when the stimulus is presented and the perception of the stimulus itself has been widely investigated. In particular, different studies (Brandt and Jansen, 1991; Busch and VanRullen, 2010; Busch et al., 2009; Ergenoglu et al., 2004; Hanslmayr et al., 2005, 2007; Jansen and Brandt, 1991; Romei et al., 2008) showed that oscillations in the alpha frequency band interfere with the processing of the visual information and modulate the gain of the visual system. The high alpha rhythm during the experiment (maybe due to drowsiness) could have kept the occipital cortex in an idle state preventing a strong modulation by the flickering of the quality change. Clearly, conclusions on these findings have to be made carefully.

The critical choice of the stimulus frequency and the *SSVEP inefficiency*

Another reason for the poorer performances of some participants could lie in the choice of the stimulus frequency. For some people not all the harmonics are modulated by attention in the same way. For example, in the study of Pei et al. (2002) stimuli at 2.4 and 3 Hz were chosen and the harmonic responses at 4.8 and 6 Hz were modulated by attention, while the responses at 9.8 and 12 Hz were not. We had a similar result showing the steady-state response present mainly at 3 and 6 Hz, while the modulation of higher harmonics was negligible. The choice of the stimulation frequency was also addressed in the study of Kelly et al. (2005). The authors suggested that the choice of the stimulus frequency should be specific to individual subjects. In this study, the frequency was chosen beforehand and kept constant for all the subjects. Some of them might have been less sensitive to that specific frequency, no matters their engagement in the task or the idle state of their alpha rhythm. Some research showed a percentage of people being not able to operate a SSVEP-based BCI at all (Allison et al., 2010, 2008; Volosyak et al., 2011). This percentage is lower than the estimation of the 'BCI-illiterates' in motor-imaging based BCIs (see Chapter 7), but it is an aspect worthwhile to study in the future.

4.5.4 Limitations and open questions

The six levels of distortions were associated to the quantization parameters (QP) which were chosen in a pilot study such that the perceived degradation at the behavioral level was the same for all the textures. In the pilot study a behavioral test was performed like described in Section 4.3.2, and finally the QP for each texture was chosen such that the average MOS value across participants was slightly below perception threshold for D2 and

slightly above for D3. In the main study the same results for a new pool of participants could not be reproduced. On average, both D2 and D3 were below the perception threshold both on the behavioral level and on the EEG level, therefore not discriminable by the classifier. Future studies could take into account more levels of the quantization parameters, especially increasing the number of the levels around the perception threshold in order to investigate the sensitivity of the EEG in relationship with the behavioral results. In these results an evidence of higher sensitivity of the EEG compared to the overt response could not be found, suggesting that there were no distortions processed unconsciously which would not result at the behavioral level. Previous ERP studies (Porbadnigk et al., 2011; Scholler et al., 2012) suggested that some people might show an unconscious neural processing of quality changes. Therefore, it is crucial to investigate more in depth the relationship between EEG response and behavioral one at the perception threshold, in order to prevent the implementation of video coding methods which would introduce distortions potentially perceived by the most sensitive people. In this study image degradation was caused by a specific compression algorithm, which introduced changes in more fundamental image features. The identification of which specific feature was responsible of the generation of the SSVEPs was beyond the scope of the presented analyses. This leads to the conclusion that the applicability of this method to other image compression algorithms has to be considered with caution, since other algorithms may control the image features in a different way.

4.6 Take-home messages

- SSVEPs were modulated over the threshold of perception by the video quality changes and could be classified with machine learning techniques both in the temporal and spectral domains;
- The SSVEP-based assessment, i.e. classification performances, significantly correlated with the standard behavioral assessment, i.e. MOS values;
- The SSVEP-based video quality assessment led to the collection of a high quantity of target trials in much shorter time than the P3-based assessment;
- More levels of degradations should be tested in order to explore the sensitivity of the neural assessment around the threshold of perception;
- The frequency of stimulation should be briefly tested before the neural assessment in order to avoid 'SSVEP inefficiency';
- The neural assessment paradigm should be revised to keep the users more engaged in the task and avoid the state of drowsiness, maybe with the introduction of online feedback.

Chapter 5

Novel BCI designs for gaze-independent spelling applications

5.1 Introduction and state of the art

The first BCI speller based on EEG and ERPs was the Matrix Speller realized by Farwell and Donchin (Farwell and Donchin, 1988). The symbols were arranged in a 6x6 matrix, and the user had to focus attention to the target symbol while the rows and the columns were intensified one after the other in random order. Typically, the ERP response to the intensification of the row and the column containing the target symbol showed an enhanced P3 component. By averaging across several repetitions of the stimuli, the enhanced P3 could be detected and the letter that the user intended to write could be inferred. Many variations of the Matrix Speller have been studied, improving communication rates and stimulus presentation (Bin et al., 2011; Donchin et al., 2000a; Guan et al., 2004; Hong et al., 2009; Townsend et al., 2010; Zhang et al., 2011). All these BCIs based on the classical matrix paradigm reached high classification accuracies and throughput. In 2010 two studies suggested that the good performance of the Matrix Speller relied to a large degree on target fixation (Brunner et al., 2010; Treder and Blankertz, 2010; van der Waal et al., 2012). In both studies, the Matrix Speller was investigated under two conditions: in one condition, the participants focused their attention on the intended letter, in the other condition they focused their gaze on the center of the display, operating the speller with covert attention only. The authors showed that the performance of the Matrix Speller was markedly lower in the covert attention condition than when the eye movements were permitted. These problems can prevent some patients at the latest stages of ALS, who lose control of eye movements, to use a BCI speller in an effective way. After these studies pointed out the limitations of the efficient Matrix paradigm, researchers were challenged in developing BCI systems independent of eye movements. Typical examples were spellers that used stimuli that relied on other sensory modalities, such as the auditory (Höhne et al., 2011; Klobassa et al., 2009; Kübler et al., 2009; Schreuder et al., 2010, 2011b) and tactile (Brouwer and Erp, 2010; Thurlings

et al., 2012). In the same years it was shown that higher accuracies could be achieved still using visual stimuli, but improving the design of the feedback in order to make it suitable for patients with oculo-motor impairments.

A first example was found in Treder and Blankertz (2010), where the two-stage visual speller Hex-o-Spell was investigated. In a follow-up study, the authors investigated three different variants of the Hex-o-Spell (Treder et al., 2011). The three variants were optimized for the gaze-independent mode by resorting to non-spatial feature attention and facilitating spatial covert attention. In the first level of selection, the 30 characters (26 letters of the English alphabet plus 4 punctuation marks) were clustered in six groups, which were presented in a serial manner in the center of the display. Each group was associated with a colored geometrical shape, in order to increase the discriminability between the stimuli. After the group's selection, the single letters of that group were presented in the same manner. So, in the second level, the user was able to select the target letter. Other variants of gaze-independent spellers were developed until year 2011 (please refer to Riccio et al. (2012) for a detailed review), when the first study of this chapter was carried out. The aim of this study was to investigate a visual BCI speller that exploited the rapid serial visual presentation (RSVP) paradigm. The RSVP paradigm was known already in psycho-physiological studies, e.g. attentional blink investigation (Raymond et al., 1992), response time variability (Gerson et al., 2005), BCI-enhanced computer vision (Sajda et al., 2010), and image classification (Bigdely-Shamlo et al., 2008). The RSVP could represent a valid alternative to the designs of previous spellers for several reasons. First, in this paradigm all the symbols were presented one-by-one in a serial manner and in the same location of display, exploiting only the foveal visual field. This characteristic makes the RSVP a suitable visual presentation for a speller based on non-spatial visual attention independent of eye movements. With this stimulus design both the N2 and the P3 components were modulated by attention and were discriminative for the classification of the attended stimulus. Besides, the RSVP speller used a direct selection of symbols and a short stimulus onset asynchrony (SOA), making it potentially faster than previous spellers that exploits a two-stage selection. As first approach in using the RSVP for mental typewriting, promising results were shown in an offline study (Acqualagna et al., 2010). In the RSVP-study described in the following sections, the design of the RSVP stimuli presentation was enhanced with more features and the performances were investigated online, with healthy participants.

All the paradigms previously introduced were addressed to as *gaze-independent*, since they did not need the user to move the eyes in order to select the target symbol. However, they still assumed fixation to the location of the display where the stimuli were shown. This characteristic can represent a limit if these spellers were used by patients who have severe oculomotor impairments that cause involuntary drifts of eye gaze. Therefore, a new approach was developed to overcome this problem, using isotropic stimuli that triggered the same visuo-attentional processes irrespective of the location of eye gaze. The second paradigm

described in this chapter is called Chroma Speller, and the stimuli consisted of colors displayed widescreen. Since the colors appeared uniformly in all the directions, the location of the gaze was irrelevant to the BCI operation. In the Chroma Speller, the symbol selection was operated in two stages like in the Center Speller, since a direct selection of symbols would mean the employment of an excessive number of colors and would be practically unfeasible. Each color was associated to a group of characters in the first level and to single characters in the second one. This abstract association colors-symbols might cause a learning workload which could affect the usability of the speller. The aim of this study was to confute such hypothesis, demonstrating that competitive performances could be achieved even with such an abstract design. In order to have a baseline for performances comparison, participants performed an online session also with the Center Speller.

Section 5.2 of this chapter describes the design, the online experiment and the main results of the RSVP Speller. Section 5.3 introduces a novel algorithm that aims at improving the classification accuracy and spelling speed of the RSVP Speller, exploiting the temporal dependencies between overlapping EEG epochs due to the fast stimuli presentation pace. The design of the Chroma Speller, together with the main results of the online experiments, is described in Section 5.4. A brief discussion about the most recent developments in the realm of visual gaze-independent spellers follows in Section 5.5.

Data and preliminary results of the RSVP Speller study were published in Acqualagna and Blankertz (2011) (©2011 IEEE) and the full study was published in Acqualagna and Blankertz (2013). Data and results of the Chroma Speller study were published in Acqualagna et al. (2013) (©2013 IEEE).

5.2 BCI spelling based on the Rapid Serial Visual Presentation paradigm

5.2.1 Methods

Participants

Twelve participants (6 males, 6 females, aged 24-55) took part in the experiment. All had normal or correct-to-normal visual acuity and none of them had a history of neurological disease or injury. Normal color vision was confirmed using the Nishihara color vision test. Three participants were volunteers and members of the BCI group and had already participated in BCI spelling experiments. The other participants were naïve in respect to BCI spelling and were paid for their participation. The study was performed in accordance with the declaration of Helsinki and all participants gave written informed consent.

Apparatus

The electrodes configuration used in this study was: Fp1,2, AF3,4, Fz, F1-10, FCz, FC1-6, FT7,8, Cz, C1-6, T7,8, CPz, CP1-6, TP7,8, Pz, P1-10, POz, PO3,4,7-10, Oz, O1,2.

Stimuli were presented on a 24" TFT screen with a refresh rate of 60 Hz and a resolution of 1920 x 1200 px^2. The experiment was implemented in Python using the open-source BCI framework Pyff (Venthur et al., 2010b) with Pygame (http://pygame.org) and VisionEgg (Straw, 2008).

Design of the stimuli

The basic idea of the BCI Speller based on the RSVP paradigm is that all symbols are presented one-by-one in random order at the center of the screen. In this study, the design of the stimuli was improved in respect to the previous one (Acqualagna et al., 2010), based on the hypotheses that more colors and different capitalizations can help the discrimination of the symbols. Three variants of presentation were implemented (Fig. 5.1). They differed with respect to the SOA and to the color of the letters (monochrome vs. five colors as explained below). In the Color condition, we used two stimulus onset asynchronies (SOAs), 83 ms and 116 ms, in order to investigate how they influence the spelling performance. Spellers that employ fast SOAs achieve higher selection speed, but less accuracy. Finding a trade-off between accuracy and selection speed is one of the challenges in BCI for spelling applications. For each condition, participants had to perform three phases: calibration, copy-spelling and free-spelling. The conditions were: 1) NoColor 116 ms; 2) Color 116 ms; 3) Color 83 ms. Their order was counterbalanced across participants. In the NoColor condition all the characters were black, in the color they were divided into 5 color groups: red, white, blue, green, black. The sequence of symbols was determined pseudo randomly, such that the order of the colors was always the same (red-white-blue-green-black-red-...). This should facilitate the task: if, e.g., the target symbol was blue-colored, only the blue stimuli, i.e., only every fifth letter, have to be checked. No blank frames were interposed between two consecutive letters. Stimuli were presented on a gray background. The visual presentation was time-locked to the screen refresh.

We used a vocabulary of 30 symbols: the 26 letters of the English alphabet, 2 punctuation marks: '`.`', '`!`', the underscore '`_`', used to separate different words, and '`<`' as backspace symbol. In our previous study (Acqualagna et al., 2010) all the letters of the alphabet were displayed in their capital form. After the experiment, we provided the participants with a questionnaire, in which they were asked to specify the letters and the colors they found more difficult to identify. It turned out that some shapes, e.g. '`O`' and '`Q`', '`K`' and '`X`', '`E`' and '`F`', were very difficult to discern between each other, especially when all the letters were black. Therefore, in order to enhance the differences between the shapes of the different letters, in this new design the letters '`A,B,E,G,H,J,L,M,N,O,R,S,T,U,X`' were uppercase and the

letters '`c,d,f,i,k,p,q,v,w,y,z`' were lowercase. The letter '`M`' had a height of 3.5 cm (1° of visual angle). In the Color conditions, the color groups were: red (`fRyGk<`), white (`pJUX!E`), blue (`iSwc_N`), green (`TBMqAH`), black (`LdvOz.`).

Experimental procedure

Participants sat on a comfortable chair at a distance of approximately 80 cm from the display. In the calibration phase, the phrase to be spelled was '`BRAiN_cOMpUTER_iNTERfAcE`'. The phrase was shown in the center of the display for 2 seconds prior the start of the trial. Then, it was displayed on the top of the screen with the current target letter highlighted. The participant had 4 seconds to identify the target letter. After that, a fixation cross appeared in the center of the screen for 3 seconds, in the location of the RSVP. Participants were asked to always look at the center, relax the face's muscles and avoid blinking during the visual presentation. The 30 symbols were randomly shuffled and presented 10 times. Between the 10 sequences there was a short break of 0.3 s. Participants had to concentrate on the target letter and were asked to silently count its number of occurrences in the RSVP. The data recorded in this phase were used to train the classifier. After the calibration phase and the training of the classifier, the on-line phases began. In the copy-spelling, for each condition the participants had to spell a different sentence : '`LET_yOUR_BRAiN_TALk`' in Color 116 ms, '`wiNTER_iS_dEpRESSiNG`' in NoColor 116, and '`doNT_wORRy_BE_HAppy`' in Color 83 ms. In this phase the instructions were the same as in the previous calibration phase. The only difference was that after each trial, the classifier selected the symbol with the best score and displayed it. In case of misclassification, the participants did not have to delete the letter or repeat the trial, but they had to proceed to the next one. In all the conditions, the last phase was free-spelling. Participants were asked to conceive a sentence containing at least 15 symbols. In this phase, if the classifier selected the wrong letter, they had to use the next trial to select the backspace symbol to erase it. All the 30 symbols were displayed on the bottom of the screen in alphabetical order. Before each trial, a 3 seconds countdown displayed in the center gave the participants the opportunity to seek the letter they wanted to spell if they did not remember its capitalization and color.

Besides the main spelling experiment, an additional recording of the ordinary oddball task was performed, aiming at finding neurophysiological predictors of the spelling performance. Finding some neurophysiological components that strongly correlate with the spelling accuracy, could shed some light on the high subject-to-subject variability in BCI performance. A short session of the standard oddball experiment was performed between the spelling conditions, for a total of two sessions. Three different colored rectangles were displayed in the center of the screen in a random order with a SOA of 1 s. Participants had to count the number of the red deviant stimuli, ignoring the yellow and blue ones.

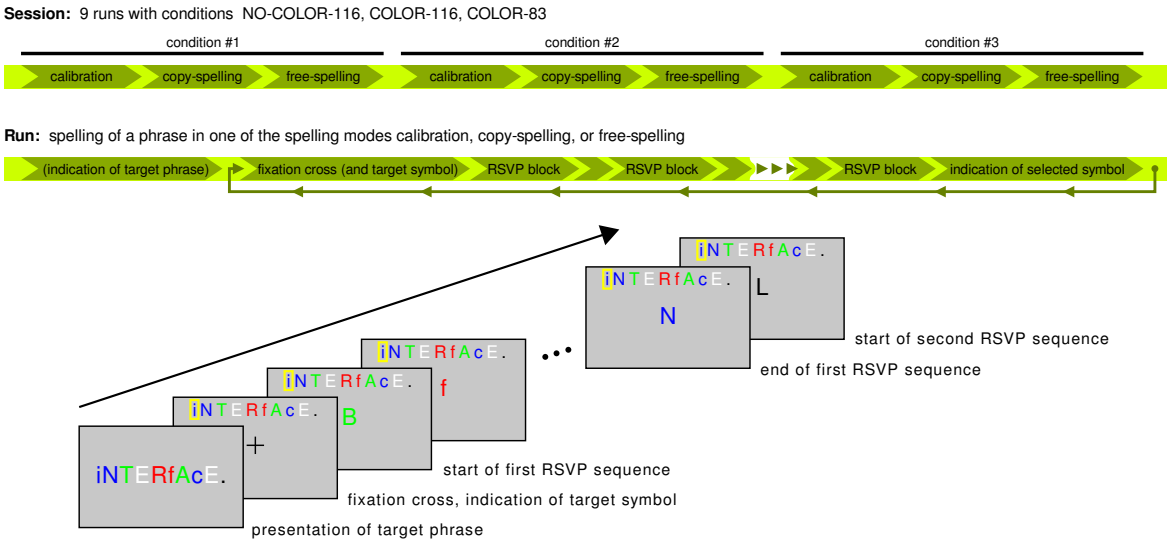


Fig. 5.1 RSVP Speller flowchart. First, the sentence is presented on the display. After the fixation cross, the RSVP of the symbols starts. The target letter is highlighted on the top of the screen. For each condition, three phases were performed: calibration, copy-spelling and free-spelling. In the last two, after the RSVP, the classifier's output was displayed.

Data analysis

For ERP analysis, EEG signals were lowpass filtered with a Chebyshev filter Type II using a passband up to 40 Hz and a stopband starting at 49 Hz, and then down-sampled to 200 Hz. For ERP analysis, continuous signals were divided into epochs ranging from -100 ms to 1200 ms relative to the stimulus onset. Baseline correction was performed on the pre-stimulus interval of either 116 ms (for 116ms SOA) or 83*2 ms (for 83 ms SOA). Epochs containing strong eye movements were detected and rejected using the min-max criterion (75 μV) on the channels F9, Fz, F10, AF3 and AF4 (Section 3.1.1). Only those non-target epochs were used, in which the three preceding and the three following symbols were also non-targets, in order to avoid the overlap from ERPs of preceding or successive targets. For the grand average, the ERP curves were averaged across all trials and participants. To compare the ERP curves of two classes (*target* and *non-target*), $sgn - r^2$-values were calculated. The N2 and P3 amplitudes and latencies were calculated for each participant extracting the negative or positive peak in the averaged epochs, for a given time interval and electrode (150–350 ms and PO7 for N2, 350–500 ms and CPz for P3). If no peak was found, the mean was taken.

For classification of attended symbols, we employed binary classifiers based on spatio-temporal features for discriminating between epochs related to targets vs. non-targets. As preprocessing, EEG signals were downsampled to 100 Hz by calculating the average for consecutive blocks of 10 samples. Epochs with an excessive power in a broadband (5–40 Hz) indicating, e.g., muscular artifacts were rejected from the calibration data. Rejection based on the min-max criterion and exclusion of non-targets as described for the ERP analysis

were not employed. For feature extraction, five temporal windows were selected individually for each participant and condition by the heuristic based on the $sgn - r^2$ in the respective calibration data (Section 3.1.2). Occasionally, the intervals determined by the heuristic were adjusted by the experimenter before starting the on-line runs. Features were calculated from 55 channels (all except for Fp1,2, AF3,4, F9,10, FT7,8) by averaging voltages within each of the five chosen time windows resulting in 55×5=275 dimensional feature vectors. For classification, shrinkage LDA was trained on calibration data (Section 3.2.1). A symbol was determined by averaging the classifier output for all 30 symbols across the 10 sequences and choosing the symbol that received the minimum averaged output (class 'target' had negative classifier values).

In order to investigate the speed/accuracy trade-off, the classifier that was used online was applied offline to the copy- and the free-spelling data, depending the symbol selection on averages across the first n sequences, where n was varied from 1 to 10. For $n = 10$ the same results as for online operation are obtained, for $n < 10$ an estimate of the potential performance when using a different number of sequences is obtained. As main measure for spelling performance, the spelling rate was employed quantified as the number of error-corrected symbols per minute. The number of error-corrected symbols is the total number symbol selections minus two times the number of wrong selections. (After each erroneous selection, first the backspace needs to be selected and then the intended letter has to be reselected). In order to determine the spelling rate for $n < 10$ sequences, the hypothetical duration of one symbol selection can be calculated using the following information: the duration of the fixation cross and of the target presentation (7 s), the SOA (either 83 ms or 116 ms), the time of displaying the classifier's output (1 s), the inter-sequence pause (300 ms) and the number of symbols (30). Whenever performance is reported as accuracy, this relates to the symbol selection accuracy (chance level $= 1/30 \approx 3.33\%$). For across-condition comparison of accuracies, it has to be kept in mind that the trial duration differed considerably between 46.5 s for 116 ms SOA and 36.6 s for 83 ms SOA.

Classification in the RSVP paradigm exploits several discriminative components, most prominently the N2 and the P3b. As an approach to disentangle the contribution of the N2 and the P3b component to classification, we determined the selection accuracy separately for a classifier trained on spatial features extracted from the time interval 300–380 (N2) and for the time interval 420–600 (P3b). Again calibration data was used to train the classifiers and copy- and free-spelling to evaluate performance.

Amplitudes and latencies of the ERPs and classification accuracies were subjected to statistical tests, using ANOVA analysis and Tukey-Kramer post hoc test.

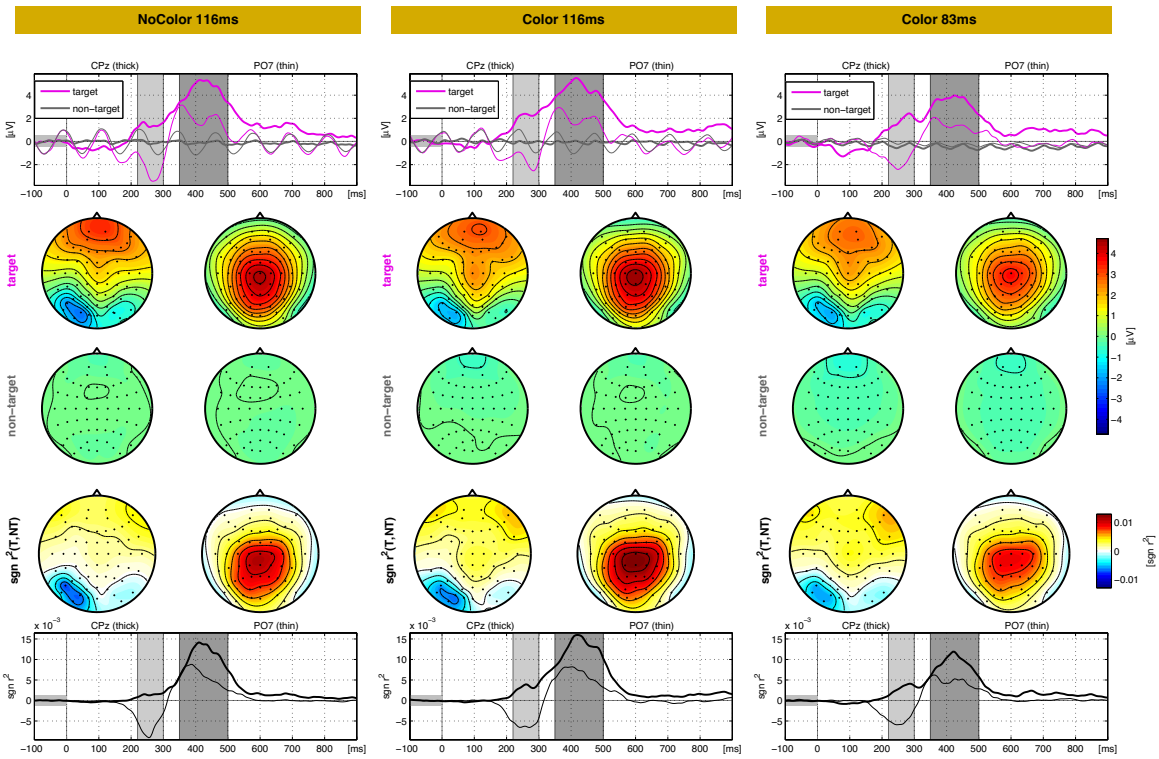


Fig. 5.2 Grand average ERPs RSVP Speller. The plots in the first row depict the grand average EEG over all the participants for targets (magenta) and non-targets (grey). The thick lines represent the channel CPz, the thin lines the channel PO7. The shadowed areas highlight the two time-intervals 220-300 ms and 350-500, in which the N2 and P3 components are elicited. The scalp plots underneath (second and third rows) refer to the shadowed areas in the ERPs plots. The EEG signal is averaged in the two time intervals and depicted as color coded maps. The fourth and fifth rows picture the $sgn - r^2$ between the classes target and non-target. The scalp plots refer to the areas shadowed in the plots in the fifth row.

Calculation of predictor of spelling performances

In order to find factors that influence spelling performance (so-called predictors (Blankertz et al., 2010a; Halder et al., 2011a,b; Hammer et al., 2012; Kübler et al., 2004; Neumann and Birbaumer, 2003; Suk et al., 2014)), we correlated standard neurophysiological measures with the P3-based classification performance (see above). As a baseline measure of individual cognitive processing, we took the P3 amplitudes in the standard oddball paradigm. Note, that in the standard oddball experiment a clearly quantifiable N2 was only found in few participants. Accordingly, this measure was not tested as predictor.

Condition	N2 amp. [μV]	N2 lat. [ms]	N2 r^2	P3 amp. [μV]	P3 lat. [ms]	P3 r^2
NoColor 116	-3.90±0.80	254±8.45	-0.043±0.014	6.85±0.76	426±10.12	0.104±0.016
Color 116	-3.44±0.74	240±10.43	-0.045±0.017	7.33±0.87	420±11.51	0.118±0.020
Color 83	-2.57±0.48	246±5.01	-0.031±0.009	5.76±0.56	416±17.20	0.088±0.012

Table 5.1 Mean and Standard Error on the Mean (SEM) of the P3 and N2 amplitudes (μV) and latencies (ms).

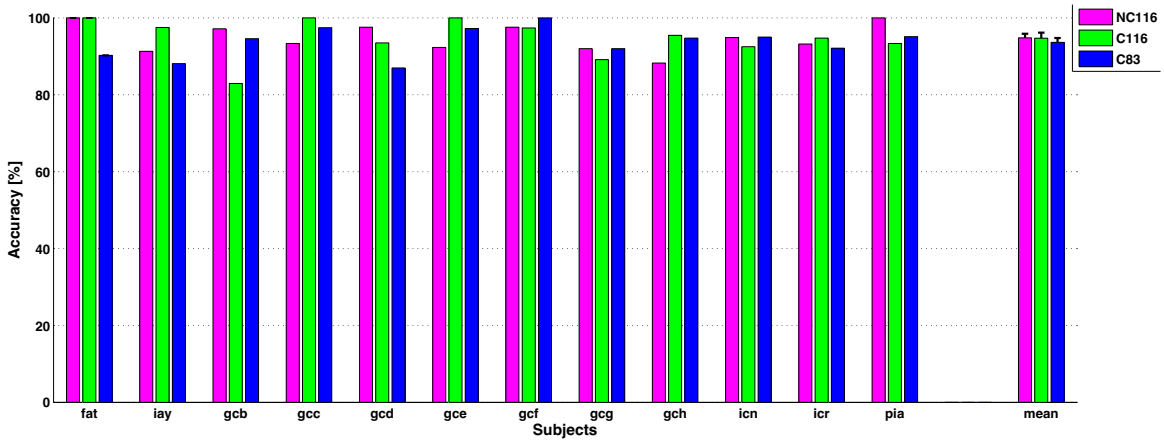


Fig. 5.3 Online classification accuracy RSVP Speller. All the participants show a high classification accuracy in all the conditions, with 5 participants reaching 100% in at least one condition. The mean accuracy is depicted in the last column.

5.2.2 Results

ERPs

Fig. 5.2 depicts the grand average ERPs in all three conditions. In the ERP plots in the top row, the N2 peaks around 350 ms after the target onset at electrode PO7, the P3a peaks around 300 ms at CPz and the P3b around 500 ms at CPz. In the time interval between 220-300 ms, the focus of the P3a is evident in the frontal cortex and the N2 in the occipital area, mainly focused on the left hemisphere. The later P3b is mainly distributed in the central-parietal area. In the non-targets scalp plots, no ERP components are elicited. The scalp plots in the fourth row and the plots underneath show the distribution of the $sgn - r^2$ values as a measure of discriminability between the two classes *target* and *non-target*. The highest values are achieved at locations and latencies of the P3a, P3b and N2 components. In Table 5.1 the amplitudes and latencies of the P3 and N2 are reported in detail. They were subjected to a one-way analysis of variance (ANOVA) with factor Condition (NoColor 116 ms, Color 116 ms, Color 83 ms). Regarding amplitudes as well as latencies, no significant effect of the condition was found, neither for the P3 nor for the N2 ($p = 0.372$, $p = 0.396$, $p = 0.348$, $p = 0.485$), even if the values show a clear decrease of the components amplitudes in the faster condition.

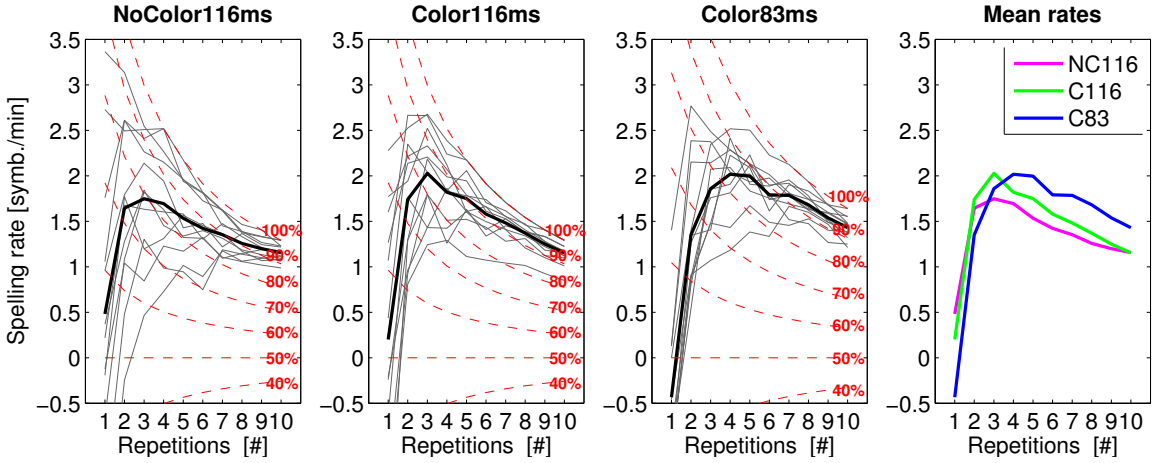


Fig. 5.4 Spelling rate RSVP Speller. The spelling rate is plotted as a function of the number of repetitions and classification accuracies (red dashed lines). The gray lines represent the different participants, the back thick lines the mean.

Classification

The feasibility of a BCI as communication device depends largely both on the accuracy achieved in the target discrimination and on the time necessary to spell the desired character. In this system, the mean on-line spelling rate was 1.16 symb/min in the NoColor 116 ms (corresponding to 5.65 bits/min), 1.15 symb/min in the Color 116 ms (5.66 bits/min), and 1.43 symb/min in the Color 83 ms condition (7.00 bits/min). The rate in the Color 83 ms was significantly higher than in the slower conditions ($F = 21.69, p < 0.001$). Mean online classification accuracies of 94.8% (NoColor 116 ms), 94.7% (Color 116 ms) and 93.6% (Color 83 ms) were achieved. The factor Condition did not show a significant effect on the online accuracy ($p = 0.756$). Fig. 5.3 depicts the results for all participants and conditions.

Offline classification analysis provides the rate/accuracy trade-off shown in Fig. 5.4, where the estimated spelling rate (and by virtue of the contour lines also the accuracy) is plotted as a function of the number of sequences. In all the three conditions, the gray lines represent individual participants and the thick black line the mean. As expected, the performance of the classifier increases significantly with the number of sequences ($F = 103.58$, $p < 0.001$). The chance level is 3.33% since classification relates to the selection of one symbol out of 30. These results give rise to an estimation of what spelling performance is possible, if an optimal number of repetitions were used: the spelling rate is 2.00 symb/min in Color 116 ms after 3 sequences (83.84 % of classification accuracy) as well as in Color 83 ms after 4 sequences (83.4% classification accuracy). In the NoColor condition, the peak mean rate is 1.75 symb/min after 3 sequences (79.2% classification accuracy).

