Towards Efficient Auditory BCI Through Optimized Paradigms and Methods

Evert-Jan Martijn Schreuder

Towards Efficient Auditory BCI Through Optimized Paradigms and Methods

vorgelegt von
M.Sc.
Evert-Jan Martijn Schreuder
geb. in Amsterdam

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
Gutachter:	Prof. Dr. Klaus-Robert Müller
Gutachter:	Prof. Dr. Benjamin Blankertz
Gutachterin:	Prof. Dr. Andrea Kübler

Tag der wissenschaftlichen Aussprache: 3. April 2014

Berlin 2014

D83

Towards Efficient Auditory BCI Through Optimized Paradigms and Methods

Evert-Jan Martijn Schreuder

epubli

Impressum

Publisher: epubli GmbH, Berlin, www.epubli.de,
and Technische Universität Berlin, Diss., 2014

ISBN: 978-3-7375-0485-0

To my son, Joaquim

Abstract

To date, the brain-computer interface (BCI) based on visual stimulation is by far the most investigated. This makes a lot of sense given the excellent visual capabilities of humans, and it has been a most successful approach. Nevertheless, a non-negligible part of the BCI end-user population with advanced paralysis is incapable of directing their eye gaze or of seeing at all. Traditional visual BCIs will rarely be a solution for these end-users and alternative paradigms are required that rely on covert attention. Furthermore, though BCIs already benefit greatly from statistical methods coming from the field of machine learning, a faster and robust performance is required for adoption by the clinical community. This thesis contributes in two key facets in an attempt to take a step towards full inclusion of all end-users.

First, an auditory alternative is proposed to address the intrinsic problems that exist with BCIs based on the visual event-related potential (ERP). It is named AMUSE (short for Auditory MUlticlass Spatial Event-related potential). Though it is not the first auditory BCI to be proposed, it differs from the mostly binary contemporary solutions in a crucial aspect. Instead of relying on pitch or other physical features of the stimuli to define the class membership, AMUSE targets the human capability of spatial localization of sound sources. With AMUSE, stimuli can contain physical differences but the main defining feature is the sound source or, more practically, the loudspeaker position. Using this feature, the stimuli can be short, stimulation speed can be high, and the number of classes can easily be increased by adding sources.

The fundamentals of AMUSE as a BCI paradigm are carefully validated, and an AMUSE-based BCI speller is proposed with minimal reliance on the visual domain. The vast majority of the tested healthy subjects is able to write a full sentence using this speller. The average performance is high compared to contemporary covert attention-based BCIs, both in terms of the information transfer rate (ITR) and the amount of written characters per minute (char/min).

Most BCI research is directed towards the final goal of successful end-user application. Accordingly, the AMUSE-based speller is tested with five end-users with advanced paralysis. Up to four online sessions are performed for each end-user, and performance is above chance on all but two sessions. Nevertheless, the performance does not suffice for the end-users to gain meaningful control over the AMUSE-based speller. This result is discussed in the context of the available attention resources, which was pathologically low for at least one end-user.

The second important contribution of this thesis is a novel algorithm for performance optimization in ERP-based BCI in general, called rank diff. ERP-based BCIs have an intrinsic trade-off between the length of a trial and the accuracy of the resulting decision; longer trials typically result in more accurate decisions. In literature, this trade-off is mostly ignored and the length of the trial is simply fixed, leading to suboptimal performance. Rank diff uses the available calibration data to estimate a threshold on the required evidence necessary for a correct decision. Online, a trial is stopped as soon as this threshold is reached, leading to trials of varying length.

By including rank diff in the AMUSE-based speller, the performance increases by about 40 % for healthy subjects. As a result, the speller is competitive with covert attention-based visual spellers. Rank diff is further validated on benchmark data along with several other methods, showing the beneficial effect not only of rank diff, but of evidence accumulation methods in general. As average performance increases as high as 78 % are found, these methods can no longer be ignored.

Zusammenfassung

Bis zum heutigen Tage hat sich die brain-computer interface (BCI) Forschung hauptsächlich auf visuelle Paradigmen fokussiert. Dieser Ansatz war bisher sehr erfolgreich, da Menschen generell über ausgezeichnete visuelle Fähigkeiten verfügen. Allerdings kämpft ein nicht vernachlässigbarer Teil der Gruppe der Endnutzer mit Problemen bei der Ausrichtung der Augen oder mit Problemen beim Sehen generell. Diese Probleme führen dazu, dass traditionelle visuelle BCI Paradigmen von diesen Endnutzern nicht verwendet werden können. Stattdessen werden alternative Paradigmen gebraucht die auf verdeckter Aufmerksamkeit beruhen. Hinzukommt dass, obwohl BCIs bereits von statistischen Verfahren aus dem Bereich Maschinelles Lernen profitieren, schnellere und robuste Anwendbarkeit erforderlich ist, um die klinische Gemeinschaft von ihrer Nützlichkeit zu überzeugen. Diese Dissertation trägt in zwei wichtigen Aspekten dazu bei, einen weiteren Schritt in Richtung vollständiger Inklusion von Endnutzern zu schaffen.

Erstens wird unter dem Namen Auditory MUlticlass Spatial Event-related potential (AMUSE) eine auditorische Alternative vorgestellt, die die intrinsischen Probleme mit visuellen, auf ereigniskorrelierten Potentialen (ERP) basierten BCIs umgeht. AMUSE unterscheidet sich von den bisher meistens binären, auditorischen BCIs in einem entscheidenden Aspekt. Anstatt sich für die Klassenzuordnung ausschließlich auf Tonhöhe oder ähnliche physikalische Aspekte der Stimuli zu verlassen, spricht AMUSE die menschliche Fähigkeit für räumliche Lokalisierung von Schallquellen an. Bei AMUSE können die Stimuli solche physikalischen Unterschiede zwar haben, der entscheidende Faktor ist aber die räumliche Position der Schallquellen bzw. der Lautsprecher. Dieser Aspekt ermöglicht die Verwendung sehr kurzer Stimuli, woraus sich eine hohe Stimulationsgeschwindigkeit ergibt. Darüberhinaus kann die Anzahl von Klassen durch Hinzufügen von weiteren Schallquellen einfach erhöht werden.

Die Grundlagen von AMUSE als BCI Paradigma werden sorgfältig validiert und eine auf AMUSE basierende Schreibanwendung wird vorgestellt welche nur minimal auf visuelle Fähigkeiten zurückgreift. Der Großteil der gesunden Versuchspersonen waren in der Lage einen vollständigen Deutschen Satz per BCI zu schreiben. Im Vergleich zu gegenwärtigen auf verdeckter Aufmerksamkeit basierenden BCIs arbeitet AMUSE sehr effizient, sowohl bezogen auf die Übertragungsgeschwindigkeit (ITR) als auch auf die Anzahl von geschriebenen Buchstaben pro Minute (CPM).

Ein wichtiges Ziel der BCI Forschung ist eine erfolgreiche Anwendung bei Endnutzern. Dementsprechend wurde die auf AMUSE basierte Schreibanwendung mit fünf Endnutzern mit fortgeschrittener Lähmung getestet. Bis zu fünf Online-Sitzungen wurde mit jedem Endnutzer durchgeführt. Mit Ausnahme von zwei Sitzungen konnte BCI-Kontrolle erreicht werden, die über Zufallsniveau lag. Eine für die Endnutzer befriedigende Kontrolle konnte dennoch nicht erreicht werden. Dieses Ergebnis wird im Kontext der zur Verfügung stehenden Aufmerksamkeitsressourcen diskutiert, welche für mindestens einen Endnutzer nachweislich pathologisch waren.

Der zweite wichtige Beitrag dieser Dissertation ist die Vorstellung eines neuen Algorithmus für die Optimierung von auf event-related potential (ERP) basierten BCIs, der als rank diff bezeichnet wird. ERP basierten BCIs ist ein Kompromiss zwischen der Länge der Stimulation und der Genauigkeit der Entscheidung inne. Längere Stimulation führt üblicherweise zu mehr Genauigkeit. In der Literatur wird dieser Kompromiss oftmals ignoriert und die Länge der Stimulation manuell gesetzt, was zu suboptimaler Leistung führt. Der hier vorgestellte Algorithmus rank diff benutzt die vorhandenen Kalibrationsdaten um einen adaptiven Schwellwert zu bestimmen, der überschritten werden muss bevor die Stimulation beendet wird.

Das Anwenden von rank diff steigert die Effizienz von AMUSE bei gesunden Versuchspersonen um etwa 40 %. Als Folge nähert sich die Leistung von AMUSE dem Niveau von auf verdeckter Aufmerksamkeit beruhenden visuellen BCI-Schreibanwendungen. Desweiteren werden rank diff und ähnliche Methoden auf Benchmark-Daten validiert. Die Ergebnisse zeigen den verbessernden Effekt nicht nur für rank diff, sondern auch für verwandte Methoden. Effizienzsteigerungen von bis zu 78 % zeigen, dass diese Art von Optimierung nicht länger ignoriert werden sollte.

Acknowledgment

With this thesis now in front of me, I realize that it would have never been finished if it was not for all the people around me to rely on, to distract me at some moments and keep me focussed at others. I would like to thank all of you in person someday if I have not already done so. But let me start here just in case life throws in one of its strange twists.

The BBCI group has been a great place to work at, not in the least because of the people that work there. I would like to sincerely thank Prof. Dr. Klaus-Robert Müller and Prof. Dr. Benjamin Blankertz. I have enjoyed the truly interdisciplinary nature of their groups, the freedom I was given to explore my own scientific interests, and the sustained support throughout personal turmoil. I am also indebted to Dr. Michael Tangermann for his engaged supervision of this thesis, often helping out hands-on with the tedious act of putting on EEG caps. His interest in the field of auditory BCI has been most helpful. I have greatly enjoyed his believe and trust in my skills and independence, not only for the thesis but in particular for the project work we did together.

But in fact I have greatly enjoyed working, conversing, laughing, and enjoying Berlin with many of my colleagues and fellow graduate students. Thank you Tobias, for putting up with me for years in the same house and office and still feeling like going for a weekend to Barcelona with me. Thank you Sven and Johannes, for challenging me both at work and on the beach volleyball court. Thank you Claudia, for enjoying so many places in this world with me. Thank you Basti, for being Basti, you nailed it! Thank you Nikolay for infecting me with your infinite curiosity. Thank you Ulf, for playing such a decisive role in my life. Thank you Anne, Carmen, Cecilia, Daniel, Felix, Florian, Frank, Irene, Jan, Janne, Javier, Kai, Katja, Laura, Leo, Matthias, Matthias SK, Marieke, Markus, Marton, Siamac, Paul, Siamac, Stanio, Stefan, Steven, and Xingwei An for having a lasting impact on me and my time in Berlin. A special thank you is reserved for Andrea, Imke and Dominik for all the organizational and technical support, often in times of despair.

Through my involvement in the European project TOBI, I have been working closely with many international researchers, some if which I am fortunate to now consider my friends. They too have made these years so much more interesting. My collaboration with Donatella Mattia and her group, in particular Angela, has been especially important and pleasant. Thank you for hosting me for a fantastic and fruitful month in Rome and giving me the opportunity to work with end-users, but also for letting me navigate you through your own wonderful city.

Every TOBI meeting was an event to look forward to. As the whole TOBI family gathered for on- and off topic discussions, I learned many important life lessons. Thank you Rod, for showing me that Glasgow's haggish is something to enjoy, not be afraid of. Thank you Lorenzo and Matteo for introducing me to Bologna's finest pasta, and for explaining that pasta bolognese is tourist trickery. Thank you Christian, for taking me to drink Graz's best glühwein. Thank you Alex, Andrea, Evert-Jan, Gernot, Luca, Matthias, Michele, Robert, Serafeim, Sonja, and Rüdiger for making the time in TOBI one to never forget. And in particular, thank you José and Nancy for steering TOBI towards the success it has been.

This thesis would have been impossible without the many experimental participants. In particular the end-users often had to go to great lengths to make it to the sessions. I am indebted to you for your efforts and sincerely hope that this thesis will aid future developments of the so necessary BCI-based assistive technology.

Time and again, my family and friends have asked me what it is that I do with genuine interest. Their unconditional support has been an unimaginable drive for me. I can only hope to repay you by returning the favor, so count on me for pushing you further.

Finally, I owe my greatest gratitude to Mia. I can not express just how thankful I am for your love and support. During the last months of writing this thesis I was particularly preoccupied, but you never seized to try and make the process as easy for me as possible. This too meant that you had to take care of our wonderful son, Joaquim, which you do with fantastic dedication. Thank you for being (in) my life.

Contents

Abstract ix

Zusammenfassung xi

Acknowledgment xiii

1. Introduction 1
1.1. Introduction to Brain-Computer Interfaces . . . 1
1.2. Applications of BCI . . . 2
1.3. The End-User of BCI Assistive Technology . . . 3
1.4. EEG-Based Brain-Computer Interfaces . . . 4
1.5. Covert Attention-Based BCI . . . 7
1.6. The Typcial Design Chain in BCI . . . 9
1.7. Scientific Hypotheses . . . 10
1.8. Outline of This Thesis . . . 11
1.9. List of Author Contributions . . . 11

2. Fundamentals 15
2.1. Human Spatial Hearing . . . 15
2.2. Neurophysiology . . . 17
2.3. Notations and Basic Definitions . . . 24
2.4. Basic BCI Pipeline . . . 25
2.5. Methods . . . 28
2.6. Metrics . . . 30

3. AMUSE Proof of Principle 35
3.1. Prior Work . . . 35
3.2. AMUSE Principles . . . 36
3.3. Experimental Validation . . . 37
Experiment 1: Behavioral Responses . . . 39
Experiment 2: Neurophysiological Responses . . . 41
Experiment 3: AMUSE in a Realistic BCI Setting . . . 45
Experiment 4: Single Direction Control Condition . . . 48
3.4. Discussion . . . 50

4. Writing With AMUSE 55
4.1. Prior Work . . . 55
4.2. Spelling Interface Principles . . . 56
4.3. Writing With AMUSE . . . 57
Experiment 5: A Fully Auditory Speller . . . 57
4.4. Discussion . . . 62

5. AMUSE End-user Tests 67
5.1. Prior Work . . . 67

5.2. AMUSE End-user Validation . . . 68
Experiment 6: End-user Validation of Online Writing . . . 68
Experiment 7: Single Subject Case-Study . . . 74
5.3. Discussion . . . 82

6. Data-Driven Performance Optimization **87**
6.1. Prior Work . . . 88
6.2. A Dynamic Stopping Framework . . . 89
6.3. Rank diff - A Dynamic Stopping Method . . . 89
6.4. Online Validation . . . 90
Experiment 8: Online Validation of Rank Diff . . . 90
6.5. Benchmark Comparison . . . 95
Experiment 9: Artificial Data . . . 97
Experiment 10: BCI Data Analyses . . . 99
Experiment 11: Approximating the Hyperparameters . . . 102
6.6. Discussion . . . 104

7. Summary and Conclusions **109**

References **111**

List of Figures **138**

List of Tables **140**

List of Terms and Abbreviations **143**

A. Appendix **151**
A.1. Technical Details for Dynamic Stopping Methods . . . 151
A.2. Additional Results Experiments 9 and 10 . . . 154

1. Introduction

Since the first report on the concept of a brain-computer interface (BCI), now 40 years ago, much progress has been made on all fronts. This thesis builds on several discoveries made in those 40 years, and before, which are the topic of this chapter. First, the basic principles of a BCI and applications thereof are introduced. Since the main goal of this thesis is to advance use of BCI as an assistive technology (AT), the potential end-users of such a technology are then defined. Given their clinical profile, the current state of the art in BCI does not provide a useful AT for at least a subset of those end-users. With the focus on electroencephalography (EEG) for brain imaging and the event-related potential (ERP) response, the current state of the art in BCI explicitly addressing the needs of this subset is discussed. This thesis aims to contribute to the understanding of these needs and the ways of addressing them. Before finally stating the scientific hypotheses of this thesis, the typical BCI design process is introduced.

1.1. Introduction to Brain-Computer Interfaces

The term BCI was first introduced by Vidal in 1973. In his landmark work, Vidal wondered if the small ERPs that result from external stimuli can be

> "put to work as carriers of information in man-computer communication or for the purpose of controlling such external apparatus as prosthetic devices or spaceships?" (Vidal, 1973)

By then, ERP components had been investigated extensively (see Section 2.2.2). However, the idea of using these in an online fashion, meaning that the brain signals should be processed and interpreted as soon as they are recorded, was striking, not in the least because of the lack of computing power in those days (see Figure 1.1). Nowadays, a BCI conveniently runs on a single computer or smartphone, but the goal of current BCI research is still exactly that: to create a direct connection between the brain and a device, bypassing the brain's natural efferent pathways altogether (Dornhege et al., 2007; Wolpaw et al., 2002). Useful application of such a connection relies on the assumption that a person can willfully alter his brain state to express an intention. This indeed is the case (Pfurtscheller et al., 1999; Sutton et al., 1965), thus allowing for direct "*communication and control*" through thought (Wolpaw et al., 2002).

Converting a user expressed intention into a meaningful control signal that drives an application requires several processing steps, roughly categorized into signal acquisition, preprocessing, feature extraction, classification, evidence accumulation, and stimulation / feedback (Blankertz et al., 2008; van Gerven et al., 2009; Mason et al., 2007; Schreuder, 2008b; Tomioka et al., 2010; Wolpaw et al., 2002). Each of these steps have been the topic of active research over the past 40 years, thereby steadily increasing the achievable accuracy level and usability. As a BCI is typically a linear system, its quality is only as high as its weakest element. Improvements at each step are thus necessary to improve the system as a whole.

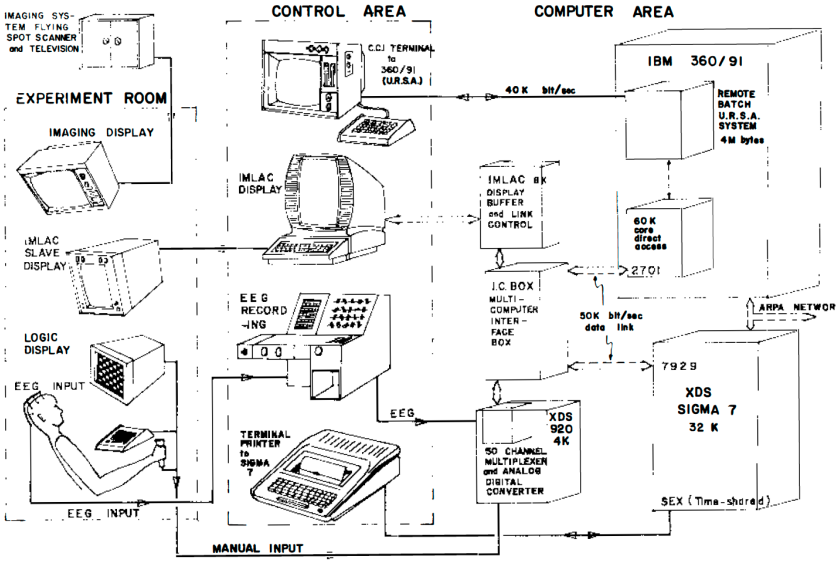


Figure 1.1.: Vidal's original approach to BCI, as proposed in Vidal (1973). The limitations of the hardware at his time are apparent from this single image, and underline the visionary nature of his proposition for BCI . Image reprinted from Vidal (1973) with permission.

1.2. Applications of BCI

Since Vidal first introduced BCI, an array of applications has been investigated. The single most prominent application that has fueled BCI research is that of an AT for the paralyzed (Birbaumer et al., 1999; Millán et al., 2010). The thought of allowing the completely paralyzed another chance at communication is compelling. With the BCI technology maturing, several other areas of application have been added. Here, they are grouped into three categories, focussing exclusively on human application.

1.2.1. BCI as an Assistive Technology

The term assistive technology (AT) refers to "*any item, piece of equipment or product system whether acquired commercially off the shelf, modified, or customized that is used to increase, maintain or improve functional capabilities of individuals with disabilities*" (Cook et al., 2008, and references therein). AT is thus used to regain some of the lost functions such as activities of daily living, sensation, communication, mobility, or manipulation. A recent survey including 77 AT users, and potential BCI end-users, evaluated the types of AT they used and the areas where they wished to increase their independence (Zickler et al., 2009). Not surprisingly, research into BCI as an AT overlaps with these user needs.

The majority of users stated they wanted more independence in mobility. Mobility has been addressed by BCI through control over a wheelchair (Galán et al., 2008; Mandel et al., 2009), or virtually, by remote controlling telepresence devices (Cincotti et al., 2008b; Escolano et al., 2012; Escolano et al., 2010; LaFleur et al., 2013; Tonin et al., 2011). Users were least satisfied with their current AT for manipulation. BCI research on manipulation, often referred to as motor restoration or motor substitution, has been particularly successful in invasive animal studies (Carmena et al., 2003; Nicolelis, 2003). Nevertheless, progress has been made for human application too,

controlling a prosthetic device or a paralyzed limb through BCI (Birbaumer et al., 2007; Collinger et al., 2013; Hochberg et al., 2006; Mattia et al., 2013b; Pfurtscheller et al., 2003; Rohm et al., 2013).

Five users stated that entertainment is amongst the most important contributions of AT, something that is only recently being addressed in the field of BCI (Blankertz et al., 2010; Holz et al., 2013a; Münßinger et al., 2010; Tangermann et al., 2009). 17 users ranked access to the internet as important; several BCI controlled browsers have been proposed in literature (Bensch et al., 2007; Karim et al., 2006; Mugler et al., 2010; Yu et al., 2012). Additionally, many studies have investigated the general BCI control over a computer cursor (Blankertz et al., 2003; Hill et al., 2006; Kübler et al., 2005; Piccione et al., 2006; Silvoni et al., 2009; Wolpaw et al., 2004). Providing such direct control over a cursor potentially gives access to many different AT applications. Another type of AT application that has been investigated is domotic control (Aloise et al., 2011; Cincotti et al., 2008b; Karmali et al., 2000), allowing end-users independent living.

However, the single most important research direction in BCI as an AT, and the one central to this thesis, is BCI aided communication (Birbaumer et al., 1999; Blankertz et al., 2006; Combaz et al., 2013; Donchin et al., 2000; Farwell et al., 1988; Kübler et al., 2009; Neuper et al., 2003; Nijboer et al., 2008b; Sellers et al., 2006a; Sellers et al., 2010; Townsend et al., 2010). Communication is the basis for several items in the AT survey (for instance "*expression of thoughts*" and "*participate in social life*") which collectively score high in areas where the AT users desired improvement. Users thus express a real need for AT- or BCI-aided communication.

1.2.2. Clinical Applications of BCI

The concept of willful generation of a specific mental state, or brain pattern, has found applications outside of the area of AT. Recently, several studies have used BCI technology to detect consciousness in vegetative state patients (Chatelle et al., 2012; Cruse et al., 2011; Lulé et al., 2013; Müller-Putz et al., 2012; Owen et al., 2006; Pokorny et al., 2013), showing that a minority of vegetative state patients retains at least some level of cognitive awareness. Furthermore, it has been used to aid in proper rehabilitation (Daly et al., 2008; Dobkin, 2007), in particular during stroke recovery (Ang et al., 2010; Broetz et al., 2010; Mattia et al., 2013a; Mattia et al., 2013b; Silvoni et al., 2011). Using BCI technology, proper brain activity can be retrained on the affected motor cortex during rehabilitation.