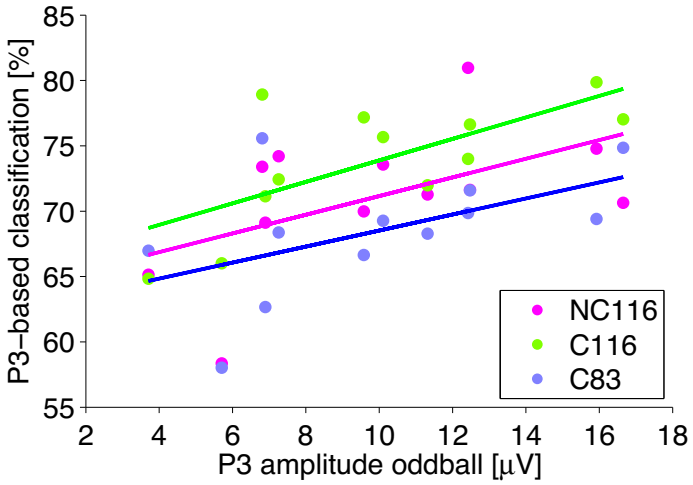


Fig. 5.5 Correlation between classification accuracy component-based and physiological parameters. The symbol selection accuracies based on the P3 component were correlated to the amplitude of the P3 in the standard oddball measurement. The results show a positive correlation between the P3 amplitude in the oddball measurement and the P3-based classification in all the conditions.

Neurophysiological predictor of spelling performances

Fig. 5.5 shows the results of the correlation between the P3-based symbol selection accuracy and the P3 amplitude in the oddball recording. All the 10 sequences for the classification are considered, for each participant and condition. Significant positive correlation between the P3 amplitude and the accuracy was found in the Color 116 ms condition ($p < 0.05$), the accuracy increasing with higher amplitudes. A ('marginally significant') positive correlation was found in the NoColor 116 ms and the Color 83 ms conditions with $p = 0.082$ and $p = 0.086$.

5.2.3 Discussion

Mean symbol selection accuracies of 94.8%, 94.7% and 93.6% and mean spelling rates of 1.16 symb/min, 1.15 symb/min and 1.43 symb/min were reached respectively in the conditions NoColor 116 ms, Color 116 ms and Color 83 ms during online spelling. These results confirmed the initial hypotheses that different capitalizations and more colors ease the discrimination of the symbols. In fact such modifications in the speller's design highly improved the classification accuracy compared to the previous offline study (90% for the NoColor 133 ms, 85% for the color 133ms and 70% for the color 83ms) (Acqualagna et al., 2010).

An important parameter to consider in the design of BCI spellers is the SOA. Shorter SOAs result in a quicker selection process (assuming a fixed number of blocks per selection). On the other hand, shorter SOAs imply short target-to-target-intervals (TTI), which are known to dampen the amplitude of evoked potentials such as P3 and N2 (Allison and Pineda, 2006; Gonsalvez and Polich, 2002; Gonsalvez et al., 2007). Two different SOAs were

investigated in this study, 116 ms and 83 ms. The results were consistent with earlier works: both the N2 and P3 amplitudes were smaller in the 83 ms condition than in the 116 ms, and the P3 peak was also smaller if compared to the mean amplitude reached in the oddball runs (10 μV). However, despite of the faster stimulus presentation, the RSVP elicited strong ERPs, comparable to other spellers in literature (Aricò et al., 2010; Hong et al., 2009; Liu et al., 2011; Treder et al., 2011).

Comparison with previous studies

The accuracies achieved online were competitive with those of the spellers based both on covert attention and non-spatial visual attention. Treder et al. (2011) investigated three variants of the two-stages speller: the Hex-o-Spell, the Cake speller and the Center speller, the first two based on covert spatial attention and the latter on feature attention. They obtained high classification accuracies in all the three spellers, the best being the Center speller with 97.1% and spelling rate of 1.35 symb/min. The authors conjectured that the higher performances could be a direct consequence of the cortical magnification, i.e. the part of the cortex processing the stimuli picked by the fovea is larger than the others connected to the surrounding parts. Also Liu et al. (2011) investigated a visual paradigm based on a central presentation of the stimuli: they arranged the columns of the classic Matrix Speller in circular shapes presented in a sequential fashion and the recognition of the target was achieved through the covert shift of attention inside the circles. At a slower presentation rate (400 ms SOA) they achieved similar high accuracies up to 96.3%.

In order to use a BCI as a realistic communication device, not only is it necessary to achieve a high classification accuracy, but it is also important to have a sufficiently high information throughput. Different approaches have been taken to improve the information throughput of the Matrix Speller (Donchin et al., 2000a; Serby et al., 2005; Townsend et al., 2010). For example, Townsend et al. (2010) estimated a practical communication rate up to 3.66 symb/min. Donchin et al. (2000a) indicated that an offline version of the Matrix-speller can communicate at the rate of 7.8 characters per minute, with 80% of accuracy. Gaze-independent spellers showed lower performances so far (Liu et al., 2011; Schreuder et al., 2010; Treder et al., 2011). In the proposed system, the maximum spelling rate was 1.43 symb/min (7 bits/min) achieved in the Color 83 ms condition, after 10 repetition and with 93.6% of accuracy. Error correction and inter-selection interval was considered in this calculation, so it reflected the actual online selection rate. This rate was higher than that one achieved by Liu et al. (2011) and competitive with that one of Treder et al. (2011). Offline analyses showed a maximum mean spelling rate of 2 symb/min (10.5 bit/min) with 83.84% of accuracy in the Color 116 ms.

All the studies previously mentioned demonstrated that presenting the stimuli in a central position was a successful choice for gaze-independent visual spellers. This is indeed the main

characteristic of the RSVP. Since in this paradigm all the stimuli were presented in a small portion of the display, they were processed mainly by the foveal region of the retina. The neurophysiological results showed high amplitudes of the ERP components and a strong target-non-target discrimination, which inevitably led to high symbol selection-accuracies. Several of the previous paradigms were based on the modulation of the P3 component (e.g. those based on covert attention (Treder and Blankertz, 2010)), which is the result of the cognitive task of counting the attended letter. Here, the presence of a robust N2 component, elicited by the target, confirmed the deep impact that the attended stimulus had on the visual cortex. This was a further proof which makes paradigms such the RSVP preferable to paradigms based on a sparse distribution of the characters, especially in case of oculo-motor impairment. For all these reasons, the RSVP speller can be considered as a promising paradigm for gaze-independent spelling.

Understanding individual differences and neurophysiological predictor

The classification results showed that all the participants were able to operate the RSVP speller with high accuracy. However, also for ERP-based systems there is a large variability in performance between participants, see Fig 5.4. In order to shed light on possible reasons, a neurophysiological measure was investigated and its correlation with spelling performance. We found a positive correlation between the P3 amplitude in the standard oddball session and the classification performances based on the P3 component. This result was in accordance with those in (Halder et al., 2011a), who found a correlation between the ERP amplitudes of the auditory standard oddball and the performance achieved in a visual P3 BCI session. Therefore, a measurement of the amplitude of the P3 peak during a preliminary oddball phase could give an indication of the P3-based classification accuracy the participants can achieve during the spelling phase.

5.3 Boosting the RSVP Speller performance accounting for dependencies in ERPs classification

The classical approach of target-non-target binary classification assumes that all the non-targets have the same feature distribution and are treated in the same way by the classifier. This assumption ignores the fact that for a fast train of stimuli (e.g. 83 ms and 116 ms in the RSVP Speller experimental design) the onset of successive symbols would occur in the time of the generation of the neural components. In fact, the elicitation of significant ERPs after the onset of the target takes several milliseconds. For example, in the online study described in the previous section, there was a pronounced N2 component peaking between 200 and 300 ms and a P3 peaking around 400 ms time-locked to the target onset. This means that the epoch time-locked to the non-target letter following the target (i.e. the epoch starting

83 or 116 ms after the target's onset) will contain part of the evoked potentials elicited by the target. The shorter is the SOA, the higher will be the number of non-targets affected by this overlap. This is true also for the non-targets preceding the target. In fact usually the length of the look up epoch varies between 600-1000 ms. Therefore, the epochs time-locked to the non-targets occurring e.g. 116 ms, 232 ms etc. before the onset of the target would overlap with the epoch referred to the target. Usually, in an ERP speller the order of the stimuli is randomized for each repetition of the same sequence. In a speller based on visual stimuli, the number of repetitions may vary between 5 and 10. In the training phase, the classifier is trained on the average of these N repetitions, being the ERPs enhanced after averaging. In the testing phase, the classifier output is given after each trial and then for each symbol the average over repetitions score is calculated. Changing the order of following symbols through the repetitions leads to different non-targets neighboring the target letter. If for example letter 'B' follows target 'A' in one repetition and the EEG epoch locked to the onset of 'B' would contain the ERPs elicited by 'A', this with high probability would not be true for the next repetition. Therefore, the process of averaging the epochs time-locked to the onset of 'B' would cancel out the contamination of the ERPs that might have occurred in one particular repetition. Interestingly, recent studies proposed the idea to not randomize the order of the symbols, but keeping the same one in all the repetitions (Tangermann et al., 2012). Usually the randomization is done in order to prevent the participants to foresee the upcoming symbol in the display, which would reduce the surprise effect typical of oddball paradigms. In this kind of designs the approach of considering the features characterizing the stimuli neighboring the targets would be potentially helpful to improve classification. Besides, even in a randomized fashion, considering the target's neighbor stimuli as different classes compared to the other non-targets introduces additional information that can be exploited by the classifier. In the novel method described in this section, which will be addressed as Maximum Likelihood with Dependencies (MLD), the different stimuli are not considered anymore independently, but the sequence is evaluated as a whole.

5.3.1 Methods

The proposed method and the binary classification with shrinkage LDA were compared offline, both applied to the RSVP speller dataset described in the previous section. The standard binary approach assumes independence between successive measurements. Given a sequence of N repetitions of 30 symbols, the outcome of the LDA is 30*N binary classification values. For each symbol the mean over the N repetitions is then calculated and, assuming negative classification outputs for the target symbol, the selected target letter is the one having the minimum mean value. As already mentioned, the proposed method accounts for temporal dependencies between consecutive stimuli considering the symbols neighboring the target as different classes. The detailed approach is described in the following part taking as example

a lookup period of 1, i.e. just the stimuli immediately preceding and following the target are treated as special non-targets. In the standard approach that considers each stimulus independently, the symbols are labeled as target 'T' or non-target 'N'. In the proposed MLD approach, each stimulus was labeled with a triplet of labels that includes the information regarding the neighbor stimuli. Thus, the target stimulus was labeled as 'NTN', the preceding stimulus as 'NNT' and the following stimulus as 'TNN'. All the other stimuli will be regarded as 'standard' non-targets and labeled with 'NNN'. Note that, for each sequence there is just one target. The letters of the triplet referred to the actual position of the stimulus that was considered (in the middle) in respect to the position of the target. In case of stimuli at the edges of the sequences, an additional N (or more for longer lookup periods) was prepended or juxtaposed, assuming that just non-target-like epochs would occur before or after the sequence due to inter-sequences intervals. Since the inter-sequence interval was 300 ms, the lookup period considered in these simulations was at maximum 3 positions (considering the SOA of 116 ms, it would be about 348 before and after the current stimulus). The features used in this approach were the same as in the previously described binary approach, i.e. the same temporal windows and number of channels were used.

Training phase

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ with $\mathbf{x}_n \in \Re^d$ be N feature vectors (single-trial EEG feature vectors as in Eq. 3.3). Let $\Omega = \langle y_n \rangle_{n=1,\ldots,N}$ be the corresponding set of class labels $y_n \in C$, with $C = \{NNT, NTN, TNN, NNN\}$. Hereinafter the labels will be indicated as $C = \langle c_j \rangle_{j=1,\ldots,4}$. The number of training samples belonging to class c_j will be indicated as $N^{(c_j)}$. For each class c_j, the mean $\hat{\mu}^{(c_j)}$ was estimated from the training data

$$\hat{\mu}^{(c_j)} = \frac{1}{N^{(c_j)}} \sum_{n:y_n=c_j} \mathbf{x}_n \tag{5.1}$$

and a common covariance matrix $\hat{\Sigma}$ was estimated with shrinkage as in Eq. 3.21.

Testing phase

In order to evaluate the test data, each sequence of 30 symbols was examined as a whole. A likelihood value for each of the 30 possible target positions was calculated as follows. For example, a given target position could be represented as:
(N)TNNNNNNNNNNNNNNNNNNNNNNNNNNNNN (i.e. the target is in first position in the sequence). This would be the first putative sequence.

Given a set $Q = \langle q_i \rangle_{i=1,\ldots,30}$ of putative sequences, each sequence was converted into overlapping triplets belonging to C, like previously described. Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ with $\mathbf{x}_k \in \Re^d$ be a set of K feature vectors of the test dataset of putative sequence q_i (note that $K = 30$). Let $\widetilde{\Omega} = \langle \widetilde{y}_k \rangle_{k=1,\ldots,K}$ be the corresponding set of assumed class labels $\widetilde{y}_k \in C$.

Assuming Gaussian distribution, the log-likelihood of each feature vector of sequence q_i was calculated[1] as

$$ln\ p(\mathbf{x}_k|\widetilde{y}_k) = -\frac{1}{2}(\mathbf{x}_k - \hat{\mu}^{(\widetilde{y}_k)})^\top \hat{\Sigma}^{-1}(\mathbf{x}_k - \hat{\mu}^{(\widetilde{y}_k)}) \tag{5.2}$$

The likelihood of the putative sequence q_i was then computed by summing the log-likelihood of the K feature vectors

$$p(\mathbf{X}|\widetilde{\Omega}) = \sum_{k=1}^{30} ln\ p(\mathbf{x}_k|\widetilde{y}_k) \tag{5.3}$$

Since the actual position of the target was unknown in the test phase, the same procedure was repeated assuming the target in the next position:
NTNNNNNNNNNNNNNNNNNNNNNNNNNNNN. The likelihood for this second putative sequence was computed. The same steps were done for each of the possible positions of the target in the sequence, i.e. for all the 30 putative sequences. After normalization, and assuming uniform prior probabilities $P(\widetilde{\Omega})$, the posterior probability for each sequence was computed through Bayes formula

$$P(\widetilde{\Omega}|\mathbf{X}) = \frac{p(\mathbf{X}|\widetilde{\Omega})P(\widetilde{\Omega})}{\sum_i p(\mathbf{X}|\widetilde{\Omega})P(\widetilde{\Omega})}. \tag{5.4}$$

A number of repetitions from one to ten was considered in order to see how classification accuracy and spelling speed would vary increasing the number of repetitions. Given the set of repetitions $R = \{1, \ldots, 10\}$, the test feature vectors can be re-written specifying the index of the j-th repetition $\mathbf{X}^{(j)} = \{\mathbf{x}_1^{(j)}, \ldots, \mathbf{x}_K^{(j)}\}$. The whole procedure was applied for each repetition of the sequence of stimuli. So, given more repetitions, the final posterior probability of the putative sequence q_i was given by multiplying the posterior probabilities of $r \in R$ repetitions (assuming independency)

$$P(\widetilde{\Omega}|\mathbf{X}) = \prod_{j=1}^{r} P(\widetilde{\Omega}|\mathbf{X}^{(j)}) \tag{5.5}$$

Note that one putative sequence corresponds to a specific target position, i.e. a specific symbol assumed as target. Therefore the symbol with the highest product probability was selected as target.

Validation

The proposed method was evaluated for conditions Color 116 ms and Color 83 ms. For each participant and for each number of repetitions four models were evaluated: shrinkage LDA, MLD with lookup period of 1 position (MLD+1), MLD+2 and MLD+3. From the training data, the model that achieved the highest accuracy using cross-validation was selected and

[1] Factor $-\frac{1}{2}ln[(2\pi)^d|\mathbf{\Sigma}^{(-1)}]$ is a constant and is omitted in this calculation.

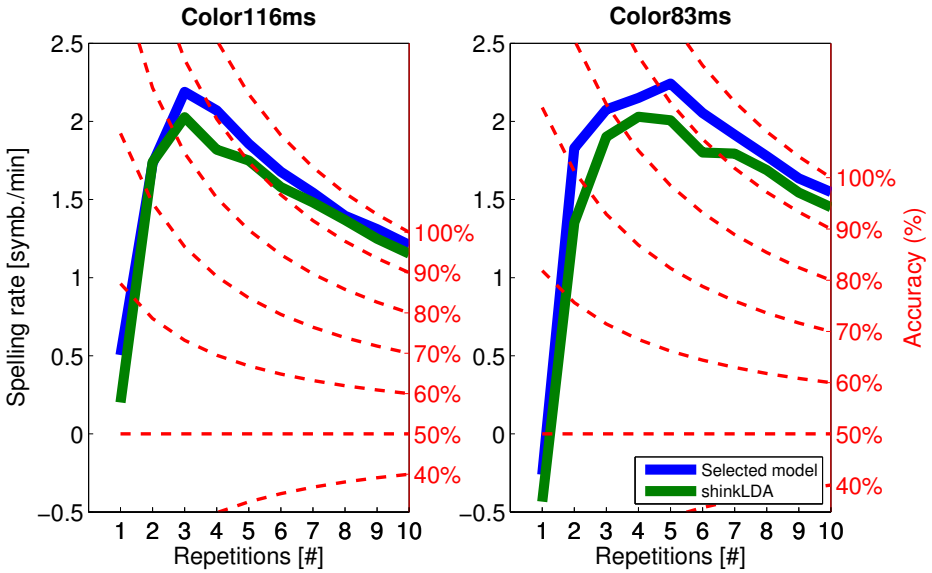


Fig. 5.6 Spelling rate RSVP Speller with MLD and shrinkage LDA methods. The spelling rates are plotted as a function of the number of repetitions and classification accuracies (red dashed lines).

used in the test data for evaluation. From the selection accuracies of the test data, the spelling speed was calculated as described in Section 5.2.1. The values obtained by the best model selected by the proposed method were then compared with those achieved by shrinkage LDA. For each number of repetitions, two-sided Wilcoxon signed-rank test was run between the spelling rates achieved by the two methods (p-values corrected for multiple comparisons, alpha=0.0025).

5.3.2 Results and discussion

Fig. 5.6 shows the spelling rate and accuracies achieved with the simulations for the proposed MLD method (blue) and the shrinkage LDA used in the online experiment (green). For each number of repetitions, the mean values over participants are depicted. In case of the MLD, the average values refer to the spelling rates achieved by the best model selected from the training data. For both conditions, the most selected model was the MLD with a lookup period of 1 position, followed by the lookup period of 2 positions. The increment of spelling rate and selection accuracy is evident for the proposed method, especially between 3 and 5 repetitions for condition Color 116 ms and between 2 and 7 repetitions for condition Color 83 ms. For Color 116 ms condition the maximum rate of 2.19 symb/min is achieved after 3 repetitions with an accuracy of 86.7 %. For Color 83 ms condition, a higher rate of 2.24 symb/min is achieved with 5 repetitions with an accuracy of 92.32 %. Especially in the latter condition, the difference with the maximum performance of the shrinkage LDA is noticeable, i.e. 2 symb/min with 87.88 % accuracy. These results indicate that there is a

positive effect in considering the temporal dependencies between consecutive epochs, even though the differences between the two methods are not statistically significant[2]. This is not surprising since the order of the stimuli was randomized between different repetitions. As already mentioned in the previous section, the fact that the neighboring stimuli were changing between repetitions makes the exploitation of the overlapping neural components less effective in test phase. Therefore, it can be hypothesized that the differences between the proposed method and the shrinkage LDA might become even more pronounced if the sequence of symbols is fixed during the stimuli presentation. Interestingly, a recent study proposed a novel method based on the similar idea to exploit the features of specific nontarget stimuli to improve classification performances (Höhne et al., 2016). The aim was to investigate additional structure in the data that can be modeled with additional label information, i.e. they identify specific subclass labels. The model computes an individual classification hyperplane for each subclass and the subclasses are used also for estimating the regularization parameters of the LDA. The method was applied to both EEG and fMRI datasets, showing that the exploitation of the subclasses structure benefits most of the datasets. However, when applied on the RSVP Speller dataset described in this chapter, the novel method does not outperform the global LDA since the high number of distinct symbols leads to the estimation of an excessive number of parameters compared to the data points available. Nevertheless, this work stresses the point that investigating underlying data structure in classes normally considered as a whole, can improve classification performances and interpretability of the results. The method proposed in this section emphasizes the same concept, exploring the dependencies between epochs in a different model. The MLD method could be further investigated with other datasets, but considering that slower SOAs (e.g. those employed by the auditory spellers paradigms analyzed by Höhne et al. (2016)) would imply a reduced overlap between epochs and a reduced advantage of the method itself.

5.4 Chroma Speller: isotropic visual stimuli for truly gaze-independent spelling

5.4.1 Methods

Participants

Nine participants (4 males, aged 20-34) participated in the experiment. All had normal or corrected-to-normal visual acuity. None of them had a history of a neurological disease or injury. The study was performed in accordance with the declaration of Helsinki and all participants gave written consent.

[2]Accuracies achieved with the MLD method is significantly higher just for condition Color 83 ms at 6 repetitions ($p<0.0025$).

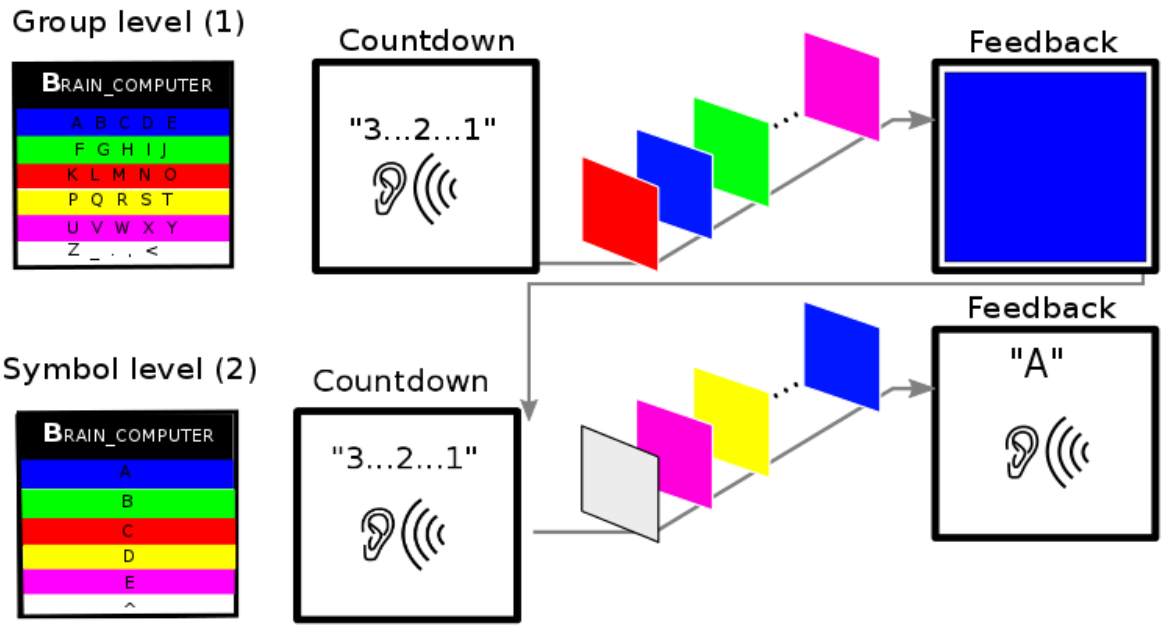


Fig. 5.7 Chroma Speller flowchart. The symbol selection occurred in two steps. Before each selection, the association between colors and either groups or single letters was displayed. The final feedback of the selected letter was provided in auditory way. ©2013 IEEE.

Apparatus

EEG was recorded with 64 electrodes according to the 10-20 system. Stimuli were presented on a 24" TFT screen with a refresh rate of 60 Hz and a resolution of 1920 x 1200 px^2 .

Design of the stimuli

In the Chroma Speller the stimuli consisted of 6 colors (red, green, blue, yellow, magenta, white), which were presented one after the other in a serial manner (Fig. 5.7). The colors were displayed widescreen for 150 ms with an Inter Stimulus Interval (ISI) of 150 ms, in which a black frame was interposed. The selection process was divided into two levels, analogous to that described in Treder et al. (2011). In the first level (group level), each color was associated to a group of characters, i.e. BLUE:'`a b c d e`', GREEN:'`f g h i j`', RED:'`k l m n o`', YELLOW:'`p q r s t`', MAGENTA:'`u v w x y`', WHITE:'`z - .  , <`'. In the second level (symbol level), each color was uniquely associated to the letters of the group that was selected. The sixth color, white, could be selected in the second level in case that the previous group level was classified wrongly. During the presentation, the order of the colors is randomized, and the whole sequence was repeated 5 times. For the detailed explanation of the design of the Center Speller, please refer to (Treder et al., 2011). In order to have a fair comparison between the performances of the two spellers, the Center Speller was operated with a SOA of 300 ms and using 5 repetitions of the stimuli.

Experimental procedure

Participants sat on a comfortable chair at a distance of approximately 80 cm from the display. In both the spellers, they had to perform calibration and online copy-spelling. The order of the spellers was counterbalanced across participants. In the calibration phase, the target phrase was '`BRAIN_COMPUTER`' and it was written on the top of the screen. Before each trial, participants had 5 seconds to identify the highlighted current target letter and to look to which color it was associated in a colored matrix displayed widescreen. The matrix was composed of 6 rows, each one being a box containing the group of letters (in the group level) or the single letter (letter level) and having as background color the corresponding one. After that, a 3 seconds auditory countdown started and the presentation of the colors followed. Participants had to silently count the number of occurrences of the target color in the presentation. In this phase, no feedback was given and the recorded data were used to train the classifier. In the copy-spelling phase, the procedure was the same as in the calibration, but the given sentence was '`THE_SUMMER_COMES_AGAIN.`' (and '`LET_YOUR_BRAIN_TALK_NOW`' for the Center Speller). In this phase, online feedback was provided to the participants, according to the classifier's output. In the group level, the feedback was the selected color. In the letter level, a voice spelled the selected letter. In case that the wrong group of letters was classified, participants were asked to focus on the white color in the second level. Note that in this study the participants could not perform a real-time error correction, so they proceeded with the next target even if a wrong letter was selected. In a real-world application, the sixth color in the second level can be used to go back to the group level, avoiding the spelling of a wrong character. Both the Chroma and Center Speller were implemented in the open-source framework Pyff (Venthur et al., 2010b) using VisionEgg (Straw, 2008).

Data analysis

For ERP analysis, EEG data were down-sampled to 200 Hz and lowpass filtered with a Chebyshev Type II filter using passband up to 40 Hz and stopband starting at 49 Hz. They were divided into epochs ranging from -200 ms to 1000 ms relative to the onset of each stimulus. Baseline correction was performed on the pre-stimulus period of 200 ms. Epochs containing eye movements were detected and rejected using a min-max criterion ($80 < \mu V$) on the channels F9, Fz, F10, AF3 and AF4. For classification, all epochs were used. For the grand average, the ERP curves were averaged across all trials and participants. To compare the ERP curves of two classes (target vs non-target), $sgn - r^2$-values were calculated. Classification was based on shrinkage LDA (Section 3.2.1). The time intervals for calculating the spatio-temporal features were determined by the heuristic searching for peaks based on the $sgn - r^2$ (Section 3.1.2). During the training of the online classifier, 5 different temporal windows were selected and occasionally adjusted by the experimenter. The online spelling-speed was calculated considering the number of correctly written symbols during

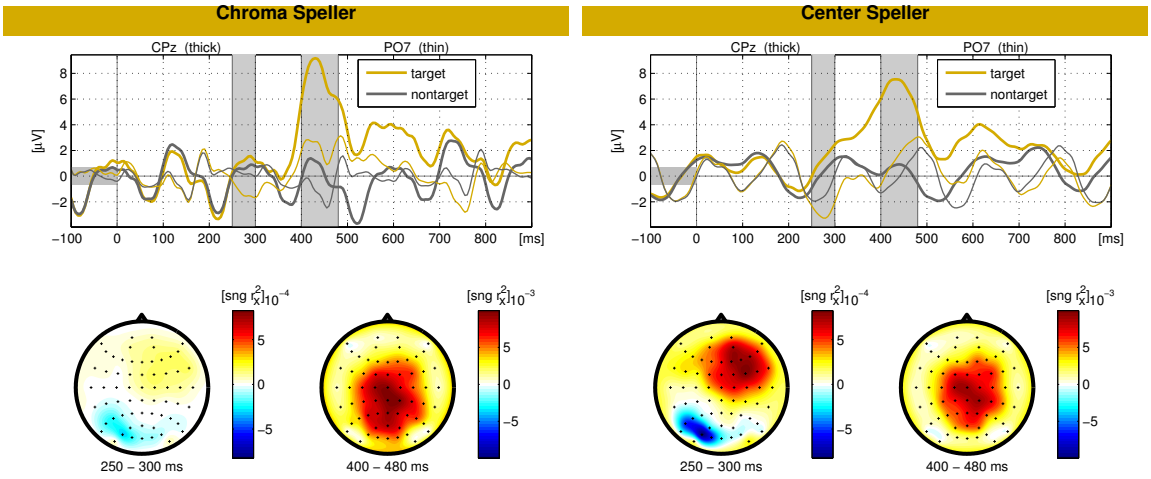


Fig. 5.8 Grand average ERPs for Chroma and Center Speller. The plots in the first row depict the grand average EEG over all the participants for targets (golden) and non-targets (grey). The thick lines represent the channel CPz, the thin lines the channel PO7. The shadowed areas highlight the two time-intervals 250-300 ms and 400-480, in which the N2 and P3 components are elicited. The scalp plots underneath refer to the shadowed areas in the ERPs plots. They picture the $sgn - r^2$ between the classes target and non-target. Note the different scales of the components between the two scalp plots. ©2013 IEEE.

the experiment and the duration of the selection of one symbol, which comprises: the time of the target's presentation and countdown (8 s for Chroma and 5 s for Center), the SOA, the time necessary to display the classifier's output (2 s) and the number of sequences. The offline theoretical spelling speed considers a more realistic scenario, in which spelling a wrong letter would cost the selection of two symbols (backspace and new symbol) and spelling the wrong group, but selecting the correct white color in the second level, would lead to no actual selection.

5.4.2 Results

ERP analysis

Fig. 5.8 shows the ERPs averaged over all participants and trials. There is a clear P3 component with a mean amplitude of 9.9 μV in the Chroma Speller and of 7.7 μV in the Center Speller. The mean latency of the P3 is about 430 ms in both the spellers. In both the spellers it is also visible a clear N2 component, more pronounced in the Center than in the Chroma Speller. The scalp plots underneath refer to the shaded intervals in the time plots and represent the scalp locations of the maximum discrimination of such components. The P3 is mainly evident on the central-parietal cortex, and the N2 around the PO7 channel. In the Center Speller, it is also visible an earlier frontal P3a subcomponent.

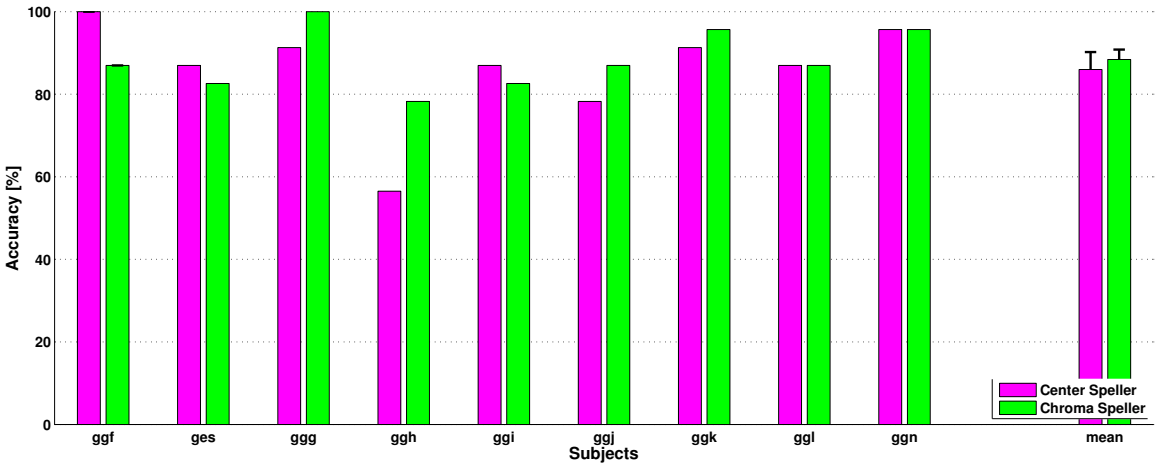


Fig. 5.9 Online classification accuracy Chroma and Center Speller. The mean accuracy is depicted in the last column. ©2013 IEEE.

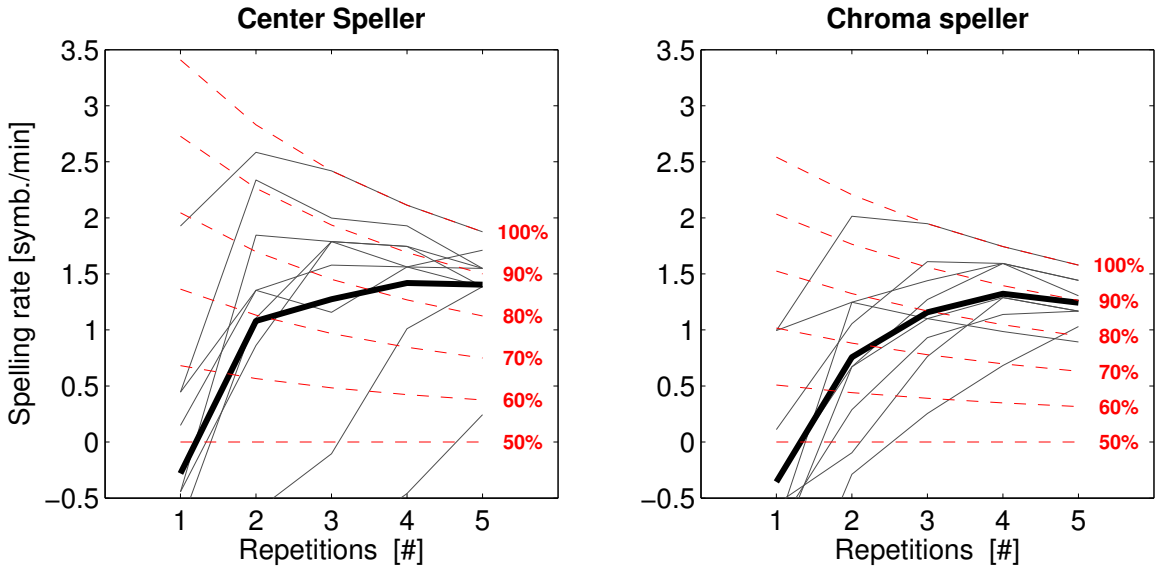


Fig. 5.10 Spelling rate Chroma and Center Speller. The spelling rate is plotted as a function of the number of repetitions and classification accuracies (red dashed lines). The gray lines represent the different participants, the back thick lines the mean. ©2013 IEEE.

Classification

Mean online classification accuracy of 88.4% for the Chroma Speller and 86% for the Center Speller were achieved (Fig. 5.9). Note that this result refers to the definition of symbol-selection accuracy adopted in the Matrix Speller (chance level 3.33%), in which if an error occurred in one of the two levels, the whole trial is considered misclassified. The mean online spelling speeds achieved were 1.61 chars/min for the Center and 1.4 chars/min for the Chroma Speller. Fig. 5.10 shows the trend of the spelling-speed calculated offline as a function of the number of repetitions and of the classification's accuracy (red dashed lines). The highest values can be found after 4 repetitions, with 1.42 chars/min for the Center and 1.32 chars/min for the Chroma Speller. Two-sample t-tests were performed on the classification accuracies and spelling-speeds and did not show significant differences between the two spellers ($p=0.49$, $p=0.24$).

5.4.3 Discussion

The Chroma Speller showed a mean online symbol-selection accuracy of 88.4%, using 5 repetitions of the stimuli. This accuracy was higher than that one achieved with the Center Speller, 86%. The achieved accuracy was competitive not only to that one of the Center Speller, but also with other gaze-independent visual spellers previously mentioned (Section 5.2; Acqualagna and Blankertz (2013); Liu et al. (2011); Treder et al. (2011)). It has to be noticed that in this study we employed just 5 repetitions of the stimuli, while the accuracies reported in literature go beyond 90% using 10 repetitions. The result of the classification reflected the trend of the ERP component P3, which showed an average amplitude higher in the Chroma than in the Center Speller. Since the amplitude of the P3 component is affected by the difficulty of the task, the ERP analysis suggested that the discrimination of the bright colors of the Chroma Speller is easier than the identification of the target symbols in the Center Speller, leading to a higher classification accuracy. Regarding the spelling speed, with the Chroma Speller an average of 1.4 chars/min was achieved online, and a maximum theoretical offline speed of 1.32 chars/min with accuracy of 88% and 4 repetitions. This result is competitive to those of the RSVP Speller and those presented by Treder et al. (2011) and Liu et al. (2011). The initial hypothesis of increased difficulty due to an extreme abstraction of the presented stimuli, was disproven. This demonstrated that once that the association color-symbol is learned, the task can be accomplished even if the letters are not explicitly printed on the display. In this study, the participants could identify the target color due to a matrix displayed before each trial. If the Chroma Speller was used in an application with patients affected by insufficient eye movement control, this association can be learned by heart before the customary use or, alternatively, the letters associated to particular colors can be presented in an auditory way before each trial until the associations are remembered. The auditory feedback was also an important characteristic of this speller, since patients could

not be able to discern clearly the spelled letters. In conclusion, this study demonstrated that it is possible to develop a truly gaze-independent visual speller, which adopts an abstract association between stimuli and symbols, without suffering in terms of performance and ease of use.

5.5 Limitations of the presented studies and recent advances in visual gaze-independent BCI spellers.

Even though the performances achieved with RSVP and Chroma Speller were competitive with other gaze-independent spellers, the spelling rates were still not comparable with those described in the matrix-based paradigms and not optimal for a practical use. This is one of the main limitations of these designs, especially if compared to technologies based on eye-trackers or Electrooculography (EOGs), which have shown to be successful in making patients communicate up to 10 words per minute (Majaranta et al., 2006). In order to make gaze-independent spellers more efficient, several options could be explored. For example, a possible approach to enhance the spelling rate would be to optimize the information transfer rate limiting the number of sequences, according to the participant's online performances (Liu et al., 2010; Schreuder et al., 2011a). A recent study investigated a similar concept using a design close to the RSVP Speller (Oken et al., 2014). Indeed, the speller employed the RSVP paradigm to show symbols on the display (named RSVP Keyboard), but with different features and longer SOA (400 ms) compared to the design of Section 5.2. The authors used a statistical language model together with the EEG input, and a classifier decision was taken if the classifier achieved a certain probability threshold after a number of sequences that could vary from 1 to 4. Interestingly, the study was performed also with patients affected by locked-in-syndrome (LIS) at different stages. Even though results were better for non-disabled participants, the RSVP Keyboard could be operated also by a small subset of people with LIS. Of course, there could be several factors that influenced the performances of the patients, like more difficulty with sustained attention and the use of EEG-altering medications (Meador, 1998; Riccio et al., 2012). These findings point out another limitation of the presented spellers, which is the evaluation with only healthy users. In fact, a BCI paradigm can be successfully operated by participants in the lab, but it could not succeed with severely impaired patients, who often suffer from cognitive dysfunctions that can negatively affect the spelling performances (Geronimo et al., 2016). Moreover, we have to consider that people at the latest stages of ALS not only may lose the control of the eye movements, but also they could not be able to focus anymore. This can prevent them to look at the center of the display, where the stimuli are presented. In this case, potential solutions could be moving the RSVP in the location of the gaze or presenting the isotropic stimuli of the Chroma Speller wider with a beamer or using video goggles. If the patient

is not even able to open the eyes anymore, then a viable option could be the employment of simple light stimuli administrated on closed eyelids as described in the recent study by Hwang et al. (2015).

5.6 Take-home messages

- Both the RSVP and Chroma Speller could be operated with high accuracies by all participants;
- Spelling rates of both systems were comparable to state-of-art gaze-independent visual spellers, but still lower than rates achieved by Matrix-based spellers;
- Unlike most of gaze-independent spellers, the RSVP paradigm allows a direct selection of symbols;
- The Chroma Speller paradigm uses isotropic stimuli that make it potentially suitable to patients with severely impaired oculomotor control;
- The performances of both RSVP and Chroma Speller could be further improved, e.g exploiting sub-class information, employing language models, dynamic stopping of the sequence presentation, etc.
- Both spellers should be investigated with patients in order to assess their performances with the real end-users and before drawing definitive conclusions about their usability.

Chapter 6

BCIs for relevance detection

6.1 Introduction and state of the art

Neurotechnology recently started to be exploited in alternative non-medical applications, e.g. Information Retrieval (IR), as complementary tool to infer the user's intentions together with the behavioral output. In this context, one central concept is that one of relevance (Mizzaro, 1997). Relevance in Information Science and IR is a broad term that has several definitions, as well as a number of dimensions or attributes, like specified by Saracevic (2007). In summary, the concept of relevance can be divided into two main classes (Borlund, 2003; Saracevic, 1975; Swanson, 1986): (1) objective or system-based relevance; and (2) subjective or human (user)-based relevance. While the first is a system-driven approach that refers to the relation between the retrieved information object and the query (e.g. algorithmic relevance, topical relevance), the second is a cognitive user-oriented approach that depends on the subjective experience. Examples are (Saracevic, 1996): cognitive relevance, related to the information need as perceived by the user; situational relevance, depending on the task interpretation; motivational and affective relevance, which is goal-oriented. Objective relevance has been widely implemented and IR systems achieve nowadays very accurate retrieval. Subjective relevance has also been investigated through the detection of the user's explicit or implicit feedback (Joachims et al., 2005; Kelly and Belkin, 2004; Koenemann and Belkin, 1996; White and Kelly, 2006), and through the use of physiological signals (Barral et al., 2015). However, these measurements are often unreliable (Kelly and Belkin, 2002; White and Kelly, 2006) and are unlikely to serve as relevance predictors. Therefore, in the last years there has been increasing interest in investigating the cognitive functions that are related to perceived relevance via the direct measurement of the brain activity (Eugster et al., 2014; Kauppi et al., 2015; Moshfeghi et al., 2013). In the study described in this chapter, subjective relevance was investigated with a BCI using both linguistic and pictorial stimuli. The next section describes the main characteristics that differentiate the 'standard' BCI experimental design to a more IR-oriented design, and which were the challenges in translating the well established BCI pipeline to this novel scenario.

6.1.1 Adapting the BCI design to the study of relevance

In a large number of BCI applications like those described in Chapter 5, the user has to choose the target stimuli that are associated to a specific action or selection (Wolpaw and Wolpaw, 2012a). If the BCI is ERP-based, the target stimuli are usually interspersed among a large number of non-target stimuli in an oddball presentation (Donchin et al., 2000b; Townsend et al., 2010). In order to make the BCI intuitive to use, the stimuli have quite simple features and the user learns them a-priori in order to recognize them at the first glance. The recognition of the stimulus is often associated with a cognitive task, e.g. mental counting, that would elicit an event-related cognitive component easy to detect and classify for machine learning predictions (Curran and Stokes, 2003; Sellers et al., 2006; Treder and Blankertz, 2010). In an IR application, the translation of the BCI pipeline is not straightforward and requires the refining of the interface design, the user's task and the prediction algorithms. In fact in an IR system the stimuli (e.g. relevant keywords) are not known a-priori by the user who types the query. In can be hypothesized that the user would scan many keywords in first place with the same attention and engagement, but then he/she would have a different cognitive response according to the relevance given to some stimuli rather than others. Moreover, the user is not expected to accomplish a stereotyped mental task while using an IR system, since it would affect the usability and pleasure of using the system itself. Therefore, it is of great interest to evaluate how a machine learning algorithm would perform with the complex and highly variable cognitive response. Another important difference between IR and standard BCI applications is the use of repeated sequences of stimuli. For example, in ERPs-based BCI spellers (Chapter 5), several repetitions of the alphabet or of the group of letters lead to an almost perfect prediction due to the averaging of the classifier outputs through the same type of stimuli. In IR this concept is more difficult to translate. It is clear that the mentioned points arise new interesting challenges both from the design of the interface and from the brain data analysis. In order to address such challenges, we extended the well established BCI paradigm to a design that addressed relevance. The simple target-non-target presentation was revised using more complex stimuli. The features of the stimuli were not know a priori by the user. This choice was done to minimize that the acknowledgement of a stimulus would be due to the characteristic of the features rather then the top-dow process of relevance assessment (Fecteau and Munoz, 2006). Moreover, the user was instructed to not perform a specific mental task associated to the recognition of the relevant stimuli, simulating a more realistic scenario of free use of the IR system. From the data analysis side, a variation of the shrinkage LDA that copes with the high inter-trial variability of the brain response was evaluated in comparison to the standard shrinkage LDA (Section 3.2.1). Both temporal and spectral features were investigated in order to have an extensive view of the brain process involved in the relevance assessment and to open the way for future

investigations and improvements. Data and results of this study were submitted as journal publication (Acqualagna and Blankertz, 2016).

6.2 Methods

6.2.1 Experimental design

Participants

Sixteen people participated in the study (8 females, age range 22-43 years). One participant was excluded from the analysis because of an inadequate preparation of the EEG cap. All participants had normal or corrected-to-normal vision and no history of neurological diseases. The study was conducted in accordance with the declaration of Helsinki and approved by the local ethics committee. Participants gave written information consent prior to the experiment and received 8 Euro per hour for their participation.

Apparatus

The electrodes configuration used in this study was: Fp1,2, AF3,4, Fz, F1-10, FCz, FC1-6, FT7,8, Cz, C1-6, T7, CPz, CP1-6, TP7,8, Pz, P1-10, POz, PO3,4,7-10, Oz, O1,2. One electrode normally placed at T8 was placed under the right eye and used as vertical EOG channel. Stimuli were shown on a 23" screen (Dell U2311H) with a native resolution of 1920x1080 pixels at a refresh rate of 60Hz. Participants sat on a comfortable chair at a distance of approximately 80 cm from the display.

Experimental design

The experiment consisted in 40 runs, 20 having as stimuli cartoon-like pictures and 20 having words. In both conditions we defined six categories which corresponded to six groups of stimuli (Fig. 6.1). Five categories referred to different stereotypes of characters, i.e. the shopping addict, the nerd, the globetrotter, the sportsman and the philosopher. Each of these categories consisted in 29 stimuli. The sixth category included general objects not specific to any stereotype of character and that could be normally found in a house, for example cutlery, plants, food, toothpaste etc. This category contained 56 pictures and 60 words. In case of the condition *pictures*, the stimuli were cartoon-like objects extracted by pictures downloaded from the online open-access database of vector graphic pictures www.freepik.com. The pictures could have been square or rectangular shaped, and they were all circumscribed in the same area of the display, i.e. a square whose side measured 0.15 times the width of the monitor and centered in the middle in order limit eye movements (5.4° visual angle). Examples of pictures for each category are depicted in Fig. 6.1. In case of condition *words*, the height of the text was 0.06 times the height of the monitor and centered in the middle.

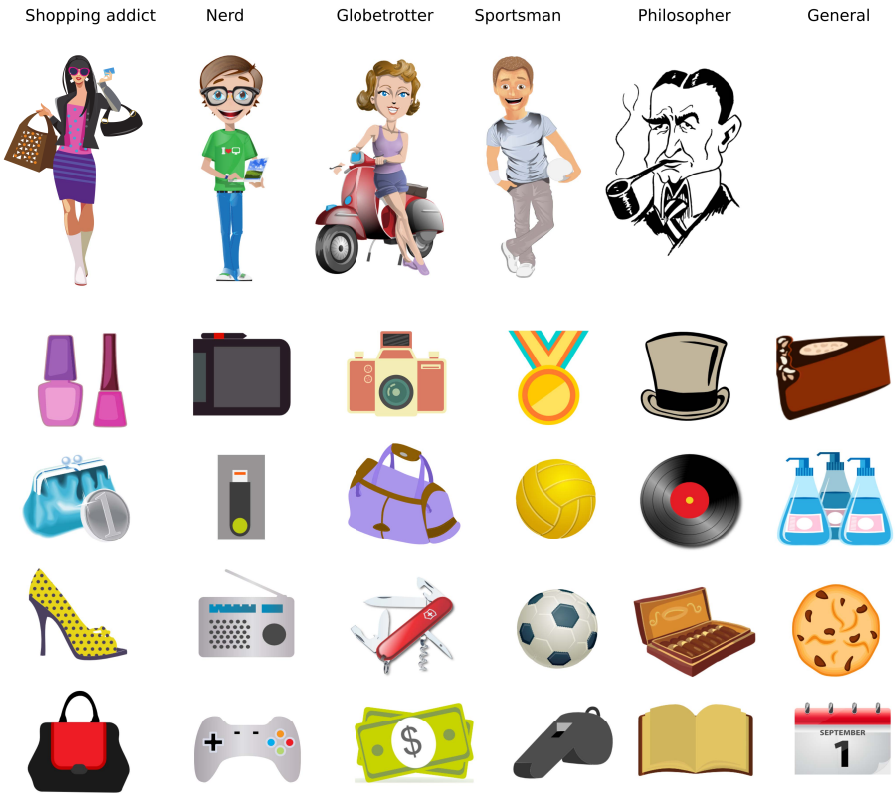


Fig. 6.1 Example of stimuli for *pictures* condition. The first row shows the six categories. In the columns below each category, an example of four pictures that were included in the stimulus presentation is depicted. Pictures were extracted from vector graphics dowloaded from www.freepik.com.

The text was displayed in German, since all participants were German mother tongue. This choice avoided that time differences in semantically processing the stimuli would be interfered by mental translation from e.g. English to the participants mother tongue (Hahne, 2001). The stimuli were displayed in a serial manner in the centre of the monitor for 1250 ms, with a stimulus onset asynchrony (SOA) of 1750 ms. For each run the total number of stimuli was 60 and comprising all the six categories as explained in the next section.

6.2.2 Experimental procedure

The participants were instructed about the task to accomplish during the preparation of the EEG cap. They had time to read the papers describing the characteristics of the six categories and ask questions in case something was not clear. In the papers also an example of stimuli was given, i.e. five or six stimuli among those ones that were displayed afterwards. In general, participants did not know exactly which and how many objects belonged to each character, but they had to have a clear understanding of the peculiarity of each category in

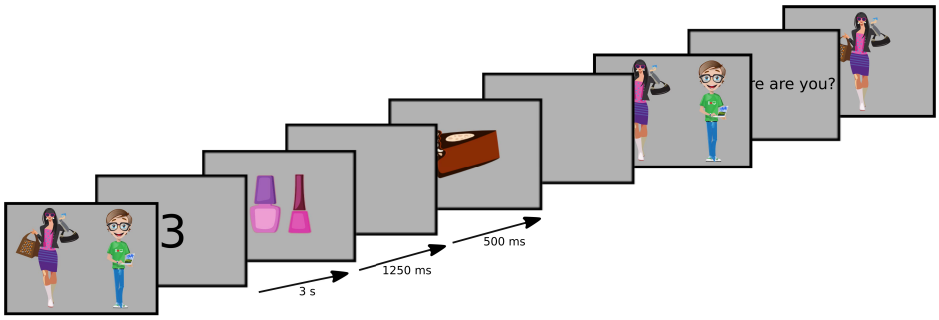


Fig. 6.2 Example of run for *pictures* condition. After displaying the two relevant categories, the presentation shows the stimuli in a serial manner. Finally, after the selection of one of the categories and the confidence question, feedback about the correct selection was given to the user.

order to discriminate the objects in the stimulus presentation. The choice of not showing all the objects before the experiment was deliberate. This would reduce the possibility that the recognition of the stimulus would be based on mere memory association and not to the cognitive association of the stimulus to the relevant characteristics of the category. Participants were told to imagine that the five characters described in previous section would live all in the same house, and therefore leave some of their belongings in the kitchen. The sequence of stimuli consisted in objects that were found in the kitchen. For each run, at the beginning two characters were displayed. The imagined setting is that it was the turn of one of those persons to do the cleaning. The task of the participant was to decide who of those two had to clean the room based on the objects found in the kitchen. These two characters were therefore 'relevant categories' and the stimuli depicting objects belonging to them were 'relevant stimuli', since they provided relevant information in order to make the participant decide who had to clean. The participant had to identify the correct character in charge of the cleaning according to the quantity of objects identified as belonging to him/her in the sequence of stimuli. The number of stimuli that corresponded to the correct character, referred as R1 category in the following part of the chapter, varied between 8 and 12. The number of the second relevant category R2 varied between 6 and 8. The rest of the stimuli consisted in: stimuli belonging to the other three characters, which were not relevant for that run, and stimuli belonging to the sixth general category, which was always non-relevant. All these stimuli will be referred as NR. The number of stimuli of each of the three non-relevant characters was the same as the number of stimuli R2, in order to check that the modulation of the EEG signal was due just to the right acknowledgment of relevance of the stimuli and not to a difference of their occurrences. The number of the stimuli of the general non-relevant category was 60 minus the number of stimuli of the other five categories. Even though the number of stimuli was indicative for the accomplishment of the task, participants were asked to avoid counting and the number of R1 and R2 stimuli was not asked in the end of the

run. They were aware that the purpose of the experiment was of paying attention to the stimuli, recognizing and associating the relevant stimuli to the correct character. In each run, the stimuli were randomly picked among the pool of pictures/words of each category. Since stimuli belonging to all categories were displayed in all the runs and participants did not have the certainty about the owner of the displayed object, the possibility that they would learn the objects, and therefore influence the EEG responses through the experiment, was reduced. As previously mentioned, at the beginning of the run the two relevant characters were shown at the left and right side of the display, either as pictures or as words, according to the type of the run (Fig. 6.2). The position was randomly assigned. When the participant was ready to start the run, she could press the 'down arrow' key in the keyboard. After a 3 seconds countdown, the stimulus presentation started. Participants were asked to look in the middle of the display, relax and limit eye movements as much as they could, in order to minimize the contamination of the EEG with artifacts. After the presentation, the two relevant characters were displayed again in same positions and participants had to express their choice pressing on the keyboard either the 'right arrow' key for the character on the right or the 'left arrow' key for the one on the left. Subsequently, the question 'How sure are you?' was displayed and the participants could press one of four keys in the keyboard that corresponded to: 1-sure, 2-almost sure, 3-almost unsure, 4-unsure. The final question was meant to be an extra control that the participant would actually do the task and not just random guessing. After the selection, the correct R1 was displayed for 1s as feedback. Before the main part of the experiment, 4 training runs were run with button press, 2 with pictures and 2 with words. During the stimulus presentation participants were asked to press the 'right arrow' key as soon as they recognized an object belonging to the relevant character that was displayed at the right at the beginning of the run, the 'left arrow' if they recognized an object belonging to the relevant character on the left, and the 'down arrow' for any other stimulus. The button press was done with right hand, which was the dominant one for all participants. Except for this, the rest of the run was exactly as described in the previous paragraph. This short test let the experimenter check that the participant could correctly identify the relevant stimuli, had a good understanding of the characteristics of the categories, and in case to give further explanations. Data recorded in these runs were not used in the analysis.

6.2.3 Preprocessing and data analysis

EEG data were down-sampled to 200 Hz and low-pass filtered at 80 Hz with a Chebyshev Type II filter of order 9, for storage purpose. Artifacts due to eye movements, muscular contractions, loose electrodes etc. were detected and corrected using Independent Component Analysis (ICA) with an automatic artifactual component classifier method (MARA) (Section 3.1.1).

ERP analysis and classification

For ERP analysis and ERP classification, continuous signals were low-pass filtered under 45 Hz with a Butterworth filter of order 5 and high-pass filtered at 0.5 Hz with a linear phase FIR filter applied forward and backward. EEG data was divided into epochs ranging from -150 to 1250 ms relative to the stimuli onset. Baseline correction was performed on the pre-stimulus interval from -150 to 0 ms. Epochs were averaged considering the three classes R1, R2 and NR. For grand-average, the ERP curves were averaged across all epochs (trials) and participants. As a measure of separability of the relevant classes vs. the non-relevant classes, the signed square of the pointwise biserial correlation coefficient ($sgn - r^2$) was calculated (Section 3.1.2). For binary classification, epochs belonging to R1 and R2 were merged into a general single class 'relevant' (R), which was classified against the class 'non-relevant' NR. Shrinkage LDA was used, with spatial-temporal features (Section 3.1.3). As spatial features 31 channels were selected, focused on the central, parietal and occipital cortex (C1-6, Cz, CP1-6, Cz, P1-4, Pz, P7-10, PO3,4, Pz, P7,8, Oz, O1,2), chosen after the visual inspection of the ERP patterns. For temporal features, 7 non-overlapping time intervals between 100 and 1000 ms after the stimulus onset were selected in the training set. The validation was done using 5-folds cross validation with 5 nested shuffles. In each run, the classifier was trained on 4 folds and tested on the remaining one. The test of the classifier was done in two ways: in the first, the classifier was applied on the test set using the same temporal features selected on the training set (i.e. mean at time intervals $\mathcal{T} = \langle \mathcal{T}_k \rangle_{k=1,\dots,K}$, with K=7), as already successfully used in previous ERP studies described in Chapters 4 and 5. In the second, the classifier was tested on sliding time windows (Marathe et al., 2014) in the following way. Let be $\Delta = \{-50, -45, \dots, +50\}$ a set of time shifts (measured in ms) with a step of 5 ms, i.e. one sample. For each time shift, the set of time intervals $\mathcal{T}$ becomes $\mathcal{T}^\delta = \langle \mathcal{T}_k^\delta \rangle$, where $\mathcal{T}_k^\delta$ is the k-th time interval shifted by time lag $\delta \in \Delta$. Recalling formula 3.15 of the general LDA model, the final classifier output is computed by

$$y_n^\Delta = \frac{1}{|\Delta|} \sum_{\delta \in \Delta} \mathbf{w}^\top \mathbf{x}(C, \mathcal{T}^\delta) \tag{6.1}$$

that is, for each trial the final classifier output is the average over the classifier outputs of each time-shift.

The choice of this kind of testing was motivated by the hypothesis that the complex stimuli used in the experiment would lead to a difference in the latencies of the ERPs between trials. A classifier set on fixed time windows would not catch these time-shifts. The classifier performances were evaluated using the AUC values (Section 3.2.2). Statistical tests on classification results between the two methods were conducted using paired t-test, with Bonferroni correction for the p-values in case of multiple comparisons.[1]. A second analysis

[1] Assumptions were checked with one-sample Kolmogorov-Smirnov test.

was performed in order to investigate how the average classifier outputs corresponding to the categories R1, R2 and NR were ranked at the end of each run. This evaluation was done with the classification method that showed the best approach on single trial, i.e. the sliding shrinkage LDA. Data were trained as described in the previous paragraph, i.e. shrinkage LDA classifier was trained considering R vs NR, with 31x7 spatio-temporal features. For the training set, epochs belonging to the even runs were used, for the testing set epochs belonging to odd runs. This choice was done in order to prevent that slow drifts of the EEG features during the experiment would affect the accuracy of the classifier. For online operation such drifts would be encountered by an unsupervised adaptation of the mean (Vidaurre et al., 2011a). On the test epochs, the sliding approach was adopted. For each run, the classifier outputs belonging to the same category ('shopping girl', 'nerd', etc.) were averaged over the epochs. The 6 mean classifier outputs were ordered in ascendant order. Each category was ranked with a value corresponding to its position, i.e. from 1 to 6. In the ideal case, the relevant categories should be in the first two positions, since negative classifier outputs referred to class R. The other 4 categories should have ideally a score between 3 and 6. The average ranking over runs was then computed. Statistical tests on ranking results were done using the Kruskal-Wallis test.

For the behavioral assessment, in order to check that participants were correctly engaged in the task, the average responses were collected over runs and participants, together with the confidence of the answer.

Spectral analysis and classification

In order to investigate in which frequency bands there was the highest discrimination between relevant and not relevant stimuli, the $sgn - r^2$ between the amplitude of the wavelet coefficients was computed. After a broad band inspection, the analysis was limited to the interval between 1 Hz and 30 Hz. Data were divided into epochs ranging between -350 ms and 1500 ms and the channels 'Cz', 'Pz' and 'Oz' were selected. The decomposition was done using a five-cycles Morlet wavelet as mother wavelet (Section 3.1.7). Baseline normalization was performed for each channel and frequency considering the average values in the interval between -350 ms and -50 ms pre-stimulus, averaged over trials. The $sgn - r^2$ between classes R1-NR and R2-NR were calculated for each participant and subsequently averaged. Significance values were derived using the z-test and p-values were Bonferroni-corrected to account for multiple comparisons. The significant time-frequency bins were highlighted. Three frequency intervals that showed significant discrimination between the classes were chosen for the extraction of the Event-Related (De)synchronization (ERD/ERS). The frequency intervals were: 2-4 Hz, 10-14 Hz, 15-20. Data were first band-pass filtered with a zero-phase Chebyschev Type II filter with the pass-band ripple of no more than 2 dB and a stop-band attenuation of at least 80 dB. The envelope of the oscillatory signal was calculated taking the absolute

value of the Hilbert transform (Section 3.1.6) and baseline correction was performed in the interval between -200 ms and 0 ms of the average activity of all trials. The $sgn - r^2$ values between classes R1-NR and R2-NR were calculated. Data were then averaged over participants. Classification was performed on the spectral features. Continuous data were first preprocessed using the spatio-spectral decomposition (SSD) (Section 3.1.5) for dimensionality reduction. The frequency bands of interest were the same three as those inspected in the ERD/ERS analysis. The filtering of the left and right side bands was implemented by applying a bandpass filter covering both side bands followed by a bandstop filter cutting out the band-of-interest. The bandpass filters were implemented in a range [-3 +3] Hz in respect to the signal bands , while the stop-band intervals were [-1 +1] Hz in respect to the signal bands. The only exception was the left side band of the first frequency band, which was [0.1 1] Hz. The number of selected SSD components were set to 15, according to results of Haufe et al. (2014a). Baseline correction was performed like previously described in the ERD/ERS analysis. As for the ERP classification, classes R1 and R2 were combined in a unique relevant class R, which was classified versus the non-relevant class NR, because the assumption in this study was that the neural modulation of both relevant categories R1 and R2 was based on the same processes. Validation was done using 5-folds cross-validation, with 5 nested shifts. Within the cross-validation, features were selected as follows. For each epoch, the time interval between 300 ms and 1200 ms after the stimulus onset was selected. CSP analysis was performed (Section 3.1.4). For each frequency band, up to 6 CSP filters were selected, i.e. up to the three CSP filters for each class. The CSP features were the log-variance (in three equally spaced intervals, to address the different latencies of the modulation of the three frequency bands) of the band-pass and CSP filtered data. The final feature vector was the concatenation of the CSP features of each frequency band, with a final dimensionality below 3x18. The classifier used was LDA. Classification performances were evaluated as AUC values.

6.3 Results

6.3.1 Behavioral data

Fig. 6.3 depicts the average number of correct and wrong runs for both conditions *pictures* and *words*. All participants achieved very good performances, selecting the correct category on average 17.7 runs out of 20 for *pictures* and 17.1 out of 20 for *words*. Among the correct answers, an average of 72% was evaluated 'sure' or 'almost sure' by the participants, for both conditions. For the wrong answers we observed a different trend, with an average 71% of wrong selected *pictures* runs and 81% of wrong selected *words* runs indicated as 'almost not sure' or 'not sure'. The results indicate that participants understood correctly the task and were properly engaged.

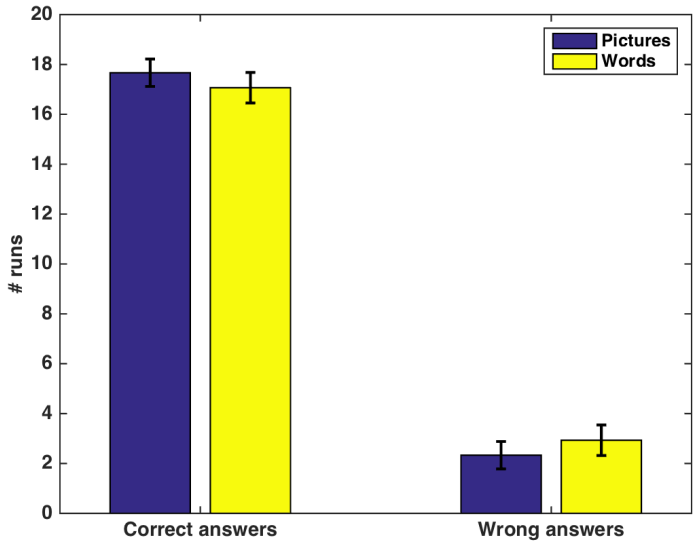


Fig. 6.3 Behavioral responses. The average number of correct and wrong runs is depicted in blue for *pictures* and in yellow for *words*. The error bars represent the Standard Error of the Mean (SEM).

6.3.2 Neurophysiological data

Time domain.

The grand averages ERPs over trials and participants are depicted in Fig. 6.4 and grand averages $sgn - r^2$ in Fig. 6.5, as a measure of discrimination between R1 and NR classes. The ERP plots refer to the epochs time-locked to R1 stimuli and NR stimuli. Plots referring to epochs time-locked to R2 stimuli and NR stimuli show similar pattern and can be found in Appendix A (Figs. A.1 and A.2). Regarding the pictorial runs, the ERP responses elicited by R1 and NR are quite complex (Fig. 6.4, top). Until 400 ms after the stimuli onset, the ERPs of relevant and non-relevant stimuli are overlapping and their scalp distribution is comparable. An early negative deflection (N1) can be distinguished between 150-180 ms, followed by a first positive peak around 250 ms (P2), whose focus is clearly distinguishable in the central and parietal cortex. Between 250-450 ms a pronounced positive visual ERP is also present in both classes. These ERP components are generally connected to cognitive processes like attention, stimulus discrimination and evaluation (Potts, 2004). The comparability of these ERP responses between the two classes R1 and NR can be attributed to the complexity of stimuli and to the necessity to evaluate each stimulus before making the judgment about its category. After 400 ms the ERPs locked to the NR stimuli decrease in amplitude, both in the occipital and parietal cortex. On the other hand, the R1 class exhibits a pronounced late positive complex (LPC), peaking around 600 ms and focusing on the parietal cortex. As

described in Section 2.3.1, the LPC is often associated to recollection of specific information and memory processes. The late ERP characteristic of the relevant category indicates the deeper cognitive elaboration of the relevant stimuli compared to the non-relevant ones. From the time course of the $sgn - r^2$ (Figure 6.5, top) is evident that the maximum discrimination occurs in the time interval 450-850 ms, i.e. at the latency of the LPC. The scalp plots visualize the distribution of the $sgn - r^2$ values over the scalp. Highest values focus mainly in the parietal cortex. Figs. 6.4 (bottom) and 6.5 (bottom) represent the ERP plots and $sgn - r^2$ for linguistic runs. Similar to the *pictures* runs, comparable early ERP responses occur for both R1 and NR classes. The N1 component is more pronounced for the R1 stimuli and has a frontal-central scalp distribution, followed by a positive peak around 300 ms focused in the occipital cortex. Interestingly, the NR class shows a pronounced negative deflection (N400) in the frontal central cortex. This ERP is known to be elicited in response to words, comprehension, semantic memory (Kutas and Iragui, 1998; Kutas and Van Petten, 1994). After 500 ms a pronounced LPC is visible for class R1, mainly focused in the parietal cortex and peaking around 700 ms. Therefore, also for the *words* runs a stronger cognitive response is elicited by the relevant stimuli. The $sgn - r^2$ values reveal that the main discrimination between the two classes R1-NR occurs in the timing of the LPC and in the central-parietal cortex. Note that, the amplitudes of the ERPs elicited by the linguistic stimuli and the relative $sgn - r^2$ are smaller if compared to the amplitudes associated to the pictorial stimuli (reflected in the different scales of the plots), and their latencies are bigger. This can indicate that the cognitive response elicited by the relevant pictures is stronger than that one elicited by relevant words, leading to higher discrimination between relevant and non-relevant stimuli.