1.2.3. Non-Medical Applications of BCI

Non-medical uses of BCI technology have recently emerged; see Blankertz et al. (2010) for a review. Amongst a large array of imaginable applications are mental state monitoring (Haufe et al., 2011; Kohlmorgen et al., 2007; Müller et al., 2008; Schultze-Kraft et al., 2013; Zander et al., 2011a), gaming (Finke et al., 2009; Krepki et al., 2007; van de Laar et al., 2010; Lalor et al., 2005; Nijholt et al., 2009) and image clustering (Gerson et al., 2006; Hild ll et al., 2010). However, the limited bandwidth of current BCIs has thus far prevented it from becoming a primary communication channel for healthy users.

1.3. The End-User of BCI Assistive Technology

BCIs hold the promise of allowing those with (near-)complete paralysis another chance at communication or (environmental) control. Such paralysis is called locked-in syndrome (LIS) if all but oculomotor functions are compromised. If this last voluntary control is lost too, the condition is called completely locked-in syndrome (CLIS) (Bauer et al., 1979). For both conditions, a BCI

may be the only form of independent expression. But even for patients with less severe paralysis, residual motor functions may fatigue quickly, and they may be augmented by BCI.

So far, BCI research has primarily targeted end-users suffering from late stage amyotrophic lateral sclerosis (ALS) (Birbaumer et al., 1999; Marchetti et al., 2013; Nijboer et al., 2010; Nijboer et al., 2008b; Riccio et al., 2013; Sellers et al., 2006a; Townsend et al., 2010), a neurodegenerative disease with short life-time expectancy after diagnosis. The median survival time after tracheostomy, the time were a BCI is most relevant, has been reported at 21 months (Vianello et al., 2011). For ALS, the course of progression varies widely between patients, but the symptoms are rather similar. In the final state, the patient loses all reliable, voluntary muscle control (Caroscio et al., 1987). It was long assumed that cognitive functions remain largely untouched, but several recent findings have challenged this concept (Phukan et al., 2012; Ringholz et al., 2005; Sarro et al., 2011). They state that a range of cognitive changes are associated with ALS, although this is difficult to assess systematically for patients in the locked-in state. Nevertheless, the symptoms largely overlap with ALS patients.

Interestingly, the largest cause of LIS is brainstem damage caused by stroke or physical injury such as traumatic brain injury (TBI) (Laureys et al., 2005). There is a large population of people suffering from the consequences of stroke or TBI, for which a BCI may be desirable (León-Carrión et al., 2002). The yearly mortality rate for TBI patients that survive the first six post-trauma months is estimated to be only 1.3 times higher than that of the general population (Brown et al., 2004). For first-time stroke patients that survive the first 30 post-incident days, the mortality rate is estimated at about 2.3 times that of the general population (Dennis et al., 1993). It is highest for immobile patients, due to an increase in secondary causes of death, such as circulatory problems. Still the life expectancy is in the order of decades once the chronic phase has been entered. In the light of this it seems odd that end-users with LIS or CLIS due to stroke or TBI have thus far only played a minor role in BCI research.

1.4. EEG-Based Brain-Computer Interfaces

Various neuroimaging techniques have been explored in BCI research. Impressive results are obtained with invasive techniques such as intracortical recordings (Carmena et al., 2003; Collinger et al., 2013; Nicolelis, 2003) and electrocorticoography (ECoG) (Leuthardt et al., 2004; Schalk et al., 2008). They feature high spatial- and temporal resolutions due to their proximity to the brain, but application is cumbersome as a surgical procedure is required. Non-invasive methods are more straight-forward in their application. Though both functional magnetic resonance imaging (fMRI) and magnetoencephalograpy (MEG) have been investigated for BCI application (Buch et al., 2008; Mellinger et al., 2007; Weiskopf et al., 2004), they are unpractical for end-user applications due to their large size and high cost. Newer imaging methods such as functional near-infrared spectroscopy (NIRS) have great end-user potential (Coyle et al., 2004; Fazli et al., 2012; Sitaram et al., 2007), but have thus far played a minor role.

Therefore, the single most widely investigated neuroimaging method in BCI research is electroencephalography (EEG) (Mason et al., 2007). It measures electrical potentials, generated in the brain, on the surface of the scalp (see Section 2.2.1). EEG has a high temporal resolution, the equipment is small and it is cost efficient. The major drawback of EEG is the relatively poor spatial resolution; focal brain activation can be picked up by distant EEG sensors. As this limitation is intrinsic to the EEG signal, it has to be dealt with methodologically (Blankertz et al., 2008; Cincotti et al., 2008a; Haufe, 2011; Sannelli, 2012).

Despite these shortcomings, EEG-based BCIs have come a long way towards practical applicability. In fact, most of the studies mentioned in Section 1.2 are EEG-based and have started to have a real impact on the end-users' life (Sellers et al., 2010; Vaughan et al., 2006). For healthy users,

EEG-based BCIs were reported with communication speeds > 100 bits per minute (bits/min) (Bin et al., 2011; Tonet et al., 2008) and > 10 characters per minute (char/min) (Volosyak, 2011), clearly demonstrating the potential of EEG as an imaging technology for BCI. Recent advances in EEG hardware towards dry electrodes and wireless data transmission further increase the potential of practical BCI applications based on EEG (Grozea et al., 2011; Popescu et al., 2007; Ruffini et al., 2007; Zander et al., 2011b).

The work presented in this thesis is based on EEG and the discussion is therefore focussed on EEG-based BCI.

1.4.1. ERP-Based BCI

BCI paradigms can roughly be divided by their dependence on exogenous stimulation. Highly investigated paradigms, independent of such stimulation, are the motor imagery (MI) paradigm and paradigms based on other imagination tasks (Blankertz et al., 2007; Cruse et al., 2011; Kübler et al., 2005; Müller-Putz et al., 2005; Wolpaw et al., 2004). They rely on the reduction in synchronicity of an active brain area, called event-related desynchronization (ERD) (Pfurtscheller et al., 1999). MI-based systems provide an asynchronous, continuous output signal that is optimal for applications such as motor control (Pfurtscheller et al., 2003; Rohm et al., 2013) and object manipulation (Galán et al., 2008; LaFleur et al., 2013; Wolpaw et al., 2004). Unfortunately, they often require elaborate training and are suboptimal for discrete tasks.

Communication, more concretely text writing, is one such tasks that allows for discretization. Even though ERD has been successfully used to drive spellers (Blankertz et al., 2006; Müller et al., 2006; Wills et al., 2006), ERP-based BCIs seem a more natural choice. They rely on exogenous stimulation to elicit ERP responses in the subject's brain that carry information on the subject's intention (see Section 2.2.2). Instead of a continuous output signal, a class decision is made after a series of such stimuli, thereby delivering a discrete command signal.

The most widely investigated ERP-based BCI application is the visual matrix speller. It was first introduced by Farwell et al. (1988), and popularized ERP-based BCI. In their classic spelling application, the letters of the alphabet were displayed in a grid on the screen. Rows and columns were (pseudo) randomly highlighted by simply increasing the brightness for a few hundred milliseconds. The subject's task was to attend to, and often gaze at, the symbol he wants to select and count the number of times the attended symbol is highlighted. Each row and column were intensified several times for one selection and the results were integrated. Though this repetition reduces the number of selections that can be made per minute, it increases the accuracy (thus exposing a trade-off that is intrinsic to ERP-based BCIs). In fact, it was estimated that just short of 90 % of healthy subjects can control this interface with an accuracy > 80 % (Guger et al., 2009).

The basic concept behind the visual matrix speller has been adapted numerous times, changing the interface (Cincotti et al., 2008b; Kaufmann et al., 2011; Kaufmann et al., 2013b; Münßinger et al., 2010; Treder et al., 2011), stimulation pattern (Bießmann, 2007; Tangermann et al., 2011; Townsend et al., 2010), or stimulation modality (Belitski et al., 2011; Furdea et al., 2009; Klobassa et al., 2009; Kübler et al., 2009). Nevertheless, the focus has been on visual applications and mostly employ overt attention; they allow the subject to gaze directly at the item they want to select.

Both rhythmic activity of the brain and ERP were already discussed in the context of BCI in Vidal's landmark paper (Vidal, 1973), but they were later popularized by other groups (Farwell et al., 1988; Pfurtscheller et al., 1999). Though more paradigms have been investigated, they are mostly derived from these two classes of paradigms.

Table 1.1.: State of the art in covert attention, ERP-based BCI. Caption continued on opposite page.

	Class	Sel. Acc.	ITR	Symb/min	Online	End-user
Visual						
Kelly et al. (2005)	2	79.5	4.60			
Acqualagna et al. (2010)	30	90.0	3.96†	0.80†		
Treder et al. (2010)	6	60.0^{+}	1.99†	0.09†		
Liu et al. (2011)	36	96.3	5.38†	1.00$^{\diamond}$	×	
Treder et al. (2011)	6	97.1^{+}	9.15†	1.88†	×	
Aloise et al. (2012)	36	77.8	5.60†	1.75$^{\diamond}$	×	
Schaeff et al. (2012)	6	96.2^{+}	8.98†	1.85	×	
Acqualagna et al. (2013)	30	93.6	7.00	1.43	×	
Marchetti et al. (2013)	4	71.4	5.11		×	×
Tactile						
Müller-Putz et al. (2006)	2	66.4	0.42†		×	
Zhang et al. (2007)	2	63.0	0.30†			
Brouwer et al. (2010)	6	70.0$^{\diamond}$	3.71		×	
van der Waal et al. (2012)	6	79.0	4.52†	0.97†	×	
Kaufmann et al. (2013a)	4	50.0	0.26†		×	×
Auditory						
Hill et al. (2004)	2	82.3	1.71†			
Sellers et al. (2006a)	4	65.9	0.19†			×
Farquhar et al. (2008)	2	64.3	0.72†			
Kanoh et al. (2008)	2	65.0	0.79†			
Furdea et al. (2009)	25	65.0^{+}	1.54	0.16	×	
Guo et al. (2009)	8	89.1	2.80†			
Kübler et al. (2009)	5	37.4	0.08†		×	×
Klobassa et al. (2009)	36	59.4^{+}	1.25†	0.11†	×	
Halder et al. (2010)	2	76.5	1.70			
Gao et al. (2011)‡	5	80.0	2.06†			
Höhne et al. (2011b)‡	9	89.4	4.23	0.89	×	
Kim et al. (2011b)	2	71.4	0.25†		×	
Hill et al. (2012b)	2	84.8	1.93†		×	
Hill et al. (2012a)	2	79.9	0.87†		×	
Lopez-Gordo et al. (2012)	2	69.0	1.49			
Cai et al. (2013)‡	8	70.5	4.40†			
Käthner et al. (2013)‡	25	66.0^{+}	2.76	0.37†	×	
Nambu et al. (2013)‡	6	89.0	0.95†			
de Vos et al. (2014)	2	71.0	1.07			

Table 1.1.: Caption continued from opposite page. Values with a † are not given in the original article and are estimated from available data. If not explicitly stated, an overhead of five seconds was assumed. Values with a ⋄ are estimated from figures. Studies marked with a ‡ include more than two spatial locations as a key feature in their paradigm. Selection accuracies (Sel. acc.) marked with a + indicate two-level spellers where the accuracy was given on the character level; selection accuracy is typically higher. Only one result per publication is given, and preference is given to end-user and online results. Results are calculated for the full number of iterations, ignoring offline optimization by picking an optimal number of iterations. Studies where neither the accuracy, the information transfer rate (ITR), nor the symbols/minute were given or could be calculated are omitted. For multiple publications of the same study only the first appearance is included. Bold printed studies were published prior to the first contribution of this thesis (Schreuder et al., 2009).

1.5. Covert Attention-Based BCI

The ERP-based BCIs mentioned so far are mostly based on the visual modality and overt attention. They are by far the most investigated type of BCI and have been successfully applied in the clinical setting. However, several recent studies have shown that they may not cover the full range of end-user needs. Two basic studies showed that the performance on the visual matrix speller breaks down severely for users that can not exert their overt attention, i.e. that can not direct their eye-gaze (Treder et al., 2010; Brunner et al., 2010). The same was shown in a single end-user case-study (Ramos Murguialday et al., 2011). The findings of a large-scale survey, held under 61 end-users suffering from ALS further indicated that at least 26 % of the end-users showed vision impairment (Huggins et al., 2011). This warrants the need for alternative BCIs, based on covert visual attention or on a different modality altogether.

Table 1.1 gives an overview of the current state of the art in covert attention, ERP-based BCI in terms of their selection accuracy (see Section 2.6.2), bitrate (see Section 2.6.4) and, if applicable, writing speed (see Section 2.6.5). As several entries were not provided by the original authors, they were estimated from other available data or figures. It therefore serves as a rough indication of the current state of the art only. The studies printed in bold represent the state of the art prior to the first contribution from this thesis (Schreuder et al., 2009), and studies marked with ‡ use more than two spatial locations in their paradigm. The majority of studies in Table 1.1 is reviewed in a recent paper by Riccio et al. (2012a) (see also Treder (2012) and Brunner et al. (2013)).

Table 1.1 exposes several trends in research on covert attention-based BCIs. First and foremost, of all studies, only four were tested with end-users. The fact that those studies reported relatively low performances is a clear indication of the necessity to design BCIs for, and test them with those end-users. Second, the performance in general is much lower than those reported for overt attention-based BCIs (see Section 1.4.1). Third, covert visual paradigms typically score higher than their tactile and auditory counterparts. Fourth, many auditory BCI studies are fundamental in nature, and have not been applied online. And last, relatively little work is done on covert attention-based BCI prior to the first contribution from this thesis (Schreuder et al., 2009).

Visual paradigms that rely on covert attention are typically achieved by placing all items around a central fixation point, each with an equally small distance to that point (Aloise et al., 2012; Aloise et al., 2013; Liu et al., 2011; Marchetti et al., 2013; Treder et al., 2010; Treder et al., 2011). That way, all stimuli are represented equally on the fovea. Though this successfully negates the need to foveate the target, it does require the user to maintain focus on a single point. This requirement was partially lifted by the presentation of all items on the same, central position on the screen

(Acqualagna et al., 2011; Acqualagna et al., 2013; Treder et al., 2011), thereby not biasing any class in particular when drifts of eye gaze occur. Alternative visual paradigms use other stimulation types such as steady-state visually evoked potentials (SSVEP) (Kelly et al., 2005; Zhang et al., 2010) and motion visual evoked potential (MVEP) (Schaeff et al., 2012).

Other researchers have moved away from the visual domain altogether, using tactile or auditory stimulation. Although artificial conditions could be conceived where overt attention is used for these modalities, they typically are inherently covert in nature. Several studies proposed tactile stimulation for eliciting ERPs (Brouwer et al., 2010; Kaufmann et al., 2013a; Ortner et al., 2013; Thurlings et al., 2012; van der Waal et al., 2012), steady-state somatosensory evoked potentials (SSSEP) (Müller-Putz et al., 2006; Zhang et al., 2007) and as feedback to other tasks (Chatterjee et al., 2007; Cincotti et al., 2007). Though they have received relatively few attention, a recent case study showed that a tactile paradigm actually outperformed visual and auditory paradigms offline for a single locked-in patient (Kaufmann et al., 2013a).

1.5.1. Auditory BCI

Another modality for covert BCIs, and the one central to this thesis, is the auditory modality. In visual spellers, the stimulus and the class (character) are equivalent; the mapping from stimulus to class is explicit. Auditory BCI, as well as tactile BCI, often have the problem that this mapping is only implicit; a character is represented by a tone, environmental sound or tactor location. In the strict case it is up to the user to keep track of this mapping, which could increase the mental demand (Riccio et al., 2012a).

Several studies used the auditory modality purely as feedback on an different task (Hinterberger et al., 2004; Nijboer et al., 2008a; Pham et al., 2005; Rutkowski et al., 2006). They are not in the focus of this thesis. Those studies that do use the auditory domain for stimulation can be categorized as streaming or sequential (Hill et al., 2012b), depending on how the stimuli are presented. In streaming paradigms, the stimuli for all classes are streamed in parallel, whereas stimuli for the different classes are presented one after the other in sequential paradigms. As a rule of thumb, ERP-based BCIs are sequential and BCIs based onsteady-state auditory evoked potentials (SSAEP) are streaming.

Streaming Auditory BCI The first auditory BCI was reported by Hill et al. and used the streaming principle (Hill et al., 2004; Hill et al., 2005). Two sequences of target and non-target stimuli were presented, one to each ear. They could successfully classify to which of the two ears the subjects were paying attention, resulting in a binary BCI. It was recently validated online (Hill et al., 2012b) and a similar principle was used to drive a continuous BCI application (Hill et al., 2012a). A comparable approach was taken by Kanoh et al. (2008). However, both streams were presented to the same ear, relying on the human capacity for stream segregation to elicit discriminative responses. Though these studies employed simple tonal stimulation, more complex stimuli such as natural speech have also been tested in this context (Lopez-Gordo et al., 2012).

Other streaming paradigms rely on the SSAEP, also referred to as auditory steady-state response (ASSR) (Desain et al., 2006; Farquhar et al., 2008; Hill et al., 2012b; Kim et al., 2011a; Kim et al., 2011b; Lopez et al., 2008). The concept of SSAEP relies on the fact that if a subject is presented with a continuous auditory tone, the frequency of that tone is enhanced in the spectrum of the EEG (similar to SSVEP and SSSEP). By presenting several such tones in parallel, the user can choose to direct his attention to one of them, thereby enhancing that frequency's trace in the EEG more in comparison to the remaining stimuli. The ability to classify this effect is fundamental to SSSEP-based BCIs.

Though Kanoh et al. (2008) suggest that human stream segregation can deal with more than two streams, to date no streaming auditory paradigm was published with more than two classes.

In the context of communication, the resulting binary BCI is potentially too slow, as a multitude of decisions is needed to effectively communicate a single character.

Sequential Auditory BCI Most of the work in auditory BCI has focused on sequential paradigms. They are a variation of the typical oddball paradigm (see Section 2.2.2.1). Two or more auditory stimuli differ on some defining property, allowing the user to pick a target and focus on appearances of that target. These stimuli typically consist of tones that differ in pitch or loudness (Halder et al., 2010; Höhne et al., 2010b; Höhne et al., 2011b; de Vos et al., 2014). The simplest sequential paradigm that is BCI applicable is the two-target oddball paradigm (see Section 2.2.2.1), leading to a binary BCI (Halder et al., 2010; de Vos et al., 2014). Binary sequential auditory BCI are in fact used in the clinical setting to probe the inaccessible minds of vegetative state patients, rather than for full communication purposes (Lopez-Gordo et al., 2011; Pokorny et al., 2013). However, the true benefit of sequential auditory BCI lies in the fact that they are readily extended to a multiclass setting by adding targets; effectively, all stimuli are then presented equally often and each one can be selected as the target. Using this principle, the visual matrix speller (Farwell et al., 1988) was transferred into the auditory domain (Furdea et al., 2009; Käthner et al., 2013; Klobassa et al., 2009; Kübler et al., 2009). Other studies made use of the same principle using a different spelling interface (Höhne et al., 2011b), or without any direct application (Cai et al., 2013; Gao et al., 2011; Guo et al., 2010; Guo et al., 2009; Nambu et al., 2013). Apart from simple tones, other defining features have been tested for the purpose of facilitating recognizability and reducing mental workload. The introduction of one such features is the central theme of this thesis.

1.6. The Typcial Design Chain in BCI

BCI research is a highly interdisciplinary field, with strong influences from engineering. Nevertheless, until recently the design of a BCI has not followed any formalized design principles. Still, there appears to be a consensus amongst researchers on a typical cascade of design steps. Depending on the target population, two or three steps can be distinguished as follows.

Step 1: Offline Healthy Subject Tests - A BCI is typically defined as a closed loop system, providing the subject with immediate feedback. Nevertheless, it is often possible to show a specific effect or the feasibility of a particular paradigm in principle, without providing such immediate feedback. It is often easier to implement and the experimental parameters can be better controlled. Such an experiment is referred to as offline, because any analysis of the data typically is done after the recording is finished.

Step 2: Online Healthy Subject Test - Though offline validation can in principle show a particular effect, the loop was not closed and the subject did not receive feedback. Any psychological effect of such feedback is thus not accounted for. For this reason, and for concerns over inproperly executed offline data analysis, a large number of BCI researchers consider an online validation the ultimate test for BCI technology. Though proper offline data analysis is of course possible, it can not account for the subject's behavior in a closed loop system.

Step 3: Online End-user Tests - The last step is testing the BCI with the envisioned target population. As a fast majority of BCIs is proposed as an AT for severely paralyzed end-users, this means that those end-users should be included in the testing protocol. Irrespective of the clinical diagnosis, end-users should be as close to the state in which they would benefit from a BCI as is practically possible. Due to the progressive nature of the disease, end-users that are in the early phases of ALS may not represent those at a later stage particularly well. Similarly, a fully able stroke or TBI patient is unlikely to use a BCI for communication, making him an inappropriate subject for end-user tests.

Step one and two could in principal be omitted, instead performing the tests with the end-users directly. However, as the time with such end-users is often limited, step one and two are used for prototyping a paradigm. Design decisions made at each step, including step three, can and should be re-evaluated to optimize the performance. The design chain is therefore recurrent; often a step is repeated several times to validated the adjusted design decisions.

Recently, the concept of user-centered design (formalized in ISO 9241-210, 2010) has entered the field of BCI (Friedrich et al., 2013; Holz et al., 2013b; Schreuder et al., 2013b; Zickler et al., 2013). Amongst other things, it states that a technology should be designed with a deep understanding of the end-user, end-users should be involved in each step of the design process, the design should be driven and refined by end-user evaluation, and the process should be iterative. This does not mean that the introduced design chain becomes obsolete; in fact it already facilitates the iterative process and the end-user can be involved in many ways, both quantitatively and qualitatively. However, the BCI specific knowledge that is required for a deep understanding of any end-user is largely missing. Step three of the above chain should therefore not only be used to test a particular system, but also to generate this much needed information.

1.7. Scientific Hypotheses

The current state of the art in BCI, as described above, is the result of 40 years of research and is starting to impact the lives of end-users. The most investigated overt attention-based, visual ERP BCIs work well for healthy subjects and a subset of end-users. Nevertheless, they exclude a particular subset of end-users that is unable to employ their overt visual attention. Thus, there is a real need for alternatives that allow inclusion of these end-users.

Reports on such covert attention-based BCIs are few and far between. They are typically binary in nature and provide a communication channel with very low capacity. Increasing the number of classes can increase the efficiency of a communication channel, but it was thus far not straightforward without relying on overt attention. Moreover, end-user tests are missing almost entirely. This thesis proposes, and carefully validates, two key contributions that potentially increase the usefulness of covert attention-based BCIs and ERP-based BCIs in general.