Spectral domain.

As for the ERP analysis, in the frequency-domain results just the neurophysiological plots referring to classes R1 vs NR are depicted. The class combination R2-NR showed in some cases (e.g. modulation of the delta band) a smaller discrimination compared to class-combination R1-NR, but the topographies and the latencies of the discriminative components were the same in all the bands of interest (Appendix A, Fig. A.3). Figs. 6.6 and 6.7 show the $sgn - r^2$ between the amplitudes of the wavelet coefficients of class R1 and class NR. The non-significant time-frequency bins were set to zero. Analogous to the time-domain case, the significant differences in the spectra mainly occur in a range between 500 ms and 1000 ms after the stimulus onset. The discrimination is higher for *pictures* than for *words*, but the affected bands are the same, that is: The delta band 2-4 Hz mainly at electrode Pz, and the upper alpha and beta bands, mainly between 10 Hz and 20-25 Hz. According to these results, ERD/ERS were calculated for the three bands of interest: 2-4 Hz, 10-14 Hz, 15-20 Hz. The $sgn - r^2$ between the classes R1 and NR are depicted in Fig. 6.8, left column for *pictures* and right for *words*. In all the frequency bands, the discrimination

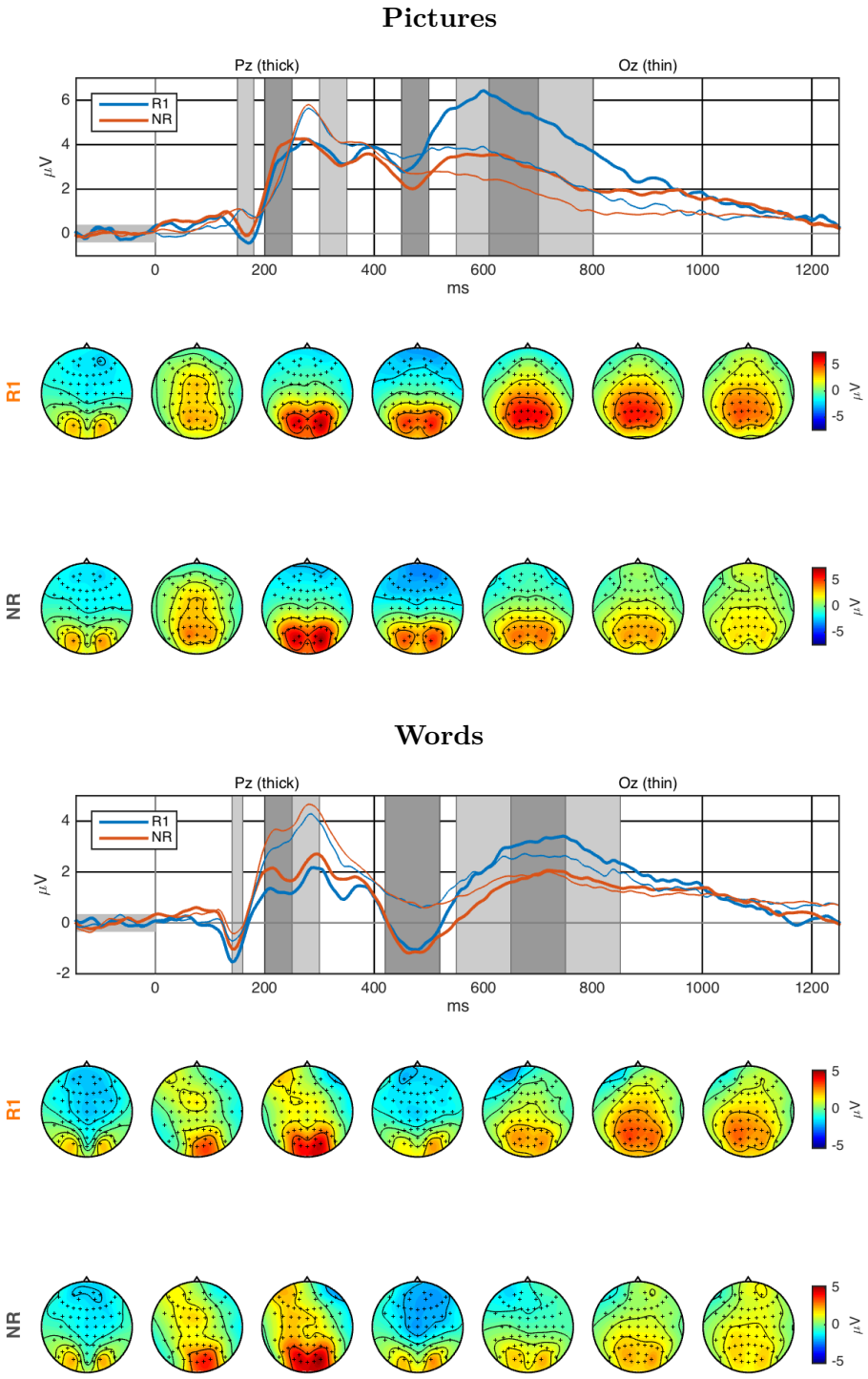
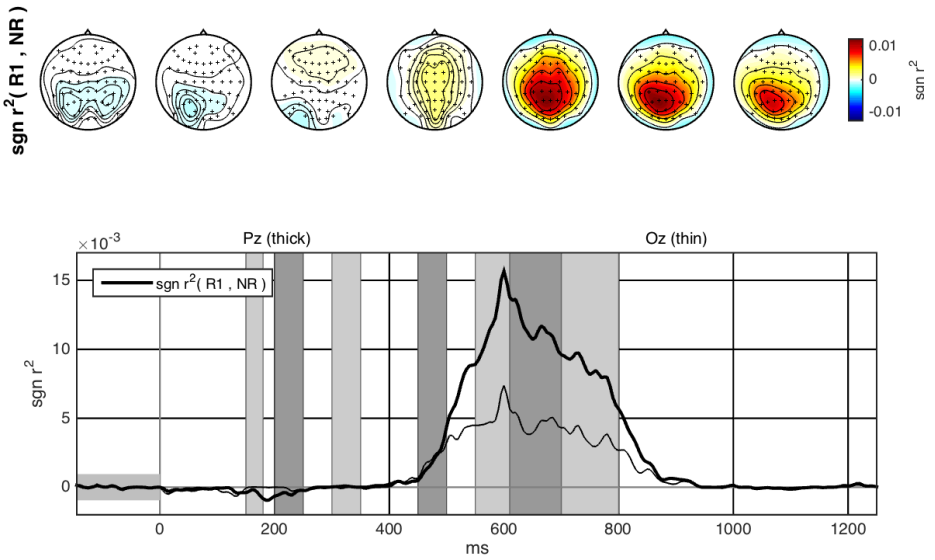


Fig. 6.4 Grand average ERP evolution time-locked to relevant (R1) and non-relevant (NR) stimuli and scalp plots, for *pictures* (top) and *words* (bottom). In the first row the time evolution of the ERPs is depicted at channels Pz (thick lines) and Oz (thin lines). In blue are the ERPs of class R1, in orange the ERPs of class NR. Time 0 correspond to the onset of the stimuli. The shaded areas highlight interesting time windows wherein the average voltage is calculated and displayed as scalp plots underneath (second row refers to R1 and third row to NR). Note that plots referring to *pictures* have different scales to those referring to *words*. Time windows are differently tuned for *pictures* and *words*.

Pictures



Words

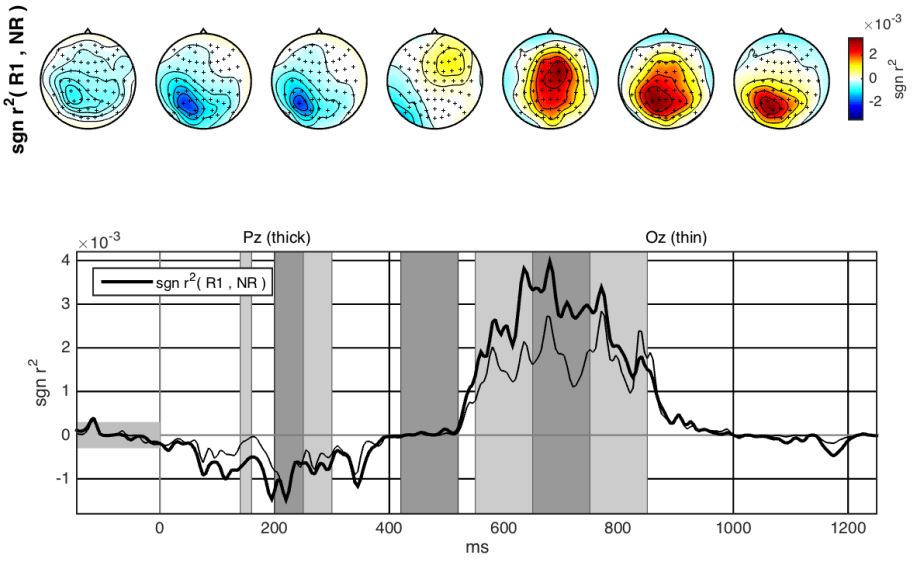


Fig. 6.5 Grand average $sgn - r^2$ of classes R1 and NR and scalp plots, for *pictures* (top) and *words* (bottom). The same time windows of Fig. 6.4 are highlighted. Note that plots referring to *pictures* have different scales to those referring to *words*.

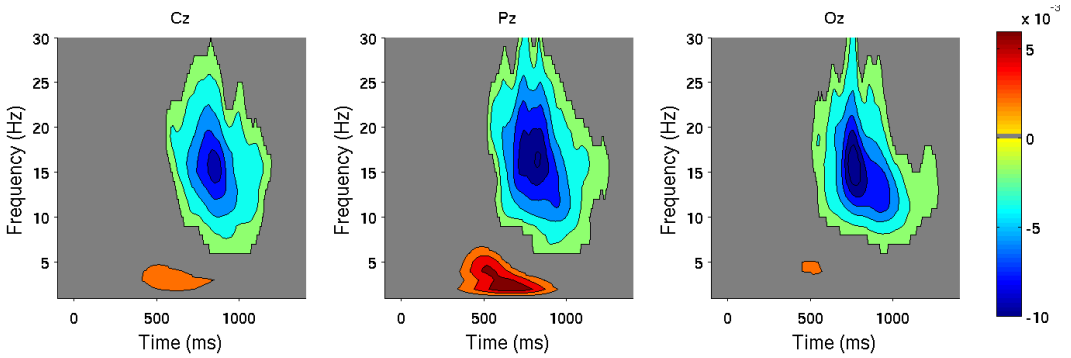


Fig. 6.6 $Sgn - r^2$ of amplitudes of the wavelet coefficients of classes R1 and NR, *pictures* condition. From left to right, plots depict channels Cz, Pz and Oz. Frequency range varies between 1 Hz and 30 Hz and time interval between -50 ms and 1500 ms. Non-significant values are set to zero.

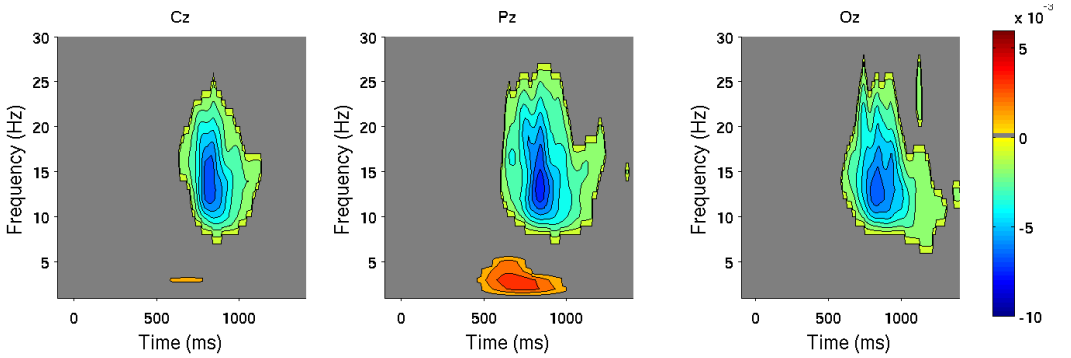


Fig. 6.7 $Sgn - r^2$ of amplitudes of the wavelet coefficients of classes R1 and NR, *words* condition. From left to right, plots depict channels Cz, Pz and Oz. Frequency range varies between 1 Hz and 30 Hz and time interval between -50 ms and 1500 ms. Non-significant values are set to zero.

between the relevant and non-relevant categories is higher for *pictures* than for *words*, but the latencies and topographies are comparable between the two conditions. In the delta band the power increases with the relevant stimuli compared to the non-relevant, that is an ERS. The maximum discrimination occurs at 600 ms in the parietal cortex, at the same latency of the maximum $sgn - r^2$ of the ERPs. The alpha and beta bands decrease their power with the relevant stimuli, i.e. an ERD. The discrimination between classes is visible already at 600 ms and peaks between 800 ms and 900 ms in the occipital cortex, but extends also to the central and parietal cortexes.

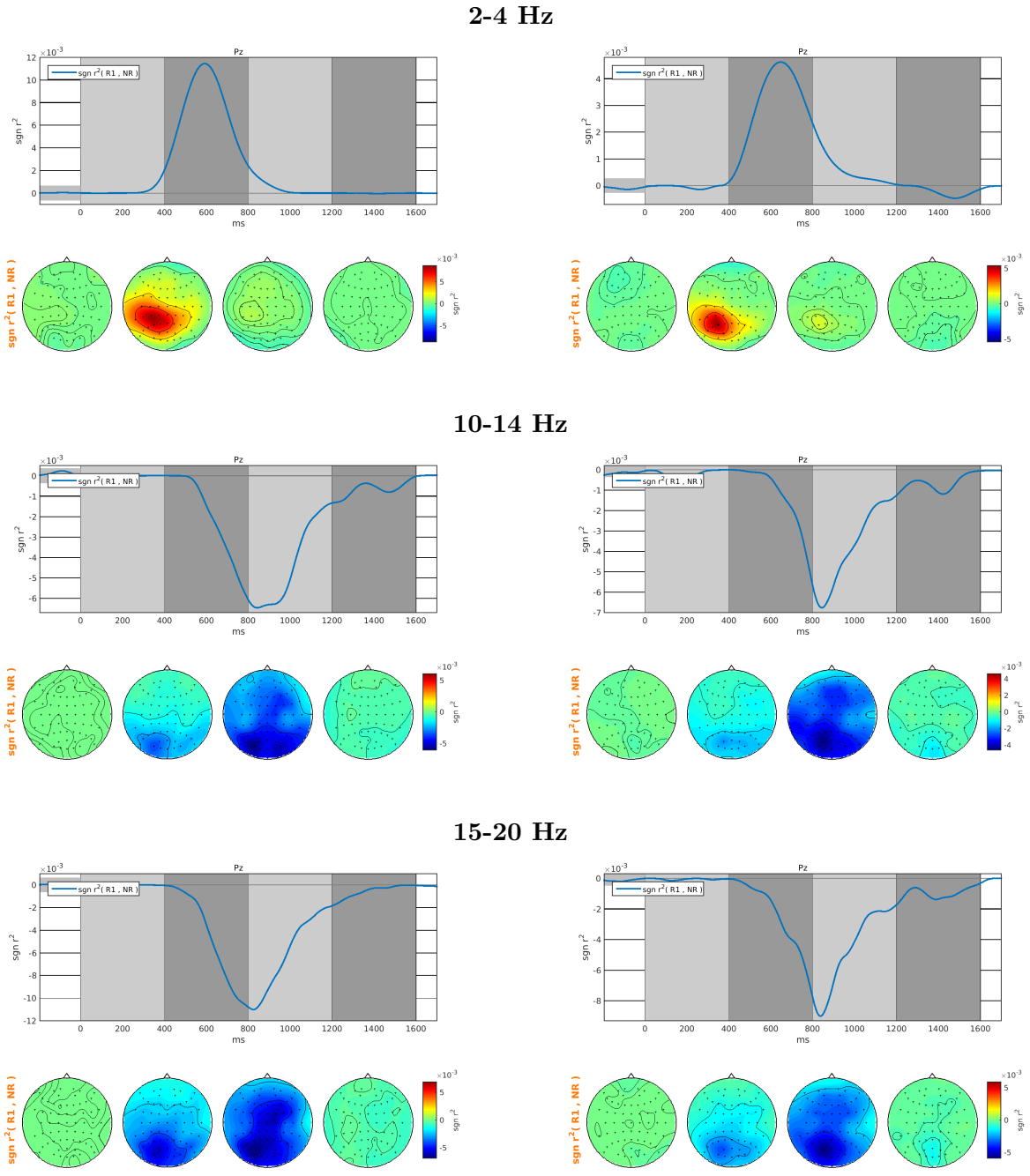


Fig. 6.8 $Sgn - r^2$ of ERD/ERS of classes R1 and NR for *pictures* (left column) and *words* (right column). From top to bottom the three investigated frequency bands: 2-4 Hz, 10-14 Hz, 15-20 Hz. Note the different scales between plots.

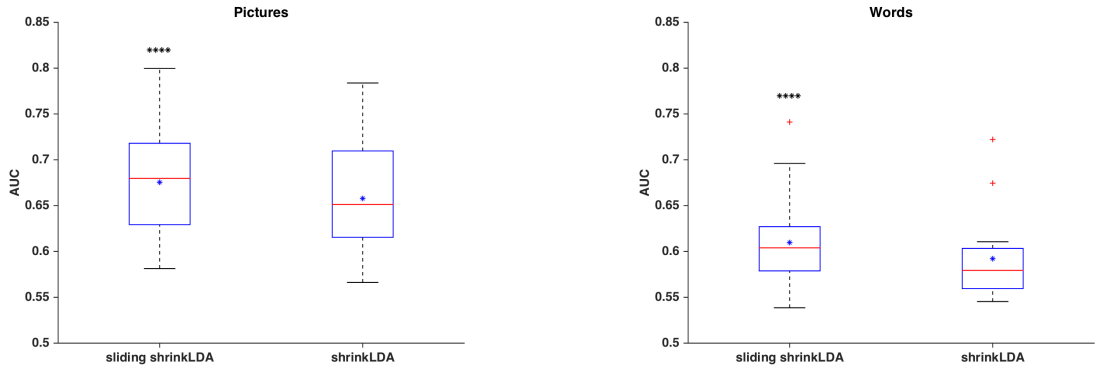


Fig. 6.9 ERP classification for *pictures* (left plot) and *words* (right plot). In both plots, results for sliding shrinkage LDA (left) and standard shrinkage LDA (right) are depicted. AUC values refer to single-trial classification performance of R1 vs NR using patio-temporal features. The central red marks are median values and the stars the mean values. The edges of the boxes are the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers and outliers are plotted individually as red crosses. Black asterisks over the box plot indicate statistical significance.

6.3.3 Classification

Single-trial classification.

The results of the two ERP classification methods are depicted in Figure 6.9 for *pictures* and *words* runs, as AUC values. The standard shrinkage LDA evaluation led to average AUCs of 0.66 for *pictures* and 0.59 for *words*. The sliding shrinkage LDA led to average AUCs of 0.68 for *pictures* and 0.61 for *words*. The performances achieved in the *pictures* runs are significantly higher than those achieved in the *words* runs (un-paired t-test, $p<0.01$). This reflect the neurophysiological results, which showed that the discriminability between relevant and non-relevant stimuli is higher in case of *pictures*. In the *pictures* condition the variability of the performances is higher than for the *words* condition, with participants achieving performances up to 0.8. Comparing the two classification approaches, the sliding-windows method brings higher performances compared to the standard method, for both conditions, and the difference between the mean AUCs is significantly different from zero ($p<0.0001$). This result suggests that the sliding testing can be considered as a successful approach in case of complex stimuli that elicit ERPs whose latencies can show a high intra-trial variability. Single trial classification on spectral features of the three frequency bands of interest led to average AUCs of 0.65 for *pictures* and 0.60 for *words* (significantly higher for *pictures* (Fig. 6.10), un-paired t-test, $p<0.05$), comparable to the ERP classification with the standard shrinkage LDA method. This result indicates that the modulation of the brain oscillation by relevant stimuli can be detected on single trial with certain degree by machine learning methods.

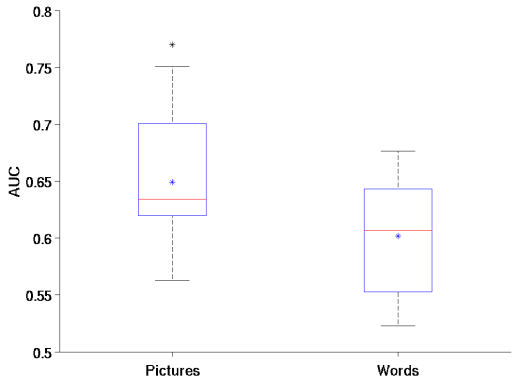


Fig. 6.10 Classification of spectral features. The left box depicts performances for *pictures*, the right for *words*. The central red marks are median values and the stars the mean values. The edges of the boxes are the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers and outliers are plotted individually as red crosses. The black asterisk over the box plot indicate statistical significance.

Relevance ranking.

Fig. 6.11 depicts the results of the classification ranking based on ERP features, for *pictures* (left) and *words* (right). Note that the ideal ranking would place the relevant categories R1 and R2 in first and second place, with an average ranking of 1.5, while the non-relevant categories would range between 3 and 6, with an average of 4.5. For *pictures* it can be noticed that the average ranking of R1 and R2 is 2.1, which is close to the ideal situation but indicates that the relevant categories were misclassified in some runs and placed in higher positions. The ranking of the NR class is on average 4.2, which is significantly higher than the ranking of R1 and R2 ($\mathbf{c}^2 = 29.46$, $p<0.0001$). In case of *words* runs the ranking of the relevant categories is less accurate. The average ranking for R1 is 2.4 and of R2 is 2.8, higher than R1 but not significantly. The ranking of the non-relevant category NR is on average 4, that is significantly higher than the ranking of the relevant categories ($\mathbf{c}^2 = 27.45$, $p<0.0001$). The results suggest that, after cumulating the classification results in the end of each run, it is possible to establish a significant ranking of relevance. The ranking is more accurate for pictorial stimuli than for linguistic stimuli, which reflects the results of the single-trial binary classification.

6.4 Discussion

The results showed that relevant stimuli elicited a cognitive response that could be discriminated from the cognitive response elicited by non-relevant stimuli. In the time domain, the discrimination of the relevant stimuli was mainly reflected in the ERP late positive complex,

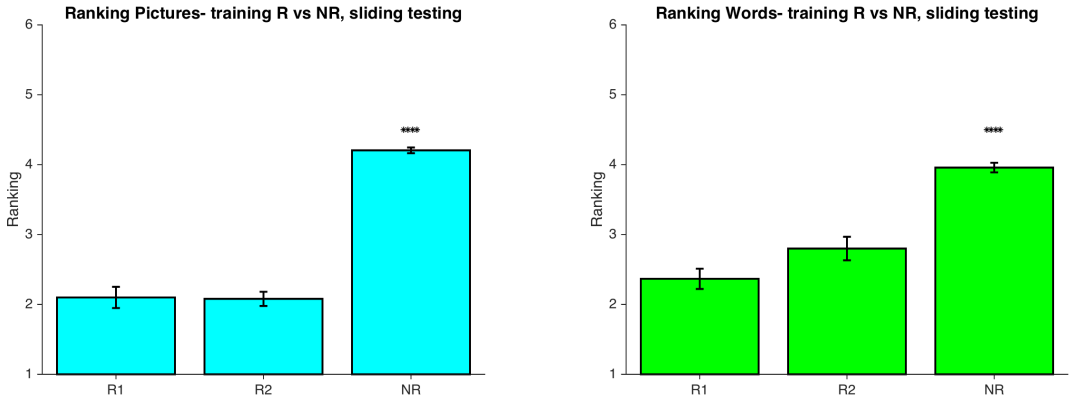


Fig. 6.11 The bar plots represent the average ranking for classes R1, R2 and NR. The ranking is based on the classification output of the sliding shrinkage LDA method. Error bars represent the SEM. Black asterisks over the bar plots indicate statistical significance.

peaking around 600-700 ms after the stimuli onset and focusing on the parietal cortex. The ERPs elicited by relevant stimuli could be further classified by multivariate machine learning methods with average performances up to 0.68 for *pictures* and 0.61 for *words*. Besides, considering the average classification in the end of each run, a significant ranking of relevance could be established. In the frequency domain, three frequency bands showed significant discrimination between relevant and non-relevant stimuli, i.e. delta, alpha and beta. While the delta band showed an increase of power related to the onset of the relevant stimuli, the alpha and beta showed a decrease of power. The different power modulation between relevant and non-relevant stimuli could also be classified with machine learning techniques with an average AUC of 0.65 for *pictures* and 0.60 for *words*. Interestingly, these classification results are higher than those of similar works based on EEG (Eugster et al., 2014). Regarding the ERP classification, the performances achieved in this study are lower than the accuracies normally achieved in ERP-BCI studies with the same feature extraction and classification methods (Chapter 5). This is not surprising for several reasons. First, the stimuli used in this study were more complex than the stimuli used, for example, in BCI spellers. The stimuli usually adopted in BCIs are designed to be well recognizable in order to make the selection of the targets (to which follows a predefined action) the easiest possible. Therefore, the stimuli often have simple features like geometric shapes and limited amount of colors. The recognition of the targets is driven both by a top-down process of attention and stimulus discrimination and by a bottom-up process of matching the features of the current stimulus with the pre-defined target. In this study, the stimuli carried richer information. The user did not know exactly which were the stimuli associated to each category, therefore the recognition was mainly based on a top-down process of relevance allocation to certain stimuli rather than others. Nevertheless, the reported classification performances looked promising considering

the challenging experimental design. Being the number of stimuli limited, it should be taken into account that some participants could recognize the occurrence of the same object in the following sequences. In this case, it cannot be excluded that for certain stimuli the ERP components would be attenuated during the development of the experiment.

The second aspect to take into account in the interpretation of the classification results relates to the task that a user had to accomplish. In BCI studies, the recognition of the target is often associated with a simple mental task that elicits well-studied ERP components like the P3. The proposed machine learning algorithms have been tailored during the recent years to achieve high performances with this kind of brain responses (Blankertz et al., 2011; Krusienski et al., 2006). In this study, the participants were instructed to detect stimuli that would be relevant to accomplish the final task (i.e. select the R1 category), but they were not performing a specific stereotyped cognitive task. In the grand average ERP plots of Fig. 6.4, the modulation of the ERP responses is quite complex for both relevant and non-relevant stimuli and the discrimination between the classes appears much later than the standard P3 response, i.e. around 600 ms. Besides, the set of stimuli was chosen to have a large diversity. For example, some stimuli could be recognized as relevant much quicker than others, and some might not have been recognized as relevant at first glance. While this design strategy makes the study closer to a real IR application than a standard BCI design, it makes the classification problem much harder due to the large trial-to-trial differences in the latency of the ERP components. Indeed the sliding shrinkage LDA classifier showed a significant improvement in relevance prediction compared to the standard shrinkage LDA, meaning that machine learning algorithms that take into account these variabilities should be further investigated. Regarding the classification of the spectral features, the main contribution was given by delta, alpha and beta bands. In this data, the theta band was excluded from the set of features for the classification. Theta modulation is considered to be involved in mental processes like attention and memory and was deeply investigated in many studies (Gomarus et al., 2006; Klimesch, 1999; Missonnier et al., 2006). Nevertheless, this dataset did not show a significant modulation of theta between classes, therefore this band was not further investigated. Rarely the modulation of the delta band is reported as a significant component in studies of relevance. Harmony (2013) defined the delta activity as an indicator of attention to internal processing during the performance of a mental task. Başar-Eroglu et al. (1992) proposed that the delta frequencies can be viewed as the frequency components of the P3 during an oddball paradigm. Alpha and beta bands also showed a significant ERD more pronounced for the relevant stimuli than for non-relevant stimuli, indicating that the frequency analysis contributed to reveal an aspect of the neural response that complements the ERP analysis. It is well reported in literature that an increase of attention and stimulus evaluation correspond to a decrease of alpha and beta power (Klimesch, 2012; Klimesch et al., 1998, 2001). Therefore, these findings suggested that both the temporal and spectral features could be used to predict relevance, meaning that future studies could employ IR

designs that exploit either domain or both. The presented design used objective labels to infer relevance, analogous to Kauppi et al. (2015). This choice made the estimator independent on the subjective evaluation and prevented the laborious task of collecting behaviorally the feedback. Besides, the classifier could be transferable to cases in which the ground-truth labels are not available, like in an online application.

6.4.1 Limitations and open questions

This chapter addressed the topic of relevance detection and evaluated it using pictorial and linguistic stimuli. The stimulus presentation followed a serial presentation, which made it suitable for the explorative investigation of the EEG components minimizing the artifacts like eye and muscular movements. However, in order to transfer the present findings to a real IR system, a free viewing case should also be investigated, like Kauppi et al. (2015) did with images and a MEG system. Classification results showed that the performances achieved in case of *words* were significantly lower than those of the *pictures*. Follow-up studies should go in depth in the investigation of the neural correlates of relevance for linguistic stimuli, since usually they are the major source of information of retrieval systems, and tailor specific algorithms to increase classification accuracies. Especially, according to these results, prediction algorithms have to take into account the trial-to-trial variably of the neural response that follows the relevance judgement. Besides, the combination of ERP and spectral features could be investigated in future analysis, since in previous works it was successful in increasing classification accuracy (Dornhege et al., 2003; Eugster et al., 2014; Li et al., 2012). In this study, classification methods approached the topic of relevance in a binary way, relevant vs non-relevant. Future studies should investigate different shades of relevance, introducing categories which have different weights according to the degree of matching with the desired topic or task.

6.5 Take-home messages

- Relevant stimuli elicited a cognitive response that could be discriminated from that one elicited by non-relevant stimuli and could be classified in both temporal and spectral domains;
- The classification method that accounted for the temporal uncertainty of the cognitive responses performed significantly better than the 'standard' classification approach;
- A significant ranking of relevance was obtained in the end of the runs;
- Classification performances were higher than previous EEG-based relevance studies, nevertheless lower than usual performances in BCI, especially for linguistic stimuli;
- Classification based on objective labels and not on subjective feedback of relevance detection (e.g. button press) prevented the knowledge of the 'ground truth' and made difficult the exploration of inter-trial variability;
- The user's freedom from specific cognitive tasks made the study closer to a realistic scenario, but also prevented the control of inter-subject variability due to potential different mental strategies;
- The presented design should be investigated with a larger pictures/words dataset and higher number of categories in order to reduce the potential effect of memorization of the stimuli;
- Being the results of the study promising with the serial presentation of stimuli, a similar paradigm should be investigated in a more realistic application that involves free viewing;
- An IR system that provides relevance feedback and stimuli with different degrees of relevance should be tested in order to extend the proposed methods to a more complex application.

Chapter 7

Testing a motor imagery-based BCI in large scale

7.1 Introduction and state of the art

The performances and usability of SMR-based BCIs have improved in the last decades: long training sessions, in which the users had to learn to control specific brain features, have been replaced by the machine-learning based approach that allows for an effective performance from the first session (Blankertz et al., 2002, 2007, 2008b). Then, further improvements have been made to compensate for the non-stationarity of the EEG signal, in particular when transitioning from calibration runs to evaluation runs. Several recent studies proposed adaptive learning for EEG signals, either based on the update of the feature space (Bamdadian et al., 2012; Samek et al., 2012; Sannelli et al., 2011; Sugiyama et al., 2007) or of the classifier (Faller et al., 2014; Shenoy et al., 2006; Vidaurre et al., 2011a,b). The bottleneck of SMR-based BCIs is that a significant percent of users (estimated between 20-50% (Guger et al., 2003; Kübler and Müller, 2007)) is not able to modulate the SMR in a volitional way (BCI inefficiency) (Blankertz et al., 2010a; Vidaurre and Blankertz, 2010). The causes of this phenomenon are unclear and vary from person to person. For example, it has been observed that for some participants the classifier cannot be successfully trained, meaning that no difference in the modulation of the SMR can be found between motor imagery tasks. In the classical case, which relies on offline calibration data, the BCI inefficiency may be due to the changes of the brain features between the offline training session and the online feedback. In this specific scenario the development of co-adaptive BCI systems has been crucial, since they let people interact with the feedback from the first trials, removing the transition from offline to online phase (Vidaurre and Blankertz, 2010; Vidaurre et al., 2011b,c).

It is important to develop advanced methods to solve the BCI inefficiency, but also is it useful to identify a priori the potential users that may have difficulties to adopt one particular approach. It would allow to skip a frustrating experience and to assign them to a different BCI system (e.g., ERP based) (Geronimo et al., 2016). Growing interest in this topic led to the investigation of neurophysiological predictors of BCI performances, based on short recordings of the brain activity in resting state before the BCI session or on the brain state

preceding the MI on single-trail (Bamdadian et al., 2014; Blankertz et al., 2010a; Burde and Blankertz, 2006; Fazli et al., 2012; Grosse-Wentrup et al., 2010; Maeder et al., 2012; Suk et al., 2014; Zhang et al., 2015a,b).