The first contribution is a new auditory BCI paradigm called AMUSE (short for Auditory MUlti-class Spatial Event-related potential). It builds on the ability of human spatial hearing, to allow for a truly multiclass, auditory BCI. The hypotheses for this contribution are as follows

i *subjects can focus on one amongst several spatial stimuli,*

ii *spatial stimuli alone are sufficient to elicit binary-class discriminative ERPs,*

iii *with spatial stimuli, a covert attention-based, auditory multiclass BCI can be realized,*

iv *spatial stimuli allow for useful BCI performance in a strictly auditory setting for healthy subjects and end-users.*

The second contribution of this thesis addresses the trade-off between speed and accuracy that is intrinsic to any ERP-based BCI (both overt and covert). A new method is proposed for dynamically optimizing this trade-off (called rank diff) and it is validated online. Furthermore, it is compared to several methods previously proposed in literature. Not only does this comparison put rank diff in perspective, but it provides evidence for the value of such methods in general. The hypotheses for this contribution are as follows

v *properly dealing with the speed / accuracy trade-off increases performance significantly,*

vi *smart methods outperform simply picking an "optimal" number of stimuli.*

These six hypotheses form the basis of this thesis and are carefully validated in several experiments.

1.8. Outline of This Thesis

Following this introduction, the thesis is structured as follows. Chaper 2 first introduces the fundamental concepts of spatial hearing and neurophysiology. It then continues by introducing some notation and the BCI processing chain, along with several methods and metrics that are used throughout the rest of the chapters. In Chapter 3, the fundamental concepts behind AMUSE are carefully validated, both behaviorally and neurophysiologically, addressing hypotheses *i* and *ii*. By measuring several conditions, the boundaries of the AMUSE paradigm are investigated, and the unique contribution of the spatial feature is determined. All experiments in this chapter are offline; subjects did not get direct feedback on their actions.

After having confirmed the usefulness of the spatial stimuli offline, AMUSE is validated online with healthy subjects in a spelling application in Chapter 4. It addresses hypothesis *iii* and hypothesis *iv* for healthy subjects. As the majority of healthy subjects is successful in communicating with the AMUSE-based speller, the next step is to test the AMUSE-based speller with end-users in Chapter 5. Five end-users are tested over several sessions, addressing the second part of hypothesis *iv*. The results are put into context through a single-subject comparison with an overt attention-based, visual BCI study.

The second contribution and hypotheses *v* and *vi* are addressed in Chapter 6. First, using a similar setup as in Chapter 4, the rank diff method is validated online, showing a substantial performance gain. Consecutively, it is tested in an offline comparison, along with several methods proposed in literature, using artificial and pre-recorded EEG data. Results clearly show the benefit of dynamic methods on the BCI performance.

Finally, the results are summarized and the impact of the work presented in this thesis is addressed in Chapter 7.

1.9. List of Author Contributions

Parts of the work that is described in this thesis was published previously in peer-reviewed journals and conferences. They are given below, split into main- and additional contributions. The number of citations to the respective articles was taken from Google Scholar on January 21st, 2014. If one or more publications were fundamental to a chapter, it is stated at the start of the chapter.

Main contributions

Schreuder, M., M. Tangermann, and B. Blankertz (2009). "Initial Results of a High-Speed Spatial Auditory BCI". In: *International Journal of Bioelectromagnetism* 11.2, pp. 105–109.

Schreuder, M., B. Blankertz, and M. Tangermann (2010a). "A New Auditory Multi-Class Brain-Computer Interface Paradigm: Spatial Hearing as an Informative Cue". In: *PLoS ONE* 5.4. Ed. by J. Yan, e9813.
DOI: 10.1371/journal.pone.0009813

Schreuder, M., J. Höhne, M. S. Treder, B. Blankertz, and M. Tangermann (2011a). "Performance Optimization of ERP-Based BCIs Using Dynamic Stopping". In: *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Boston, USA: IEEE EMBS, pp. 4580–4583.
DOI: 10.1109/IEMBS.2011.6091134

Schreuder, M., T. Rost, and M. Tangermann (2011c). "Listen, You are Writing! Speeding Up Online Spelling with a Dynamic Auditory BCI". In: *Frontiers in Neuroscience* 5.112.
DOI: 10.3389/fnins.2011.00112

Schreuder, M., J. Höhne, B. Blankertz, S. Haufe, T. Dickhaus, and M. Tangermann (2013a). "Optimizing Event-Related Potential Based Brain-Computer Interfaces: a Systematic Evaluation of Dynamic Stopping Methods". In: *Journal of Neural Engineering* 10.3, p. 36025.
DOI: 10.1088/1741-2560/10/3/036025

Schreuder, M., A. Riccio, M. Risetti, S. Dähne, A. Ramsay, J. Williamson, D. Mattia, and M. Tangermann (2013b). "User-Centered Design in Brain-Computer Interfaces – A Case Study". In: *Artificial Intelligence in Medicine* 59.2, pp. 71–80.
DOI: 10.1016/j.artmed.2013.07.005

Additional relevant contributions

Dähne, S., J. Höhne, M. Schreuder, and M. Tangermann (2011a). "Band Power Features Correlate With Performance In Auditory Brain-Computer Interface". In: *Frontiers in Human Neuroscience - XI International Conference on Cognitive Neuroscience (ICON XI)*. Vol. 108. 109. Palma de Mallorca, Spain: Frontiers Research Foundation.
DOI: 10.3389/conf.fnhum.2011.207.00109

Höhne, J., M. Schreuder, B. Blankertz, K.-R. Müller, and M. Tangermann (2010a). "Novel Paradigms for Auditory P300 Spellers with Spatial Hearing: Two Online Studies". In: *Frontiers in Computational Neuroscience - Bernstein Conference on Computational Neuroscience*. 44. Berlin, Germany: Frontiers Research Foundation.
DOI: 10.3389/conf.fncom.2010.51.00044

– (2011a). "Novel Paradigms for Auditory P300 Spellers with Spatial Hearing: Two Online Studies". In: *Proceedings of the 2nd TOBI Workshop / International Journal of Bioelectromagnetism*. Vol. 13. 2, pp. 96–97.

Höhne, J., M. Schreuder, B. Blankertz, and M. Tangermann (2010b). "Two-Dimensional Auditory P300 Speller with Predictive Text System". In: *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Buenos Aires, Argentina: IEEE EMBS, pp. 4185–4188.
DOI: 10.1109/IEMBS.2010.5627379

– (2011b). "A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System". In: *Frontiers in Neuroscience* 5.99.
DOI: 10.3389/fnins.2011.00099

Höhne, J., M. Schreuder, and M. Tangermann (2011c). "Auditory ERP Speller Applications as a Tool for BCI End-Users". In: *Frontiers in Human Neuroscience - XI International Conference on Cognitive Neuroscience (ICON XI)*. Vol. 108. Palma de Mallorca, Spain: Frontiers Research Foundation.
DOI: 10.3389/conf.fnhum.2011.207.00108

Kindermans, P.-J., M. Schreuder, B. Schrauwen, K.-R. Müller, and M. Tangermann (2014a). "True Zero-Training Brain-Computer Interfacing – An Online Study". In: *PLoS ONE* 9.7, e102504.
DOI: 10.1371/journal.pone.0102504

Kleih, S. C., A. Riccio, D. Mattia, M. Schreuder, M. Tangermann, C. Zickler, and A. Kübler (2011b). "Motivation Affects Performance in a P300-Brain-Computer Interface". In: *Proceedings of the 2nd TOBI Workshop / International Journal of Bioelectromagnetism*. Vol. 13. 1. Rome, Italy, pp. 46–47.

Quek, M., D. Boland, J. Williamson, R. Murray-Smith, M. Tavella, S. Perdikis, M. Schreuder, and M. Tangermann (2011). "Simulating the Feel of Brain-Computer Interfaces for Design, Development and Social Interaction". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. New York City: ACM Press, pp. 25–28.
DOI: 10.1145/1978942.1978947

Riccio, A., M. Schreuder, S. Dähne, A. Ramsay, M. Quek, A. Crossan, J. Höhne, S. C. Kleih, A. Kübler, R. Murray-Smith, M. Tangermann, and D. Mattia (2012b). "Bridging the Gap between BCI Technology Design and Severely Disabled End-Users: Experience from a Photo Browser Validation". In: *Proceedings of the 3rd TOBI Workshop*. Ed. by S. Kleih, T. Kaufmann, B. Hörning, and A. Kübler. Würzburg, Germany: Universität Würzburg, pp. 105–106.

Rost, T., M. Schreuder, and M. Tangermann (2010). "Optimized Spatial Auditory Stimuli for an ERP-Based BCI Paradigm". In: *Proceedings of the 1st TOBI Workshop*. Graz, Austria: Technische Universität Graz, p. 45.

Schreuder, M. (2008a). *Classification of the P300 Response to Auditory Cues Differing in Spatial Location*. Master Report, Vrije Universiteit Amsterdam, the Netherlands.

Schreuder, M., A. Riccio, F. Cincotti, M. Risetti, B. Blankertz, M. Tangermann, and D. Mattia (2011b). "Putting AMUSE to Work: an End-User Study". In: *Proceedings of the 2nd TOBI Workshop / International Journal of Bioelectromagnetism*. Vol. 13. 3, pp. 139–140.

Schreuder, M., A. Riccio, A. Ramsay, S. Dähne, J. Höhne, M. Quek, A. Crossan, D. Mattia, R. Murray-Smith, and M. Tangermann (2012a). "End User Performance in a Novel Social BCI Application: The Photobrowser". In: *Proceedings of the 3rd TOBI Workshop*. Ed. by S. Kleih, T. Kaufmann, B. Hörning, and A. Kübler. Würzburg, Germany: Universität Würzburg, pp. 16–17.

Schreuder, M. and M. Tangermann (2010b). "Online Spelling Using the Brand New Spatial Auditory P300 Paradigm". In: *Proceedings of the 1st TOBI Workshop*. Graz, Austria: Technische Universität Graz, p. 22.

– (2010c). "Online Spelling Using the New Spatial Auditory BCI". In: *Proceedings of the 4th International BCI Meeting*. Monterey, USA.

Schreuder, M., M. E. Thurlings, A.-M. Brouwer, J. B. F. van Erp, and M. Tangermann (2012b). "Exploring the Use of Tactile Feedback in an ERP-Based Auditory BCI". In: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. San Diego, USA: IEEE EMBS, pp. 6707–6710.
DOI: 10.1109/EMBC.2012.6347533

Tangermann, M., J. Höhne, and M. Schreuder (2013a). "Fixed-Sequence Stimulus Presentation in ERP-BCI". In: *Proceedings of the 4th TOBI Workshop*. Sion, Switzerland: École Polytechnique Fédéderal de Lausanne, pp. 141–142.

Tangermann, M., J. Höhne, M. Schreuder, M. Sagebaum, B. Blankertz, A. Ramsay, and R. Murray-Smith (2011a). "Data Driven Neuroergonomic Optimization of BCI Stimuli". In: *Proceedings of the 5th International BCI Conference, Graz*. Graz, Austria: Verlag der Technischen Universität Graz, pp. 160–163.

Tangermann, M., J. Höhne, H. Stecher, and M. Schreuder (2012). "No Surprise — Fixed Sequence Event-Related Potentials for Brain-Computer Interfaces". In: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. San Diego, USA: IEEE EMBS, pp. 2501–2504.
DOI: 10.1109/EMBC.2012.6346472

Tangermann, M., P.-J. Kindermans, M. Schreuder, B. Schrauwen, and K.-R. Müller (2013b). "Zero Training for BCI – Reality for BCI Systems Based on Event-Related Potentials". In: *Biomedical Engineering / Biomedizinische Technik*. Vol. 58. S1-1. Graz, Austria: De Gruyter.
DOI: 10.1515/bmt-2013-4439

Tangermann, M. and M. Schreuder (2010). "Optimized Stimulus Design for a Spatial Auditory BCI Based on ERP". In: *Proceedings of the 4th International BCI Meeting*. Monterey, USA.

Tangermann, M., M. Schreuder, S. Dähne, J. Höhne, S. Regler, A. Ramsay, M. Quek, J. Williamson, and R. Murray-Smith (2011b). "Optimized Stimulation Events for a Visual ERP BCI". In: *Proceedings of the 2nd TOBI Workshop / International Journal of Bioelectromagnetism*. Vol. 13. 3, pp. 119–120.

Williamson, J., A. Ramsay, M. Quek, M. Tangermann, M. Schreuder, C. Vidaurre, and R. Murray-Smith (2010). “A P300-Based Image Browser with Automatic Stimulation Optimization”. In: *Proceedings of the 1st TOBI Workshop*. Graz, Austria: Technische Universität Graz, p. 21.

2. Fundamentals

This chapter introduces some of the fundamental principles that are used throughout this thesis. First, two concepts that are essential to this thesis are explained in more detail, starting with the human ability of spatial hearing and continuing with the fundamentals of (human) neurophysiology. The latter is highly focussed on EEG as a brain imaging tool and includes an introduction of the classical oddball paradigm.

After these conceptual descriptions, the concrete tools for interpreting the remaining chapters are provided. First, a set of notations and terminology is introduced that is used throughout this thesis. Subsequently, details are given on the typical BCI processing pipeline that forms the basis for all experiments. In an attempt to reduce redundancy, only changes to this archetypical setup are described in the methods section of each individual experiment. At the end of this chapter, several methods and metrics are introduced that are used throughout the rest of this thesis. At such points where these are used, the reader is referred back to this chapter for details.

2.1. Human Spatial Hearing

The heavy bias in BCI research towards the visual modality in BCI research (see Section 1.4.1) is understandable, as humans are highly visual creatures. Our 'nominal' visual resolution, also called acuity, is about 1' (or 1/60th of a degree) (Westheimer, 1965). This translates to distinguishing two lines that are roughly 1.75 mm apart, at 6 meters distance; a remarkable capability. In comparison, the minimum audible angle (MAA) is about two orders of magnitude worse, with the lowest recognizable differences being $>1°$ (Blauert, 1996). This spatial resolution is found only under optimal conditions and for displacements around the frontal direction, where the spatial resolution is highest. Towards the lateral positions, this resolution worsens when tested behaviorally (Makous et al., 1990; Mills, 1958).

Nevertheless, localization of sounds in space is one of the processes that our brain does continuously and with minimal mental effort, and it has been the topic of extensive research (for in-depth reviews, see Middlebrooks et al. (1991) or Blauert (1996)). Furthermore, it was shown that the spatial resolution might be improved if the user is cued about the location. When subjects focus on a particular direction, their attentional resources appear to be distributed in a gradient around that direction, with decreasing alertness when moving away from the attended direction (Mondor et al., 1995; Teder-Sälejärvi et al., 1998). Furthermore, a recent EEG study shows that the brain might register smaller differences for lateral positions, even when this does not show behaviorally (Bennemann et al., 2013).

The location of a sound source in the spatial field around a listener (origin) is defined by three dimensions (see Figure 2.1). The horizontal plane runs through both ear canal entrances and the lower part of the eye sockets. The location on this plane is defined by the azimuth angle away from the frontal direction, with the far right being defined as $90°$ azimuth. The vertical axis is called elevation, which is measured by the angle between the source and the horizontal plane. The third factor is the distance from the observer. In this thesis, all sources are on the horizontal plane (zero elevation) and have equal distance within an experiment. They can thus be uniquely described by their azimuth angle.

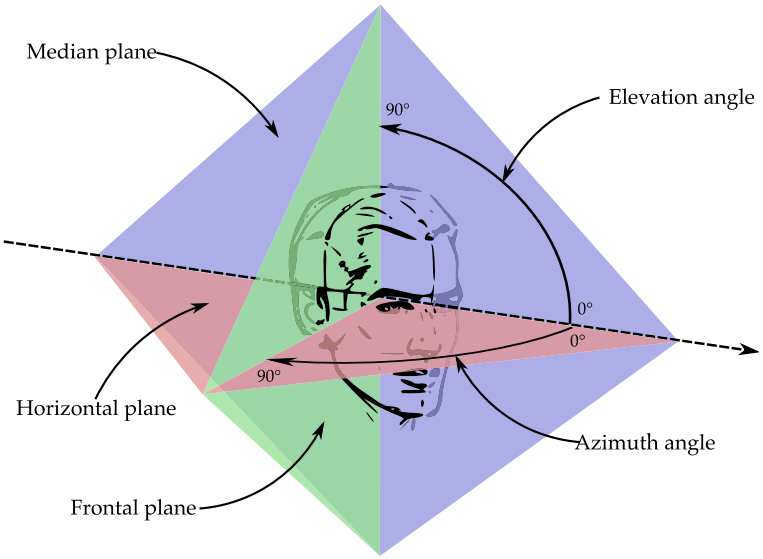


Figure 2.1.: Sound source locations are divided by three planes, crossing halfway between the listener's ear. Any source can be defined by its azimuth angle on the horizontal plane, its elevation angle off the horizontal plane and its distance from the listener.

The localization of a source on the horizontal plane depends on bi- and monaural cues. Binaural cues are typically most salient for lateralization of a source (left vs. right). The interaural timing differences (ITD) cue is based on the fact that the distance between a source and the ear is different for both ears, for all sources that are off the median plane. This difference causes a phase-shift in the signal perceived at each ear, which can be exploited mainly for low-frequent sounds (up to 1-1.3 KHz). At the same time, for sources that are off the median plane, either one of the ears is (partially) shadowed by the head, causing a reduced intensity of the sound, called interaural level differences (ILD). The ILD effect provides the localization information for higher frequencies, roughly >3 KHz. This is often referred to as the "duplex theory" of sound localization and was introduced as early as 1907 by Lord Rayleigh (Rayleigh, 1907).

Sources that have similar angular displacement around -90 ° or 90 ° azimuth (the frontal plane) provide ambiguous binaural cues. For instance, sources at 80 ° and 100 ° azimuth have similar ITD and ILD; the paths to both ears are each equal in length and the shadowing of the head is similar. In fact, this ambiguity extends for positive- and negative elevations on a circle trough these points and parallel to the median plane. The effect of this so-called "cone of confusion" (Woodworth, 1938) is particularly strong for positions close to the frontal plane (Carlile et al., 1997; Makous et al., 1990). Sources on the cone of confusion are resolved in different ways. First, using slight head movements, the path lengths are changed and the ITD and ILD effects become unambiguous (Thurlow et al., 1967; Wallach, 1940; Wightman et al., 1999). But even without such head movements, broadband sound sources on the cone of confusion that contain high frequency components can be disambiguated by differences in the reflections on the outer ear (monaural cues) for different source positions (Batteau, 1967; Hofman et al., 2003). Particularly important for resolving front-back ambiguities are the components in the high frequency range (HFR) above 8 kHz (Langendijk et al., 2002).

Sounds with a complex spectrum containing high frequent noise are typically better localized (Roffler et al., 1968), in particular when they have a sharp onset, causing attack transients (Hartmann, 1983; Rakerd et al., 2005). The latter was shown to trigger "the precedence effect" (Wallach

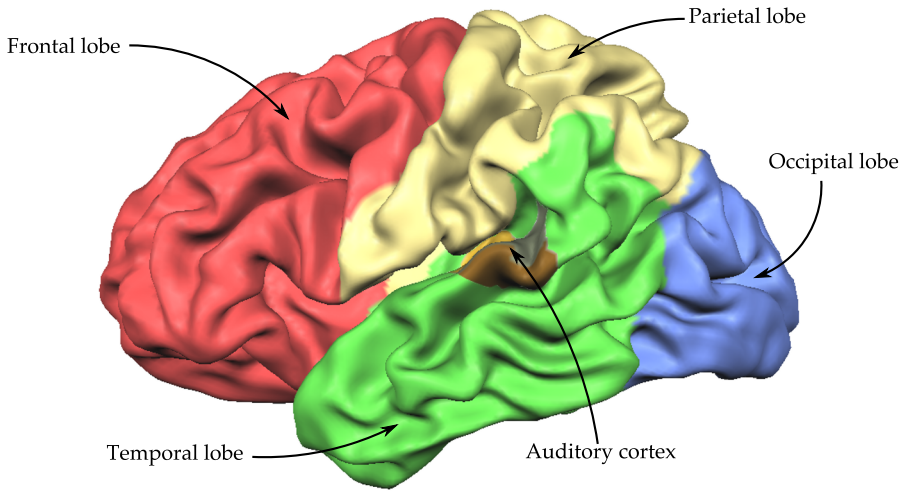


Figure 2.2.: View on the left hemisphere of a model of the human brain. Indicated are the rough division into four lobes and the cortical areas involved in primary auditory processing. Image generated using BrainVoyager Brain Tutor (available at http://www.brainvoyager.com/).

et al., 1949) for localization in reverberating environments.

In psychophysics, the goal is to understand the relationship between a physical stimulus and the resulting perceptions and sensations. For human spatial hearing, the paradigms are traditionally designed to find the limits of the localization capacity, typically tested under artificially optimal conditions such as anechoic chambers. For determining the MMA, two sounds are played concurrently or in close temporal proximity and the subject is asked to judge if their sources overlap or not (Mills, 1958; Perrott, 1984; Perrott et al., 1990). By varying a source around a reference source, the MMA is determined for the reference source. In other paradigms, sounds are played from varying sources and the subject is requested to indicate the perceived location of the source. The angle between the true- and reported source is the measure of the spatial resolution of human sound localization (Makous et al., 1990; Oldfield et al., 1984).

2.2. Neurophysiology

Neurophysiology is the study of the functioning of the brain and the nervous system at large. The brain is a vastly complex organ that is involved in all but the most automatic tasks and emotions. Here, only the fundamentals of neurophysiology are introduced that are necessary for understanding EEG recordings. Due to the textbook nature of this section and Section 2.2.1, the reader is referred to Kandel et al. (2012) or ten Donkelaar et al. (2011) for details and individual references.

The nervous system can roughly be divided into seven parts, including the spinal cord, the cerebellum and the brain stem. Here, the focus is on the cortical layer of the cerebrum, or simply cortex, as the EEG almost exclusively reflects signals originating here. The cerebrum has two hemispheres, and its cortical sheet is anatomically divided into four lobes (see Figure 2.2). These four lobes also indicate a first rough functional division. For instance, the occipital lobe is almost

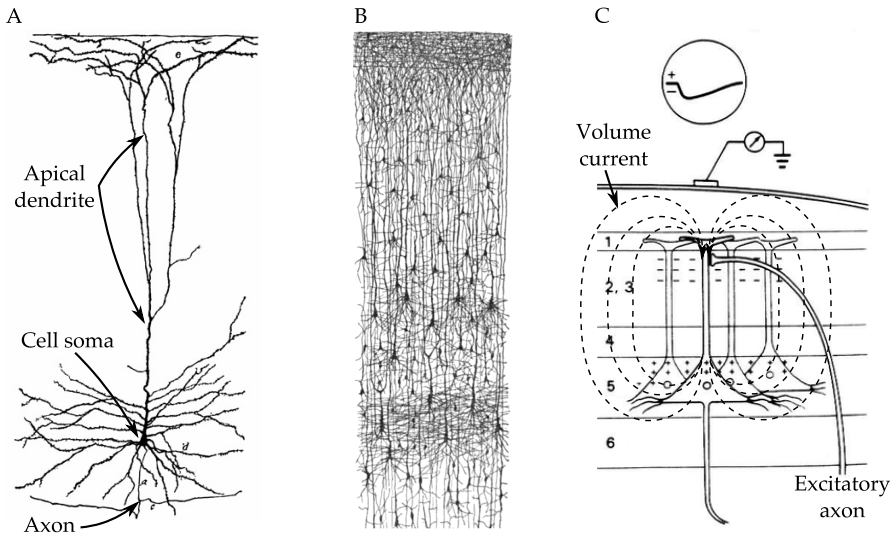


Figure 2.3.: Fundamentals underlying the generation of the EEG signal. A and B are early drawings of Santiago Ramón y Cajal, showing an individual pyramidal neuron (A) and the laminar organization of the cerebral cortex (B) roughly on the same scale. The apical dendrites of the pyramidal cells are oriented perpendicular to the scalp (top), explaining their contribution to EEG. C shows a schematical representation of the origins of EEG resulting from an excitatory PSP. The pyramidal cell body lies in layer V of the cortex and the apical dendrites extend into layer I. C is adapted from http://quizlet.com/13012152.

exclusively involved in the processing of visual input. The remaining three lobes have more diverse functions, often coinciding with differences in cytoarchitectual organization of the cortex as reported by Brodmann (1909). The cortical areas responsible for the processing of auditory events are located on the superior part of the temporal lobe (see Figure 2.2). The cortex forms the outer layer of the cerebrum, and is closest to the skull. It is highly wrinkled, thereby increasing its surface.