7.1.1 A fully-automatic SMR-BCI tested on large scale

These studies that address the problem of BCI inefficiency, as well as 'standard' BCI studies for other applications like described in previous chapters, are usually conducted with a relatively low number of participants. In particular, it has not been assessed yet whether an adaptive system using the state of the art machine learning technology can be successfully applied on a large scale, without the direct intervention of BCI experts on sensible parameters. This chapter tackles this point and describes a large-scale study conducted in two different locations, the Technical University of Berlin and University of Würzburg. From the BCI perspective, the two main goals of the study were: first, to achieve the successful results of the previous studies based on co-adaptive calibration in naive un-categorized users and with a fully automatic system. Second, to test the neurophysiological predictor in a new larger pool of participants. Therefore, a sample of 168 participants was recruited to conduct a co-adaptive EEG-based SMR-BCI, in single-session study. The BCI session immediately started with online feedback, applying the co-adaptive techniques described in Vidaurre and Blankertz (2010) and Vidaurre et al. (2011b). However, differently from the mentioned studies, the proposed system was fully automatized, i.e. the experimenters did not manipulate the selection of the features or the parameters of the classifier or the adaptation scheme. This approach was meant to simulate a realistic case in which an adaptive system would be in-home applied by users not familiar with the sophisticated machine learning algorithms. Nevertheless, since the participants were totally naive in respect of BCI technology, it was necessary that BCI experts provided them with the instructions about how to correctly perform motor imagery. The brain resting state activity of all participants was acquired prior to the BCI task and allowed a later offline investigation of the neurophysiological predictor of BCI performances, according to the methods described in Blankertz et al. (2010a). Moreover, two different training strategies were adopted which correspond to the psychological factors of degree of concentration and visuo-motor coordination that were found in Hammer et al. (2012) to correlate with BCI performance. In the current study, these training strategies were used as interventions prior to the BCI session in order to assess their ability to improve BCI performance (the third goal of this study). The results are discussed taking into consideration the different training groups but a more detailed investigation of the effect of the interventions goes beyond the purpose of the following analyses.

Data and results of this study were previously published in Acqualagna et al. (2016).

7.2 Methods

7.2.1 Participants

Hundred-sixty-eight (168) participants were recruited for the study. 83 experiments were conducted at the Technical University of Berlin and 85 at the University of Würzburg. Participants were all naive in respect to BCIs. Eight participants were excluded from the analysis for taking Central Nervous System (CNS)-affecting medication or reporting a psychiatric disorder, e.g. ADHD or depression. Four participants were excluded because of technical problems in the EEG setup and 5 for not complying with the instructions given. So, the final sample was reduced to 151 participants (96 females, mean age 24.9, SD=6.5). The study was conducted in accordance with the declaration of Helsinki, approved by the Ethical Review Board of the Medical Faculty of the University of Tübingen, and written informed consent was obtain prior to the experiment. Participants received 8 Euro per hour for their participation.

7.2.2 Apparatus

The electrodes configuration used in this study was: Fp1,2, AF3,4, F1-10, Fz, FC1-6, FCz, FT7,8, CFC1-6, T7,8, C1-6, Cz, CCP1-6, CP1-6, CPz, TP7,8, P1-6, Pz, PO3,4, O1,2, A2. The stimuli were shown on a 23" screen with a native resolution of 1920x1080 pixels at a refresh rate of 60Hz. Participants sat on a comfortable chair at a distance of approximately 80 cm form the display.

7.2.3 Design and procedure

Interventions and pre-measurements

Participants were randomly assigned to one of three groups, in which different kinds of intervention were performed before the BCI session. The aim of the interventions was to investigate whether different pre-training could enhance SMR-BCI performances. The first group performed a progressive muscular relaxation (PMR), following the instructions of a Jacobson Progressive Muscle relaxation audio recording of 23 minutes. The second group practiced a two-hand visuo-motor coordination task (inspired by the 2HAND test (Schuhfried GmbH), in which they had to steer a ball along virtual paths, controlling with the right hand the knob that set the horizontal position and with the left hand the knob of the vertical position (Hammer et al., 2012). In Würzburg this task was performed for 5 paths of increasing difficulties, with an average duration of 7.6 min (SD=2.6). In Berlin, the paths were cycled for a total duration of 23 min. The third group was the control group, and participants had to spend 23 min reading a text about BCI technology and answer to related questions. After the intervention, participants started the BCI session. The first measurement consisted in

recording the brain activity during resting state. A recorded voice instructed the participants to close the eyes for 15 s and open the eyes for an other 15 s. While with open eyes, they had to look at a geometrical moving shape at the middle of the display. The cycle 'eyes open-closed' was repeated for 10 times. The SMR BCI motor imagery (MI) part followed the resting state measurement. The total duration of the experiment was approximately 3.5-4 hours, including the preparation of the EEG cap (about 1 hour), intervention and EEG recording (about 1.5 hours), pauses and hair wash in loco.

Design of the BCI feedback

The BCI feedback used in the study consisted of a cursor with cross shape, which was displayed in the center of the screen for 2 seconds, followed by a white arrow-shaped cue, those direction indicated the MI to perform: right direction for MI of the right hand, left direction for MI of the left hand and bottom direction for MI of the feet. The user was instructed to perform the MI as soon as the cue was displayed and the cross turned color from black to magenta. After 1 second, the feedback was displayed for 3 seconds. In this period, the output of the classifier was translated into cursor movement in a rate control manner, i.e. every 40 ms a fraction of the classifier output was added to the current cursor position. When the 3 seconds of feedback were over, the color of the cursor turned black again and a new trial started. The final position of the cross determined the success or failure of the trial. For the first three runs, a positively biased feedback was employed. This feedback manipulation was not present in previous co-adaptive studies (Vidaurre and Blankertz, 2010; Vidaurre et al., 2011b) and it will be explained in the next section. Here, the cursor could only move from the center position towards the indicated target direction. The number of trials varied in the different parts of the experiment, but in all the runs after 20 trials there was a short break of 15 seconds.

Online motor imagery BCI

The motor imagery session consisted of 3 parts: the first part comprised 3 runs (1-3), of 40 trials each. The second and the third part had 2 runs (here referred as runs 4-5 and runs 6-7), of 80 trials each (Fig. 7.1). Between the runs there was a short pause to let the participants relax. In runs 1-3, EEG was preprocessed using three Laplacian derivations (McFarland et al., 1997) over C3, Cz and C4, calculated from four surrounding channels, equally weighted and subtracted from the central one. The signal was also frequency-filtered in the alpha (8-15 Hz) and beta (16-32 Hz) bands using two Butterworth filters of order 10. Three binary subject-independent classifiers were adopted, one for each pair of classes, *left-right*, *left-feet* and *feet-right*. They were trained on a dataset of 48 people whose performance was above 70% in a previous study (Blankertz et al., 2010a). The classifier was based on LDA (Section 3.2.1). In this phase, the LDA classifier was adapted in a supervised manner after

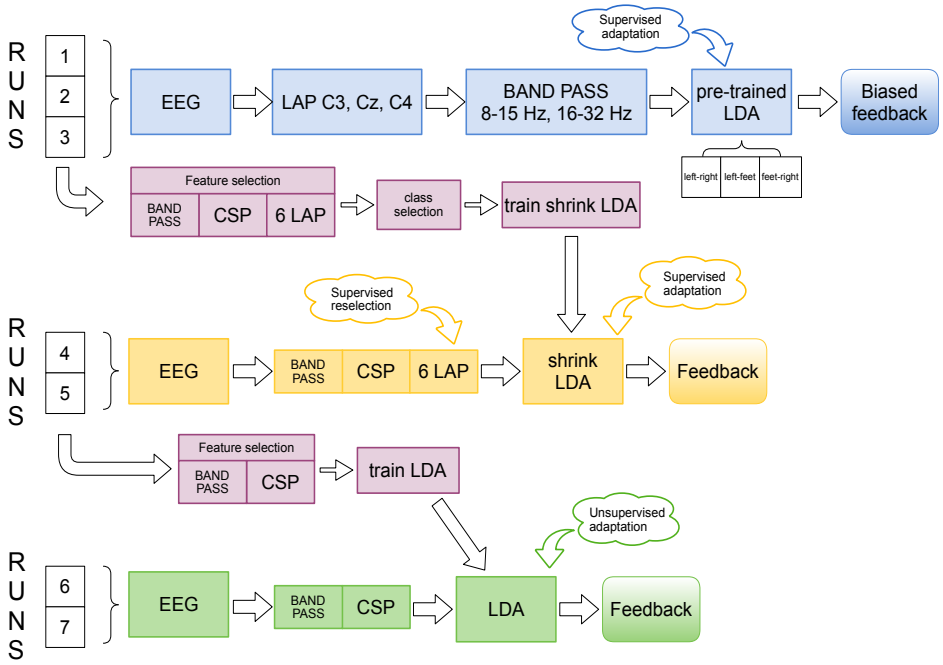


Fig. 7.1 Schematic flowchart of the online protocol. The EEG processing and adaptation protocol during runs 1-3 with positive biased feedback are depicted in blue, in yellow the processing and adaptation during runs 4-5 with real feedback, in green the processing and adaptation during runs 6-7 also with real feedback. The adaptation applied in runs 1-3 and 4-5 uses supervised methods, the adaptation of runs 6-7 uses unsupervised methods. In magenta are depicted the phases of subject-specific features selection (e.g. frequency band, CSP filters, etc) and training of the classifier that happened two times, i.e. after runs 1-3 and after runs 4-5.

every trial using the Adaptive Mean Estimation and Adaptive Inverse Covariance Matrix Estimation algorithms (see Vidaurre and Blankertz (2010); Vidaurre et al. (2011b) for a detailed description). The inverse of the covariance matrix and class mean values were updated after every trial using the class label (type of motor imagery task) only for the mean values of the past trial. Feedback was provided to the participants slightly differently from the procedure previously described. In this phase, the movement of the cursor was positively biased in the following way. The two binary classifiers including the target class were evaluated. If one of them had a positive output for the target class, the cursor was moved proportionally towards the target direction. Otherwise, the cursor would slowly move back towards the center position, but it was never moved in direction of one of the two wrong classes. The use of a positively biased feedback was motivated by the fact that in previous studies some users were on average not able to reach significant good performance (over 70% of accuracy) in the initial phase (Vidaurre and Blankertz, 2010; Vidaurre et al., 2011b). This lack of control could therefore irritate the user (Kübler et al., 2001b). Positive feedback was already found to be useful in BCI training (Barbero and Grosse-Wentrup, 2010). Data recorded in runs 1-3 were used to select subject-specific settings (according to the tutorial in Blankertz et al. (2008c)), which were used in runs 4-5. The frequency band in which the classes were better discriminated was selected. A second shrinkage LDA classifier was trained on log-band power features extracted from the Laplacian and CSP filters (CSF) (Section 3.1.4) based on 24 EEG channels. In this training, the pair of classes that showed the highest classification accuracy (using cross-validation) was chosen and used for the rest of the experiment. This classifier was applied in runs 4-5. The CSF were selected automatically for each participant using the bandpass filtered data of runs 1-3. The number of selected CSF varied between 2 and 6 and they remained fixed during runs 4-5. In order to allow some spatial adaptation, six Laplacian derivations were selected for each participant based on their discriminative power (quantified by the biserial correlation coefficient) and concatenated to the CSP channels, leading to a feature vector of dimension between 8 and 12. Two Laplacian channels were selected from the left hemisphere, 2 from the center and 2 from the right hemisphere. In runs 4-5 the adaptation was performed in supervised manner. After every trial, the selection of the six Laplacian channels was updated using the last 100 trials and the classifier recalculated. Data recorded in runs 4-5 (160 trials) was used to recalculate the frequency band and new CSP features based on 47 EEG channels, and thereafter to train the LDA. In runs 6-7 no additional Laplacian channels were used. During these last 2 runs, after each trial the linear classifier was adapted in unsupervised manner, using the adaptation of the pooled mean. The adaptation of the pooled mean produces a shift of the classifier's hyperplane, tracking the position of the mean of the features. The value of the update coefficient used in this study was 0.05, selected according to Vidaurre et al. (2008). In runs 4-5 and 6-7, the feedback was not biased and the cursor position was update according

to the classifier output as described in the previous section. Thus, the final position of the cursor reflected the success or failure of the trial.

7.2.4 Post hoc data analysis

EEG signal was lowpass filtered from 0 to 40 Hz with a Chebyshev Type II filter of order 10 (3 dB of ripple in the pass-band and 40 dB of attenuation in the stop-band) and down-sampled to 100 Hz. For ERD/ERS visualization, the class combination chosen in the first calibration after runs 1-3 was selected for each participant. The chosen pair of MI classes did not change during the following part of the experiment. Data were filtered using the specific band selected during this calibration. The grand average ERD/ERS was done over the participants having the same combination of selected classes for each adaptation scheme, i.e. runs 1-3, runs 4-5 and runs 6-7. Online classification performances referred to the results of runs 4-7, in which the actual classifier output was translated into cursor feedback. The percentage of feedback accuracy refers to the number of trials in which the final position of the cursor was in line with the target class. Mean feedback accuracies over 54.69% were considered significantly higher than chance level (Mueller-Putz et al., 2008). This threshold was calculated with the binomial inverse cumulative distribution function (cdf), which returns the minimum number of trials such that the binomial cdf is equal to or exceeds 0.95. Given that the total amount of feedback trials was 320 with 0.5 probability of success, the result of the the binomial inverse cdf was 175, i.e. 54.69% of successful trials. Two-sample t-test was run between the performances of participants in Berlin and Würzburg, in order to check if the difference between the average accuracies of the two laboratories were statistically significant.

Statistical tests run between different intervention groups and between different runs were done using one way-ANOVA with multiple comparisons to test which means were significantly different.[1]

Calculation of the neurophysiological predictor and transfer of the prediction model between datasets

The measurement of the brain in resting state was used to extract the neurophysiological predictor of BCI performances. Two Laplacian derivations, C3 and C4, were calculated from 9 monopolar channels. For each Laplacian derivation, the time intervals in which the participants had eyes opened were concatenated and the continuous data were divided into 2 s epochs. The power spectral density (PSD) was calculated for the two derivations (between 2 Hz and 34 Hz, with a step of 0.5 Hz) and smoothed with a moving average filter (window 3 Hz). The PSDs were averaged over epochs. The difference between the maximum PSD of each derivation and the fit of the 1/f noise spectrum was calculated. The average of the two

[1]The assumption of normality distribution of the data was checked with the one-sample Kolmogorov-Smirnov test. The assumption of equal variances was tested using Bartlett's test.

values determined the predictor. These values represent the estimate of the strength of the SMR rhythm over the motor areas. For a more detailed description about the modeling of the PSD and 1/f curves, please refer to Blankertz et al. (2010a). Three participants were excluded in this analysis because of the bad quality of the resting state recording. Therefore, the predictor was evaluated on a pool of 148 participants.

In order to test whether the prediction model could be transferred between different datasets, the model derived from the data of the previous study (Blankertz et al., 2010a) was used to estimate the feedback accuracies of the present study. The coefficients of the linear regression were calculated from the previous dataset, after outlier rejection. The estimated accuracies were derived by applying the linear regression model with the new predictor values extracted by the resting state EEG recording.

7.3 Results

7.3.1 Online performances

The following results refer to classification performances of runs 4-7, in which the feedback was representing the actual classifier output. Correct trials are those in which the final position of the cross was on the display at the correct side of the target class (right for right MI, left for left MI and down for feet MI) with respect to the starting point at the center of the display. The classification accuracies represent the percentage of correct accomplished trials. Fig. 7.2 depicts the feedback accuracies for each participant, sorted in ascending order. The black crosses represent the average accuracies of the overall feedback phase and the colored stars the accuracy of each run. The values of the participants tested at TU Berlin are depicted in magenta, those of the participants of Würzburg in green. A total of 135 participants had mean classification performance significantly better than chance. Among these, 90 had mean accuracy over the threshold considered necessary for efficient BCI control (i.e. 70%, Kübler et al. (2001b)), with a mean of 85.65% (SEM=0.90), and 45 participants had mean accuracy of 62.47% (SEM=0.70). The other 16 participants had mean accuracy not significantly higher than chance level, 52.19% (SEM=0.44). From the plot it is visible that many participants show variable performances between runs, especially those who have an average accuracy between 60% and 80%. Considering the whole sample of participants, the variability of average feedback accuracy between runs was not significant ($p=0.67$). However, the participants who performed on average at chance level had a significant increase of accuracy at run 6 compared to runs 4-5 ($F=6.35$, $p<0.001$), performing significantly better than chance in the last two runs. The plot also shows a different distribution between the average accuracies of the participants recruited in the two universities. Unpaired t-test run between the accuracies of the two groups showed a significant effect of the location ($p<0.05$), with participants of Berlin having higher mean accuracy than participants of Würzburg.

Therefore, the following results are presented considering not only the different intervention groups, but also the location of the study.

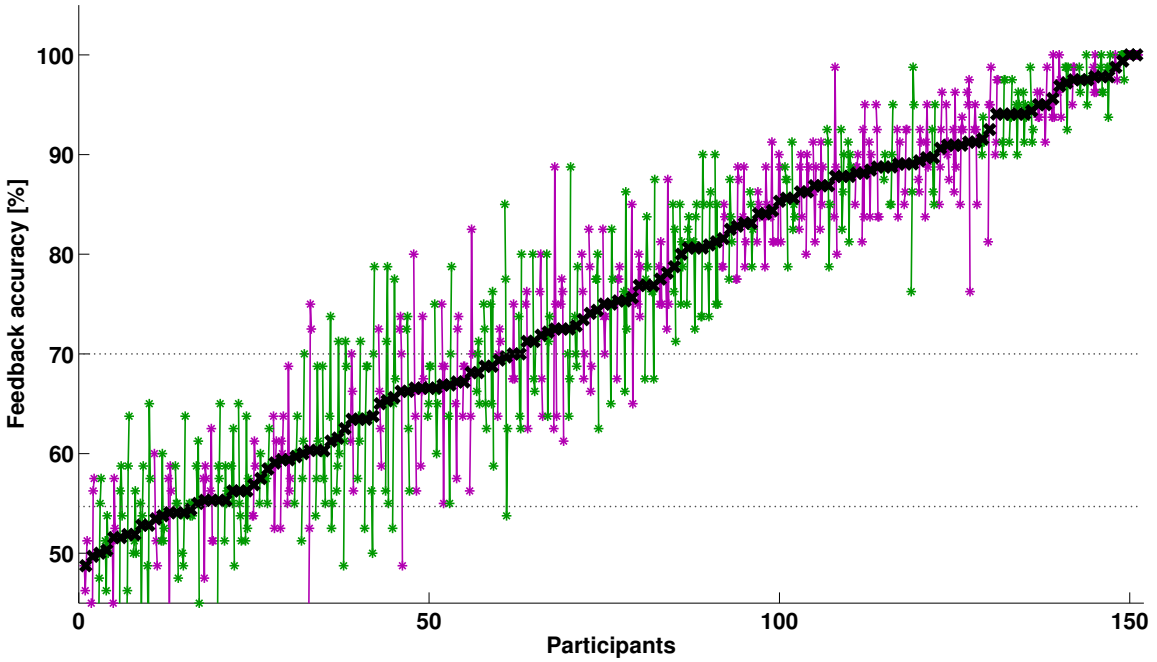


Fig. 7.2 Overview of BCI performances sorted in ascending order. For each participant, the black crosses indicate the overall mean feedback accuracy and the colored stars the accuracies of the four feedback runs (4-7). The participants of Berlin are displayed in magenta and the ones of Würzburg in green. The accuracies of each participant are connected by colored lines that emphasize the variance of the accuracies between runs. The dotted gray lines represent the accuracy considered necessary for BCI control (70%) and the threshold for accuracy significantly higher than chance (54.69%).

Fig. 7.3 (left) depicts the mean feedback accuracies for the different intervention groups. The groups are labeled according to the intervention performed (Control, PMR, 2HAND) and to the location of the experiment ('B' for Berlin and 'W' for Würzburg). Participants who performed progressive muscular relaxation show the highest average BCI performances, followed by participants who did motor coordination training and control group, in the respective locations. In Würzburg, the mean performances are significantly different ($F=3.64$, $p<0.05$), and the difference is significant for the PMR group as compared to the control group, but not for the 2HAND group. Note that the motor coordination training was performed in Würzburg for about 7.6 min (SD=2.6) versus the 23 minutes in Berlin. The control group of Berlin has an average accuracy higher than the control group in Würzburg, so even though the PMR group has higher performances, the difference is not significant ($p=0.82$). Fig. 7.3 (right) depicts in detail the trend of the classification accuracy run-wise for the corresponding groups. Note that after run 5 the classifier was re-trained and in the following runs a different adaptation scheme was adopted. In general, the groups that begin at run 4 with average

classification higher than 70% do not show a great variation of performances between runs. Both PMR groups and Control B have the same trend, with slightly higher accuracy at run 6 after the second training of the classifier, and decreasing in the last run. Group 2HAND B instead had the opposite trend, with decreasing accuracy at run 6. Control W, which had the lowest classification performances, had a substantial increase of average accuracy after run 5, reaching the threshold of 70% in the last two runs. Note that the average final performance of all groups is at the threshold of control or higher.

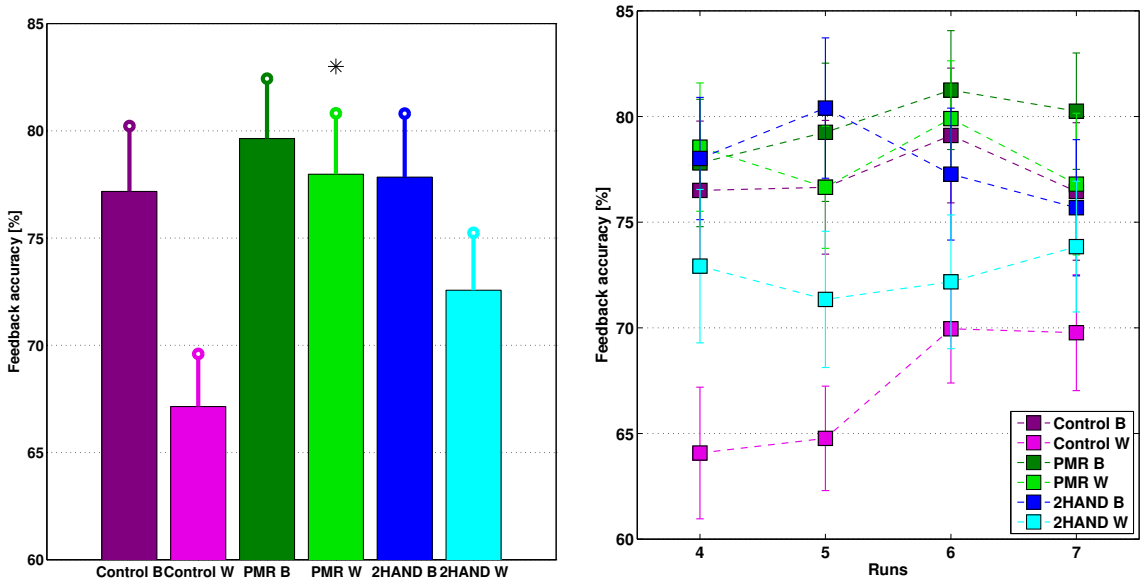


Fig. 7.3 Mean BCI feedback accuracies for each intervention group (Control, PMR, 2HAND) over runs 4-7 (left) and run-wise (right). 'B' refers to participants of Berlin and 'W' to participants of Würzburg. PMR W has significantly higher mean BCI performance compared to Control W ($p<0.05$), marked with a star in the bar plot.

7.3.2 Grand Average ERD/ERS

Fig. 7.4 shows the grand average ERD/ERS across participants whose chosen class combination was *left-right*, for the different intervention groups tested in Berlin and different types of adaptation used in the experiment (the other class combinations are in Appendix A, Figs. A.4 and A.6). In each column, in the first row is depicted the time evolution of the ERD/ERS between the onset of the cue on the display indicating the class of the motor imagery to perform (0 ms) and 6000 ms after. Feedback was displayed between 1000 ms and 4000 ms. The shaded areas divide the whole time course in three areas of interest, one short interval referring to the 500 ms before the start of the feedback and the other two during the motor imagery performed in presence of feedback. The average ERD/ERS values in these three intervals are calculated for each channel and depicted as scalp plots in the second and third rows. The second row refers to trials in which left motor imagery was done, the third to

right motor imagery. The fourth row pictures the values of the $sgn - r^2$ as a measure of discriminability between the two classes. The participants of the PMR group show more pronounced class-wise ERD as compared to the other two groups, which results in high discriminability between the two classes. The control group shows ERD topographies more localized than the PMR and visible ERS in the ispilateral hemispheres, especially during the right motor imagery. This leads to a high discrimination between classes also in this group. In both groups, the interval between 1 s and 2 s of feedback is that one in which the ERD/ERS are more pronounced. Nevertheless the modulation of the SMR in both classes is already visible 500 ms before the beginning of the feedback. The 2HAND group shows different patterns. The ERD begins mainly after the beginning of the feedback, and no clear pattern is visible in the preceding 500 ms interval. After 1 s the ERD becomes pronounced, especially in the last interval of the feedback. Even though deeper ERD is visible compared to the control group, especially in runs 1-3 and 6-7, the desynchronization appears less localized. Especially in runs 6-7, the ERD of the left motor imagery is present in both hemispheres, and in the last part of the trial the ERD pattern is similar in the two motor imageries. This leads to lower discrimination between the two classes compared to the PMR and control groups.

Fig. 7.5 shows the ERD/ERS for the different intervention groups tested in Würzburg, for class combination *left-right* (the other class combinations are in Appendix A, Figs. A.5 and A.7). The PMR group shows clear discrimination between the 2 classes, depicted in the $sgn - r^2$ scalp plots. We can also notice that in in this group the ERS arises 500 ms before the beginning of the feedback in both classes, in the ipsilateral hemisphere in runs 1-3 and in the central hemisphere in the other runs. The difference between the $sgn - r^2$ plots of the control and PMR groups is evident and reflects the significant difference between the respective classification accuracies. Besides, in the control group the class-wise discrimination is influenced by muscular artifacts 500 ms before the start of the feedback. The 2HAND group shows great ERD in all the runs of the experiment for both classes. Analogous to the 2HAND group in Berlin, the ERD are spread through a wider region than the motor cortex. This is visible especially in the scalp plots of the first interval in all the runs. In the intervals 2 s-4 s instead, the ERD is more focused and also the ERS is visible in the ispilateral motor cortex. The $sgn - r^2$ scalp plots show clear discrimination between the two classes. Comparing the control groups of Würzburg and Berlin, Control W shows smaller ERD if compared to the Control B and also the discrimination between the two classes is less pronounced. This can be an explanation for the lower classification accuracy of Control W.

7.3.3 Neurophysiological predictor of BCI performances

The average amplitude of the mu rhythm at rest measured at electrodes C3 and C4 was correlated with the mean BCI performances of each participant. Results are shown in the scatter plot in Fig. 7.6. The values of the predictor significantly correlate with the feedback

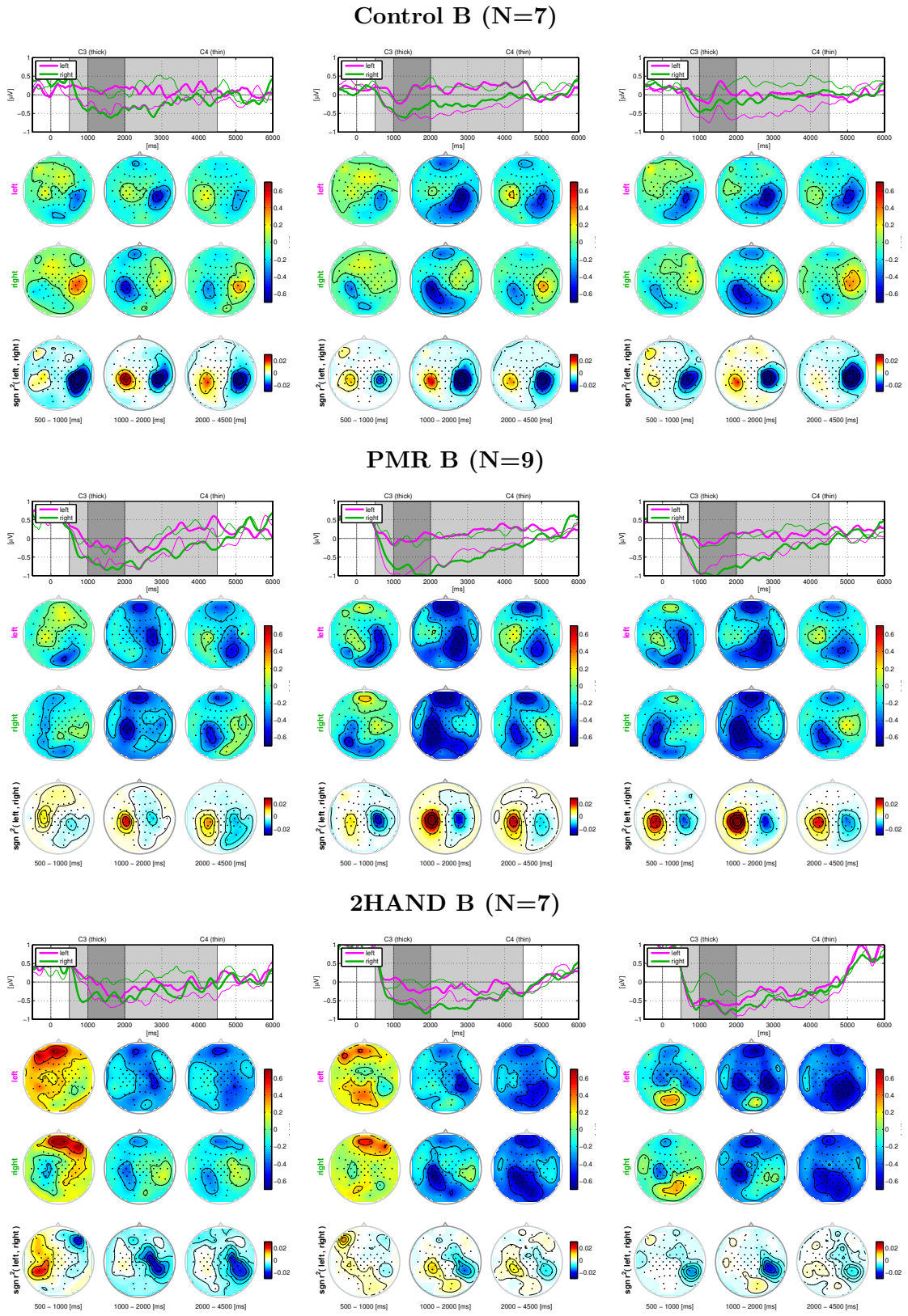


Fig. 7.4 Grand average ERD/ERS for class combination *left-right* and intervention groups in Berlin. 'N' is the number of participants of each group. From left to right: runs 1-3, runs 4-5, runs 6-7. The time plots in the first rows picture the evolution of the ERD/ERS for about 6000 ms at C3 (thick lines) and C4 (thin lines). At time 0 is the onset of the cue, at times 1000-4000 the display of the feedback. Magenta lines refer to *left* MI trials, green lines to *right* MI trials. The scalp plots underneath refer to the shaded areas of the time plots and show the distribution of the ERD/ERS. In the second rows, the scalp plots of the *left* MI trials, in the third rows the scalp plots of the *right* MI trials and in the fourth the scalp plots of the $sgn - r^2$.

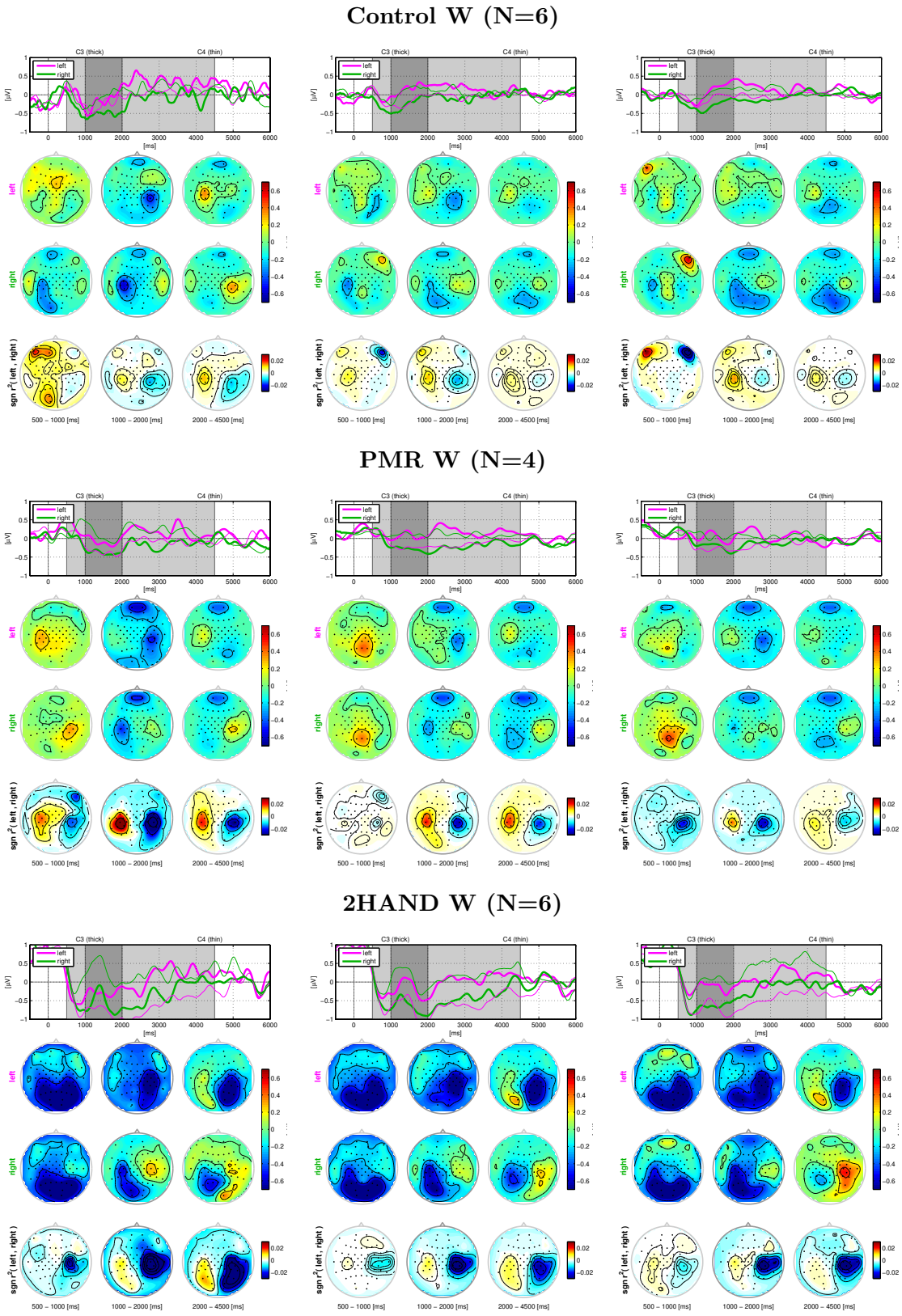


Fig. 7.5 Grand average ERD/ERS for class combination *left-right* and intervention groups in Würzburg. 'N' is the number of participants of each group. From left to right: runs 1-3, runs 4-5, runs 6-7. The time plots in the first rows picture the evolution of the ERD/ERS for about 6000 ms at C3 (thick lines) and C4 (thin lines). At time 0 is the onset of the cue, at times 1000-4000 the display of the feedback. Magenta lines refer to *left* MI trials, green lines to *right* MI trials. The scalp plots underneath refer to the shaded areas of the time plots and show the distribution of the ERD/ERS.In the second rows, the scalp plots of the *left* MI trials, in the third rows the scalp plots of the *right* MI trials and in the fourth the scalp plots of the $sgn - r^2$.

accuracies ($p<0.01$), with a Person correlation coefficient of 0.53. Participants having the 10% largest Mahalanobis distances to the data center were considered as outliers (Huber and Ronchetti, 1975) and the linear regression was re-calculated. After the outliers' removal the correlation increased to 0.66. The same method was applied separately to the data of the two laboratories, resulting in a coefficient of 0.55 for participants in Berlin (0.68 after outliers rejection) and of 0.52 for participants in Würzburg (0.60 after outliers rejection).

Classification accuracies of the present study were estimated using the prediction model derived from Blankertz et al. (2010a) and it resulted in a RMSE of 16%. Pearson linear correlation calculated between the estimated accuracies and the actual ones led to a high significant correlation of $r=0.53$ ($p<0.01$).

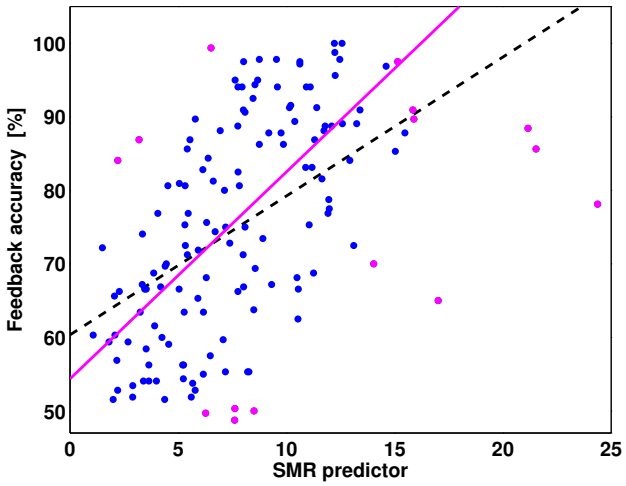


Fig. 7.6 Predictor of BCI performance. The predictor is calculated from 2.5 minutes of recording of the brain in resting state with eyes open and correlated with the mean feedback accuracy for each participant (blue dots). The back dashed line pictures the linear regression between the predictors and the accuracies ($r=0.53$). The magenta dots are the values detected as outliers. After the exclusion of the outliers, a higher correlation of 0.66 is reached (magenta line).

7.4 Discussion

The results of this large-scale study run with 151 participants operating a fully co-adaptive SMR-BCI with online feedback since the first trial showed overall mean accuracies of 78.23% in Berlin and 72.44% in Würzburg. In particular, among the 72 data sets considered in these analyses and recorded in Berlin, 50 participants reached mean BCI performances over the threshold of efficient BCI control (>70%), 17 reached mean BCI performances significantly higher than chance level and 5 had performances at chance level. Among the 79 data sets of Würzburg, 40 participants performed over 70%, 28 significantly higher than chance and 11 at

chance level. All participants were novel to BCI in general and at their first session with the system. These findings show that there is still a not-negligible portion of users who did not reach the 70% threshold of BCI performances. The results refer to the average performance of runs 4-7. Reporting the percentage of inefficient BCI users after averaging the accuracies of the whole feedback session is rather pessimistic. In fact, this measure would consider participants having poor performances at the beginning of the session and reaching the 70% at the end as suffering from inefficiency even if they learnt and reached control (e.g. participants in group Control W, Fig. 7.3 (right)). Besides, in previous studies (Vidaurre and Blankertz, 2010; Vidaurre et al., 2011b) the performance achieved either at the last level or the last run was used to calculate the rate of BCI inefficiency, since participants showed a clear learning trend. These results instead showed high intra-runs variability for many participants (Fig. 7.2) and the interpretation of the average performances is not straight-forward. Therefore, it is important to distinguish between the percentage of participants whose performance was significantly lower than the control level of 70% (threshold of significance calculated with the binomial test, i.e. 59%) and those who were close to the threshold of control but did not achieve it on average. Using this statistical threshold, the total number of users suffering from inefficiency in the beginning of the session (run 4) was 41 (27%) and decreased to 34 (22.5%) in the end of the session (run 7).

7.4.1 Comparison with previous studies

The previous large-scale study run with 80 participants (Blankertz et al., 2010a) identified some causes of failure in the BCI operation (Vidaurre et al., 2011c). Follow up studies developed promising new co-adaptive methods in which user and machine mutually learn from each other since the very first trial, in order to broaden the applicability of BCI technology. In this study we applied the co-adapative methods to a large sample of BCI novices. The reasons why a not-negligible portion of users did not reach the control level before the end of the session can only be speculated. While it is difficult to explore single-subject reasons for this phenomenon, we can formulate some general hypotheses. One reason can be due to the fact that in the previous studies the selection of the subject specific reactive frequency bands and time intervals after runs 1-3 was done in semi-automatic way, while in our study it was totally automatic. Semi-automatic means that after the automatic selection the BCI expert could visually explore the power spectrum and ERD to check that no wrong parameters were selected and, in case, force the choice of the parameters. While the fully automatic parameters selection would not be a problem for people who are naturally skilled with motor imagery BCIs, it could bring to a sub-optimal choice for those users who do not show clear motor imagery patterns. Another difference can be the frequency of subject-specific parameters update. For example, in Sannelli et al. (2010) the parameters were checked and selected after every run when necessary, while in this study it was only performed after

every co-adaptive level. If the update of parameters is not performed regularly enough, the system might lose the chance to identify a change in the parameters that leads to successful performance. Another possible reason could lie in the choice of the positive biased feedback of run 1-3. The feedback given to the user plays an important role in BCI performances. Recent studies demonstrated that visual BCI feedback clearly modulates sensorimotor EEG rhythms (Neuper et al., 2009). Barbero and Grosse-Wentrup (2010) investigated the effect of feedback design on BCI performances, biasing the belief that the participants had on their level of control of the BCI system. They concluded that people who were capable to operate a BCI might be impeded or get frustrated because of inaccurate feedback, while people who performed at chance level may actually benefit form it. In this design, the feedback had two aspects which made it positively biased in runs 1-3: (1) From two binary classifiers that one was chosen (at each time point) which had a better output for the true target. (2) The cursor was never moved from the center towards a wrong location. As a consequence of the way the feedback was biased, it could occur most of the time that the cursor moved in the correct direction just due to chance, in absence of any appropriate activation in the sensorimotor areas. We assume that the classifier was too positively biased and could happen that the person had the feeling to perform correct motor imagery even when that was not the case. Even though participants were aware that the following runs reflected the actual classifier output, the transition between the biased and real feedback could have led to confusion and frustration in some users. Incompetence fear and mastery confidence are key factors to be considered when operating a BCI system, as it was related to BCI performances (Kleih et al., 2010; Nijboer et al., 2010).

7.4.2 Notes on the effects of psychological intervention

In this data an unexpected significant effect of the study location on BCI performance was found, that is Control B performed significantly better than Control W. This is quite surprising, since all the phases of the BCI experiment and psychological interventions were identical (except the duration of 2HAND training) and automatic in the two locations and instruction given in the same way (written form and oral form by BCI experts). The ERD/ERS plots shown in Fig. 7.4 and 7.5 illustrate that the Control B group already had clear class-wise desynchronization since the first runs. Based on this observation, we speculate that the lab effect might, despite the large number of participants, be just due to chance, i.e. in Berlin a larger number of participants naturally skilled with MI were recruited in the control group. The PMR groups had higher accuracy compared to the control groups, and this difference became significant in Würzburg. This trend reflects the differences in the ERD plots, with the PMR groups having greater desynchronization and better discrimination between classes in respect to control. Regarding the groups performing the 2HAND test we did not find any significant difference in performances compared to the control groups.

However, the ERD patterns of these groups differed from the patterns of the other two groups. We could notice a deep desynchronization that appears to be less focused and more spread over the motor cortices. Since the motor coordination training implied the control of the knobs with synchronized coordination of both hands, it caused a simultaneous modulation of the SMR in both motor areas. This effect might temporary influence the following motor imagery and be the reason for the spread ERD patterns.

7.4.3 Consolidation of the neurophysiological predictor

One goal of this study was to replicate the correlation between the neurophysiological indicator developed by Blankertz et al. (2010a) and BCI performances on a new large pool of participants. We found a significant positive correlation of 0.53 ($p<0.01$), which rose to 0.66 after outliers rejection, confirming the results of the previous study. Moreover, results show that the proposed model can be transferred between two SMR-BCI studies that employ different designs. Indeed, the BCI performances of new participants could be predicted with high significant correlation ($p<0.01$) by the prediction model derived from the previous dataset (Blankertz et al., 2010a) and with a RMSE of 16%. Rarely the replication of a prediction model is reported in BCI literature. An example is described in (Hammer et al., 2014), in which the authors aimed at replicating the two psychological predictors of SMR-BCI performances previously found in Hammer et al. (2012), in a different BCI setup. They found that the psychological variable 'visuo-motor coordination ability' explained a moderate amount of the variance of the SMR feedback performance and could be consolidated as a small predictor of BCI performances. Besides, they applied the regression model of the previous study to the new dataset, predicting performances with an average error of 12.07% and achieving significant correlation with the real performances.

The replication of the neurophysiological predictor is a major achievement of the study, since the participants tested were different and also the BCI procedure changed with the introduction of the online feedback since the first trial. These results consolidate the validity of the developed prediction model across different participants and experimental procedures.

7.4.4 Limitations of the study and open questions

The findings showed that a relatively large number of users were close to 70% feedback accuracy without reaching it on average over the session. This trend made it difficult to estimate the rates of BCI efficiency. Using a statistical threshold level, the rate of people who could not learn to control the system decreased from 27% to 22.5% in the end of the session. However, this level is arguable low (59%) and the average performance of the session had also to be reported. It is important to note that the comparison with previous studies (Vidaurre and Blankertz, 2010; Vidaurre et al., 2011b), in which only the final performance was reported, is then not straight-forward. Future developments should pay closer attention

to the aspects that showed the weakness of this study. For example, potential users not able to operate successfully the BCI after some runs should get closer assistance from a BCI expert in the calibration of the system and in advising about the motor imagery strategies. Moreover, the positive biased feedback should be revised. For example, we could only bias the feedback not letting the cursor moving towards the wrong direction, which avoids demotivation, while avoiding to choose the best output for the true class, as it more often falsely suggests good performance. (Assuming two random binary classifiers, at least one of them will give the correct output in 75% of the cases on average). The identification of the users who would potentially need more assistance could be assessed a-priori by the calculation of the neurophysiological predictor. For these participants alternative approaches could be explored. One option could be the use of another type of BCI, e.g. based on ERPs (Geronimo et al., 2016). An alternative can be also the integration of multiple features with features extracted form the ERD/ERS effect, which was shown to be beneficial for SMR-BCI (Fazli et al., 2015). For example, the integration can involve multiple spatial locations, multiple frequency bands and different physiological processes (e.g. Slow Cortical Potentials, Error Potentials) (Dornhege et al., 2004; Höhne et al., 2014b; Sannelli et al., 2011; Schalk et al., 2000). These kinds of BCI belong to the class of systems called *hybrid BCI* (Pfurtscheller et al., 2010). Please refer to Fazli et al. (2015) for a recent review about data fusion techniques for SMR-BCI.

7.5 Take-home messages

- A fully automatic co-adaptive SMR-BCI that used online feedback since the first trial was investigated for the first time on a large scale and was successfully operated by the majority of the healthy novice users;
- The neurophysiological predictor of BCI performances developed in the previous study, and based on just few minutes of EEG recording at rest, was successfully replicated. Moreover, it was shown that the model could be transferred between different datasets;
- Even though the number of participants who were not able to control the BCI system decreased from the beginning to the end of the session, better performances could be achieved given previous results on co-adaptation;
- The employment of a positively biased feedback of the first runs should be revised, in order to avoid confusion in the transition between the initial phases and the following runs;
- In cases of 'BCI inefficiency', the BCI expert should check that correct parameters were selected during the training of the classifier and, if necessary, manually tune them after the inspection of ERD/EDS and spectra plots;
- The update of the parameters could be done more frequently, since in previous studies it was shown to be beneficial;
- The present co-adaptation framework should be validated with patients, whose neural responses show even higher inter-subject variability and who usually have more difficulties in the BCI operation.

Chapter 8

Summary and conclusions

This thesis approached the topic of BCI from several points of view. First of all, BCI technology was broadened to new fields that can benefit from brain-machine interaction as complementary tool to the standard behavioral output. Second, in case of established applications like BCI for spelling, this work contributed in demonstrating the feasibility of novel BCI designs for limit cases, like the potential applicability to patients with impaired oculomotor control. Finally, the transfer of the BCI on a large scale was investigated, pointing out limitations and further improvements that this technology still necessitates before being operated outside of the lab.

The first part of the thesis focused on novel applications and interface designs. In Chapter 4, a BCI for video quality assessment based on SSVEPs was described. The flickering effect in the stimulus presentation was caused by changing the levels of degradation of texture images with a video codec and alternating them with the original undistorted textures. The main brain responses elicited by the flicker at 3 Hz were visual evoked potentials focused on the occipital cortex. SSVEPs could be successfully classified in the temporal and spectral domain, with performances that increased linearly with the the distortion levels and significantly correlated with the behavioral responses. The employment of a SSVEP-based paradigm let the collection of features useful for classification in a much faster way compared to previous ERP studies. Results also pointed out that even though the brain modulation was the direct reflection of the stimuli processing and no cognitive responses were elicited, the user was nevertheless expected to overtly allocate attention to the presentation. An excessive state of drowsiness or relaxation could lead the visual cortex in a idle state, diminishing the size of the SSVEPs and consequently affecting classification performances. Chapter 5 described novel designs for gaze-independent BCI spellers, i.e. to account for extreme cases in which potential end-users are unable to overtly shift gaze to the location of the stimuli. Discriminative visual and cognitive ERPs were elicited by the stimuli and the mental task accomplished by the users, who were actively engaged in the operation of the BCI with the screening and selection of the target symbol. Both the RSVP and Chroma Speller could be operated with high accuracy, and spelling rates were competitive with other visual gaze-independent spellers

in literature. Performances with the RSVP Speller could be further improved (in offline simulation) with a new classification approach that accounted for dependencies between overlapping epochs. The standard BCI design like that of the RSVP Speller was then modified to suit a specific HCI application, i.e. for the investigation of relevance in the context of Information Retrieval. This kind of novel application opened new challenges not only from the design's point of view, but also from the data analysis. In fact, stimuli carried a richer variety of information and the modulation of the brain response was driven by top-down processes of relevance judgment. Here, specific features of the stimuli played a minor role. Indeed, the main discrimination between relevant and non-relevant stimuli was based on the modulation of the cognitive LPC, while the early visual potentials were not influenced. The complex task and stimuli evaluation arose the problem of uncertain time location of the ERP components in respect to the stimulus onset. Therefore, a shrinkage LDA with sliding windows was evaluated together with the standard classification approach, leading to a significant increase in classification performances. However, especially for linguistic stimuli, average performances were much lower than the standard accuracies achieved in classical BCI applications.

The second part of this thesis tackles a major challenge of BCI research, which is the exploitation of the technology for everybody, coping with the BCI inefficiency problem and the use of an automatic system on a large scale. Sophisticated adaptive algorithms were employed to let the user operate the BCI since the first trial and account for non-stationarity of the signals between runs. Besides, the neurophysiological predictor developed in a previous study was replicated on this new dataset. The large-scale motor imagery study of Chapter 7 shows that, despite the brain-machine co-adaptation, a not-negligible portion of users was not able to operate the BCI with high accuracy before the end of the session. Results points out the major limitations of extending the well-know SMR paradigm to a wide audience of naive users. Apart from the minor technical choices in the online feedback that should be revised, the main point is that a fully-automatic BCI system can be successful with SMR-skilled users, but still requires assistance by a BCI expert in case of BCI inefficiency. This aspect would be amplified in case of in-home use with patients in impaired physical conditions, who normally have more difficulties in the BCI operation (Geronimo et al., 2016; Höhne et al., 2014a; Kleih et al., 2011; Kübler et al., 2015; Oken et al., 2014).

To conclude, this thesis showed that BCI can be considered as a powerful technology for tailoring computer-based systems to the user's perception, for enhancing man-machine interaction through the direct investigation of user's brain activity and for potentially giving the chance of communication to severely impaired users. Even though further research is still necessary to broaden the applicability of the existing BCI paradigms on large scale, the advances in machine learning and signal processing already made this prospective more realistic for a near future.

References

L. Acqualagna and B. Blankertz. A gaze independent speller based on rapid serial visual presentation. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2011, pages 4560–4563, 2011. doi: 10.1109/IEMBS.2011.6091129.

L. Acqualagna and B. Blankertz. Gaze-independent BCI-spelling using rapid visual serial presentation (RSVP). *Clin Neurophysiol*, 124(5):901–908, May 2013. URL http://dx.doi.org/10.1016/j.clinph.2012.12.050.

L. Acqualagna and B. Blankertz. Classification of subjective relevance through a brain-computer interface. 2016. submitted.

L. Acqualagna, M. S. Treder, M. Schreuder, and B. Blankertz. A novel brain-computer interface based on the rapid serial visual presentation paradigm. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2010, pages 2686–2689, 2010. URL http://dx.doi.org/10.1109/IEMBS.2010.5626548.

L. Acqualagna, M. S. Treder, and B. Blankertz. Chroma speller: Isotropic visual stimuli for truly gaze-independent spelling. In *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, pages 1041–1044, 2013. doi: 10.1109/NER.2013.6696115. URL http://dx.doi.org/10.1109/NER.2013.6696115.

L. Acqualagna, S. Bosse, A. K. Porbadnigk, G. Curio, K.-R. Müller, T. Wiegand, and B. Blankertz. Eeg-based classification of video quality perception using steady state visual evoked potentials (ssveps). *Journal of neural engineering*, 12(2):026012, 2015.

L. Acqualagna, L. Botrel, C. Vidaurre, A. Kübler, and B. Blankertz. Large-scale assessment of a fully automatic co-adaptive motor imagery-based brain computer interface. *PloS one*, 11(2):e0148886, 2016.

B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser. Bci demographics: How many (and what kinds of) people can use an ssvep bci? *Neural Syst Rehab Eng, IEEE Trans*, 18(2):107–116, 2010.

B. Z. Allison and J. A. Pineda. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: implications for a BCI system. *Int J Psychophysiol*, 59(2):127–140, Feb 2006. doi: 10.1016/j.ijpsycho.2005.02.007. URL http://dx.doi.org/10.1016/j.ijpsycho.2005.02.007.

B. Z. Allison, D. J. McFarland, G. Schalk, S. D. Zheng, M. M. Jackson, and J. R. Wolpaw. Towards an independent brain–computer interface using steady state visual evoked potentials. *Clin Neurophysiol*, 119(2):399–408, 2008.

B. Z. Allison, S. Dunne, R. Leeb, J. D. R. Millán, and A. Nijholt. *Towards practical brain-computer interfaces: bridging the gap from research to real-world applications*. Springer Science & Business Media, 2012.

K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. Filter bank common spatial pattern (fbcsp) in brain-computer interface. In *Neural Networks, IEEE Int J Conf*, pages 2390–2397, June 2008. doi: 10.1109/IJCNN.2008.4634130.

J.-N. Antons, R. Schleicher, S. Arndt, S. Moller, A. Porbadnigk, and G. Curio. Analyzing speech quality perception using electroencephalography. *J Sel Top Signal Processing, IEEE*, 6(6):721–731, Oct 2012. ISSN 1932-4553. doi: 10.1109/JSTSP.2012.2191936.

P. Aricò, F. Aloise, F. Schettini, A. Riccio, S. Salinari, F. Babiloni, D. Mattia, and F. Cincotti. GeoSpell: an alternative P300 based speller interface towards no eye gaze required. In *Proceedings on the TOBI Workshop II*, 2010.

S. Arndt, J.-N. Antons, R. Schleicher, S. Moller, and G. Curio. Using electroencephalography to measure perceived video quality. *J Sel Top Signal Processing, IEEE*, 8(3):366–376, June 2014. ISSN 1932-4553. doi: 10.1109/JSTSP.2014.2313026.

C. Babiloni, F. Vecchio, A. Bultrini, G. L. Romani, and P. M. Rossini. Pre-and poststimulus alpha rhythms are related to conscious visual perception: a high-resolution EEG study. *Cereb Cortex*, 16(12):1690–1700, 2006.

T. Ball, M. Kern, I. Mutschler, A. Aertsen, and A. Schulze-Bonhage. Signal quality of simultaneously recorded invasive and non-invasive eeg. *Neuroimage*, 46(3):708–716, 2009.

A. Bamdadian, C. Guan, K. K. Ang, and J. Xu. Online semi-supervised learning with kl distance weighting for motor imagery-based bci. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 2732–2735, Aug 2012. doi: 10.1109/EMBC.2012.6346529.

A. Bamdadian, C. Guan, K. K. Ang, and J. Xu. The predictive role of pre-cue {EEG} rhythms on mi-based {BCI} classification performance. *Journal of Neuroscience Methods*, 235(0): 138 – 144, 2014. ISSN 0165-0270. doi: http://dx.doi.org/10.1016/j.jneumeth.2014.06.011. URL http://www.sciencedirect.com/science/article/pii/S0165027014002179.

A. Barbero and M. Grosse-Wentrup. Biased feedback in brain-computer interfaces. *J Neuroeng Rehabil.*, 7(34), 2010. doi: 10.1186/1743-0003-7-34.

O. Barral, M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, N. Ravaja, S. Kaski, and G. Jacucci. Exploring peripheral physiology as a predictor of perceived relevance in information retrieval. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 389–399. ACM, 2015.

E. Başar, M. Schurmann, C. Başar-Eroglu, and S. Karakaş. Alpha oscillations in brain functioning: an integrative theory. *International Journal of Psychophysiology*, 26(1-3):5–29, 1997.

E. Başar, C. Başar-Eroglu, S. Karakaş, and M. Schürmann. Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, 39(2): 241–248, 2001.

C. Başar-Eroglu, E. Başar, T. Demiralp, and M. Schürmann. P300-response: possible psychophysiological correlates in delta and theta frequency channels. a review. *International Journal of Psychophysiology*, 13(2):161–179, 1992.

H. Berger. Über das elektrenkephalogramm des menschen. *European Archives of Psychiatry and Clinical Neuroscience*, 87(1):527–570, 1929.

C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau, G. Davis, V. T. Zivkovic, R. E. Olmstead, P. D. Tremoulet, and P. L. Craven. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78 (Supplement 1):B231–B244, 2007.

N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig. Brain activity-based image classification from rapid serial visual presentation. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16(5):432–441, 2008.

G. Bin, X. Gao, Y. Wang, Y. Li, B. Hong, and S. Gao. A high-speed BCI based on code modulation VEP. *J Neural Eng*, 8:025015, Apr 2011. URL http://dx.doi.org/10.1088/1741-2560/8/2/025015.

N. Birbaumer and L. G. Cohen. Brain–computer interfaces: communication and restoration of movement in paralysis. *The Journal of physiology*, 579(3):621–636, 2007.

N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398:297–298, 1999.

N. Birbaumer, A. R. Murguialday, and L. Cohen. Brain–computer interface in paralysis. *Current opinion in neurology*, 21(6):634–638, 2008a.

N. Birbaumer, A. Murquialday, and L. Cohen. Brain-computer interface in paralysis. *Curr Opin Neurol*, 21:634–8, 2008b.

C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

B. Blankertz, G. Curio, and K.-R. Müller. Classifying single trial EEG: Towards brain computer interfacing. In T. G. Diettrich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Inf. Proc. Systems (NIPS 01)*, volume 14, pages 157–164, 2002.

B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, G. Curio, and K.-R. Müller. The Berlin Brain-Computer Interface: Machine learning based detection of user specific brain states. *J Universal Computer Sci*, 12(6):581–607, 2006.

B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio. The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. *Neuroimage*, 37(2):539–550, 2007. URL http://dx.doi.org/10.1016/j.neuroimage.2007.01.051.

B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K.-R. Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 113–120. MIT Press, Cambridge, MA, 2008a.

B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Müller. The Berlin Brain-Computer Interface: Accurate performance from first-session in BCI-naive subjects. *IEEE Trans Biomed Eng*, 55(10):2452–2462, 2008b. URL http://dx.doi.org/10.1109/TBME.2008.923152.

B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process Mag*, 25(1):41–56, Jan 2008c.

B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus. Neurophysiological predictor of SMR-based BCI performance. *Neuroimage*, 51(4):1303–1309, 2010a.

B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio, and K.-R. Müller. The Berlin Brain-Computer Interface: Non-medical uses of BCI technology. *Frontiers in Neuroscience*, 4:198, 2010b.

B. Blankertz, S. Lemm, M. S. Treder, S. Haufe, and K.-R. Müller. Single-trial analysis and classification of ERP components – a tutorial. *Neuroimage*, 56(2):814–825, 2011.

P. Bocquillon, J.-L. Bourriez, E. Palmero-Soler, B. Molaee-Ardekani, P. Derambure, and K. Dujardin. The spatiotemporal dynamics of early attention processes: a high-resolution electroencephalographic study of n2 subcomponent sources. *Neuroscience*, 271:9–22, 2014.

G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44:58–75, 2014.

P. Borlund. The concept of relevance in ir. *Journal of the American Society for information Science and Technology*, 54(10):913–925, 2003.

S. Bosse, L. Acqualagna, A. K. Porbadnigk, B. Blankertz, G. Curio, K.-R. Muller, and T. Wiegand. Neurally informed assessment of perceived natural texture image quality. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1987–1991. IEEE, 2014.

M. E. Brandt and B. H. Jansen. The relationship between prestimulus alpha amplitude and visual evoked potential amplitude. *Int J Neurosci*, 61(3-4):261–268, 1991.

A.-M. Brouwer and J. B. F. V. Erp. A tactile p300 brain-computer interface. *Frontiers in Neuroscience*, 4(0), 2010. ISSN 1662-453X. doi: 10.3389/fnins.2010.00019.

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

P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk. Does the "P300" speller depend on eye gaze? *J Neural Eng*, 7:056013, 2010. doi: 10.1088/1741-2560/7/5/056013.

W. Burde and B. Blankertz. Is the locus of control of reinforcement a predictor of brain-computer interface performance? In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course, Verlag der Technische Universität Graz, Graz*, pages 76–77, 2006.

N. A. Busch and R. VanRullen. Spontaneous eeg oscillations reveal periodic sampling of visual attention. *Proc Nat Acad Sci*, 107(37):16048–16053, 2010.

N. A. Busch, J. Dubois, and R. VanRullen. The phase of ongoing eeg oscillations predicts visual perception. *J Neurosci*, 29(24):7869–7876, 2009.

G. Buzsáki. *Rhythms of the Brain*. Oxford University Press, 2006.

G. Buzsáki and A. Draguhn. Neuronal oscillations in cortical networks. *science*, 304(5679): 1926–1929, 2004.

G. E. Chatrian, M. C. Petersen, and J. A. Lazarte. The blocking of the rolandic wicket rhythm and some central changes related to movement. *Electroencephalography and clinical neurophysiology*, 11(3):497–510, 1959.

M. X. Cohen. *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press, 2014.

T. F. Collura. History and evolution of electroencephalographic instruments and techniques. *Journal of clinical neurophysiology*, 10(4):476–504, 1993.

E. Courchesne, S. A. Hillyard, and R. Galambos. Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and clinical neurophysiology*, 39(2): 131–143, 1975.

S. M. Coyle, T. E. Ward, and C. M. Markham. Brain–computer interface using a simplified functional near-infrared spectroscopy system. *Journal of neural engineering*, 4(3):219, 2007.

E. A. Curran and M. J. Stokes. Learning to control brain activity: a review of the production and control of eeg components for driving brain–computer interface (bci) systems. *Brain and cognition*, 51(3):326–336, 2003.

I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *Information Theory, IEEE Transactions on*, 36(5):961–1005, 1990.

J. del R. Millán, R. Rupp, G. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, R. Leeb, C. Neuper, K.-R. Müller, and D. Mattia. Combining brain-computer interfaces and assistive technologies: State-of-the-art and challenges. *Frontiers in Neuroprosthetics*, 4, 2010. doi: 10.3389/fnins.2010.00161.

A. Delorme and S. 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.

T. Demiralp, Z. Bayraktaroglu, D. Lenz, S. Junge, N. A. Busch, B. Maess, M. Ergen, and C. S. Herrmann. Gamma amplitudes are coupled to theta phase in human eeg during visual perception. *Int J Psychophysiol*, 64(1):24–30, 2007.

J. Dien, K. M. Spencer, and E. Donchin. Parsing the late positive complex: mental chronometry and the erp components that inhabit the neighborhood of the p300. *Psychophysiology*, 41(5):665–678, 2004.

E. Donchin, K. M. Spencer, and R. Wijesinghe. The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. *IEEE Trans Rehabil Eng*, 8(2):174–179, June 2000a.

E. Donchin, K. M. Spencer, and R. Wijesinghe. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *Rehabilitation Engineering, IEEE Transactions on*, 8(2):174–179, 2000b.

G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller. Combining features for BCI. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Inf. Proc. Systems (NIPS 02)*, volume 15, pages 1115–1122, 2003.

G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Trans Biomed Eng*, 51(6):993–1002, June 2004. URL http://dx.doi.org/10.1109/TBME.2004.827088.

G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *Biom Eng, IEEE Trans*, 53(11):2274–2281, Nov 2006. ISSN 0018-9294. doi: 10.1109/TBME.2006.883649.

G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, editors. *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA, 2007.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2012.

C. C. Duncan-Johnson and E. Donchin. On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, 14(5):456–467, 1977.

A. K. Engel, P. Fries, and W. Singer. Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, 2(10):704–716, 2001.

T. Ergenoglu, T. Demiralp, Z. Bayraktaroglu, M. Ergen, H. Beydagi, and Y. Uresin. Alpha rhythm of the eeg modulates visual detection performance in humans. *Cogn Brain Res*, 20 (3):376–383, 2004.

M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski. Predicting term-relevance from brain signals. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 425–434. ACM, 2014.

M. Falkenstein, J. Hohnsbein, and J. Hoormann. Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(2):148–160, 1994.

J. Faller, R. Scherer, U. Costa, E. Opisso, J. Medina, and G. R. Müller-Putz. A co-adaptive brain-computer interface for end users with severe motor impairment. *PLoS ONE*, 9(7): e101168, 07 2014. doi: 10.1371/journal.pone.0101168. URL http://dx.doi.org/10.1371%2Fjournal.pone.0101168.

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

T. Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

S. Fazli, J. Mehnert, J. Steinbrink, and B. Blankertz. Using NIRS as a predictor for EEG-based BCI performance. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2012, pages 4911–4914, 2012.

S. Fazli, S. Dahne, W. Samek, F. Bieszmann, and K.-R. Muller. Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based brain–computer interfaces. *Proceedings of the IEEE*, 103(6):891–906, 2015.

J. H. Fecteau and D. P. Munoz. Salience, relevance, and firing: a priority map for target selection. *Trends in cognitive sciences*, 10(8):382–390, 2006.

J. R. Folstein and C. Van Petten. Influence of cognitive control and mismatch on the n2 component of the erp: a review. *Psychophysiology*, 45(1):152–170, 2008.

J. H. Friedman. Regularized discriminant analysis. *Journal of the American statistical association*, 84(405):165–175, 1989.

P. Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9(10):474–480, 2005.

K. Fukunaga. *Introduction to statistical pattern recognition*. Academic press, 2013.

K. Fukunaga and W. L. Koontz. Application of the karhunen-lo? ve expansion to feature selection and ordering. *IEEE Transactions on Computers*, C–19(4):311–318, 1970.

H. Gastaut, R. Naquet, and Y. Gastaut. A study of mu rhythm in subjects lacking one or more limbs. *Electroencephalography and Clinical Neurophysiology*, 18(7):720, 1965.

A. Geronimo, Z. Simmons, and S. Schiff. Performance predictors of brain–computer interfaces in patients with amyotrophic lateral sclerosis. *Journal of Neural Engineering*, 13(2):026002, 2016.