The primary functional units of the brain are neurons (see Figure 2.3-A). Neurons are electrically charged cells that maintain a resting potential of roughly -70 mV by actively pumping charged ions across their cell membrane. Both the intra- and extracellular environment contain high amounts of charged ions in different relative concentrations. The cell membrane, which is otherwise impermeable to such ions, functions as a capacitor with ions of opposite polarity lining up along the cell membrane due to their electrical attraction. A large portion of the neurons in the human brain are located in the cortex in a laminar fashion (see Figure 2.3-B).

Neurons are interconnected through synapses, both inhibitory and excitatory. Communication at the synapse is most commonly chemical; the presynaptic cell releases neurotransmitters into the synaptic cleft, an area with close proximity to the postsynaptic cell. These neurotransmitters activate ion channels in the cell membrane of the postsynaptic cell, allowing the exchange of charged ions and causing a small, transient change in polarization of the postsynaptic cell called postsynaptic potential (PSP). The PSP travels away from the synapse in the cell's interior. A single PSP lasts between 10 and 100 ms. Importantly, PSPs can be inhibitory (causing hyperpolarization of the postsynaptic cell) and excitatory (causing depolarization of the postsynaptic cell) and they are spatially and temporally additive. Through these connections, the neurons form large-scale networks, both local and distant.

Neurons are roughly divided into three functional segments (see Figure 2.3-A). The dendrites serve as the neurons antennas for incoming signals; the majority of synapses terminates on them. All dendrites ultimately lead to the cell soma, or cell body, were all incoming PSPs are integrated. Several excitatory PSPs in close temporal proximity increasingly depolarize the cell body until the threshold potential of approximately -40 to -55 mV is exceeded and an action potential (AP) is fired. The AP rapidly travels through the axon by means of active propagation and lasts roughly 1 ms. The axon terminals again form synapses, releasing neurotransmitters upon an incoming AP and causing a PSP in the postsynaptic cell.

2.2.1. The Origins of Electroencephalography

The electrical nature of the brain was already investigated and described by Canton (1875). Using electrodes at the surface of rabbit- and monkey brains, he showed the presence of voltage deflections, sparking a series of animal experiments. The recording of electroencephalography (EEG) relies on these electrical potentials, but measures them at the surface of the scalp. The first human EEG recording was described in a landmark paper by Berger (1929).

The generation of the EEG signal can be summarized as follows (Baillet et al., 2001; Gloor, 1985; Wolters et al., 2007). The high impermeability of the neuron's cell membrane to charged ions divides the brain into the intracellular- and extracellular compartments. Arrival of an excitatory AP at a distant synapse on the apical dendrite results in a local influx of positively charged ions into the cell. As a result, the cell's interior is locally less polarized than the cell body, causing a primary current within the cell. By the same reasoning, the extracellular compartment locally becomes more negatively charged, as positive ions are extracted. The primary current is not measurable with EEG, but it causes a secondary, volume current. Because of the primary current, the capacitive attraction over the cell membrane along the dendrite temporarily becomes less strong, thereby freeing extracellular positively charged ions. The existence of both a negatively charged (source) and a positively charged (sink) area in a conducting volume causes the volume current. Though some of this volume current takes the shortest path between source and sink, it also spreads throughout the entire conducting volume and onto the scalp (see Figure 2.3-C).

The volume current resulting from a single PSP is tiny and can probably not be measured at the surface of the scalp. However, as can be seen in Figure 2.3-B, the cortex is neatly organized. Many neurons, in particular pyramidal neurons of layer V, have an apical dendrite that is oriented perpendicularly to the scalp and extends up to layer I. By the principle of co-activation, many of these neurons receive input roughly at the same time, and their volume currents add up spatially. Similarly, and due to the relatively long time course of the PSP, PSPs from consecutive APs terminating on the same cell will add up temporarily. The EEG reflects the concerted activation of several thousands to millions of neurons.

The combined activation of a patch of cortex is often summarized as a single dipole. The electrical field on the scalp resulting from a single dipole has a positive and a negative peak. For parts of the cortex that are on the surface and thus parallel to the surface of the scalp, the volume current projects directly outwards to the scalp, causing one of the poles to be focal and directly over the generator. On the other hand, cortical areas that are in a gyrus project to the scalp at an angle, thus resulting in peak potential fields that are more diffuse and away from the generator. Furthermore, the EEG always measures a linear mixture of many active sources. Therefore, a particular scalp activation pattern (see Figure 2.4) can not simply be interpreted as originating directly underneath the area of peak activation and inverse methods are required to make claims about the origin of a signal (Haufe, 2011).

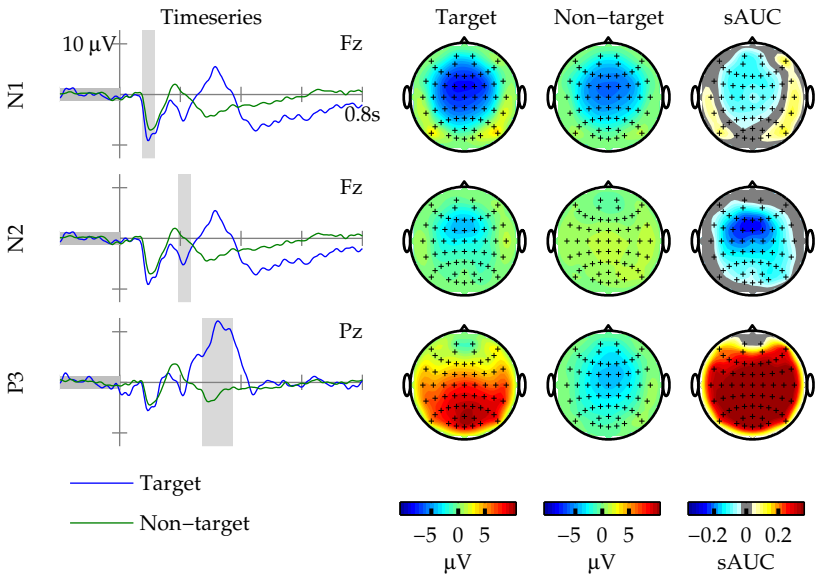


Figure 2.4.: Basic ERP components of the EEG. Data are taken from a single subject from Experiment 5. The first column displays the temporal features of the single channel indicated in the top-right corner. The second and third columns show the spatial potential distribution for target (second column) and non-target (third column) stimuli, averaged over the shaded interval in the first column. Intervals are selected according to Section 2.5.1. The fourth column shows the spatial distribution of binary-class discriminative information for the same interval, as expressed in sAUC (see Section 2.6.3). Columns two to four are plotted in top-down view, with the nose showing upwards. In this particular plot, the rows represent different components. However, in the following chapters they may also represent individual subjects, grand averages or a mixture of both.

Practically, EEG is measured as the differential signal between two electrodes and thus bipolar. However, EEG is often recorded unipolar, by recording all channels against a single (pair of) reference electrode(s). For reproducibility, recording electrodes are placed on the scalp according to a fixed scheme, for which the extended 10-20 international system is commonly used. Traditional recording equipment requires a conductive paste to establish an electrical connection between the electrodes and the scalp, but recently dry electrode systems without the need for gel have been proposed (Grozea et al., 2011; Popescu et al., 2007; Ruffini et al., 2007; Zander et al., 2011b). EEG equipment is cheap in comparison to other imaging techniques. Because the volume current is instantaneously, EEG can be recorded with high temporal resolution.

2.2.2. Event-Related Potentials

The term event-related potential (ERP) refers to a complex of deflections in the EEG following an event. Many different endogenous or exogenous events can elicit an ERP response, for instance a tone, visual cue, or motor action. ERPs are both time- and phase-locked to the event, meaning that each component of the complex has a latency and polarity that is approximately the same for repetitions of the event. Because of this, it is common practice to increase the signal-to-noise ratio (SNR) of ERPs by simply averaging over the responses to multiple events, thereby canceling out event unrelated background EEG.

The different components of the ERP complex are typically named according to their polarity (P for positive, N for negative) and their latency to the event onset in milliseconds (P300 or N200) or ordinal number in de ERP complex (P3 or N2). Each component is characterized by its spatial pattern and temporal dynamics (amplitude, polarity and latency), and each component is typically associated with a different step in the processing of the event.

Under special recording conditions, several components can be measured already in the first eight milliseconds after the onset of a sharp auditory click (Picton et al., 1974a). These very early brainstem auditory evoked responses (BAER) are generated in the first relays of the auditory pathway in the brainstem and are often used to asses the functioning of the auditory pathways (Starr et al., 1975). They are substrates of the physical processing of the stimulus features, and are mainly modulated by those features, and not attention. The corresponding components in the visual modality are referred to as visually evoked potentials (VEP).

Where the early components reflect the processing of physical properties of the stimulus, later components are increasingly cognitive in nature; they can be modulated by top-down processes such as attention (Picton et al., 1974b). This is of paramount importance for any ERP-based BCI, as they depend on the user being able to express his intention by focusing his attention to one of several stimuli.

Some major ERP components are described below. It is important to note that each of these components in fact has been divided into several subcomponents, depending on the experimental conditions, stimulus modality and task. Each component in itself could warrant a full review; here the most general understanding of these components is introduced.

Given the latency of the P3, an important component of the ERP complex, a stimulus onset asynchrony (SOA) of at least one second is normally recommended to properly determine the amplitude and latency of ERP components (Duncan et al., 2009). In the context of BCI the SOA is almost always much smaller, and components clearly overlap (see Box 2.3 on page 23).

Box 2.1: Conventions
It is customary in psychology literature to reverse the y-axis for ERP plots, displaying negative deflections upwards. As this convention is less common in BCI literature, positive deflections are plotted upwards throughout this thesis.
Similarly, in psychology the absolute difference between target and non-target traces are often taken, to emphasize the attention effect. Since absolute differences are less important for BCI than discriminative differences, the absolute difference is substituted for a metric that measures the difference, or separation, in terms of the variance on both traces. Examples of such metrics are the pointwise biserial correlation coefficient (Blankertz et al., 2011) and the area under the receiver-operator characteristic (ROC) curve (see Section 2.6.3). The temporal and spatial characteristics of discriminative differences may differ from those of the absolute differences.

2.2.2.1. The Oddball Paradigm

Any perceived event is likely to elicit an ERP of some sort, given that they represent the processing of such events. Experimentally, a particular paradigm has been used to specifically target individual steps of the processing pathway, called the oddball paradigm (introduced by Squires et al., 1975). The name refers to the fact that all versions of this paradigm introduce an infrequent stimulus, called the *"oddball"* or *"deviant"*. The simplest oddball paradigm presents a single stimulus at relatively long (> 1 second) intervals. The most typical version, normally referred to simply as the oddball paradigm, presents a stream of standard stimuli where infrequently, and randomly, a standard is swapped with a deviant. A third version, the three-stimulus oddball, adds an additional deviant to the stream. Through task instructions, one of these deviants is to be focussed, and the other is considered a distractor. Given such an active task, the deviant

is often referred to as target and the standard stimuli are then referred to as non-targets. The stimulus sequence is typically created at random to increase the surprise effect of the deviant. However, it was recently shown that this may not be strictly necessary for oddball paradigms with a greater number of classes (Tangermann et al., 2013; Tangermann et al., 2012). ERP components are typically present without the need for training other than initial introduction necessary to understand the task.

When the subject is instructed to passively perceive the stimuli, or even participate in a competing task, a subset of ERP components is elicited. If the subject is instructed to actively pay attention to the stimuli, for instance by counting the targets, an additional subset of components is elicited, in particular by the target stimulus. Even when the instructions are to count the frequent non-targets, it is the infrequent targets that elicit these additional components (Squires et al., 1975). In the case of the three-stimulus oddball, the distractor stimulus elicits yet another subset of ERP components.

Typically, the deviant and standard stimuli differ between them on some defining property, but the stimuli are identical within each class. The task relevant difference is thus on the physical level. However, it is also possible for the deviant and standard stimuli to differ on a higher cognitive level. The "deviantness" may be defined purely by context; they may for instance be infrequently presented male names in a stream of otherwise female names (Kutas et al., 1977). Though each stimulus is unique in its physical properties, they differ on the semantic level.

To completely eliminate the confound of stimulus property choice on the attention effect, the stimuli for deviants and standards are sometimes swapped on half of the trials.

Box 2.2: The "*P300-speller*" fallacy
The P3 component is often the largest and most discriminative of all components, leading to the historical, but somewhat misleading name "P300 BCI" for ERP-based BCIs. It is now accepted that other components contribute significantly to the classification task (Allison et al., 2003; Bianchi et al., 2010). For instance, in visual spellers such as the one described by Farwell et al. (1988), early components are very prominent and binary-class discriminative. A more appropriate name for these BCIs is thus simply "ERP-based BCI", a term that is used throughout this thesis.

2.2.2.2. P3

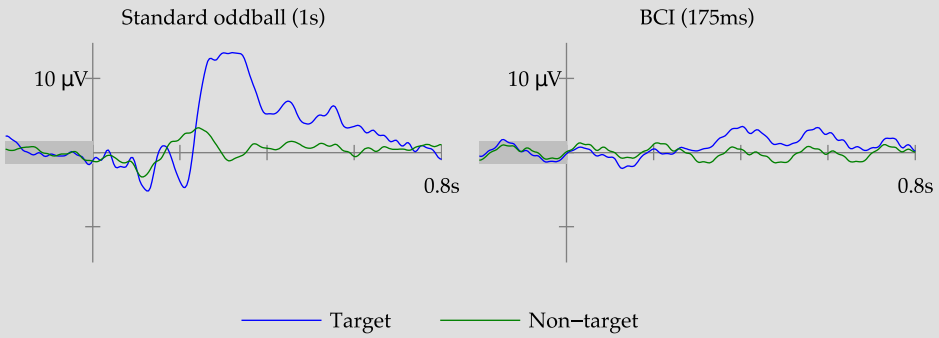
The P3 (or P300) is a positive going component that is measurable on the vertex (Conroy et al., 2007). It was first reported in 1965 by Sutton et al. (1965). The latency of the P3 response is roughly between 250 and 500 ms, and depends on "*stimulus evaluation timing, is sensitive to task processing demands, and varies with individual differences in cognitive capability*" (Polich, 2007). In fact, it depends on many experimental factors, such as stimulus intensity, duration and target-to-target interval (Allison et al., 2003; Gonsalvez et al., 2007; Picton, 1992; Polich, 1989). The P3 response is the most prominent peak of the ERP complex, and has sparked the investigation of ERP-based BCI (Farwell et al., 1988), to which it initially gave its name (see Box 2.2).

In psychophysiology literature, the P3 is typically subdivided in the P3a and P3b components, a distinction first introduced by Squires et al. (1975). They showed that both components were elicited by infrequent deviant stimuli. However, the frontal/central P3a was present after a deviant stimulus, regardless of any task relevance, as was the case in their "ignore" condition – the subject was reading a book whilst passively perceiving auditory stimuli. The P3a thus reflects an automatic response, that is not attention dependent. The second subcomponent, the parietal P3b, was only present when the deviants were relevant to the task at hand. Both when subjects were told to count the deviant and the standard tones did the deviants elicit a P3b (Squires et al., 1975). The P3b component is thus attention dependent. The P3b response is also consistently elicited by

semantical deviants, such as the male names in the study described earlier (Kutas et al., 1977). In BCI research, and in fact in psychophysiology in general, the P3b is typically meant when referring to the P3, a common practice that is used in this thesis too.

In early work, the ERP components were typically measured with three electrodes (*Fz*,*Cz* or vertex, and *Pz*). Given those channels, the peak amplitude of the P3b is typically largest over *Pz*, and smallest for the frontal *Fz* channel (Johnson, 1993; Squires et al., 1975). P3 latency typically increases from frontal channels (*Fz*) to parietal channels (*Pz*) (Mertens et al., 1997; Polich, 2007).

Box 2.3: Cranking up the Pace



BCI research differs from neurophysiological studies in its goals, in that it pursues the pragmatic vision of efficient brain-based control. Several adjustments are therefore made to the oddball paradigm, the traditional experimental paradigm of the neurophysiologists. The most prominent change is the decrease in SOA, to allow for the recording of many epochs in limited time. In the above figure, the left plot shows the ERP response to a standard oddball task as employed in classical neurophysiology. With one second SOA, all peaks are clearly distinguishable, allowing for analyses on peak amplitude and latency. The right figure represents the ERP response to a BCI oddball paradigm, taken from the same subject. At 175 ms, the SOA is almost six times shorter, allowing for six times the amount of epochs per second. Classical peak analyses is impossible, as the ERP components clearly overlap (but see Section 2.5.1). Furthermore, the target and non-target responses are less distinct. As this leads to less discriminative information per epoch, several repetitions are necessary in BCI to increase the SNR. The obvious trade-off between the number of repetitions and the SNR is the topic of Chapter 6.

Habituation The stability of the P3 response within a trial, and over trials, has been investigated extensively. Using an auditory oddball paradigm,Lammers et al. (1989) showed that the amplitude of the deviant P3 decreased significantly over blocks of stimulation. But even within each block, the P3 reduced as time passed. The P3 latency showed a similar effect, with an increase in latency as time passed. A somewhat different result from Murphy et al. (2004) showed that with short trials with a fixed number of iterations, the P3 response to the final targets increased again. This rebound effect was diminished when the number of iterations was unknown to the subject, in which case a negative linear trend was found, similar to that of Lammers et al. (1989).

Interestingly, several studies did not find such a habituation effect (Courchesne et al., 1978; Ravden et al., 1998; Woods et al., 1986). These contradicting results are often explained by differences in the experiment parameters. In particular, a longer SOA tends to cancel out the habituation effect (Lammers et al., 1989), as do large breaks in between blocks (Ravden et al., 1998). Both interrupt the internal task automation which would reduce the required attentional

resources and thereby reduce P3 amplitude. But also the stimulation modality matters. Simple visual tasks typically require more attentional resources than simple auditory tasks, and thereby suffer less from habituation (Posner et al., 1976).

In BCI research, both the SOA and the breaks between blocks tend to be small, thus introducing the risk of P3 habituation. As the P3 component is the primary contributor to the discrimination task, such systematic changes in the P3 response could impact the classification performance; discriminative features may loose strength, or worse, change their position or polarity over time. This known feature of ERPs was the reason for developing a new method for evidence accumulation (see Chapter 6).

2.2.2.3. N2

The negative going N2 (or N200) component, found around 200 ms after event onset, was first described in the same paper as the P3 (Sutton et al., 1965). It is elicited by an infrequent, unexpected stimulus and it is thought to reflect a decision process, or target classification (Ritter et al., 1979). The N2 can be subdivided into an automatic N2a, an attentional N2b and a classification N2c component, depending on the task at hand (Näätänen et al., 1978; Pritchard et al., 1991).

The central N2a subcomponent, more commonly known as mismatch negativity (MMN), is elicited by stimuli that are physically different from their predecessor, even when the stimuli are perceived passively. It "*is generated by the brain's automatic response to any change in auditory stimulation exceeding a certain limit roughly corresponding to the behavioural discrimination threshold*" (Näätänen et al., 2007). The central N2b subcomponent is elicited if the subject is engaged in a discrimination task. Non-target stimuli elicit a larger N2b than do target stimuli (Näätänen et al., 1982). It is typically followed by a P3a. Finally, the frontocentral auditory N2c subcomponent is elicited by target stimuli and associated with the processing thereof (Pritchard et al., 1991; Ritter et al., 1979). It is typically followed by a P3b, and it is of primary interest for BCI applications (Halder et al., 2013).

2.2.2.4. N1

Components as early as the central N1 (or N100), around 100 ms after event onset, have been found to be attention dependent (Hillyard et al., 1973). Amongst other things, they are believed to represent a gating process by which task relevant stimuli can be further processed, and non-relevant stimuli are suppressed (Coull, 1998). Reports of such a component date back to Davis (1939). The attention effect was explored in the context of BCI (Allison et al., 2003; Bianchi et al., 2010), concluding that though it is less strong than that of the P3, it is significant and can aid in the classification task. The amplitude and latency of the N1 are highly dependent on stimulus properties such as stimulation frequency and in particular intensity (Näätänen et al., 1987). Though Näätänen et al. (1987) describe a multitude of N1 subcomponents, they are highly overlapping and will be considered as a single complex, collectively named N1.

2.3. Notations and Basic Definitions

Throughout this thesis, scalars are denoted by italic upper- and lower-case letters, vectors by bold lower-case letters and matrices by bold upper-case letters. Unless stated otherwise, vectors are understood to be in columnar shape. The i^{th} element of vector $\boldsymbol{x}$ is denoted by x_i. The i^{th} row of matrix $\boldsymbol{M}$ is denoted as $\boldsymbol{m}_i$ and the j^{th} column of the same matrix as $\boldsymbol{m}_{,j}$. The matrix element at the i^{th} column and the j^{th} row is denoted as $m_{i,j}$. Matrices can be stacked in the third dimension for time or epoch, forming a tensor which can be indexed by a third index ($\boldsymbol{M}_{,,k}$).

Some variables are used consistently throughout this thesis. Their names are not introduced at every occurrence, but rather introduced in Box 2.4. Notations that are only relevant in individual cases are introduced in-situ. The usage of the word "*class*" is ambiguous, as it may refer to the binary target and non-target classes or the task level classes. Examples of the latter are the rows and columns in the visual matrix speller (Farwell et al., 1988). Throughout this thesis, *binary-class* refers to the first and simply *class* refers to the second use.

Box 2.4: Basic notations

C: Number of classes; c indexes classes

E: Number of epochs; e indexes epochs

J: Number of iterations; j indexes iterations

T: Number of trials; t indexes trials

$\boldsymbol{D} \in \mathbb{R}^{C \times J \times T}$: Tensor containing classifier outputs

$\boldsymbol{D}^{train}$: Tensor containing the training classifier outputs

$\boldsymbol{D}^{test}$: Tensor containing the test classifier outputs

$\boldsymbol{b} \in \mathbb{Z}^{E}$: Vector with the true binary-class labels

$\boldsymbol{l} \in \mathbb{Z}^{T}$: Vector with the true task level class labels

2.3.1. Experimental Terminology in ERP-based BCI

As in any field of research, BCI comes with its own jargon, derived from psychology and neurophysiology. However, in BCI it is often used inconsistently. Throughout this thesis, the following terminology will be used.