A. D. Gerson, L. C. Parra, and P. Sajda. Cortical origins of response time variability during rapid discrimination of visual objects. *NeuroImage*, 28(2):342 – 353, 2005. ISSN 1053-8119. doi: DOI:10.1016/j.neuroimage.2005.06.026. URL http://www.sciencedirect.com/science/article/pii/S1053811905004180.

H. K. Gomarus, M. Althaus, A. A. Wijers, and R. B. Minderaa. The effects of memory load and stimulus relevance on the eeg during a visual selective memory search task: an erp and erd/ers study. *Clinical Neurophysiology*, 117(4):871–884, 2006.

C. Gonsalvez and J. Polich. P300 amplitude is determined by target-to-target interval. *Psychophysiology*, 39(03):388–396, 2002. doi: 10.1017.S0048577201393137. URL http://dx.doi.org/10.1017/S0048577201393137.

C. J. Gonsalvez, R. J. Barry, J. a. Rushby, and J. Polich. Target-to-target interval, intensity, and P300 from an auditory single-stimulus task. *Psychophysiology*, 44(2):245–50, Mar. 2007. ISSN 0048-5772. doi: 10.1111/j.1469-8986.2007.00495.x. URL http://www.ncbi.nlm.nih.gov/pubmed/17343708.

J. Groß, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, and R. Salmelin. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, 98(2):694–699, 2001.

M. Grosse-Wentrup, B. Scholkopf, and J. Hill. Causal influence of gamma oscillations on the sensorimotor rhythm. *Neuroimage*, May 2010. doi: http://dx.doi.org/10.1016/j.neuroimage.2010.04.265.

C. Guan, M. Thulasidas, and J. Wu. High performance p300 speller for brain-computer interface. *Biomedical Circuits and Systems, 2004 IEEE International Workshop, on*, Dec 2004.

C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller. How many people are able to operate an EEG-based Brain-Computer Interface (BCI)? *IEEE Trans Neural Syst Rehabil Eng*, 11(2):145–147, 2003.

C. Guger, G. Krausz, B. Z. Allison, and G. Edlinger. Comparison of dry and gel based electrodes for p300 brain–computer interfaces. *Frontiers in neuroscience*, 6, 2012.

A. Hahne. What's different in second-language processing? evidence from event-related brain potentials. *Journal of psycholinguistic research*, 30(3):251–266, 2001.

S. Halder, A. Furdea, B. Varkuti, R. Sitaram, M. Bogdan, W. Rosenstiel, N. Birbaumer, and A. Kübler. Auditory standard oddball and visual P300 brain-computer interface performance. *International Journal of Bioelectromagnetism*, 13(1):5–6, 2011a.

S. Halder, M. Spühler, E. Hammer, S. Kleih, M. Bogdan, W. Rosenstiel, A. Kübler, and N. Birbaumer. Prediction of visual P300 BCI aptitude using spectral features. *Proceedings of the 5th International Brain-Computer Interface Workshop and Training Course*, pages 144–145, 2011b.

E. M. Hammer, S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, K.-R. Müller, and A. Kübler. Psychological predictors of SMR-BCI performance. *Biol Psychol*, 89(1): 80–86, 2012.

E. M. Hammer, T. Kaufmann, S. C. Kleih, B. Blankertz, and A. Kübler. Visuo-motor coordination ability predicts performance with brain-computer interfaces controlled by modulation of sensorimotor rhythms (SMR). *Frontiers in human neuroscience*, 8:574, 2014.

J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

S. Hanslmayr, W. Klimesch, P. Sauseng, W. Gruber, M. Doppelmayr, R. Freunberger, and T. Pecherstorfer. Visual discrimination performance is related to decreased alpha amplitude but increased phase locking. *Neurosci Lett*, 375(1):64–68, 2005.

S. Hanslmayr, A. Aslan, T. Staudigl, W. Klimesch, C. S. Herrmann, and K.-H. Bäuml. Prestimulus oscillations predict visual perception performance between and within subjects. *Neuroimage*, 37(4):1465–1473, 2007.

T. Harmony. The functional significance of delta oscillations in cognitive processing. *Frontiers in integrative neuroscience*, 7, 2013.

S. Haufe, S. Dähne, and V. V. Nikulin. Dimensionality reduction for the analysis of brain oscillations. *Neuroimage*, 2014a.

S. Haufe, J.-W. Kim, I.-H. Kim, A. Sonnleitner, M. Schrauf, G. Curio, and B. Blankertz. Electrophysiology-based detection of emergency braking intention in real-world driving. *Journal of neural engineering*, 11(5):056011, 2014b.

S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87:96–110, 2014c. URL http://dx.doi.org/10.1016/j.neuroimage.2013.10.067.

H. Hayashi, H. Shirai, M. Kameda, S. Kunifuji, and M. Miyahara. Assessment of extra high quality images using both eeg and assessment words on high order sensations. *Syst, Man, and Cybern, 2000 IEEE Int Conf*, 2:1289–1294 vol.2, 2000. ISSN 1062-922X. doi: 10.1109/ICSMC.2000.886031.

C. S. Herrmann. Human eeg responses to 1–100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp Brain Res*, 137(3-4): 346–353, 2001.

S. A. Hillyard, E. K. Vogel, and S. J. Luck. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 353(1373):1257–1270, 1998.

J. Höhne, M. Schreuder, B. Blankertz, and M. Tangermann. A novel 9-class auditory ERP paradigm driving a predictive text entry system. *Front Neuroscience*, 5:99, 2011. ISSN 1662-453X. doi: 10.3389/fnins.2011.00099. URL http://www.frontiersin.org/Journal/Abstract.aspx?s=763&name=neuroprosthetics&ART_DOI=10.3389/fnins.2011.00099.

J. Höhne, E. Holz, P. Staiger-Sälzer, K.-R. Müller, A. Kübler, and M. Tangermann. Motor imagery for severely motor-impaired patients: evidence for brain-computer interfacing as superior control solution. *PloS one*, 9(8):e104854, 2014a.

J. Höhne, E. M. Holz, P. Staiger-Sälzer, K.-R. Müller, and M. Kübler, Andrea Tangermann. Motor imagery for severely motor-impaired patients: Evidence for brain-computer interfacing as superior control solution. *PLoS ONE*, 9(8):e104854, 2014b.

J. Höhne, D. Bartz, M. N. Hebart, K.-R. Müller, and B. Blankertz. Analyzing neuroimaging data with subclasses: a shrinkage approach. *NeuroImage*, 124:740–751, 2016.

B. Hong, F. Guo, T. Liu, X. Gao, and S. Gao. N200-speller using motion-onset visual response. *Clin Neurophysiol*, 120:1658–1666, Sep 2009.

P. J. Huber and E. M. Ronchetti. Robustness of design. *Robust Statistics, Second Edition*, pages 239–248, 1975.

H.-J. Hwang, V. Y. Ferreria, D. Ulrich, T. Kilic, X. Chatziliadis, B. Blankertz, and M. Treder. A gaze independent brain-computer interface based on visual stimulation through closed eyelids. *Scientific reports*, 5, 2015.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

ITU. Methodology for the subjective assessment of the quality of television pictures. *Rec. ITU-R BT 500-11*, 2002.

ITU. Subjective video quality assessmentmethods formultimedia applications. *Rec. ITU-T P.910*, 2008.

B. H. Jansen and M. E. Brandt. The effect of the phase of prestimulus alpha activity on the averaged visual evoked response. *Electroencephalogr Clin Neurophysiol/Evoked Potentials Section*, 80(4):241–250, 1991.

H. Jasper and W. Penfield. Electrocorticograms in man: effect of voluntary movement upon the electrical activity of the precentral gyrus. *Archiv für Psychiatrie und Nervenkrankheiten*, 183(1-2):163–174, 1949.

N. Jayant, J. Johnston, and R. Safranek. Signal compression based on models of human perception. *Proc IEEE*, 81(10):1385–1422, 1993.

JCT-VC. Subversion repository for the hevc test model reference software. 2014.

T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 154–161. ACM, 2005.

T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, T. J. Sejnowski, et al. Extended ica removes artifacts from electroencephalographic recordings. *Advances in neural information processing systems*, pages 894–900, 1998.

J.-P. Kauppi, M. Kandemir, V.-M. Saarinen, L. Hirvenkari, L. Parkkonen, A. Klami, R. Hari, and S. Kaski. Towards brain-activity-controlled information retrieval: Decoding image relevance from meg signals. *NeuroImage*, 112:288–298, 2015.

D. Kelly and N. J. Belkin. A user modeling system for personalized interaction and tailored retrieval in interactive ir. *Proceedings of the American Society for Information Science and Technology*, 39(1):316–325, 2002.

D. Kelly and N. J. Belkin. Display time as implicit feedback: understanding task effects. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 377–384. ACM, 2004.

S. P. Kelly, E. C. Lalor, R. B. Reilly, and J. J. Foxe. Visual spatial attention tracking using high-density ssvep data for independent brain-computer communication. *Neural Syst Rehab Eng, IEEE Trans*, 13(2):172–178, 2005.

S. Kleih, F. Nijboer, S. Halder, and A. Kübler. Motivation modulates the p300 amplitude during brain–computer interface use. *Clinical Neurophysiology*, 121(7):1023–1031, 2010.

S. Kleih, T. Kaufmann, C. Zickler, S. Halder, F. Leotta, F. Cincotti, F. Aloise, A. Riccio, C. Herbert, D. Mattia, and A. Kübler. Out of the frying pan into the fire – the p300-based bci faces real-world challenges. *Jens Schouenborg, Martin Garwicz, Nils Danielsen; Elsevier: Progress in Brain Research*, pages 27–46, 2011.

G. H. Klem, H. O. Lüders, H. Jasper, and C. Elger. The ten-twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol*, 52(suppl.):3, 1999.

W. Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2):169–195, 1999.

W. Klimesch. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*, 16(12):606–617, 2012.

W. Klimesch, M. Doppelmayr, H. Russegger, T. Pachinger, and J. Schwaiger. Induced alpha band power changes in the human eeg and attention. *Neuroscience letters*, 244(2):73–76, 1998.

W. Klimesch, M. Doppelmayr, H. Wimmer, W. Gruber, D. Röhm, J. Schwaiger, and F. Hutzler. Alpha and beta band power changes in normal and dyslexic children. *Clinical Neurophysiology*, 112(7):1186–1195, 2001.

D. S. Klobassa, T. M. Vaughan, P. Brunner, N. E. Schwartz, J. R. Wolpaw, C. Neuper, and E. W. Sellers. Toward a high-throughput auditory P300-based brain-computer interface. *Clin Neurophysiol*, 120:1252–1261, Jul 2009.

J. Koenemann and N. J. Belkin. A case for interaction: a study of interactive information retrieval behavior and effectiveness. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 205–212. ACM, 1996.

J. Kohlmorgen, G. Dornhege, M. Braun, B. Blankertz, K.-R. Müller, G. Curio, K. Hagemann, A. Bruns, M. Schrauf, and W. Kincses. Improving human performance in a real operating environment through real-time mental workload detection. *Toward Brain-Computer Interfacing*, pages 409–422, 2007.

Z. J. Koles, M. S. Lazar, and S. Z. Zhou. Spatial patterns underlying population differences in the background eeg. *Brain topography*, 2(4):275–284, 1990.

M. Krauledat, G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller. The Berlin brain-computer interface for rapid response. *Biomed Tech*, 49(1):61–62, 2004.

E. Kroupi, P. Hanhart, J.-S. Lee, M. Rerabek, and T. Ebrahimi. EEG correlates during video quality perception. *European Sign Process Conf.*, Sept 2014.

D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. A comparison of classification techniques for the p300 speller. *Journal of neural engineering*, 3(4):299, 2006.

A. Kübler and K.-R. Müller. An introduction to brain computer interfacing. In *Toward brain-computer interfacing*, pages 1–25. MIT press, 2007.

A. Kübler, B. Kotchoubey, J. Kaiser, J. Wolpaw, and N. Birbaumer. Brain-computer communication: Unlocking the locked in. *Psychol Bull*, 127(3):358–375, 2001a.

A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. P. Birbaumer. Brain-computer communication: self-regulation of slow cortical potentials for verbal communication. *Archives of physical medicine and rehabilitation*, 82(11):1533–1539, 2001b.

A. Kübler, N. Neumann, B. Wilhelm, T. Hinterberger, and N. Birbaumer. Predictability of brain-computer communication. *Int J Psychophysiol*, 18(2-3):121–129, 2004.

A. Kübler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer, and J. R. Wolpaw. Patients with als can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–1777, 2005.

A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey. A brain-computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients. *Annals of the New York Academy of Sciences*, 1157:90–100, Mar 2009.

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

M. Kutas and V. Iragui. The n400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 108(5): 456–471, 1998.

M. Kutas and C. Van Petten. Psycholinguistics electrified. *Handbook of psycholinguistics*, pages 83–143, 1994.

G. Kylberg. The kylberg texture dataset v. 1.0. *External report (Blue series)*, 35, 2011.

O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller. Spatio-spectral filters for improving the classification of single trial eeg. *Biomed Eng, IEEE Trans*, 52(9):1541–1548, 2005.

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

E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran. A brain–computer interface using electrocorticographic signals in humans. *Journal of neural engineering*, 1(2):63, 2004.

J. Li, Y. Wang, L. Zhang, and T.-P. Jung. Combining erps and eeg spectral features for decoding intended movement direction. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1769–1772. IEEE, 2012.

L. Lindemann and M. Magnor. Assessing the quality of compressed images using EEG. *18th IEEE Int. Conf. Image Process.*, pages 3109 – 3112, 2011. doi: 10.1109/ICIP.2011.6116324.

L. Lindemann, S. Wenger, and M. Magnor. Evaluation of video artifact perception using event-related potentials. *Proc ACM SIGGRAPH Symp Appl Percept in Graph and Visualiz*, pages 53–58, 2011. doi: 10.1145/2077451.2077461.

D. Little. The common european framework of reference for languages: Perspectives on the making of supranational language education policy. *J Mod Lang*, 91(4):645–655, 2007.

T. Liu, L. Goldberg, S. Gao, and B. Hong. An online brain-computer interface using non-flashing visual evoked potentials. *J Neural Eng*, 7(3):036003, 2010.

Y. Liu, Z. Zhou, and D. Hu. Gaze independent brain-computer speller with covert visual search tasks. *Clin Neurophysiol*, 122(6):1127–1136, 2011. URL http://dx.doi.org/10.1016/j.clinph.2010.10.049.

F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, et al. A review of classification algorithms for eeg-based brain–computer interfaces. *J Neural Eng*, 4(2):R1–R13, 2007.

S. J. Luck. *An Introduction to the Event-Related Potential Technique*. The MIT Press, 2005.

S. J. Luck and S. A. Hillyard. Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31(3):291–308, 1994.

C. Maeder, C. Sannelli, S. Haufe, and B. Blankertz. Pre-stimulus sensorimotor rhythms influence brain computer interface classification performance. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 20(5):653–662, Sept 2012. ISSN 1534-4320. doi: 10.1109/TNSRE.2012.2205707.

P. Majaranta, S. MacKenzie, A. Aula, and K. Räihä. Effects of feedback and dwell time on eye typing speed and accuracy. *Univ Access Inf Soc*, 5(2):199–208, 2006.

S. Makeig, A. J. Bell, T.-P. Jung, T. J. Sejnowski, et al. Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, pages 145–151, 1996.

S. Makeig, G. Leslie, T. Mullen, D. Sarma, N. Bigdely-Shamlo, and C. Kothe. First demonstration of a musical emotion bci. In *Affective Computing and Intelligent Interaction*, pages 487–496. Springer, 2011.

A. R. Marathe, A. J. Ries, and K. McDowell. Sliding hdca: single-trial eeg classification to overcome and quantify temporal variability. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 22(2):201–211, 2014.

D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. Spatial filter selection for eeg-based communication. *Electroencephalography and Clinical Neurophysiology*, 103(3): 386 – 394, 1997. ISSN 0013-4694. doi: http://dx.doi.org/10.1016/S0013-4694(97)00022-2. URL http://www.sciencedirect.com/science/article/pii/S0013469497000222.

D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain topography*, 12(3): 177–186, 2000.

K. J. Meador. Cognitive side effects of medications. *Neurologic clinics*, 16(1):141–155, 1998.

J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kübler. An meg-based brain–computer interface (bci). *Neuroimage*, 36(3):581–593, 2007.

E. Miranda and A. Brouse. Toward direct brain-computer musical interfaces. In *Proceedings of the 2005 conference on New interfaces for musical expression*, pages 216–219. National University of Singapore, 2005.

P. Missonnier, M.-P. Deiber, G. Gold, P. Millet, M. G.-F. Pun, L. Fazio-Costa, P. Giannakopoulos, and V. Ibáñez. Frontal theta event-related synchronization: comparison of directed attention and working memory load effects. *Journal of Neural Transmission*, 113 (10):1477–1486, 2006.

S. Mizzaro. Relevance: The whole history. *JASIS*, 48(9):810–832, 1997.

A.-N. Moldovan, I. Ghergulescu, S. Weibelzahl, and C. Muntean. User-centered eeg-based multimedia quality assessment. *Broadband Multimedia Syst and Broadcasting (BMSB), 2013 IEEE Int Sym*, pages 1–8, June 2013. ISSN 2155-5044. doi: 10.1109/BMSB.2013.6621743.

Y. Moshfeghi, L. R. Pinto, F. E. Pollick, and J. M. Jose. Understanding relevance: An fmri study. In *Advances in Information Retrieval*, pages 14–25. Springer, 2013.

G. Mueller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller. Better than random: A closer look on bci results. *International Journal of Bioelectromagnetism*, 10 (EPFL-ARTICLE-164768):52–55, 2008.

K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Neural Networks*, 12(2):181–201, May 2001.

K.-R. Müller, C. W. Anderson, and G. E. Birch. Linear and nonlinear methods for brain-computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):165–169, 2003.

K.-R. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz. Machine learning techniques for brain-computer interfaces. *Biomed Tech*, 49(1):11–22, 2004.

K.-R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz. Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *J Neurosci Methods*, 167(1):82–90, 2008. URL http://dx.doi.org/10.1016/j.jneumeth.2007.09.022.

A. Muller-Gass, M. Macdonald, E. Schröger, L. Sculthorpe, and K. Campbell. Evidence for the auditory p3a reflecting an automatic process: elicitation during highly-focused continuous visual attention. *Brain research*, 1170:71–78, 2007.

M. Mustafa, S. Guthe, and M. Magnor. Single-trial EEG classification of artifacts in videos. *ACM Trans. Appl. Percept.*, 9(3):12:1–12:15, Aug. 2012. ISSN 1544-3558. doi: 10.1145/2325722.2325725.

R. Näätänen and T. W. Picton. N2 and automatic versus controlled processes. *Electroencephalography and clinical neurophysiology. Supplement*, 38:169, 1986.

N. Neumann and N. Birbaumer. Predictors of successful self control during brain-computer communication. *J Neurol Neurosurg Psychiatry*, 74(8):1117–1121, 2003.

C. Neuper, G. Müller, A. Kübler, N. Birbaumer, and G. Pfurtscheller. Clinical application of an eeg-based brain–computer interface: a case study in a patient with severe motor impairment. *Clinical neurophysiology*, 114(3):399–409, 2003.

C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller. Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial eeg. *Cognitive Brain Research*, 25(3):668–677, 2005.

C. Neuper, R. Scherer, S. Wriessnegger, and G. Pfurtscheller. Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain–computer interface. *Clinical neurophysiology*, 120(2):239–247, 2009.

F. Nijboer, E. Sellers, J. Mellinger, M. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. Krusienski, T. Vaughan, et al. A p300-based brain–computer interface for people with amyotrophic lateral sclerosis. *Clinical neurophysiology*, 119(8):1909–1916, 2008.

F. Nijboer, N. Birbaumer, and A. Kübler. The influence of psychological state and motivation on brain–computer interface performance in patients with amyotrophic lateral sclerosis–a longitudinal study. *Frontiers in neuroscience*, 4, 2010.

V. V. Nikouline, K. Linkenkaer-Hansen, H. Wikström, M. Kesäniemi, E. V. Antonova, R. J. Ilmoniemi, and J. Huttunen. Dynamics of mu-rhythm suppression caused by median nerve stimulation: a magnetoencephalographic study in human subjects. *Neuroscience letters*, 294(3):163–166, 2000.

V. V. Nikulin, G. Nolte, and G. Curio. A novel method for reliable and fast extraction of neuronal eeg/meg oscillations on the basis of spatio-spectral decomposition. *NeuroImage*, 55(4):1528–1535, 2011.

A. Norcia, J. Ales, E. Cooper, and T. Wiegand. Measuring perceptual differences between compressed and uncompressed video sequences using the swept-parameter visual evoked potential. *Journal of Vision*, 14(10):649–649, 2014.

J. V. Odom, M. Bach, C. Barber, M. Brigell, M. F. Marmor, A. P. Tormene, and G. E. Holder. Visual evoked potentials standard (2004). *Documenta ophthalmologica*, 108(2): 115–123, 2004.

T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen. Outex-new framework for empirical evaluation of texture analysis algorithms. In *Pattern Recog 2002. Proc 16th Intern Conf*, volume 1, pages 701–706. IEEE, 2002.

B. S. Oken, U. Orhan, B. Roark, D. Erdogmus, A. Fowler, A. Mooney, B. Peters, M. Miller, and M. B. Fried-Oken. Brain–computer interface with language model–electroencephalography fusion for locked-in syndrome. *Neurorehabilitation and neural repair*, 28(4):387–394, 2014.

L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda. Linear spatial integration for single-trial detection in encephalography. *Neuroimage*, 17(1):223–230, 2002.

F. Pei, M. W. Pettet, and A. M. Norcia. Neural correlates of object-based attention. *J Vision*, 2(9):1, 2002.

J. Perez and E. Delechelle. On the measurement of image quality perception using frontal eeg analysis. *Smart Comm in Netw Tech (SaCoNeT), 2013 Int Conf*, 01:1–5, June 2013. doi: 10.1109/SaCoNeT.2013.6654581.

G. Pfurtscheller. Event-related synchronization (ers): an electrophysiological correlate of cortical areas at rest. *Electroencephalography and clinical neurophysiology*, 83(1):62–69, 1992.

G. Pfurtscheller and A. Aranibar. Event-related cortical desynchronization detected by power measurements of scalp eeg. *Electroencephalography and clinical neurophysiology*, 42(6): 817–826, 1977.

G. Pfurtscheller and W. Klimesch. Event-related synchronization and desynchronization of alpha and beta waves in a cognitive task. In *Induced rhythms in the brain*, pages 117–128. Springer, 1992.

G. Pfurtscheller and C. Neuper. Motor imagery activates primary sensorimotor area in humans. *Neuroscience letters*, 239(2):65–68, 1997.

G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123–1134, 2001.

G. Pfurtscheller, C. Neuper, and G. Krausz. Functional dissociation of lower and upper frequency mu rhythms in relation to voluntary limb movement. *Clinical neurophysiology*, 111(10):1873–1879, 2000.

G. Pfurtscheller, C. Neuper, C. Brunner, and F. L. da Silva. Beta rebound after different types of motor imagery in man. *Neuroscience letters*, 378(3):156–159, 2005.

G. Pfurtscheller, B. Z. Allison, G. Bauernfeind, C. Brunner, T. S. Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, and N. Birbaumer. The hybrid bci. *Frontiers in neuroscience*, 4:3, 2010.

J. A. Pineda, B. Allison, and A. Vankov. The effects of self-movement, observation, and imagination on μ rhythms and readiness potentials (rp's): toward a brain-computer interface (bci). *Rehabilitation Engineering, IEEE Transactions on*, 8(2):219–222, 2000.

J. Polich. P300, stimulus intensity, modality, and probability. *International Journal of Psychophysiology*, 23(1-2):55–62, Sept. 1996. ISSN 01678760. doi: 10.1016/0167-8760(96)00028-1. URL http://dx.doi.org/10.1016/0167-8760(96)00028-1.

J. Polich. Updating p300: an integrative theory of p3a and p3b. *Clinical neurophysiology*, 118(10):2128–2148, 2007.

F. Popescu, S. Fazli, Y. Badower, B. Blankertz, and K.-R. Müller. Single trial classification of motor imagination using 6 dry eeg electrodes. *PloS one*, 2(7):e637, 2007.

A. K. Porbadnigk, J.-N. Antons, B. Blankertz, M. S. Treder, R. Schleicher, S. Möller, and G. Curio. Using ERPs for assessing the (sub)conscious perception of noise. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2010, pages 2690–2693, 2010.

A. K. Porbadnigk, S. Scholler, B. Blankertz, A. Ritz, M. Born, R. Scholl, K.-R. Müller, G. Curio, and M. S. Treder. Revealing the neural response to imperceptible peripheral flicker with machine learning. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2011, pages 3692–3695, 2011.

A. K. Porbadnigk, N. Görnitz, M. Kloft, and K.-R. Müller. Decoding brain states during auditory perception by supervising unsupervised learning. *Journal of Computing Science and Engineering*, 7(2):112–121, 2013a.

A. K. Porbadnigk, M. S. Treder, B. Blankertz, J. N. Antons, R. Schleicher, S. Möller, G. Curio, and K.-R. Müller. Single-trial analysis of the neural correlates of speech quality perception. *J Neural Eng*, 10(5):056003, 2013b.

A. K. Porbadnigk, N. Görnitz, C. Sannelli, A. Binder, M. Braun, M. Kloft, and K.-R. Müller. Extracting latent brain states–towards true labels in cognitive neuroscience experiments. *NeuroImage*, 120:225–253, 2015.

G. F. Potts. An erp index of task relevance evaluation of visual stimuli. *Brain and cognition*, 56(1):5–13, 2004.

H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial eeg during imagined hand movement. *Rehabilitation Engineering, IEEE Transactions on*, 8(4):441–446, 2000.

J. E. Raymond, K. L. Shapiro, and K. M. Arnell. Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of experimental psychology: Human perception and performance*, 18(3):849, 1992.

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

D. Regan. *Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine*. Elsevier, 1989.

A. Riccio, D. Mattia, L. Simione, M. Olivetti, and F. Cincotti. Eye-gaze independent eeg-based brain–computer interfaces for communication. *Journal of neural engineering*, 9(4): 045001, 2012.

V. Romei, T. Rihs, V. Brodbeck, and G. Thut. Resting electroencephalogram alpha-power over posterior sites indexes baseline visual cortex excitability. *Neuroreport*, 19(2):203–208, 2008.

S. Romero, M. Mañanas, and M. J. Barbanoj. Ocular reduction in eeg signals based on adaptive filtering, regression and blind source separation. *Annals of biomedical engineering*, 37(1):176–191, 2009.

W. T. Roth, G. H. Blowers, C. M. Doyle, and B. S. Kopell. Auditory stimulus intensity effects on components of the late positive complex. *Electroencephalography and Clinical Neurophysiology*, 54(2):132–146, 1982.

P. Sajda, E. Pohlmeyer, J. Wang, L. C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. K. Singh, and S.-F. Chang. In a blink of an eye and a switch of a transistor: cortically coupled computer vision. *Proceedings of the IEEE*, 98(3):462–478, 2010.

W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe. Stationary common spatial patterns for brain-computer interfacing. *J Neural Eng*, 9(2):026013, 2012. doi: 10.1088/1741-2560/9/2/026013.

W. Samek, M. Kawanabe, and K.-R. Muller. Divergence-based framework for common spatial patterns algorithms. *Biomedical Engineering, IEEE Reviews in*, 7:50–72, 2014.

C. Sannelli, C. Vidaurre, K.-R. Müller, and B. Blankertz. Common spatial pattern patches - an optimized filter ensemble for adaptive brain-computer interfaces. In *Conf Proc IEEE Eng Med Biol Soc*, volume 2010, pages 4351–4354, 2010. URL http://dx.doi.org/10.1109/IEMBS.2010.5626227.

C. Sannelli, C. Vidaurre, K.-R. Müller, and B. Blankertz. Common spatial pattern patches - an optimized filter ensemble for adaptive brain-computer interfaces. *J Neural Eng*, 8(2):025012 (7pp), 2011. doi: 10.1088/1741-2560/8/2/025012.

T. Saracevic. Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26(6):321–343, 1975.

T. Saracevic. Relevance reconsidered. In *Information science: Integration in perspectives. In Proceedings of the Second Conference on Conceptions of Library and Information Science*, volume 1, pages 201–218, 1996.

T. Saracevic. Relevance: A review of the literature and a framework for thinking on the notion in information science. part ii: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13):1915–1933, 2007.

M. Sarter, B. Givens, and J. P. Bruno. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain research reviews*, 35(2):146–160, 2001.

J. Schäfer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1), 2005.

G. Schalk, J. R. Wolpaw, D. J. McFarland, and G. Pfurtscheller. Eeg-based communication: presence of an error potential. *Clinical Neurophysiology*, 111(12):2138–2144, 2000.

N. Schaul. The fundamental neural mechanisms of electroencephalography. *Electroencephalography and clinical neurophysiology*, 106(2):101–107, 1998.

S. Scholler, S. Bosse, M. S. Treder, B. Blankertz, G. Curio, K.-R. Müller, and T. Wiegand. Towards a direct measure of video quality perception using EEG. *Image Processing, IEEE Transactions on*, 21(5):2619–2629, 2012. doi: 10.1109/TIP.2012.2187672.

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

M. Schreuder, J. Höhne, M. S. Treder, B. Blankertz, and M. Tangermann. Performance optimization of ERP-based BCIs using dynamic stopping. In *Conf Proc IEEE Eng Med Biol Soc*, pages 4580–4583, 2011a. doi: 10.1109/IEMBS.2011.6091134.

M. Schreuder, T. Rost, and M. Tangermann. Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. *Front Neuroscience*, 5(112), 2011b. ISSN 1662-453X. doi: 10.3389/fnins.2011.00112. URL https://www.frontiersin.org/Journal/Abstract.aspx?s=763&name=neuroprosthetics&ART_DOI=10.3389/fnins.2011.00112.

M. Schultze-Kraft, D. Birman, M. Rusconi, C. Allefeld, K. Görgen, S. Dähne, B. Blankertz, and J.-D. Haynes. The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences*, page 201513569, 2015.

E. W. Sellers and E. Donchin. A p300-based brain–computer interface: initial tests by ALS patients. *Clinical neurophysiology*, 117(3):538–548, 2006.

E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. A p300 event-related potential brain–computer interface (bci): the effects of matrix size and inter stimulus interval on performance. *Biol Psychol*, 73(3):242–252, 2006.

H. Serby, E. Yom-Tov, and G. Inbar. An improved P300-based brain-computer interface. *IEEE Trans Neural Syst Rehabil Eng*, 13:89–98, Mar 2005.

K. Seshadrinathan and A. C. Bovic. Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Trans Image Process*, 19(2):335–350, 2010.

P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller. Towards adaptive classification for BCI. *J Neural Eng*, 3(1):R13–R23, 2006. URL http://dx.doi.org/10.1088/1741-2560/3/1/R02.

R. B. Silberstein, A. Pipingas, et al. Steady-state visually evoked potential topography during the wisconsin card sorting test. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 96(1):24–35, 1995.

R. Sitaram, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer. Fmri brain-computer interface: a tool for neuroscientific research and treatment. *Computational intelligence and neuroscience*, 2007, 2007.

J. O. Smith. *Mathematics of the discrete Fourier transform (DFT): with audio applicaitons*. Julius Smith, 2007.

K. M. Spencer, J. Dien, and E. Donchin. Spatiotemporal analysis of the late erp responses to deviant stimuli. *Psychophysiology*, 38(02):343–358, 2001.

N. K. Squires, K. C. Squires, and S. A. Hillyard. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, 38(4):387–401, 1975.

R. Srinivasan, F. A. Bibi, and P. L. Nunez. Steady-state visual evoked potentials: distributed local sources and wave-like dynamics are sensitive to flicker frequency. *Brain topography*, 18(3):167–187, 2006.

A. D. Straw. Vision egg: an open-source library for realtime visual stimulus generation. *Frontiers in Neuroinformatics*, 2(0), 2008. doi: 10.3389/neuro.11.004.2008.

P. Suffczynski, J. P. M. Pijn, G. Pfurtscheller, F. Lopes da Silva, et al. Event-related dynamics of alpha band rhythms: a neuronal network model of focal erd-surround ers. In *Event-related desynchronization. Handbook of Electroencephalography and Clinical Neuorphysiology, Revised Series*, volume 6, pages 67–85. Elsevier Science, 1999.

M. Sugiyama, M. Krauledat, and K.-R. Müller. Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8:1027–1061, 2007.

H.-I. Suk, S. Fazli, J. Mehnert, K.-R. Müller, and S.-W. Lee. Predicting bci subject performance using probabilistic spatio-temporal filters. *PLoS ONE*, 9(2):e87056, 02 2014. doi: 10.1371/journal.pone.0087056.

G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand. Overview of the high efficiency video coding (hevc) standard. *Circuits Syst Video Tech, IEEE Trans*, 22(12):1649–1668, 2012.

S. Sutton and D. S. Ruchkin. The late positive complex: Advances and new problems. *Annals of the New York Academy of Sciences*, 425(1):1–23, 1984.

S. Sutton, M. Braren, J. Zubin, and E. John. Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700):1187–1188, 1965.

D. R. Swanson. Subjective versus objective relevance in bibliographic retrieval systems. *The Library Quarterly*, pages 389–398, 1986.

C. Tallon-Baudry and O. Bertrand. Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences*, 3(4):151–162, 1999.

M. Tangermann, J. Höhne, H. Stecher, and M. Schreuder. No surprise-fixed sequence event-related potentials for brain-computer interfaces. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 2501–2504. IEEE, 2012.

M. E. Thurlings, J. B. van Erp, A.-M. Brouwer, B. Blankertz, and P. Werkhoven. Control-display mapping in brain-computer interfaces. *Ergonomics*, 55(5):564–580, 2012. doi: 10.1080/00140139.2012.661085. URL http://www.tandfonline.com/doi/abs/10.1080/00140139.2012.661085.