A study is the largest time window, and refers to a set of experiments sufficient to test a hypothesis. A study may consist of several sessions, for instance to show an effect that is robust over time or to test for a learning effect. A session refers to the time from putting on the cap until taking it off again. During a session, an entire experimental protocol is performed, covering all conditions. Typically, only one session is performed per subject and per day. A session consists of several runs for different experimental conditions. Runs of different conditions are often interleaved to prevent obvious time influences. In the case of online spelling, a run refers to the writing of a single word.

A run consists of several trials. Each trial is a time window of uninterrupted task engagement, preceded and followed by a short break. In BCI, a trial typically refers to performing a single selection. Each trial consists of several iterations. During an iteration, all classes are stimulated exactly once in a pseudo-random order, before proceeding to the next iteration. The number of iterations is often fixed beforehand, but can be variable (see Chapter 6). Finally, both a single stimulus event and the resulting EEG data are called an epoch. They form the smallest time entity in an ERP-based BCI.

2.4. Basic BCI Pipeline

This section describes the typical processing steps that are performed in an ERP-based BCI. The process of single-trial analysis of ERP data is aptly described in Blankertz et al. (2011), and the methods used here closely follow the methods proposed in that paper.

2.4.1. Preprocessing

The raw data is acquired and stored at 1 KHz using BrainAmp amplifiers (BrainProducts). After acquisition, the data are filtered between .33 Hz and 40 Hz with causal filters (Chebyshev low-pass filter, Butterworth high-pass filter). To cancel out any phase shifts resulting from the filtering, the filters are applied in both directions strictly for plotting purposes only. Reverse filtering is omitted for (offline) classification purposes, as it does not transfer to an online setting; the effect of the filter is no longer causal. After filtering, the data are typically downsampled to 100 Hz for efficiency, respecting the Nyquist rate. These preprocessed data are represented by $\boldsymbol{X} \in \mathbb{R}^{M \times S}$, where M is the number of recorded channels and S the number of recorded samples.

The data are then segmented for a pre-stimulus samples and b post-stimulus samples and channel-wise baselined on the pre-stimulus interval, according to

$$\boldsymbol{x}^{\dagger}_{c,,e} = \boldsymbol{x}_{c,m_e-a,\ldots,m_e+b} - \frac{1}{a}\sum_{i=1}^{a} x_{c,m_e-i} \tag{2.1}$$

where $\boldsymbol{m}$ is a vector containing the stimulus onset timestamps and $\boldsymbol{X}^{\dagger} \in \mathbb{R}^{M \times S \times E}$. The third dimension E on $\boldsymbol{X}^{\dagger}$ thus indexes over the epochs. For simplicity, the third dimension may not be mentioned explicitly, indicating the spatio-temporal features of a single trial. Then, $\boldsymbol{x}^{\dagger}_{i}$ refers to a row-vector with the temporal evolution of the i^{th} channel, and $\boldsymbol{x}^{\dagger}_{,j}$ refers to a column-vector with the spatial pattern over all channels at time j.

Additionally, mean subtraction and unit variance scaling are performed for each feature dimension separately. For cross-validation or online studies, the scaling factors are estimated from the calibration data and stored for application to the unseen data.

2.4.2. Artifact Rejection

Two simple artifact rejection methods are used. The first rejects epochs with a peak-to-peak difference greater than 100 μV over frontal channels. The second disregards epochs with high variance on more than 20 % of the channels. High variance is typically defined as three times the distance between the 10^{th} and the 90^{th} percentile away from the 10^{th} or 90^{th} percentile. Both are calculated on the raw data, and the union of artifactual epochs is removed afterwards.

For visual presentation, artifacts are removed from the entire dataset. During cross-validation and in sessions with online runs, artifacts were only removed from the calibration data. Artifact removal from online epochs is not straight-forward; omitting artifact rejection from the test data in cross-validation thus increases the generalizability of the results.

2.4.3. Feature Extraction

The time- and phaselocked properties of ERPs allow for rather simple spatio-temporal feature extraction (see Section 2.2.2). In fact, the data are often simply down-sampled further to around 20 Hz by averaging over regular intervals $\mathcal{T}$, and the remaining post-stimulus samples of each channel are concatenated into a single feature vector. In literature, further pruning of features is sometimes done by the sparse step-wise linear discriminant analysis (SWLDA) classifier (an extension of regular linear discriminant analysis (LDA), as described in Section 2.4.4), which determines a predefined number of features that contribute significantly to the classification task, whilst setting the rest to zero (Farwell et al., 1988).

Here, two slightly different methods are used for interval selection, as visualized in Figure 2.5. Though the intervals in $\mathcal{T}$ are typically regularly distributed over time, there is no particular reason why they should be. Instead, discriminative intervals can be determined beforehand,

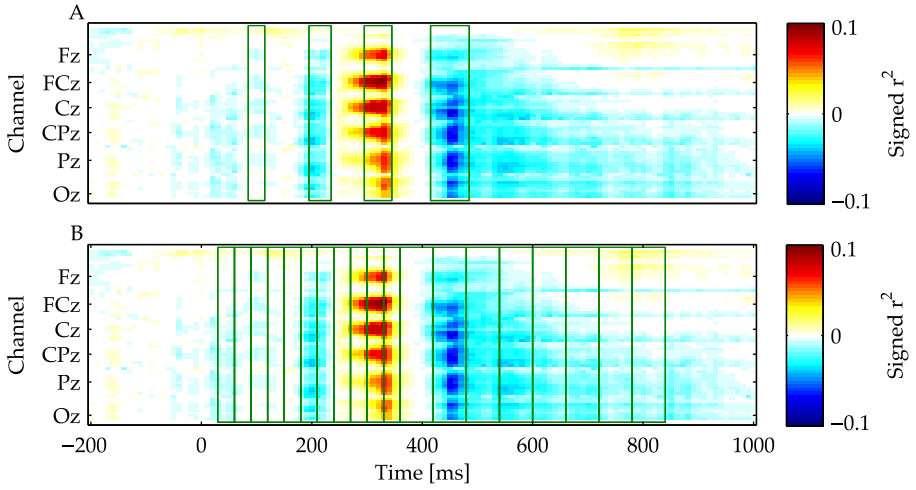


Figure 2.5.: Two alternative ways of subsampling the EEG data for feature extraction are visualized. A) a simple algorithm searches for the most discriminative intervals on the training data, following some constraints, and B) using prior knowledge, a denser sampling is taken for early components than for later components. Note that, although the intervals in B) are plotted on top of a single subject's discriminative data for visualization purposes, they are the same for all subjects.

either by a-priori knowledge or by means of an algorithm. The first selects the intervals using a-priori knowledge of the ERP time evolution, such that the faster early components are sampled with a higher resolution (33 Hz, between 30 and 350 ms) than the slower late components (17 Hz, between 360 and 800 ms), effectively resulting in 19 features per channel (see Figure 2.5-B). These intervals are subject independent.

The obtain an even lower number of features per channel, a subject dependent set of intervals can be pursued. For this, the signed pointwise biserial correlation coefficient (signed r^2) (Blankertz et al., 2011) is calculated for each dimension in $\boldsymbol{X}^{\dagger}$ over all epochs. The result is a matrix of size $\mathbb{R}^{M\times S}$, which can be depicted as in Figure 2.5-A (background image). The algorithm then searches for a number of stable peak signed r^2 values (positive and negative) as starting points. After this initialization, the algorithm widens the intervals over the time dimension. This widening is subject to some constraints, the most important of which is that the spatial patterns remain constant, i.e. the sign of the signed r^2 remains the same on most channel. In contrast to SWLDA, the resulting intervals are defined over all channels.

The result of any such method is a set of intervals $\mathcal{T} = [\mathcal{T}_1,\ldots,\mathcal{T}_n]$. The column vector containing the features for a single epoch is thus constructed as $\boldsymbol{v}_{,e} = \left[\boldsymbol{x}^{\dagger}_{,\mathcal{T}_1}{}^{\top},\ldots,\boldsymbol{x}^{\dagger}_{,\mathcal{T}_n}{}^{\top}\right]^{\top}$, where $\boldsymbol{x}^{\dagger}_{,\mathcal{T}_1}{}^{\top}$ contains the channel-wise averages over interval $\mathcal{T}_1$. Feature matrix $\boldsymbol{V} \in \mathbb{R}^{G\times E}$ is then defined as $\boldsymbol{V} = \left[\boldsymbol{v}_{,1},\ldots,\boldsymbol{v}_{,E}\right]^{\top}$, were G is the number of dimensions of a single feature vector $\boldsymbol{v}$.

2.4.4. Classification

Based on $\boldsymbol{V}$ and the true binary-class labels $\boldsymbol{b}$, a linear model can be trained that defines an optimal hyperplane for dividing the elements of $\boldsymbol{V}$ into targets and non-targets (Müller et al., 2008), according to

$$w = \hat{\Sigma}_c^{-1}\left(\hat{\mu}_2 - \hat{\mu}_1\right) \tag{2.2}$$

where w is the projection vector, μ_i is the mean of the i^{th} binary-class as calculated using the true labels in $\boldsymbol{b}$, and $\hat{\Sigma}$ is the pooled covariance matrix. Some form of regularization is typically advisable when the number of epochs is lower than, or on the same order as the number of dimensions in the feature vector (Tomioka et al., 2010). Here, $\hat{\Sigma}$ is regularized using analytical shrinkage to prevent under- and overestimation of the small- and large eigenvalues of the empirical covariance matrix (Bartz et al., 2013; Blankertz et al., 2011; Ledoit et al., 2004).

Using the trained classifier, a decision value can then be assigned to unseen samples according to

$$d_e = w^\top v_{,e} + b \tag{2.3}$$

where b is a constant bias and the binary-class membership is determined by sign(d_e). The separating hyperplane is defined by $w^\top v + b = 0$. Throughout this thesis, the classifier was trained to assign more negative scores to target epochs than to non-target epochs. Though more sophisticated discriminant functions exist (Müller et al., 2003), the assumptions of LDA appear to be satisfied for ERP data (Blankertz et al., 2011).

2.4.5. Evidence Accumulation

For further processing, the classifier scores are sorted according to their class membership and iteration, such that $d_{c,j,t}$ is the classifier score belonging to class c in iteration j of trial t. Thus, $\boldsymbol{d}_{c,,t}$ is the vector of all classifier scores within a trial for a single class. The result is the matrix $\boldsymbol{D} \in \mathbb{R}^{C \times J \times T}$. Alternatively, $\boldsymbol{D}$ may be indexed as $\boldsymbol{D}^{train}$ or $\boldsymbol{D}^{test}$, where $train$ and $test$ represent disjoint sets of epoch indices, as is used during cross-validation (see Section 2.5.3).

The winning class $c^w_{j,t}$ for trial t at iteration j is then defined as

$$c^w_{j,t} = \underset{c \in [1,\ldots,C]}{\arg\min} \left(\text{median} \left(\boldsymbol{d}_{c,1\ldots j,t} \right) \right) \tag{2.4}$$

finding the class that has the lowest median classifier score over all iterations of a trial. Though it is typical in BCI literature to calculate $c^w_{j,t}$ at a fixed number of iterations j, there are good reasons to dynamically determine the iteration at which to stop a trial (see Chapter 6).

2.5. Methods

Several of the methods and metrics that are used repeatedly in this thesis are introduced here. In the following chapters, the reader is referred back to this section when needed.

2.5.1. Peak Picking

ERP components are typically defined by their polarity, latency to event onset, amplitude and spatial distribution (see Section 2.2.2). Obtaining the latency and amplitude of a component is done using prior knowledge (Duncan et al., 2009). Given a predefined latency window for each component, the most positive- or negative voltage deflection is determined for a given channel. This peak defines the latency and amplitude of the component. For obtaining a robust estimation of the spatial distribution, several samples around the peak are averaged according to a preset window width. The latency windows, window width, polarities and channels that are used for finding the components of interest can be found in Table 2.1.

When the SOA is lower than approximately one second, components start to overlap (see Box 2.3 on page 23), and components can no longer be defined in the traditional way. In this case the peaks are found on the basis of the discriminative information (see Section 2.6.3). Using the

Table 2.1.: Settings used for finding the peak values of ERP components.

Component	Latency window	Polarity	Channel	Window width
P3	280 - 650 ms	+	*Pz*	100 ms
N2	210 - 350 ms	-	*Fz*	40 ms
N1	80 - 140 ms	-	*Fz*	40 ms

same latency window, polarity and channel, the peak in discriminative information is taken to be the "component". After obtaining the latency, the same procedure is followed for obtaining the spatial distribution. For BCI purposes, this definition is a pragmatic solution, as the discriminative information is of primary interest. However, they can not be interpreted as peaks in the traditional sense, as the underlying assumptions are different.

2.5.2. Questionnaires

To truly asses the usefulness of a BCI, objective data measures are only one part of the story. With a human in the loop who ultimately would operate the BCI for extended periods of time, the subjective user experience is at least as important (Holz et al., 2013b; Kübler et al., 2013; Pasqualotto et al., 2011). To asses this, and the overall cognitive state of the subjects, several questionnaires are administered during each session of all online experiments (healthy subjects and end-users).

Motivation and mood are scored at the start of each online session by means of a visual analogue scale (VAS) between one and ten. Motivation is furthermore rated by means of the questionnaire for current motivation in BCI (QCM-BCI), consisting of 18 statements to be rated on a 7-point Likert scale and including four factors: *mastery confidence*, *incompetence fear*, *interest*, and *challenge* (Nijboer et al., 2010). It is an adaptation of the original QCM questionnaire (Rheinberg et al., 2001) to fit the BCI context. After each online session with feedback, the multidimensional NASA Task Load Index (NASA-TLX) questionnaire is administered, assessing perceived workload on a scale from 0 to 100 by means of six factors: *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort* and *frustration* (Hart et al., 1988; Riccio et al., 2011). All questionnaires are applied in German (for healthy subjects) and Italian (for end-users).

To investigate condition differences, both VAS scores and all factors of the QCM-BCI and the NASA-TLX questionnaires are compared independently. Visual inspection of the data warrants the use of non-parametric statistics. The Kruskal-Wallis H test is used for between group comparisons and the Wilcoxon singed rank test for paired comparisons. In the case of repeated measures within a condition, a non-parametric Spearman's rank correlation test is used to assess the impact of time (session number) on the questionnaire factors.

2.5.3. Cross-validation

Cross-validation is a technique that is commonly used to estimate the predictive value of a model in a way that generalizes to unseen data (Lemm et al., 2011; Müller et al., 2001). In order to achieve an N-fold cross-validation, the available data is divided into N mutually exclusive sets. At each fold, N-1 sets are used to estimate the parameters of the model, and the remaining set is subsequently used to estimate the model's prediction performance. As this performance estimate, or test performance, is based on unseen data, it generalizes better to new data.

In the case of offline analyses, the above cross-validation scheme can not fully avoid over-fitting. When using the same data over and over again, whilst changing or optimizing the model, the best solution that is found might well over-fit the test data. Therefore, it is common consensus in BCI

literature that an online experiment is the only ultimate proof of the validity and relevance of any method.

There is, however, a way to properly prevent over-fitting on the test data by introducing an additional dataset. This so-called validation set should be left untouched until the final model is chosen, and its parameters trained. By applying the model to the validation set once, and only once, an unbiased performance estimate can be obtained. For offline analyses of BCI data it is good practice to use the calibration data for cross-validation, leaving the online data out as the validation set.

2.6. Metrics

The following sections describe the metrics that are used throughout this thesis. Most of these metrics can be calculated from the trial-level confusion matrix $CM \in \mathbb{Z}^{C\times C}$ with additional (experimental) parameters such as the SOA and time overhead. The confusion matrix provides information about the prediction accuracy for the different classes.

2.6.1. Spatial Accuracy Quantification

From CM, the pairwise F-scores are calculated to assess the confusion between each pair of classes. This exposes any systematic preference, if present.

Since not every class is necessarily the target equally often, each row of the confusion matrix is first normalized according to $CM^{\dagger}_{i,j} = \frac{CM_{i,j}}{\sum_{k=1}^{C} CM_{i,k}}$. Then, the sensitivity and precision for pairs of classes $i, j \in \{1,\dots,C\}$ are defined as follows

$$precision(i,j) = \frac{CM^{\dagger}_{i,i}}{CM^{\dagger}_{i,i} + CM^{\dagger}_{j,i} * (C-1)} \tag{2.5}$$

$$sensitivity_{(i,j)} = \frac{CM^{\dagger}_{(i,i)}}{CM^{\dagger}_{i,i} + CM^{\dagger}_{i,j} * (C-1)} \tag{2.6}$$

with the constant $C-1$ compensating for the pairwise comparison. The matrix $F \in \mathbb{R}^{C\times C}$ of pairwise F-scores is then defined as

$$F_{i,j} = 2 \times \frac{precision(i,j) * sensitivity_{(i,j)}}{precision(i,j) + sensitivity_{(i,j)}} \tag{2.7}$$

It is thus the harmonic mean of sensitivity (the ability to identify a target) and precision (the ability to exclusively identify targets as targets). Since both sensitivity and precision range from 0 to 1, the F-score also ranges from 0 to 1.

Diagonal values are always 1. Other pairs (elements in F) are marked as location neighbors (l), sound neighbors (s) and locations that are symmetric in the front-back plane (fb) (see Section 3.2). All are potential sources for a reduced separability.

2.6.2. Classification Accuracy

Two different accuracy measures are used throughout this thesis. First, the binary accuracy assesses the performance of the classifier on the individual epochs. It assesses the ability to separate target epochs from non-target epochs. Typically, binary accuracy refers to the overall percentage of correctly classified epochs, irrespective of their binary-class membership. With

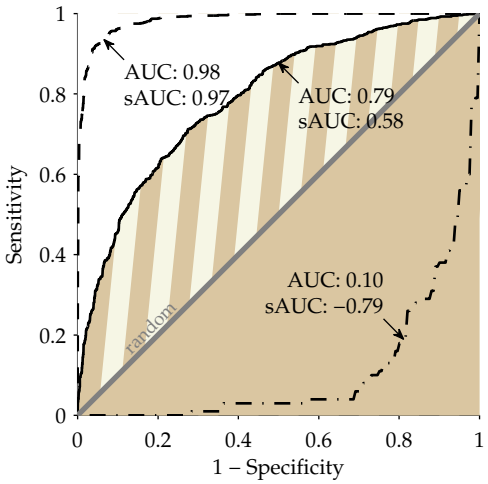


Figure 2.6.: The ROC curve describes the ratio between sensitivity and 1-specificity for different decision thresholds. The ROC curve can be summarized by a single value in different ways. In this thesis the typical area under the ROC curve (AUC) is used to describe the quality of a classifier, and the sAUC is used to describe the separability of individual feature dimensions.

a similar number of epochs per binary-class, this is a valid approach. However, in the case of unbalanced binary-classes sizes, this can lead to an overestimation of the actual performance. For example, with one out of six epochs belonging to binary-class one, an accuracy of 83.3 % can be achieved by assigning all epochs to binary-class two. This is clearly a bad classifier. In this case, the accuracy should be class-wise normalized. The binary accuracy is then calculated for each binary-class separately, and the results are averaged. In the above case, this would lead to an accuracy of 50 % (0 % on binary-class one and 100 % on binary-class two). The chance level is thus at 50 %.

The second accuracy measure, the selection accuracy, is evaluated on the trial-level after accumulating the evidence from the individual epochs (see Section 2.4.5). It measures how often the class that was predicted by the BCI matches the correct class, as specified by the task. The accuracy of an unbiased, random classifier (chance level) is on average $100\,\%/C$. With a six class BCI, a random classifier would thus on average be correct on 16.7 % of all trials.

Caution is to be taken in stating that a selection accuracy that is marginally larger than chance level is meaningful. For example, in a six class BCI, an unbiased, random classifier would score a selection accuracy > 30 % more than 22 % of the time, when tested on 10 trials per try; a performance that is almost double the chance level. As a rule of thumb, it is typically safe to assume a meaningful result if consecutive measures consistently score above chance level.

2.6.3. Area Under the ROC Curve

The receiver-operator characteristic (ROC) curve is a graphical representation of the quality of a binary classifier that is often used in signal detection theory (Fawcett, 2006), as shown in Figure 2.6. Sensitivity is plotted as a function of 1−specificity, with each point on the line representing a shift of the classifier bias. The ROC curve of perfectly mixed distributions is (approximately) the diagonal line (gray line), and the ROC curve of perfectly separated distributions is a right angle going from (0,0) either through (1,0) or through (0,1) to (1,1). In this thesis, the ROC curve is summarized by the area under the ROC curve (AUC) in two ways.

For assessing the quality of a classifier (in particular in Chapter 6), the classical AUC metric is used, simply integrating the complete area under the ROC curve. In Figure 2.6, this refers to the combined striped and plain colored areas for the solid black line. It ranges from 0 to 1, with 1 indicating perfect separation, 0 indicating perfect but inverted separation, and 0.5 indicating complete binary-class overlap.

The AUC is also powerful in estimating the predictive power of individual features (Green et al., 1966). In this case, one dimension of the feature vector is used to estimate the ROC curve and the corresponding AUC. To include additional information on the order of target and non-target means, the AUC is transformed into the sAUC, the striped area in Figure 2.6, as follows

$$\text{sAUC} = 2 * (\text{AUC} - .5) \tag{2.8}$$

The sAUC ranges from -1 to 1, where 0 indicates complete binary-class overlap, -1 indicates perfect separability with the mean of the targets being smaller than that of the non-targets, and 1 indicates perfect separability with reversed means order.

The sAUC can be plotted as a spatial map to visualize the distribution of discriminative information over the scalp for a given time interval (see Figure 2.4, right column). For comparability, all such scalp maps are plotted on a common scale throughout this thesis. Additionally, sAUC values between -0.04 and 0.04 are colored gray uniformly to emphasize the zero-crossing. These thresholds were chosen empirically to produce meaningful results on all plots.

The advantage of the AUC metrices over methods like Fisher score (Duda et al., 2000; Müller et al., 2004), Student's t-statistic (Müller et al., 2004; Student, 1908), or pointwise biserial correlation coefficient (Blankertz et al., 2011) is that it is non-parametric and does not rely on the assumption that the distributions are Gaussian. Furthermore, though the smoothness of the ROC curve does depend on the number of epochs (compare the dashed line with 2000 epochs in Figure 2.6 with the dash-dotted line with 200 epochs), the AUC estimate does not systematically depend on this number. It thus allows for a robust comparison across subjects with different number of trials.

2.6.4. Information Transfer Rate

In BCI, as in information theory, the amount of information transferred over a noisy channel can be quantified by the information transfer rate (ITR) (Shannon et al., 1950). Various BCI specific variations of the ITR were proposed, which were recently reviewed by Schlögl et al. (2007). In 2000, Wolpaw et al. (2000) proposed a version, which has since widely been used in literature. Though its assumptions may not always hold, and more general versions have been proposed (Nykopp, 2001), this original version is used here for the sake of comparison. It is defined as

$$R = \log_2 C + P \cdot \log_2 P + (1-P) \cdot \log_2 (\frac{1-P}{C-1}) \tag{2.9}$$

where R is the ITR in bits/selection and P is the selection accuracy. As a BCI has overhead on each selection, and for a more interpretable result, the ITR is often expressed in terms of the bits per minute (bits/min), defined as

$$\text{bits/min} = V \cdot R \tag{2.10}$$

where V is the speed in selections/minute. Throughout this thesis, the latter definition of the ITR is used. It is obvious that the ITR depends both on the accuracy and the selection speed.

2.6.5. Writing Speed

Though the ITR is a good theoretical performance measure, interpreting the practical relevance of a given value remains difficult. For particular cases, more straight forward metrics exist. In

an online speller setting, simply calculating the written characters per minute (char/min) is a robust and interpretable metric. In the strictest version, the subject is instructed to copy a text letter for letter, and any committed mistakes have to be corrected by the subject. In this case, the char/min is simply defined as the number of letters written, divided by the elapsed time. For less strict protocols where the subject is notified of, but does not need to correct any mistakes, the char/min is often defined as the number of correct letters divided by the elapsed time. This may overestimate the actual char/min, as correcting an error would add an additional time penalty.

For offline analyses, the char/min can be approximated by the symbols per minute (symb/min) metric, which integrates BCI performance with application information such as overhead and error handling strategy. For single-level spellers such as the visual matrix speller (Farwell et al., 1988), a correct selection counts as +1 symbol. An erroneous selection counts as -1 symbol, to account for the added effort of a necessary subsequent corrective action. Symb/min is then defined as

$$\text{effective correct} = P - E \tag{2.11}$$

$$\text{symb/min} = V * \text{effective correct} \tag{2.12}$$

where V and P are equal to Equations 2.9 and 2.10, and E is the classification error ($1-P$ for single level spellers).

In two-level spellers, errors can be made on the group level (level one error) and on the character level (level two error). In case of a level one error, the second level contains the possibility to select an "undo" action, which cancels the previous selection. As no actual symbol is selected, it counts as a 0 symbol selection. Level two errors always count as -1 symbol. P is now defined as the selection accuracy on both levels, E as the percentage of erroneous selections on the second level only, and V as the classification speed in selection pairs/minute. Note that the simplifying assumption is made that the accuracy is the same for each symbol (similarly to the assumption of the ITR).

The validity of symb/min as an approximation of char/min is investigated in Experiment 5.

3. AMUSE Proof of Principle

The first scientific contribution described in Section 1.7 introduces spatially distributed stimulus sources, an aspect new to covert attention-based BCI that needs careful offline validation (see Section 1.6). This chapter gives a detailed account of the fundamental concept of AMUSE and an experimental validation of these principles is presented. Concretely, hypotheses *i* (*subjects can focus on one amongst several spatial stimuli*) and *ii* (*spatial stimuli alone are sufficient to elicit binary-class discriminative ERPs*) will be addressed. The results are discussed in the context of contemporary covert attention-based BCIs.

Parts of this chapter are based on prior publications Schreuder (2008a), Schreuder et al. (2010a), and Schreuder et al. (2009).

3.1. Prior Work

Covert attention-based BCIs have been realized in various domains and are summarized in Section 1.5 (and in particular in Table 1.1). Early research on these BCIs was mainly binary in nature, and the achieved accuracies were modest. This severely limits their usefulness in real applications. Since then, the covert attention-based visual BCIs have been particularly successful in increasing the number of classes and the achieved accuracy, resulting in high communication rates of up to 9.15 bits/min (Schaeff et al., 2012; Treder et al., 2011). Nevertheless, other modalities need to be investigated, as many end-users suffer from loss of sight (Huggins et al., 2011).

In a recent study by van der Waal et al. (2012), tactile stimuli to six fingers were used to successfully elicit ERP responses. The achieved selection accuracy (79 %) and re-estimated ITR (4.52 bits/min) mark an improvement over other tactile BCI. Kaufmann et al. (2013a) recently showed the feasibility of patient application for such a BCI offline. Both studies relied on spatial attention to tactile stimuli.

Recent years have seen a sharp increase in the number of studies on auditory BCI, which were introduced extensively in Section 1.5.1. Selection accuracy is often lower than covert attention-based visual BCI, but higher than tactile BCIs. From Table 1.1 it is apparent that studies that simply used natural sounds to transfer the visual matrix speller directly to the auditory domain have relatively low selection accuracy and modest ITR (Furdea et al., 2009; Klobassa et al., 2009; Kübler et al., 2009). Studies with less (non-spatial) classes did at times result in a higher selection accuracy, but the ITR is reduced accordingly (Guo et al., 2009; Hill et al., 2004; Hill et al., 2012b). As also suggested in Kübler et al. (2009), adding spatial information might be one way to increase the number of classes whilst maintaining a high selection accuracy and ITR.

Several psychophysiological experiments have been described that investigate the influence of the human capacity for sound localization on ERPs. Teder-Sälejärvi et al. (1998) essentially presented seven oddball paradigms to the subject. An array of seven speakers, with 9° azimuth angle between them, each presented a random stream of targets and non-targets. Subjects were requested to attend left, front or right and only targets coming from the attended stream reliably elicited a P3 response. In a similar study, Arnott et al. (2002) used three virtual sources with varying azimuth angle between them, resulting in attention modulated N1 and P3 peaks. In these studies, the spatial location was not used to separate non-targets from targets, but rather to separate different streams, similar to Hill et al. (2004).

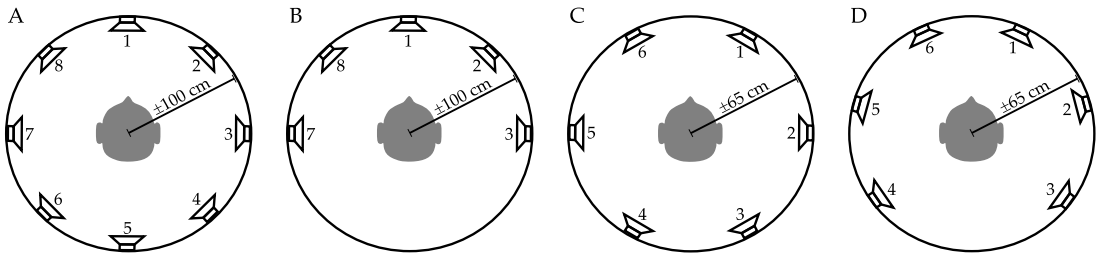


Figure 3.1.: Schematic representation of the different setups used in this thesis. A) High density setup for the proof of concept, B) frontal half only setup, used for offline BCI testing purposes, C) setup used in online spelling with healthy subjects, D) setup used in online spelling with end-users, with extra space in the back to prevent obstruction by the wheel-chair. From B) to C), the radius of the circle was reduced to improve the discrimination.

A more recent study did use the spatial separation (albeit virtually through stereo headphones) to set aside the non-targets (0°, straight ahead) from the targets (± 30° and ± 90° azimuth) (Sonnadara et al., 2006). Because the subjects were engaged in a passive listening task whilst watching a movie, no P3 responses were elicited; the goal was rather to elicit MMNs. However, it does show that the spatial location can be a stimulus determining factor. A similar experiment was performed in free-field with only 10° azimuth spatial separation (Deouell et al., 2006). An oddball paradigm purely based on spatial location has been used in Rader et al. (2008), but merely as a training for detecting stimuli from different locations in a later task. No behavioral- or neurophysiological data for this condition are reported.

To date of the first contribution of this thesis (Schreuder et al., 2009), no auditory BCI employed more than two spatial sources. However, in that same year, Rutkowski et al. (2009) proposed a spatial auditory BCI paradigm similar to AMUSE, albeit in a preliminary state. It appears that these paradigms were developed independently.

3.2. AMUSE Principles

The AMUSE paradigm builds on the classical oddball paradigm (see Section 2.2.2.1) and psychophysiological paradigms (see above), which were altered in several ways.

To overcome some of the shortcomings in auditory BCI, and to create a high speed, multiclass and auditory BCI, stimuli consisted of sounds played from different sources around the subject. Stimuli from these sources are presented in rapid succession. Due to the human ability of spatial hearing (see Section 2.1), this gives an intuitive task to the subject: to focus the attention to one of several spatially distinct sound sources. The non-target binary-class then contains all non-focussed stimuli, and the binary target and non-target classes are defined purely by the task instructions (similar to the “names” oddball paradigm from Section 2.2.2.2).

An AMUSE setup can consist of any number of sources placed around the subject in free field, keeping in mind the human MMA (see Section 2.1). Though not necessary, in this thesis they were placed on a circle around the subject at ear-height. Along this circle, placement can be equidistant or irregular, and the number of sound sources can vary. Sound sources can all produce the same sound or each source can be associated with a unique sound, the latter of which adds another identifying property to the stimuli. The flexibility of the AMUSE paradigm is explored throughout this thesis (see Figure 3.1). The AMUSE paradigm is defined by this spatial principle, with no connection to any user application in specific.

As in any ERP-based paradigm, stimulation is typically performed in random order. Following Section 2.3.1, all sound sources are stimulated exactly ones in what is called an iteration, before moving to the next iteration. After a certain number of iterations, either fixed or determined dynamically (see Chapter 6), the trial is terminated. In online modus, the collected classifier scores are then converted into a class decision according to Equation 2.4, before the next trial is started.

As discussed in Section 2.2.2.1, formally the target and non-target stimuli should be swapped to remove the confounding effect of stimulus properties from the attention effect. Since in AMUSE the target is defined by the task instead of physical features alone, and the task requires a changing target source, such a reversal is intrinsic. Any binary-class difference that is found is thus due to the attention effect.

3.2.1. General Methodology

The following describes the basic AMUSE setup as it was used throughout this thesis; it is extended by the methods description of individual experiments.

Subjects sit in a comfortable chair, facing a screen with a fixation cross. EEG is recorded in a room that is not electromagnetically shielded and no special sound attenuation precautions are taken. Speakers are calibrated to a common stimulus intensity of ~58 dB. If requested by the subject, an initial round of stimuli is given before the actual experimental trials, to familiarize the subject with the setup. Before the experiments, subjects are asked to minimize eye movements and other muscle contractions during the experiment.

The stimulus presentation, the online Berlin Brain-Computer Interface (BBCI) system, and the offline analyses are implemented in Matlab (Mathworks), making use of the Psychophysics Toolbox for multichannel audio presentation (Brainard, 1997). In the online setting, the class labels and BCI decision results are read to the subject using the open-source speech synthesis system Mary (Schröder et al., 2003). A multichannel, low-latency firewire soundcard from M-Audio (M-Audio Firewire 410) is used to individually control the low-budget, off-the-shelf computer speakers (Genius SP-S100 for Experiments 1 to 4, and Sony SRS-A201 for Experiments 5 to 7).

EEG is recorded using a varying number of Ag/AgCl electrodes and channels are referenced to the nose. The signals are amplified using BrainAmp amplifiers (Brain Products), sampled at 1 kHz and filtered by an analog bandpass filter between 0.1 and 250 Hz before being digitized and stored for offline analysis. Further analyses are done in Matlab (The Mathworks), according to Section 2.4.

3.3. Experimental Validation

Though the concept of ERP-based BCIs is well known and the basis of many studies, the proposal of the AMUSE paradigm introduces some new factors that warrant careful investigation (see Section 1.6). To disentangle hypotheses *i* and *ii* from other effects, several experiments were performed in two blocks. The following describes the methods that are used for all experiments in this chapter, extending Section 3.2.1. Experiment specific methods and settings are described in the respective sections.

3.3.1. Participants

The first block (Experiments 1 and 2) included seven healthy volunteers (five male, mean age 29.1 years, range 25-34 years). All subjects were volunteering group members and had some previous experience with BCI, mainly based on MI tasks. The second block (Experiments 3 and 4) included

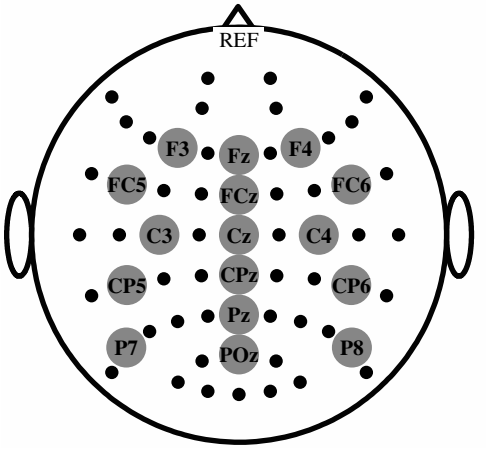


Figure 3.2.: A subset of the available channels was used for classification purposes in Experiments 3 and 4. The subset was fixed for all subjects and is highlighted by the gray circles.

five healthy volunteers (three male, mean age 32.4 years, range 22-55 years), out of which two were paid subjects with no previous experience in BCI. They were compensated for their time with eight euro per hour. Of the other three volunteering group members, two also participated in the first round. All subjects reported to be free of neurological symptoms and to have normal hearing. However, two subjects (*VPip* and *VPig*) reported having difficulty with spatial localization of sounds in natural situations and subject *VPzq* reported a high-pitched tinnitus in the right ear.

3.3.2. General Protocol

Procedures were positively evaluated by the Ethics Committee of the Charité University Hospital (number EA4/073/09). All subjects provided verbal informed consent and subsequent analyses and presentation of data were anonymized.

Subjects were surrounded by eight speakers at ear height. The speakers were spaced evenly with 45° azimuth angle between them, at approximately one meter distance from the subject's ears (see Figure 3.1-A). At the start of the session, subjects were asked to subjectively judge the equality of the loudness of all sources and alter these if necessary.

The first of two blocks consisted of two experiments to test for the feasibility of the paradigm behaviorally (Experiment 1) and neurophysiology (Experiment 2), under high SOA conditions. The second block consisted of two experiments under conditions that are realistic for BCI usage. With lower SOAs, data were classified as they would be in a BCI setting in Experiment 3. Experiment 4 is the control condition, disentangling the influence of the spatial property from other effects.

3.3.3. Data Acquisition

After acquisition, the raw data was low-pass filtered with an order 10 Chebyshev II filter (40 Hz pass-frequency, 49 Hz stop-frequency, 50 dB damping), down sampled to 250 Hz and high-pass filtered with an order 4 Butterworth filter (.33 Hz pass-frequency, .1 Hz stop-frequency, 50 dB damping). Finally, the data were epoched between -150 ms and 800 ms relative to stimulus onset, using the first 150 ms as baseline. Estimates of the ERP component latencies and amplitudes thus have a maximum resolution of 4 ms. The data were further down-sampled to 100 Hz for

Table 3.1.: Behavioral performance for Experiment 1. Because the majority of stimuli are non-targets, true negatives (no response to a non-target) are not reported and also not considered in Equation 3.1. The total number of targets is equal to the sum of hits and misses.

Subject	Hits	False alarms	Misses	Error
VPja	72	3	0	4 %
VPiz	71	0	1	1.4 %
VPip	72	1	0	1.4 %
VPzq	71	5	1	7.8 %
VPig	88	17	8	22.1 %
VPjf	96	4	0	4 %
VPjb	95	3	1	4 %
Average	-	-	-	6.4 %

classification purposes, to resemble the typical settings used in the BBCI. Artifacts were removed according to Section 2.4.2.

3.3.4. Classification

Data in Experiments 3 and 4 were classified to estimate the usefulness of AMUSE for BCI. In order to obtain a fair comparison, all datasets were reduced to the maximum possible common number of epochs (approximately 1500 epochs) and a subset of the available channels (see Figure 3.2). Irregular feature intervals were selected using prior knowledge (see Figure 2.5-B) and classification of the data was done using a 10-fold chronologically sampled cross-validation.

3.3.5. Experiment 1: Behavioral Responses

A potential reason for any BCI paradigm to not elicit the expected responses in the EEG is that the task is simply too difficult. If the subject can not distinguish target from non-target, no large differences are to be expected in the EEG, if any at all. To test that subjects are able to perform the task at hand, the first experiment is a purely behavioral task and does not involve EEG.

3.3.5.1. Setting

Using an eight speaker setup (see Figure 3.1-a), the stimuli only differed in their spatial location. The stimulus consisted of 75 ms of band-pass filtered white noise (150-8000 Hz) with 3 ms rise and fall, and was the same for all directions. Per trial, one of the eight sources was the target (probability 12.5 %), leaving the others as non-targets (probability 87.5 %). The target was explicitly indicated to the subject before the start of a trial, both visually on the screen and by presenting the stimulus from that source.

Seven healthy subjects were asked to respond to the indicated target stimulus by pressing a key on the keyboard. The SOA was set to 2000 ms with a random jitter (mean 25 ms, standard deviation (SD) 14.4 ms). This relatively high SOA allowed for the ERP response to be completed before the following stimulus. Any response that fell within the 2000 ms window after a target stimulus was considered a true positive; the failure to produce a response was considered a false negative. Any response not directly preceded by a target stimulus was considered a false positive.

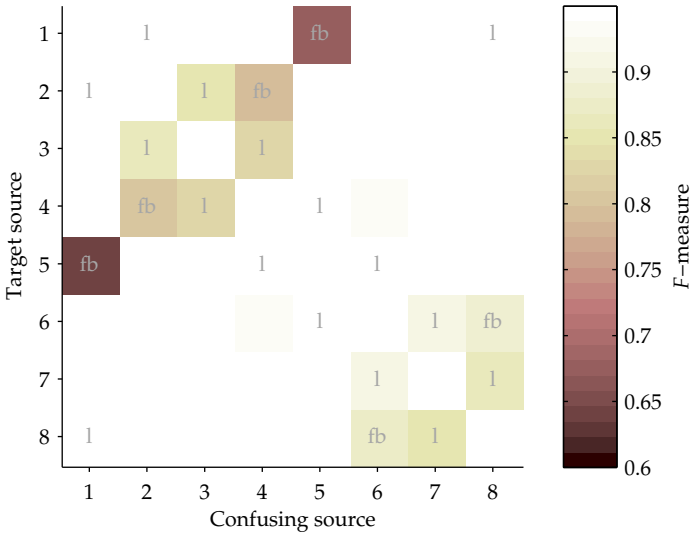


Figure 3.3.: The grand average source discrimination performance is depicted as a confusion matrix containing *F*-scores (see Section 2.6.1). Lower *F*-scores indicate worse performance. Sources with a front/back confusion (fb) are clearly more difficult to separate. Sources with no degrading factor (no label) have near-perfect recognition performance.

Error scores were calculated according to

$$\text{error} = \frac{\text{miss} + \text{false alarms}}{\text{hits} + \text{miss} + \text{false alarms}} * 100\,\% \tag{3.1}$$

The number of true negatives is excluded from this equation because its large number would mask the error size. Between 576 and 768 epochs per subject were recorded. Blocks of this experiment were interleaved with blocks from Experiment 2, as they took place in the same session.

3.3.5.2. Results

Recognition Performance Individual performance results can be found in Table 3.1. Although no subject showed a perfect recognition, the number of errors was under 10 % for all subjects but *VPig*. Subject *VPig*, with an error rate of 22.1 %, was one of two subjects who reported to have difficulty with sound localization in natural settings. Subject *VPip*, who also reported this, had an excellent performance, with an error rate of 1.4 %.

Directional Preference The grouped performances of subjects on different directions can be found in Table 3.2; the corresponding *F*-score matrix in Figure 3.3. The first observation to be made is the confusion of the front speaker with the rear speaker. All eight false alarms on the front trials were due to confusion with the rear speaker. Vice versa, the only false alarm on the rear trials is a confusion with the front speaker. Several subjects also reported to have difficulty with this distinction. The median *F*-score for speakers that share a front-back plane is low (median: .80; inter-quartile range (IQR): 0.20; N: 6), compared to neighboring sources (median: .93; IQR: 0.15; N: 16) and all remaining pairs (median: 1, IQR: -, N: 34). The latter two show (near) perfect recognition rates. Bonferroni-corrected non-parametric Kruskall-Wallis H

Table 3.2.: Grand average performance per direction. Errors for all subjects were averaged according to the direction of the target stimulus. Types of errors made can be found in Figure 3.3. Due to the random target assignment, not all directions are designated as a target equally often. Source numbers correspond to the speaker locations given in Figure 3.1-A.

Source	Hits	False alarms	Misses	Error
1	67	8	1	11.8 %
2	40	1	4	11.1 %
3	84	3	0	3.5 %
4	83	10	1	11.7 %
5	57	1	3	6.6 %
6	51	4	1	8.9 %
7	108	5	0	4.4 %
8	75	1	1	2.6 %
Average	-	-	-	7.6 %

tests ($\alpha = 0.008$) confirmed that errors on front-back pairs were significantly more frequent than on neighboring sources ($\chi^2(1) = 7.19$; $p = 0.007$) and all remaining pairs ($\chi^2(1) = 20.57$; $p < 0.001$). Furthermore, recognition for neighboring sources was significantly worse than for the remaining pairs ($\chi^2(1) = 11.82$; $p = 0.001$).

3.3.5.3. Summary

AMUSE was shown to be behaviorally feasible, with good overall recognition rates. Special care is to be taken with stimuli that lie on the same cone of confusion as the target stimulus (see Section 2.1); without any additional cues they may lead to front-back confusions.

3.3.6. Experiment 2: Neurophysiological Responses

The fact that a task is behaviorally feasible does not imply that it is useful for BCI. For it to be applicable in the domain of BCI, target and non-target stimuli must result in ERP components in the EEG with significantly different time courses. This difference would allow for the classification between target and non-target, which drives any BCI. The following experiment was designed to verify that indeed such ERP components are elicited by the proposed task.

3.3.6.1. Setting

The subjects that participated in Experiment 1 also performed the current experiment, to test for the existence of binary-class related differences in the EEG. Runs of this experiment were interleaved with those of Experiment 1.

The settings for this experiment were largely the same as in Experiment 1, with some notable differences. Subjects were now asked to mentally count the amount of target stimuli instead of responding by key press, and the SOA was reduced to 1000 ms, maintaining the same jitter. EEG data were recorded for offline analyses. With 10 iterations per trial, 80 epochs were available per trial. For each subject, 32 trials were recorded, changing the target with every trial, for a total of 2560 epochs (320 target epochs). The number of trials left after artifact rejection far exceeded the recommended minimum number (Duncan et al., 2009).