R. Tomioka and K.-R. Müller. A regularized discriminative framework for eeg analysis with application to brain–computer interface. *Neuroimage*, 49(1):415–432, 2010.

G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. E. Frye, C. K. Hauser, N. E. Schwartz, T. M. Vaughan, J. R. Wolpaw, and E. W. Sellers. A novel p300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. *Clin Neurophysiol*, 121(7):1109–1120, 2010.

M. S. Treder and B. Blankertz. (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behav Brain Funct*, 6:28, May 2010. URL http://www.behavioralandbrainfunctions.com/content/6/1/28.

M. S. Treder, N. M. Schmidt, and B. Blankertz. Gaze-independent brain-computer interfaces based on covert attention and feature attention. *J Neural Eng*, 8(6):066003, 2011. URL http://dx.doi.org/10.1088/1741-2560/8/6/066003. Open Access.

M. S. Treder, H. Purwins, D. Miklody, I. Sturm, and B. Blankertz. Decoding auditory attention to instruments in polyphonic music using single-trial eeg classification. *Journal of neural engineering*, 11(2):026009, 2014.

M. van der Waal, M. Severens, J. Geuze, and P. Desain. Introducing the tactile speller: an ERP-based brain–computer interface for communication. *Journal of Neural Engineering*, 9 (4):045002, 2012.

M. van Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, A. Nijholt, N. Ramsey, P. Haselager, L. Vuurpijl, S. Gielen, and P. Desain. The brain–computer interface cycle. *Journal of neural engineering*, 6(4):041001, 2009.

F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie. The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience*, 2(4):229–239, 2001.

H. G. Vaughan and W. Ritter. The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and clinical neurophysiology*, 28(4):360–367, 1970.

B. Venthur, B. Blankertz, M. F. Gugler, and G. Curio. Novel applications of bci technology: psychophysiological optimization of working conditions in industry. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 417–421. IEEE, 2010a.

B. Venthur, S. Scholler, J. Williamson, S. Dähne, M. S. Treder, M. T. Kramarek, K.-R. Müller, and B. Blankertz. Pyff – a pythonic framework for feedback applications and stimulus presentation in neuroscience. *Front Neuroscience*, 4:179, 2010b. URL http://dx.doi.org/10.3389/fnins.2010.00179.

R. Verleger, P. Jaśkowski, and B. Wauschkuhn. Suspense and surprise: On the relationship between expectancies and p3. *Psychophysiology*, 31(4):359–369, 1994. ISSN 1469-8986. doi: 10.1111/j.1469-8986.1994.tb02444.x. URL http://dx.doi.org/10.1111/j.1469-8986.1994.tb02444.x.

F.-B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Prog Neurobiol*, 90(4):418–438, 2010.

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

C. Vidaurre and B. Blankertz. Towards a cure for BCI illiteracy. *Brain Topogr*, 23:194–198, 2010. URL http://dx.doi.org/10.1007/s10548-009-0121-6. Open Access.

C. Vidaurre, A. Schlögl, B. Blankertz, M. Kawanabe, and K.-R. Müller. Unsupervised adaptation of the lda classifier for brain–computer interfaces. In *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course*, volume 2008, pages 122–127. Citeseer, 2008.

C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl. Time domain parameters as a feature for EEG-based Brain Computer Interfaces. *Neural Networks*, 22:1313–1319, 2009. URL http://dx.doi.org/10.1016/j.neunet.2009.07.020.

C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K.-R. Müller. Toward unsupervised adaptation of lda for brain-computer interfaces. *IEEE Trans Biomed Eng*, 58(3):587–597, 2011a. URL http://dx.doi.org/10.1109/TBME.2010.2093133.

C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz. Co-adaptive calibration to improve BCI efficiency. *J Neural Eng*, 8(2):025009 (8pp), 2011b. doi: 10.1088/1741-2560/8/2/025009. URL http://dx.doi.org/10.1088/1741-2560/8/2/025009.

C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz. Machine-learning-based coadaptive calibration for brain-computer interfaces. *Neural computation*, 23(3):791–816, 2011c.

R. N. Vigário. Extraction of ocular artefacts from eeg using independent component analysis. *Electroencephalography and clinical neurophysiology*, 103(3):395–404, 1997.

F. C. Viola, J. Thorne, B. Edmonds, T. Schneider, T. Eichele, and S. Debener. Semi-automatic identification of independent components representing eeg artifact. *Clinical Neurophysiology*, 120(5):868–877, 2009.

E. K. Vogel and S. J. Luck. The visual n1 component as an index of a discrimination process. *Psychophysiology*, 37(02):190–203, 2000.

I. Volosyak, D. Valbuena, T. Malechka, J. Peuscher, and A. Gräser. Brain–computer interface using water-based electrodes. *Journal of neural engineering*, 7(6):066007, 2010.

I. Volosyak, D. Valbuena, T. Luth, T. Malechka, and A. Graser. Bci demographics ii: how many (and what kinds of) people can use a high-frequency ssvep bci? *Neural Syst Rehab Eng, IEEE Trans*, 19(3):232–239, 2011.

N. Weiskopf, K. Mathiak, S. W. Bock, F. Scharnowski, R. Veit, W. Grodd, R. Goebel, and N. Birbaumer. Principles of a brain-computer interface (bci) based on real-time functional magnetic resonance imaging (fmri). *Biomedical Engineering, IEEE Transactions on*, 51(6): 966–970, 2004.

R. W. White and D. Kelly. A study on the effects of personalization and task information on implicit feedback performance. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 297–306. ACM, 2006.

A. Widmann and E. Schröger. Filter effects and filter artifacts in the analysis of electrophysiological data. *Frontiers in psychology*, 3:233, 2012.

I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld, and M. Tangermann. Robust artifactual independent component classification for bci practitioners. *Journal of neural engineering*, 11(3):035013, 2014.

J. Wolpaw and E. W. Wolpaw. *Brain-computer interfaces: principles and practice*. Oxford University Press, 2012a.

J. R. Wolpaw and E. W. Wolpaw, editors. *Brain-computer interfaces : principles and practice*. Oxford University press, 2012b. ISBN-13: 978-0195388855.

J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clin Neurophysiol*, 113(6):767–791, 2002.

T. Womelsdorf, J.-M. Schoffelen, R. Oostenveld, W. Singer, R. Desimone, A. K. Engel, and P. Fries. Modulation of neuronal interactions through neuronal synchronization. *science*, 316(5831):1609–1612, 2007.

G. Wübbeler, A. Ziehe, B.-M. Mackert, K.-R. Muller, L. Trahms, and G. Curio. Independent component analysis of noninvasively recorded cortical magnetic dc-fields in humans. *Biomedical Engineering, IEEE Transactions on*, 47(5):594–599, 2000.

T. O. Zander and C. Kothe. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering*, 8(2):025005, 2011.

D. Zhang, X. Honglai, W. Wu, S. Gao, and B. Hong. Integrating the spatial profile of the n200 speller for asynchronous brain-computer interfaces. In *Conf Proc IEEE Eng Med Biol Soc*, pages 4564–45767, 2011.

R. Zhang, P. Xu, R. Chen, F. Li, L. Guo, P. Li, T. Zhang, and D. Yao. Predicting inter-session performance of smr-based brain–computer interface using the spectral entropy of resting-state eeg. *Brain topography*, pages 1–11, 2015a.

R. Zhang, D. Yao, P. A. Valdés-Sosa, F. Li, P. Li, T. Zhang, T. Ma, Y. Li, and P. Xu. Efficient resting-state eeg network facilitates motor imagery performance. *Journal of neural engineering*, 12(6):066024, 2015b.

Appendix A

Supplementary figures

Supplementary figures of Chapter 6

Grand-average ERPs and sgn-r^2 R2-NR

Figs. A.1 and A.2 show the grand average ERPs and sgn-r^2 for class combination R2-NR. Note that the highlighted time intervals are the same as in Figs. 6.4 and 6.5.

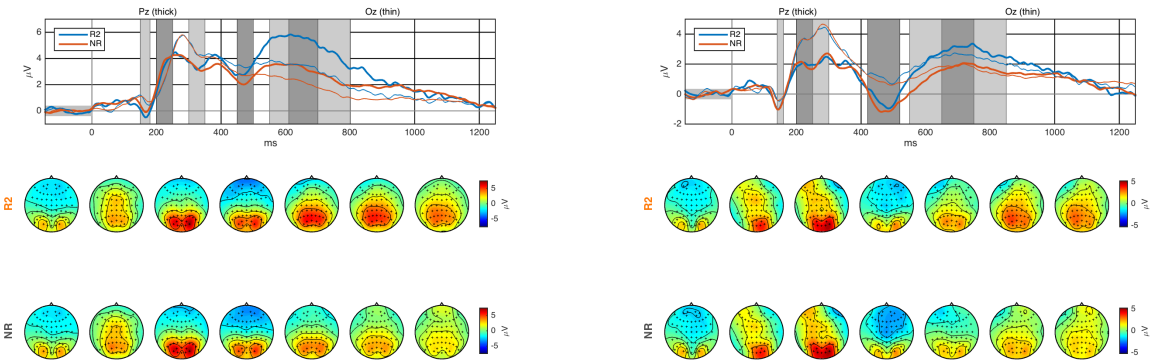


Fig. A.1 Grand average ERP evolution time-locked to relevant (R2) and non-relevant (NR) stimuli and scalp plots, for *pictures* (left) and *words* (right). In the first row the time evolution of the ERPs is depicted at channels Pz (thick lines) and Oz (thin lines). In blue are the ERPs of class R2, in orange the ERPs of class NR. Time 0 correspond to the onset of the stimuli. The shaded areas highlight interesting time windows wherein the average voltage is calculated and displayed as scalp plots underneath (second row refers to R1 and third row to NR). Note that plots referring to *pictures* have different scales to those referring to *words*. Time windows are differently tuned for *pictures* and *words*.

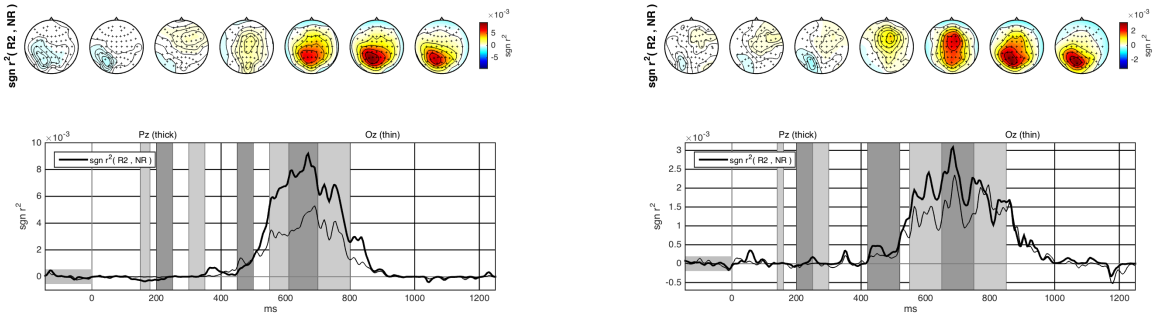


Fig. A.2 Grand average sgn-r^2 of classes R2 and NR and scalp plots, for *pictures* (left) and *words* (right). The same time windows of Fig. 6.4 are highlighted. Note that plots referring to *pictures* have different scales to those referring to *words*.

Grand-average ERD/ERS R2-NR

Fig. A.3 shows the ERD/ERS for class combination R2-NR, for the selected frequency bands 2-4 Hz, 10-14 Hz, 15-20 Hz. Note that the selected time intervals are the same as Fig. 6.8.

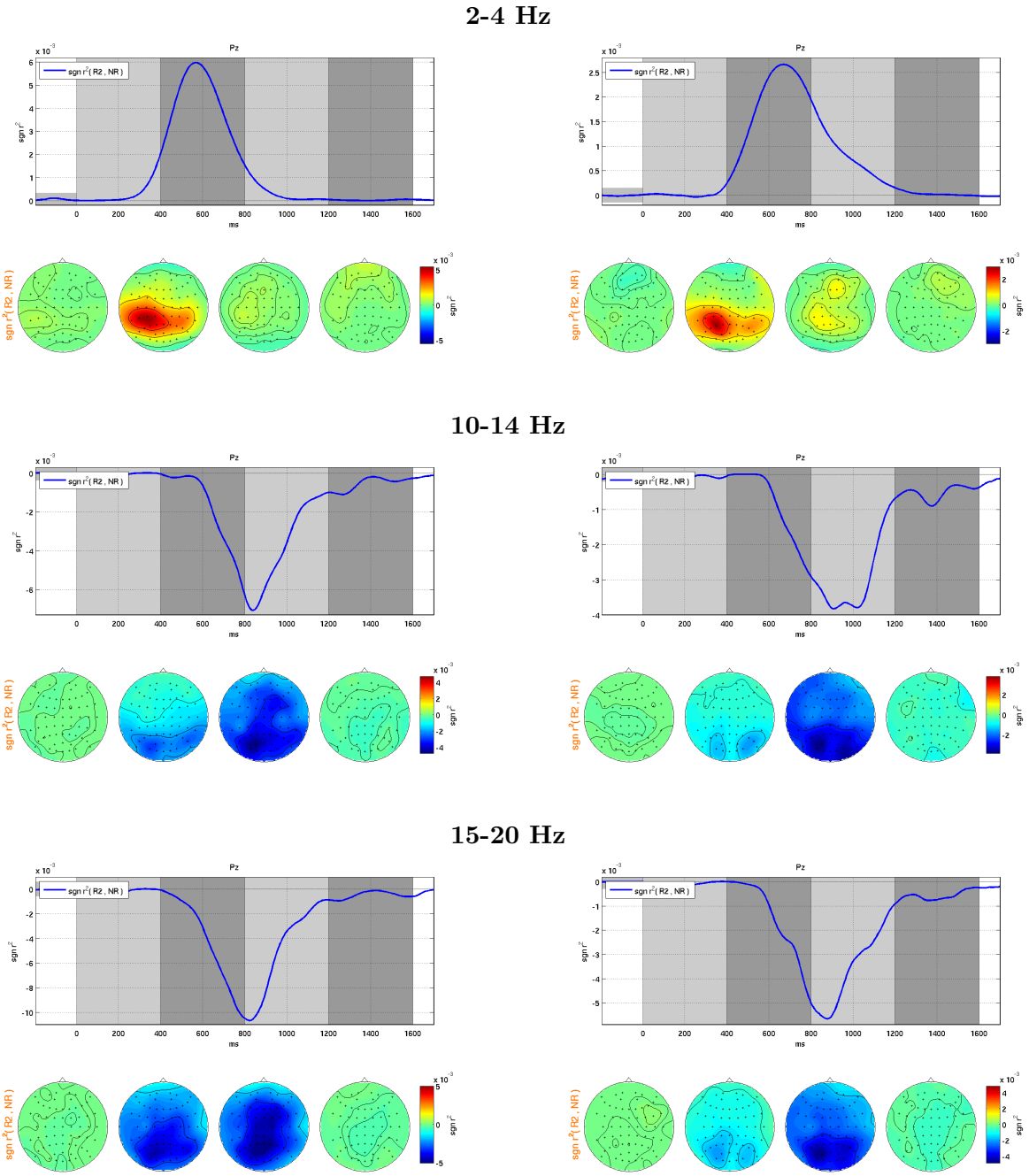


Fig. A.3 Sgn-r^2 of ERD/ERS of classes R2 and NR for *pictures* (left column) and *words* (right column). From top to bottom the three investigated frequency bands: 2-4 Hz, 10-14 Hz, 15-20 Hz. Note the different scales between plots.

Supplementary figures of Chapter 7

Figs A.4 and A.5 show the grand average ERD/ERS and sgn-r^2 for class combination *left-foot*, for participants in Berlin and Würzburg, respectively. Figs A.6 and A.7 show the grand average ERD/ERS and sign-r^2 for class combination *foot-right*, for participants in Berlin and Würzburg, respectively.

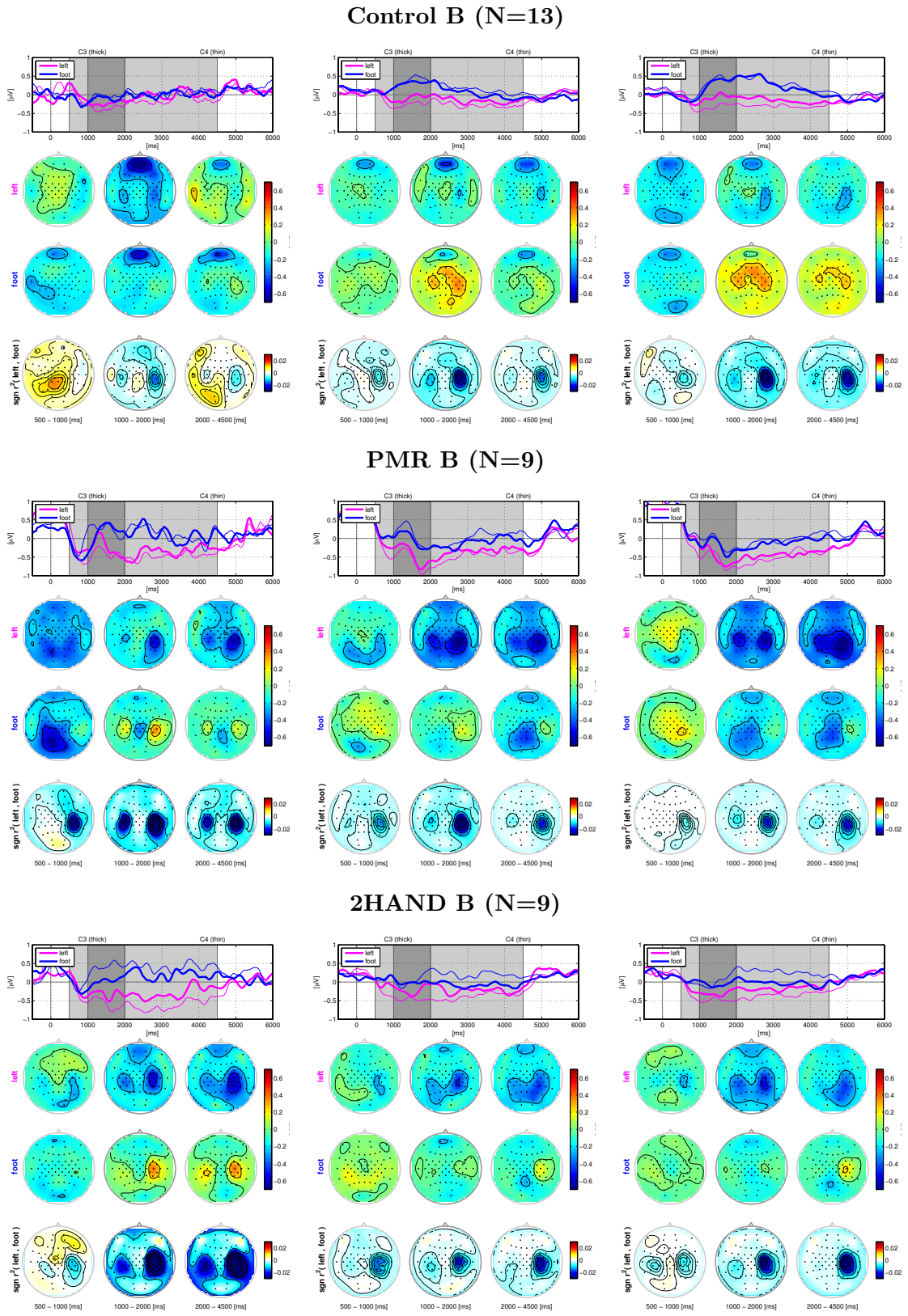


Fig. A.4 Grand average ERD/ERS for class combination *left-foot* and intervention groups in Berlin. 'N' is the number of participants of each group. From left to right: runs 1-3, runs 4-5, runs 6-7. The time plots in the first rows picture the evolution of the ERD/ERS for about 6000 ms at C3 (thick lines) and C4 (thin lines). At time 0 is the onset of the cue, at times 1000-4000 the display of the feedback. Magenta lines refer to *left* MI trials, blue lines to *foot* MI trials. The scalp plots underneath refer to the shaded areas of the time plots and show the distribution of the ERD/ERS. In the second rows, the scalp plots of the *left* MI trials, in the third rows the scalp plots of the *foot* MI trials and in the fourth the scalp plots of the sgn-r^2.

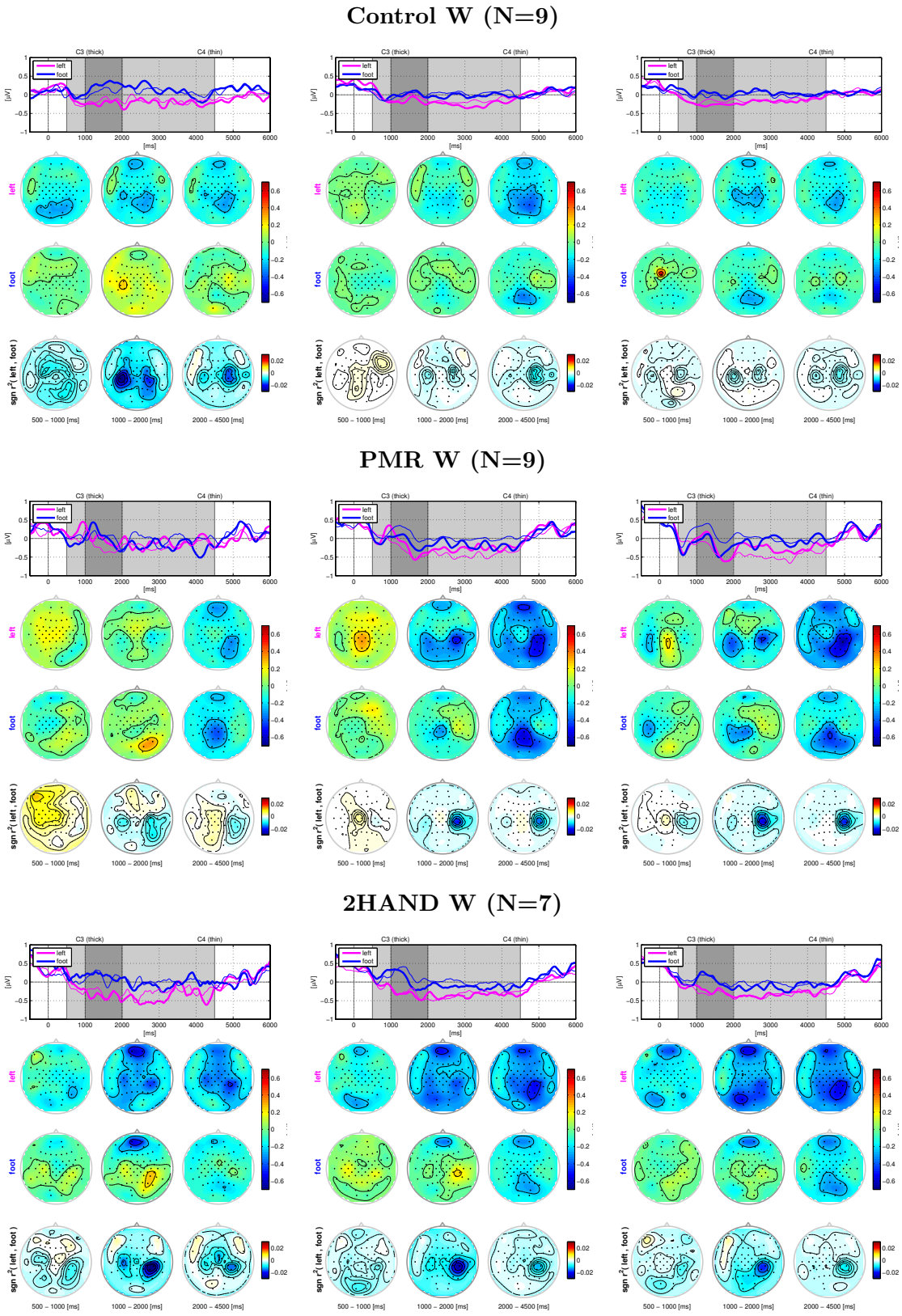


Fig. A.5 Grand average ERD/ERS for class combination *left-foot* and intervention groups in Würzburg. 'N' is the number of participants of each group. From left to right: runs 1-3, runs 4-5, runs 6-7. The time plots in the first rows picture the evolution of the ERD/ERS for about 6000 ms at C3 (thick lines) and C4 (thin lines). At time 0 is the onset of the cue, at times 1000-4000 the display of the feedback. Magenta lines refer to *left* MI trials, blue lines to *foot* MI trials. The scalp plots underneath refer to the shaded areas of the time plots and show the distribution of the ERD/ERS.In the second rows, the scalp plots of the *left* MI trials, in the third rows the scalp plots of the *foot* MI trials and in the fourth the scalp plots of the sgn-r^2.

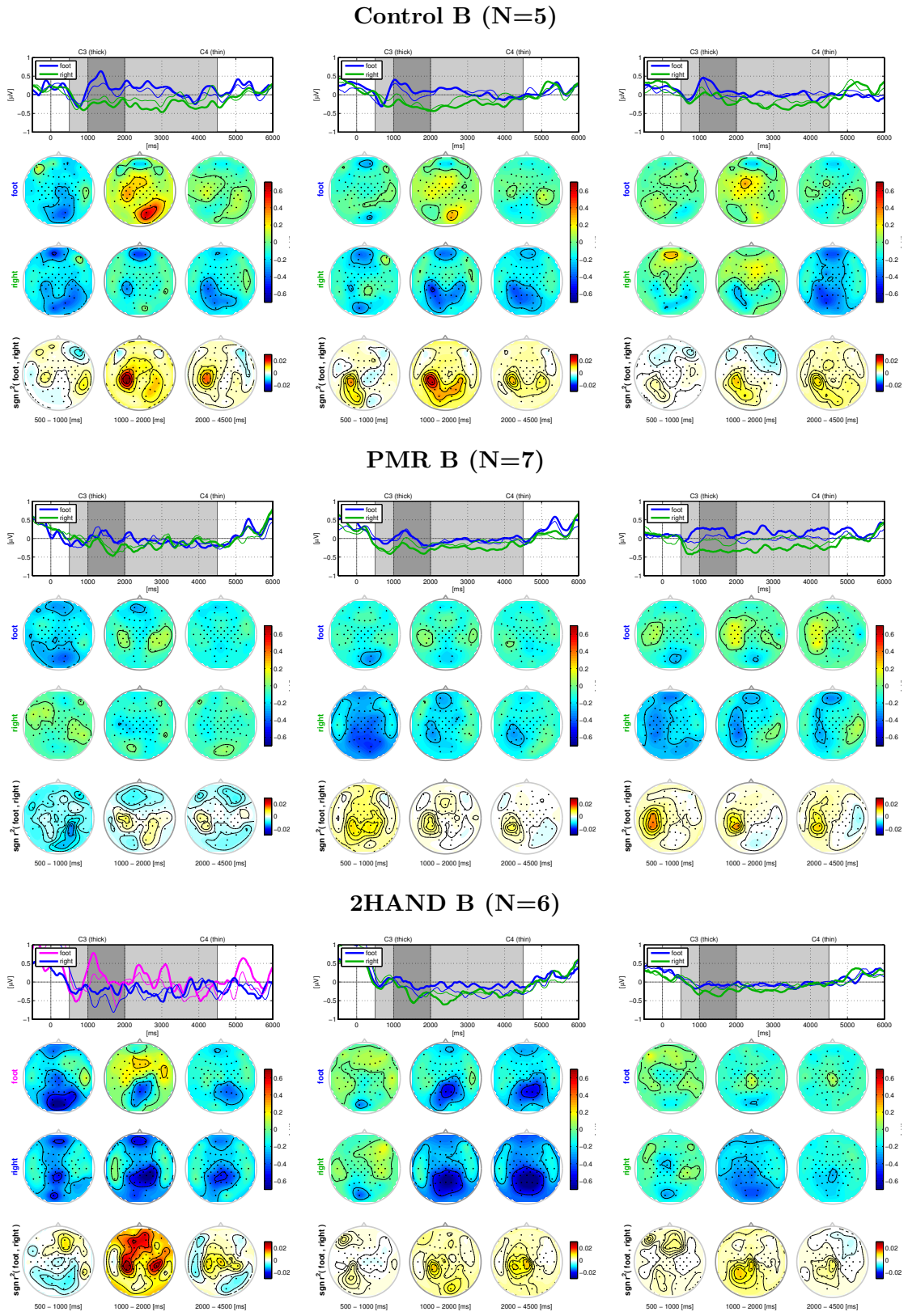


Fig. A.6 Grand average ERD/ERS for class combination *foot-right* and intervention groups in Berlin. 'N' is the number of participants of each group. From left to right: runs 1-3, runs 4-5, runs 6-7. The time plots in the first rows picture the evolution of the ERD/ERS for about 6000 ms at C3 (thick lines) and C4 (thin lines). At time 0 is the onset of the cue, at times 1000-4000 the display of the feedback. Blue lines refer to *foot* MI trials, green lines to *right* MI trials. The scalp plots underneath refer to the shaded areas of the time plots and show the distribution of the ERD/ERS. In the second rows, the scalp plots of the *foot* MI trials, in the third rows the scalp plots of the *right* MI trials and in the fourth the scalp plots of the sgn-r^2.

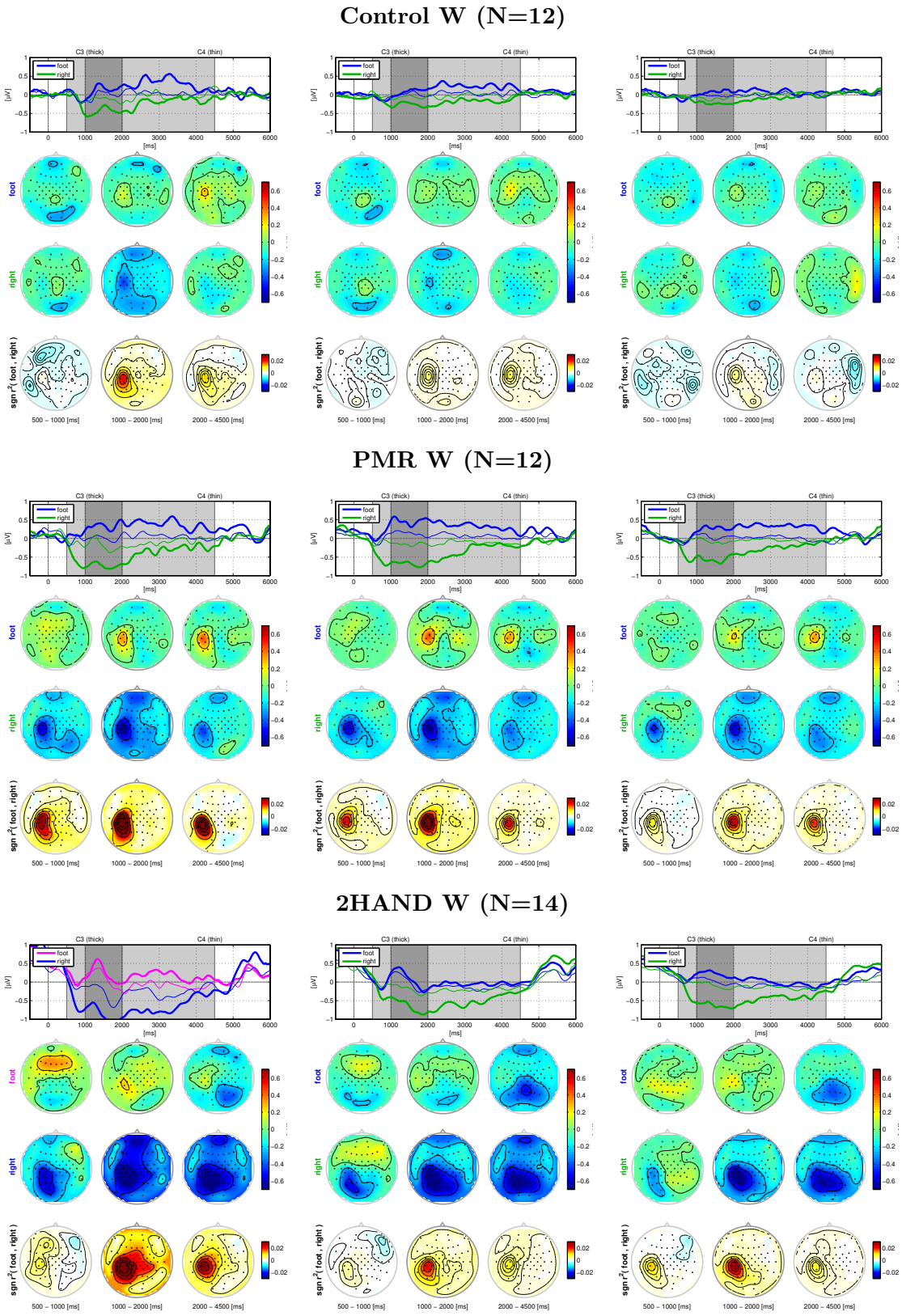


Fig. A.7 Grand average ERD/ERS for class combination *foot-right* and intervention groups in Würzburg. 'N' is the number of participants of each group. From left to right: runs 1-3, runs 4-5, runs 6-7. The time plots in the first rows picture the evolution of the ERD/ERS for about 6000 ms at C3 (thick lines) and C4 (thin lines). At time 0 is the onset of the cue, at times 1000-4000 the display of the feedback. Blue lines refer to *foot* MI trials, green lines to *right* MI trials. The scalp plots underneath refer to the shaded areas of the time plots and show the distribution of the ERD/ERS.In the second rows, the scalp plots of the *foot* MI trials, in the third rows the scalp plots of the *right* MI trials and in the fourth the scalp plots of the sgn-r^2.