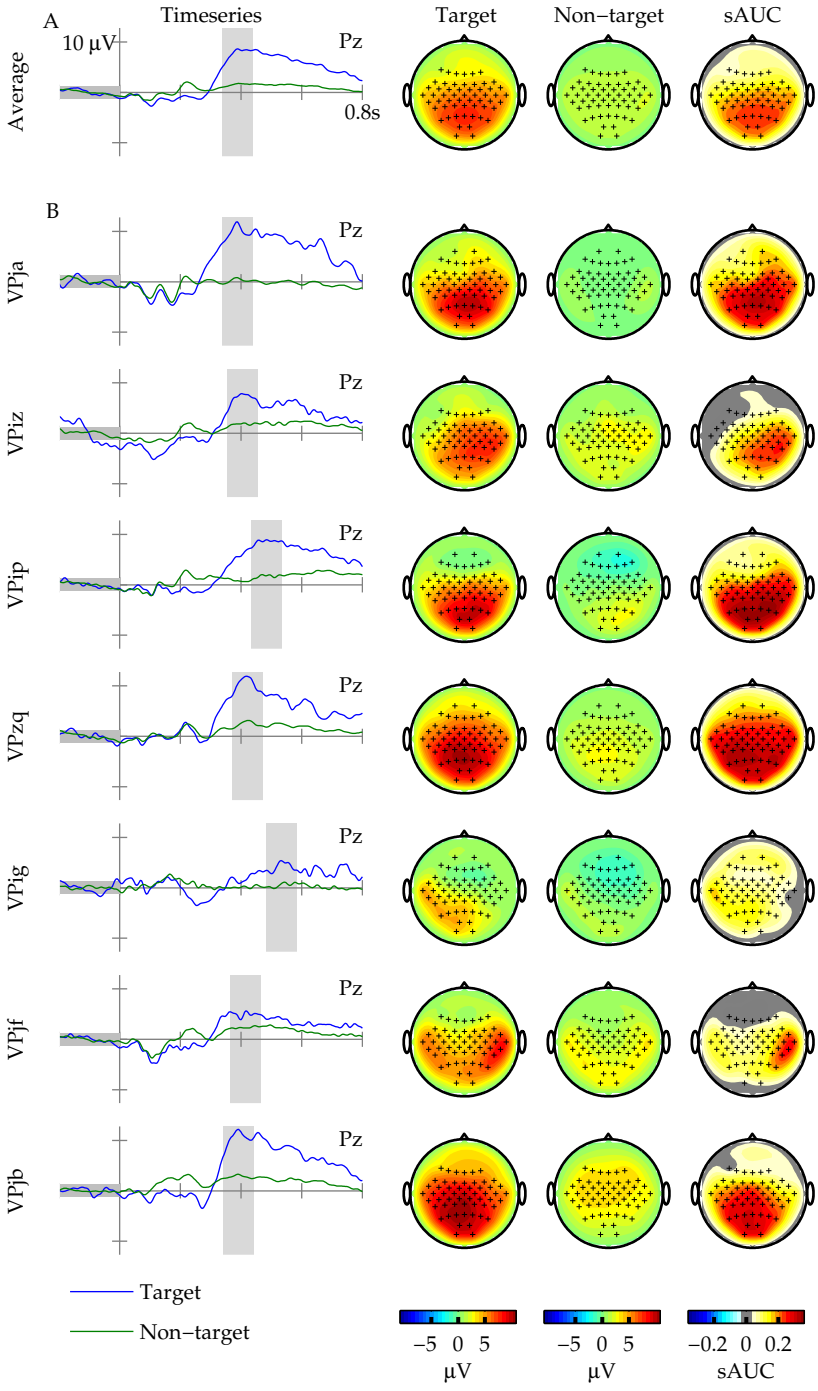


Figure 3.4.: The P3 response is plotted in grand average (A) and for each individual subject (B). The P3 response is largely typical and robust over subjects. See Figure 2.4 for a full explanation of the figure elements.

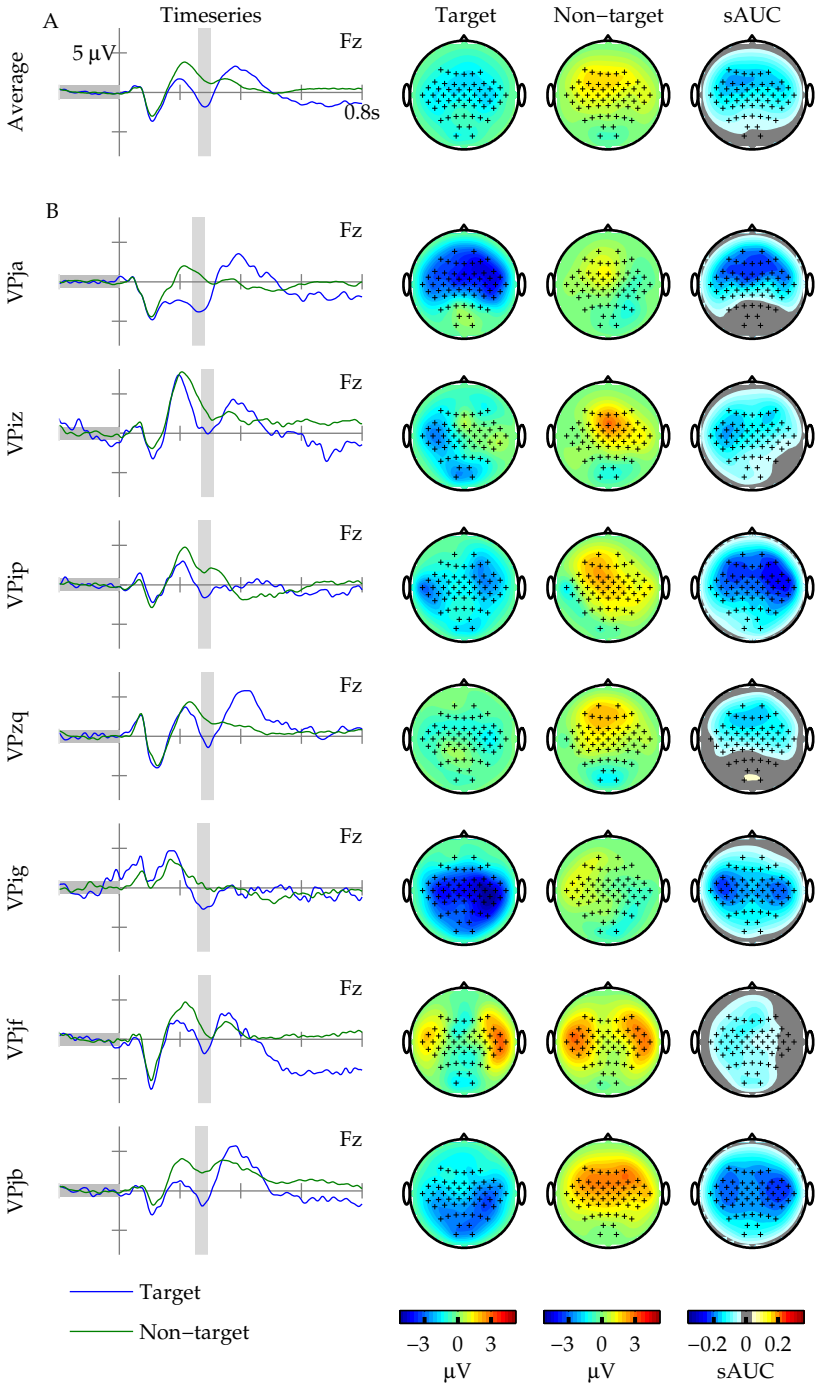


Figure 3.5.: The N2 response is plotted in grand average (A) and for each individual subject (B). The N2 response is largely typical and robust over subjects. See Figure 2.4 for a full explanation of the figure elements.

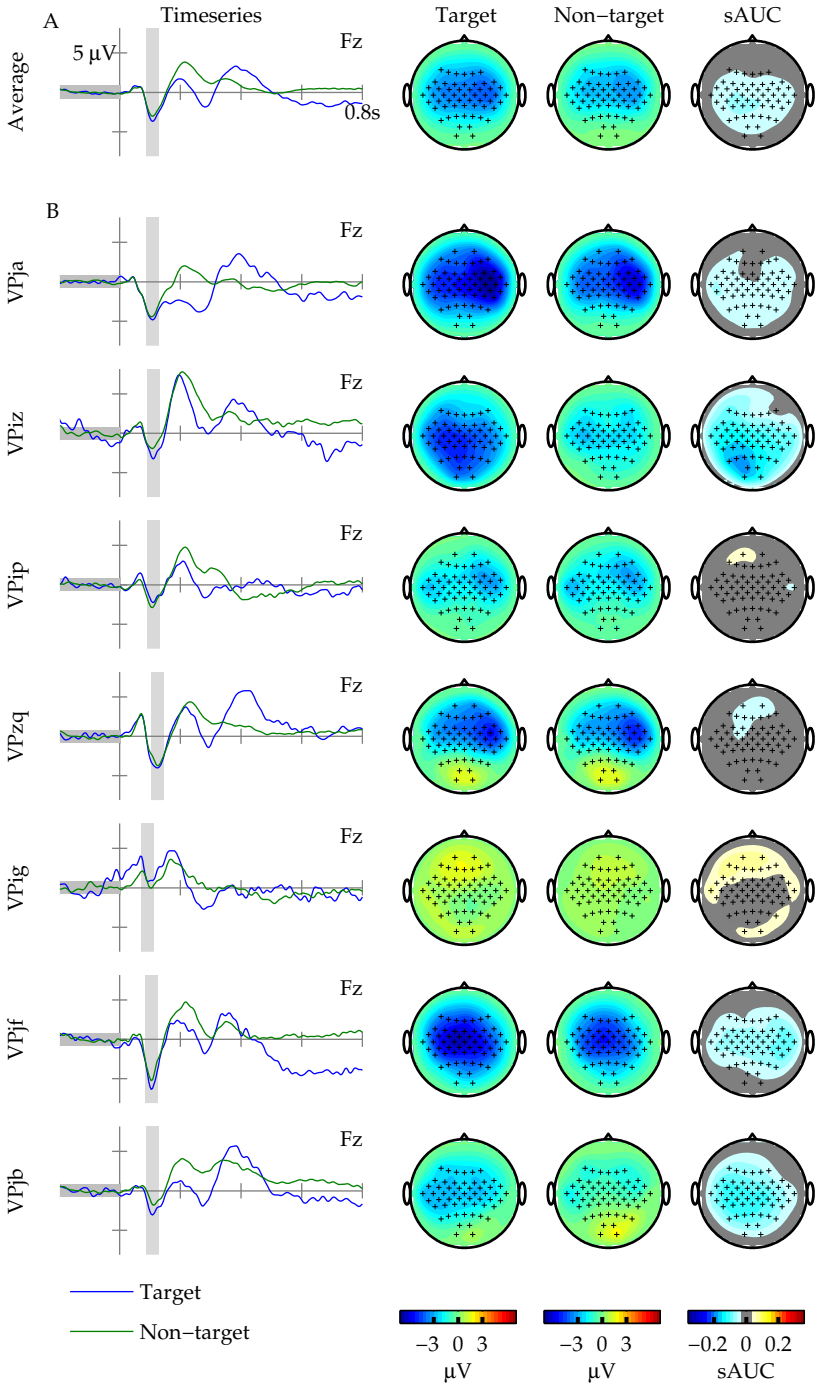


Figure 3.6.: The N1 response is plotted in grand average (A) and for each individual subject (B). The N1 response is largely typical and robust over subjects. See Figure 2.4 for a full explanation of the figure elements.

Table 3.3.: Stimulus properties for Experiments 3 and 4. Source numbers refer to the speaker labels given in Figure 3.1-B. The stimuli consist of a tone (Base tone) with harmonics and band-pass filtered white noise (Noise range).

Source	Base tone (Hz)	Noise range [KHz]
1	554 (cis)	0.54 - 4.2
2	622 (dis)	0.70 - 5.5
3	699 (f)	0.91 - 7.1
7	440 (a)	0.32 - 2.5
8	494 (b)	0.42 - 3.3

3.3.6.2. Results

Grand average and subject specific ERP responses and scalp topographies for all subjects can be found in Figures 3.4, 3.5, and 3.6. Figure 3.4 focusses on the cognitive P3 response (see Section 2.2.2), which typically contributes highly to the BCI performance. In grand average (top row), the peak voltage deflection for target stimuli at channel *Pz* is 9.73 µV, with 417 ms latency to stimulus onset. A prominent positive deflection is found over parietal channels for target stimuli, resulting in binary-class discriminative information in the same region (right column). This general pattern is found for most individual subjects, but two; *VPig* shows a much higher than average peak latency (568 ms), and *VPjf* shows an a-typical, off-center spatial distribution of the discriminative information.

Figure 3.5 focusses on the earlier N2 component, in particular over channel *Fz*. In grand average, the peak voltage deflection for target stimuli is -1.79 µV with a latency of 279 ms after stimulus onset. Spatially, the discriminative binary-class information is found in frontal regions and around channels *F5* and *F6*, which lie close to the primary- and secondary auditory cortices (but see Section 2.2.1). Both the latency and spatial distribution of the N2 are robust over subjects.

The N1 peak is shown in Figure 3.6, in particular over channel *Fz*. In grand average, the peak voltage deflection for target stimuli is -3.96 µV with a latency of 111 ms after stimulus onset. Both the target and non-target responses are most prominent over central electrodes, with low discriminative information. However, for at least three subjects, the N1 has prominent discriminative power. Latency and spatial distribution are mostly robust over subjects.

A one-way analysis of variance (ANOVA) revealed a significant effect of channel location (factors *Fz*, *Cz*, *Pz*) on the P3 amplitude ($F(2,15) = 11.01$, $p<0.01$). A Tukey-Kramer post-hoc comparison showed that channel *Fz* (mean: 3.91, SD: 1.85) significantly differs from channels *Cz* (mean: 7.68, SD: 1.90) and *Pz* (mean: 9.73, SD: 2.68). Channels *Cz* and *Pz* did not differ significantly between them. Average P3 latency increased from frontal- to parietal electrodes, peaking at 417 ms (SD = 35.02), but this effect was not significant. Subject *VPig* was excluded from this analysis due to the lack of a clear P3 peak. Neither the latency nor the peak of the N1 and N2 components interacted significantly with electrode site.

3.3.6.3. Summary

Though individual difference exist, the grand average latency and location of all three components are in line with what is to be expected from literature. AMUSE thus leads to typical ERP components with good binary-class separability for most subjects; a fundamental requirement for any BCI paradigm.

3.3.7. Experiment 3: Realistic BCI parameters

Up to this point the parameters did not resemble a typical BCI system, but were rather chosen to investigate the feasibility of the paradigm. BCI systems deal with a trade-off between the speed of the system and the amount of discriminative information that is contained in each individual epoch. One factor in this trade-off is the SOA. In Experiments 1 and 2, this was set to high values, favoring the maximum discriminative information per epoch. For online BCI systems, the SOA is typically set to values between 150 and 300 ms. With such a low SOA, the discriminative components that are important for classification start to overlap and (at least partially) cancel out, leading to a reduction in SNR (see Box 2.3 on page 23). The current experiment investigates the performance of the paradigm under realistic BCI conditions.

3.3.7.1. Setting

For this experiment, the paradigm was altered in several ways, based on findings from Experiments 1 and 2 and to change to the typical settings of a BCI system. First, the amount of sources was reduced to the front five to make the task easier and to avoid front-back confusions (see Figure 3.1-B). Thus, the target was presented with 20 % probability and non-targets with 80 % probability. It has been shown that this is rare enough to produce a robust P3 response (Sellers et al., 2006b). All five speakers were each given a unique, complex 40 ms stimulus, build from band-pass filtered white noise and a tone overlay (see Table 3.3). The stimuli thus had two discriminating features each: the physical sound properties and the source location. Latency jitter was omitted. In order to explore the possibilities of the paradigm, two conditions with different SOA were tested (condition C300: 300 ms, and condition C175: 175 ms).

A trial consisted of 75 epochs, 15 for each source. For both conditions, 40 of such trials were recorded, making a total of 3000 epochs. Stimulus order now had an additional constraint: there were at least two other stimuli between stimulation from any particular source to prevent overlap of target ERPs.

Experiments 3 and 4 were recorded in the same session, and runs of both experiments were interleaved.

3.3.7.2. Results

Physiology The grand average neurophysiological responses for both conditions are presented in Figure 3.7 (P3 response) and Figure 3.8 (N1 response). Due to the superposition of different components, the N2 lost most of its discriminative power and is not shown here.

As expected, the P3 response was superimposed on the early components that were rhythmically evoked by the physical processing of the stimuli. It caused the negative deflections to miss a cycle in condition C175 (see Figure 3.7-B). Spatially, the discriminative information for the P3 in condition C175 was found more frontally, centered over channel *Cz*. In condition C300, the P3 response had more time to develop, which resulted in a more typical positive going potential (see Figure 3.7-A). It also resulted in stronger, and more parietally centered discriminative information. Clearly, the P3 carried more discriminative information for condition C300 than for condition C175.

For both conditions, the N1 component was found over frontal electrodes, spreading to the temporal electrodes covering the primary auditory cortices (see Figures 3.8-a and b). Interestingly, the N1 component was similarly binary-class discriminative in both conditions.

Classification Performance Individual and averaged binary-classification performances can be found in Table 3.4, along with the selection accuracy and the ITR. When considering all 15

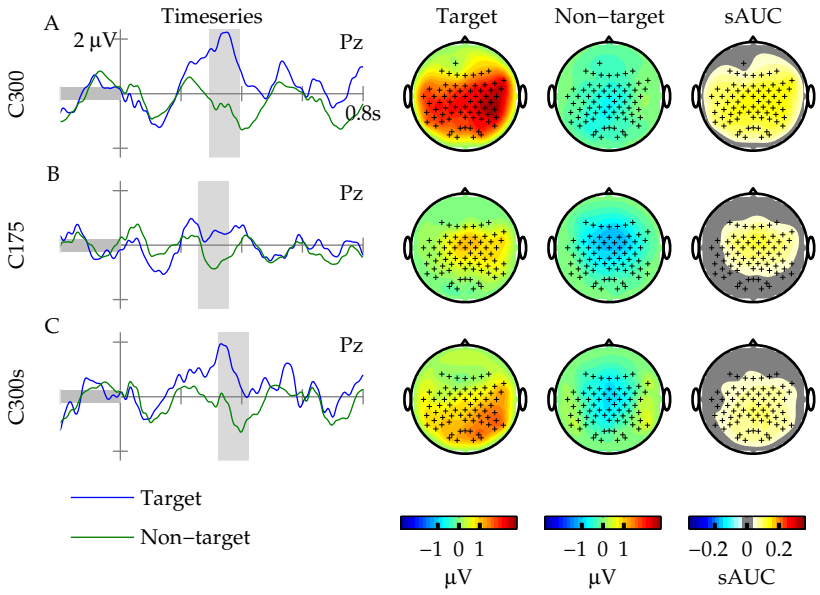


Figure 3.7.: The grand average P3 response is plotted for each condition of Experiments 3 and 4. All three conditions result in discriminative information. The temporal response for condition C175 is reduced due to the interruption by early components. See Figure 2.4 for a full explanation of the figure elements.

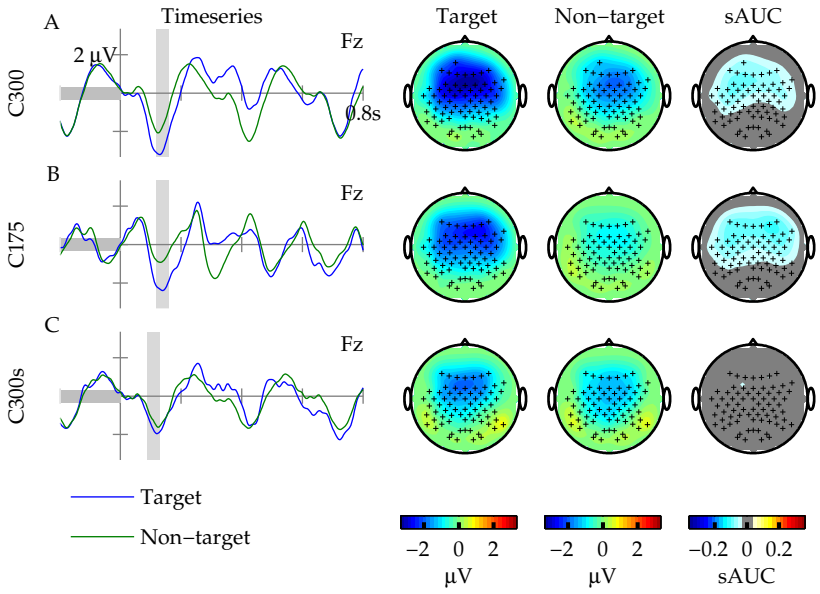


Figure 3.8.: The grand average N1 response is plotted for each condition of Experiments 3 and 4. Discriminative information only remains for the conditions using spatial stimuli (A and B). See Figure 2.4 for a full explanation of the figure elements.

Table 3.4.: BCI performance results for Experiments 3 and 4. All scores were calculated using the maximum of 15 iterations. All datasets were reduced to the size of the smallest dataset.

		Subject					
		1	2	3	4	5	Average
Experiment 3							
C300	Binary accuracy	66.4 %	76.5 %	71.3 %	63.1 %	71.4 %	69.7 %
	Selection accuracy	90 %	100 %	90 %	65 %	90 %	87 %
	ITR [bits/min]	3.6	5.1	3.6	1.5	3.6	3.5
C175	Binary accuracy	65.9 %	77 %	64.5 %	67.8 %	67.2 %	68.5 %
	Selection accuracy	90 %	100 %	80 %	95 %	75 %	88 %
	ITR [bits/min]	5.5	7.7	4	6.4	3.3	5.4
Experiment 4							
C300s	Binary accuracy	55.3 %	68.4 %	57 %	53.2 %	57.3 %	58.2 %
	Selection accuracy	35 %	90 %	40 %	45 %	55 %	53 %
	ITR [bits/min]	0.2	3.6	0.3	0.5	0.9	1.1

iterations, subjects performed similarly on condition C300 (mean binary accuracy: 69.7 %; mean selection accuracy: 87 %) and condition C175 (mean binary accuracy: 68.5 %; mean selection accuracy: 88 %). On the contrary, in terms of ITR, condition C175 (mean: 5.4 bits/min) outperformed condition C300 (mean: 3.5 bits/min) by about 54 %.

Figure 3.9-A and B show the performance evolution over iterations in terms of ITR for conditions C300 and C175. The performance for condition C300 peaked at iteration four with an average of 7.7 bits/min. The performance for condition C175 also peaked at iteration four, but with an average of 8.0 bits/min. From these initial peaks, performance dropped with additional iterations for both conditions. The time added by performing more iterations (reducing V in Equation 2.10) outweighed the slight increase in selection accuracy from that point on. Condition C175 outperformed condition C300 on all but the first two iterations in terms of ITR, with similar selection accuracies.

One subject outperformed all others on condition C175, with a peak ITR of 13.66 bits/min (*VPzq*). This subject reported to regularly sing in a choir and to have a trained musical ear.

3.3.7.3. Summary

With the quicker BCI settings, the amount of discriminative information that is contained in each single epoch is reduced. The relative contribution of the early N1 component increases, with the P3 being interrupted by early components of following stimuli. Nevertheless, the task relevant selection accuracy is good for both conditions (average >70 %). The quicker C175 condition performs best in terms of ITR on almost any number of iterations.

3.3.8. Experiment 4: Single Direction Control Condition

In Experiments 1 and 2, all stimuli consisted of the same band-pass filtered noise. The spatial location was thus the only possible cue to elicit the ERPs. However, in the quicker Experiment 3, a unique tone was added to each direction, providing a second discriminative cue. The ERPs were thus elicited by a mixture of both cues. The current experiment investigates the contribution of the spatial cue to the ERPs and classification performance.

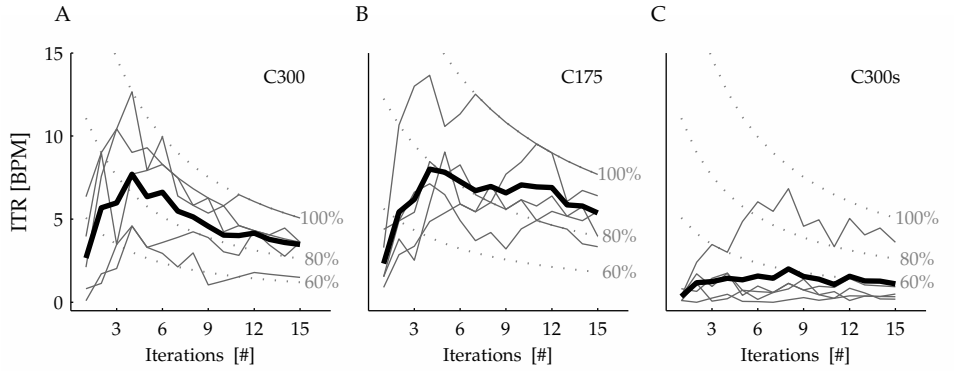


Figure 3.9.: The performance in terms of ITR is plotted against the number of iterations for conditions C300 (A), C175 (B) and C300s (C) and for all subjects (grey lines). The black lines represent the average performances. For conditions C300 and C175, using the AMUSE principles, a sharp rise in ITR is present peaking at four iterations. Such a rise is not visible for the control condition. The high performing subject in conditions C175 is the same as in C300s.

3.3.8.1. Setting

For this experiment, a single condition was tested (C300s). Like condition C300 of Experiment 3, it had a SOA of 300 ms. However, all stimuli were presented through a single speaker (front), thereby leaving the pitch properties of the stimuli as the only discriminating cue. All other parameters remained the same as in condition C300. 20 trials of 75 epochs were recorded for this condition, making a total of 1500 epochs. Experiments 3 and 4 were recorded in the same session, and blocks of all conditions were interleaved.

3.3.8.2. Results

Physiology The grand average physiological responses are shown alongside those of Experiment 3 in Figure 3.7-C and Figure 3.8-C, for the P3 and N1 component, respectively.

The temporal shape of the P3 component was similar to that of condition C300, with similar latency and amplitude. Spatially, the discriminative information was located over the parietal region. The discriminative power was largely reduced compared to that of condition C300. This detrimental effect was even greater for the early N1 component. The temporal traces for target and non-target epochs nearly overlapped, resulting in an almost complete loss of discriminative information.

Classification Binary- and selection accuracy, as well as ITR, can be found in Table 3.4. With an average binary accuracy of 58.2 % and selection accuracy of 53 %, condition C300s performed considerably worse than the spatial condition C300. Even though the binary accuracy was only marginally above the chance level of 50 %, the selection accuracy rose well above the chance level of 20 % due to the accumulation of evidence. However, it is typically not sufficient for controlling a BCI application. The ITR reflected this, with only 1.1 bits/min on average.

Despite the low average ITR, one subject managed to complete the task with very high accuracy and a peak ITR of 6.8 bits/min. Not surprisingly, this was the same subject as the high performer in condition C175 of Experiment 3 (*VPzq*), who had a trained musical ear. The spatial cue was of less importance for this subject.

3.3.8.3. Summary

By changing condition C300 on one key aspect, removing the spatial cue from the stimuli, the contribution of the spatial property on the performance was investigated. All performance measures dropped dramatically in comparison to condition C300. Physiologically, in particular the early N1 component lost most of its discriminative power, with the P3 component being strongly reduced. The resulting selection accuracy is unlikely to drive a BCI. The spatial feature is thus key in enabling the multiclass auditory paradigm to control a BCI application.

3.4. Discussion

In this chapter, a new auditory BCI paradigm is proposed and its basic principles are carefully validated with healthy subjects. In contrast to most contemporary auditory BCI paradigms (Bentley et al., 2008; Farquhar et al., 2008; Furdea et al., 2009; Guo et al., 2009; Hill et al., 2005; Hong et al., 2009; Kanoh et al., 2008; Klobassa et al., 2009; Kübler et al., 2009; Lopez et al., 2008; Sellers et al., 2006a), AMUSE results in high performances, and at the same time it is intuitively multiclass and can readily vary in the number of classes.

3.4.1. On Spatial Stimulation

From Experiment 1 it is clear that the task is behaviorally feasible, with an average error rate over all subjects of 6.4 % on an eight class setup (see Table 3.1). Using an equidistant source distribution, there is a 45 ° azimuth angle between any two neighboring speakers. The discrimination for such neighboring sources was generally very good, with a median F-score of .93. As several studies have shown that the spatial acuity (MMA) of the human hearing system is about an order more precise than 45 °, even for azimuth positions that typically have a lower than optimal resolution (Blauert, 1996; Mills, 1958), this finding may seem trivial. However, important experimental differences exist between the early psychophysics studies (see Section 3.1) and AMUSE.

For one, AMUSE was tested in a normal room without extensive sound-attenuating measures, whereas most psychophysical MMA experiments are performed in anechoic, sound dampened chambers with minimal ambient noise (Bennemann et al., 2013; Brungart et al., 1999; Mills, 1958; Wightman et al., 1999). Stimulus echo and in particular ambient noise can degrade localization performance (Giguère et al., 1993; Good et al., 1996; Hartmann, 1983). Since they are both an integral part of any out-of-the-lab BCI, AMUSE has to deal with them. Results of Experiment 1 clearly show that AMUSE is robust to such unfavorable conditions. The choice of stimuli (broadband noise with sharp onset) possibly aided in the localization task by eliciting the precedence effect (Wallach et al., 1949, or see Section 2.1), though the frequency spectrum could be broadened into the HFR to facilitate monaural cues (Langendijk et al., 2002).

A second important difference is the stimulation sequence. In typical psychophysics studies, a single (pair of) sound(s) is presented and subjects are requested to localize these. In ERP-based BCI, and thus in AMUSE, stimulation consists of a rapid train of successive stimuli. This requires extended attention and working memory resources, as the target needs to be kept in memory during the entire sequence. This can result in reduced amplitude of the P3 component (Kotchoubey et al., 1996), which was argued to lead to reduced performance in non-spatial auditory BCI (Kübler et al., 2009). Though this claim can not formally be tested for AMUSE without a control condition, the presence of strong grand average ERP components for Experiment 2 suggests that these required resources are available for most of the tested healthy users (see Figures 3.4 to 3.6).

Experiment 1 clearly exposes an adverse effect for sources that lie on a cone of confusion (see Figure 3.3). The cone of confusion is caused by ambiguities in binaural features. The user thus

only has the less salient monaural features available for localization, and both stimuli may appear to come from a common source (see Section 2.1). This is the case for source pairs 2 & 4 and 6 & 8, and in particular for 1 & 5 (see Figure 3.1-A). The severity of the latter pair was also reported verbally by several subjects and contradicts previous findings (Carlile et al., 1997; Makous et al., 1990). The F-score for pairs on the cone of confusion (median: .80) is significantly lower than that for neighboring pairs. A simple remedy is to restrict all speakers to the frontal half, as done in Experiment 3 and Belitski et al. (2011). When strictly adhering to the horizontal plane, this leads to a decreased angle between neighboring sources. If a large number of sources is required, sources with different elevation angles could be added to sustain the number of classes without compromising the spatial resolution of human hearing. Even so, care has to be taken in placing the sources, as the cone of confusion extends to elevated sources (see Section 2.1). Another remedy is to introduce additional cues to the stimuli, as done in Experiment 3 and the next chapters.

Results of the control condition (Experiment 4) show a dramatic effect of removing the spatial component, both on the resulting ERPs and on all performance metrics. Most subjects are unable to accurately distinguish the stimuli without the spatial cue. Clearly, the spatial property adds vital information to the quick, complex tone stimuli, allowing for a fast multiclass BCI. Other studies have shown that a multiclass, covert attention-based BCI can also be achieved without spatial cues, using natural sounds (Furdea et al., 2009; Guo et al., 2010; Guo et al., 2009; Klobassa et al., 2009; Kübler et al., 2009; Sellers et al., 2006a; Yoshimoto et al., 2012). After the introduction of AMUSE, yet others have investigated a combination of both spatial cues and natural sounds, combining the best of both worlds (Gao et al., 2011; Höhne et al., 2012a; Ruf et al., 2013).

3.4.2. Neurophysiology

Stimulation in Experiment 2 lead to strong and mostly typical ERP components. The significant reduction in P3 amplitude from parietal- to frontal channels that was found, has consistently been reported throughout literature (Allison et al., 2006; Gonsalvez et al., 2007; Gonsalvez et al., 1999; Squires et al., 1975). In grand average, the three investigated components were typical: a parietal P3 (Squires et al., 1975), a fronto-central N2 (Halder et al., 2013; Pritchard et al., 1991), and a central N1 (Davis, 1939). Though individual differences exist, the components were largely stable over subjects.

In Experiment 2, target stimuli elicited P3 components with high amplitude and just over 400 ms latency. The P3 component was highly binary-class discriminative. P3 latency appears lower in both conditions of Experiment 3 (compare Figures 3.4 and 3.7). P3 latency is known to be correlated with task difficulty, with more difficult tasks causing longer P3 latency (Polich, 1987). It could thus indicate the beneficial effect of double cueing –as both spatial location and pitch were now indicating a class– and the reduced class number. However, the interruption of normal P3 development by the early components of following stimuli, in particular the N1, is another likely explanation for this latency shift. It is impossible to disentangle these causes based on the current data . Interestingly, though the N1 shows no attention effect in condition C300s (see below), the P3 component is still binary-class discriminative, albeit much reduced compared to condition C300. Binary-class separation is thus largely a cognitive effort in Experiment 4.

As with the P3, Experiment 2 resulted in pronounced and highly discriminative N2 component. Due to the latency, distribution, target dependency and experimental task it is most likely the N2c subcomponent (Halder et al., 2013; Pritchard et al., 1991), reflecting target recognition and processing. It was almost completely unrecoverable at the faster stimulation speeds in Experiments 3 and 4, possibly due to the overlap with the P3 or N1 components from stimuli directly before and after the epoch.

In Experiment 2, the N1 component is clearly visible. It is binary-class discriminative, though less so than the later components. The N1 depends highly on physical stimulus properties such

as pitch and in particular stimulus intensity (Näätänen et al., 1987). However, both pitch and intensity were equal for all stimuli in Experiment 2, and the target changed with each trial. The selective attention on spatial location alone thus induces binary-class discriminative differences as early as the N1. Interestingly, much unlike the later component, the N1 keeps most of its discriminative power for the faster conditions of Experiment 3, thus becoming increasingly important for the classification task. The positive effect of faster stimulation on the selective attention modulation of the N1 component was previously described (Näätänen et al., 1987; Schwent et al., 1976). Though pinpointing this effect on faster stimulation is impossible here due to several confounds (stimulus type, number of classes), the current findings are in line with these early results. The N1 has completely lost its discriminative power in Experiment 4. In a multiclass setting, spatial location thus provides a stronger physical feature for early selective attention than does pitch alone (see also Höhne et al., 2012b).

It should be noted that in Experiment 2 the peaks were selected based on the timeseries themselves, whereas for Experiments 3 and 4 the discriminative data were used (see Section 2.5.1). Furthermore, though the N1 and P3 components have similar discriminative power for the selected intervals in Experiment 3, the P3 component extends beyond the interval, whereas the N1 component is sharp. Averaging over broader intervals could lead to more robust discrimination for the P3.

3.4.3. BCI Performance

After careful validation of the behavioral and physiological responses to AMUSE, the settings were changed in Experiment 3, to resemble those typically found in a BCI. Conditions C300 and C175 resulted in similar performances, both in terms of binary- and selection accuracy. Average binary accuracy is clearly above chance for both conditions (C300: 69.7 %; C175: 68.5 %). The evidence accumulation introduced in Section 2.4.5 (specifically in Equation 2.4) increases the SNR and improves the selection accuracy further (C300: 87 %; C175: 88 %). The selection accuracy is not only above the chance level of 20 %, but a selection accuracy >70 % can likely support a BCI speller application (Kübler et al., 2001). The SOA of condition C175 is almost half that of condition C300, which results in a much higher ITR for condition C175 (C300: 3.5 bits/min; C175: 5.4 bits/min).

When AMUSE was introduced in 2009 (Schreuder et al., 2009), the results for condition C175 marked a sharp increase compared to contemporary BCIs, both in selection accuracy and ITR compared to contemporary auditory BCI (Farquhar et al., 2008; Hill et al., 2004; Kanoh et al., 2008; Sellers et al., 2006a). In fact, at the time of writing it is still competitive with current auditory BCI (see Table 1.1 and Hill et al., 2012a; Hill et al., 2012b; Kim et al., 2011b; Lopez-Gordo et al., 2012; de Vos et al., 2014). Interestingly, a number of studies included spatial location in their stimuli after the introduction of AMUSE, most of which show higher than average ITR performances (Belitski et al., 2011; Cai et al., 2013; Gao et al., 2011; Höhne et al., 2010b; Höhne et al., 2011b; Käthner et al., 2013; Nambu et al., 2013; Ruf et al., 2013). This again indicates the vital information that can be added to auditory stimuli by including spatial separation.

Given the selection accuracy and ITR, condition C175 appears to be most suited for BCI applications. Interestingly, several subjects reported that the lower SOA helped them focus on the task at hand. This effect is known in psychophysics literature as stream segregation or streaming, and is part of the human capacity of auditory scene analysis (Bregman, 1990). When confronted with a repetitive stream of tones (for instance: high-low-high-high-low-high) at moderate speed, the listener will perceive a single coherent “galloping” rhythm. However, by increasing the speed, the listener will start perceiving two segregated streams, one of high and one of low tones (Bregman et al., 2000). The same phenomena is likely to play a role for AMUSE; at increased speeds the listener can start perceiving the targets as a single stream, making the task easier.

By looking at the selection accuracy at different number of iterations, the theoretical potential

of AMUSE was estimated at 7.7 and 8.0 bits/min for conditions C300 and C175, respectively (see Figure 3.9). Viewing the data in this manner is common practice in BCI literature (for some examples, see Acqualagna et al., 2010; Furdea et al., 2009; Guo et al., 2009; Schaeff et al., 2012; Treder et al., 2010), and even though it can over-estimate the true performance (see Chapter 6), it clearly shows that less than 15 iterations could be sufficient. In fact, an ITR of 8.0 bits/min is close to what is currently achievable with covert attention-based visual BCIs (though the latter would likely benefit similarly from a post-hoc selection of the number of iterations).

The selection accuracy, the ITR and the number of classes of AMUSE are all major improvements over contemporary covert attention-based BCIs (see bold printed studies in Table 1.1). Prior to AMUSE, covert attention, ERP-based BCIs were mainly binary in nature, with the exception of the four-class BCI in Sellers et al. (2006a). Here, AMUSE was proven to work under BCI conditions for five classes, and clearly discriminable components were found for up to eight classes with higher SOA. The highest achieved ITR for covert attention-based BCI before AMUSE was estimated to be around 1.71 bits/min (Hill et al., 2004). The current results thus show a three-fold improvement in terms of ITR. It is important to note again that the majority of ITR values in Table 1.1 were recalculated to include overhead, something that is typically ignored in offline studies.

4. Writing With AMUSE

The previous chapter shows the beneficial effect of spatially separated sources in principle. The next step in the BCI design chain (see Section 1.6) is that of an online validation study. Therefore, AMUSE is combined with a speller interface in online mode, to create an AT application for communication. It allows a subject to write words by attending to spatially distributed stimuli, mapped to the characters of the alphabet. Concretely, it addresses hypotheses *iii* (*with spatial stimuli, a covert attention-based, auditory multiclass BCI can be realized*), and *iv* (*spatial stimuli allow for useful BCI performance in a strictly auditory setting for healthy subjects and end-users*) for healthy subjects only.

This chapter is based on previous publications, in particular Höhne et al. (2010a), Höhne et al. (2011a), Höhne et al. (2011c), Schreuder et al. (2011c), Schreuder et al. (2010b), and Schreuder et al. (2010c)

4.1. Prior Work

However important communication is, not all BCI paradigms are designed for this purpose. Several studies have focussed on much simpler BCIs, aiming to elicit a low-bandwidth but robust response (Halder et al., 2010; Hill et al., 2012b; Sellers et al., 2006a). Given their different purposes, they will not be considered in this chapter. Furthermore, the focus will be on covert attention-based BCIs.

The covert attention-based visual BCIs described before are mostly designed for spelling purposes and tested online (Acqualagna et al., 2013; Aloise et al., 2012; Liu et al., 2011; Schaeff et al., 2012; Treder et al., 2011). In accordance with their high selection accuracy and ITR, they achieve writing performances of up to 1.88 char/min (Aloise et al., 2012; Schaeff et al., 2012; Treder et al., 2011). Though this is low when compared to current overt attention-based BCIs (see Section 1.4.1), it is unparalleled by auditory or tactile spellers.

The only spelling application operated by a tactile BCI was recently reported on, achieving a writing speed of roughly 1 char/min (van der Waal et al., 2012). Thus far, the result for auditory BCI is still below that. Using natural sounds to represent the rows and columns of a spelling matrix, Furdea et al. (2009) and Klobassa et al. (2009) achieved writing speeds of around 0.16 and 0.11 symb/min, respectively. Recently, Käthner et al. (2013) proposed a portable BCI-based on virtual spatial cues, achieving a writing speed of roughly 0.37 symb/min. Clearly, the auditory spellers perform worse than their tactile and covert attention-based visual contemporaries.

Höhne et al. (2011b) proposed an auditory BCI paradigm based on stimuli differing both in virtual spatial location and pitch. Both features represented orthogonal dimensions in a 3×3 matrix, for a total of nine classes. Using a T9-based application intelligence for word prediction, a writing speed of 0.89 char/min was achieved.

It is worth noting that the early work on BCIs based on slow cortical potentials (SCP), not relying on overt attention, two end-users with advanced LIS were able to select a symbol in approximately two minutes, or 0.5 char/min. Though they required between 220 and 260 training sessions, and three other end-users did not reach sufficient control, it shows that most current non-visual covert attention, ERP-based BCIs have progressed only marginally beyond this.

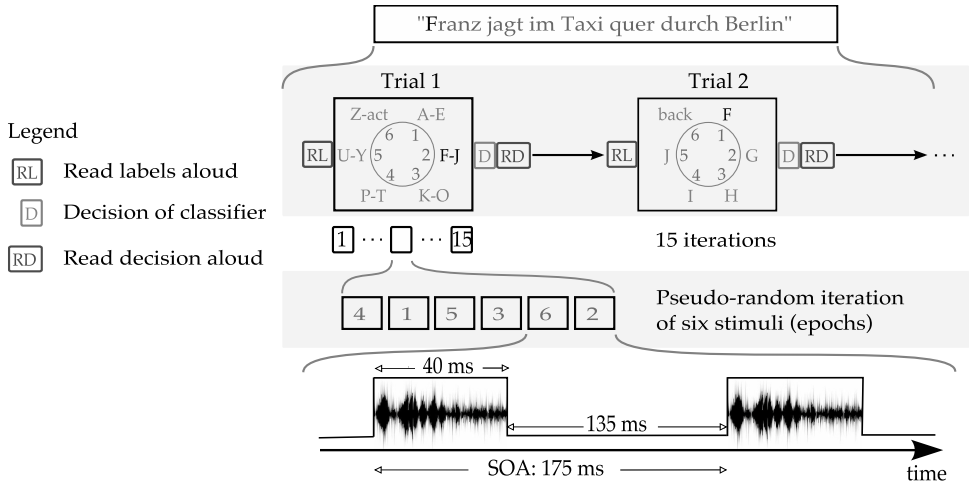


Figure 4.1.: Visualization of the experimental protocol for online writing. Subjects wrote one of two German sentences. The speller consisted of two levels. In the first level, a set of characters was selected; an individual character of that set could be selected in the second level. A fixed number of 15 iterations was used. The class labels and classifier decisions were communicated to the subject verbally, using speech synthesis software by Schröder et al. (2003).

4.2. Spelling Interface Principles

For online writing with AMUSE, an adapted version of the six class hex-o-spell speller (Blankertz et al., 2006; Müller et al., 2006) was created, driven by six sound sources (see Figure 3.1-C). Characters can be selected in two steps. First, a group of characters is selected (group 'F-J' in Figure 4.1) by focusing on the corresponding source. In the second step, the characters of this group are divided over five of the sources and an individual letter can be selected. Choosing the sixth source takes the subject back to the first selection step, which can be used to undo an erroneous decision in the first step.

The copy text and writing progress were presented visually as support (see Figure 4.2). Which (group of) letter(s) corresponded to each source, and which source had been selected, was read out to the subject using speech synthesis software (Schröder et al., 2003). The application could thus be operated with minimal reliance on the visual modality. The letters were distributed over the sources in alphabetical order. From an information theory perspective there are more optimal distributions imaginable, but they would increase the cognitive demand on the subject, as the ordering would not be "logical".

In contrast to most ERP-based BCI studies, the application did not operate in copy-spelling mode, but rather in quasi-free mode. Subjects were given a sentence to write and any committed errors had to be corrected. An erroneously selected group (level one error) could be undone by a single *'undo'* selection on level two. On the other hand, an incorrectly written letter (level two error) required two additional selections for deletion. All selections are taken into account for calculating the selection accuracy. However, the task relevant char/min metric only measures how many correct letters are spelled per time unit. Any corrective selections add time and thereby decrease the char/min score, as no additional correct letters are written. This protocol is rather conservative compared to the traditional copy-spelling mode, where errors are not corrected and no corrective selections are considered for calculating the spelling speed.

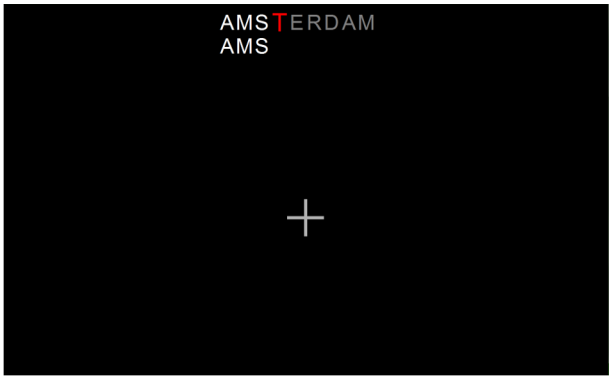


Figure 4.2.: Shown is the screen as the subject would see it during a run. Subjects did not receive visual guidance during online writing, other than the word they had to write and their progress. Additionally, a fixation cross is show; subjects were requested to focus the cross during a trial.

4.3. Writing With AMUSE

The ultimate validation of a BCI is its online application. As AMUSE refers to the mere concept of the spatially distributed stimuli, it was combined with a spelling interface to form a spelling application. This compound system was validated online.

4.3.1. Experiment 5: A Fully Auditory Speller

An AMUSE-based spelling application was tested in online mode with 21 healthy subjects. Subjects received immediate feedback on the classifier's interpretation of their intention, expressed through their ERP responses. Such an online validation is important for at least two reasons: 1) as all experimental parameters have to be fixed based on the calibration data, over-fitting on the online run is prevented, and 2) it considers the subject in the loop, who adjusts his mental state according to the feedback he receives. For these reasons, all reported performance measures are taken from the online runs unless stated otherwise.

4.3.1.1. Setting

In addition to the basic settings described in Chapter 2 and Section 3.2.1, the experimental settings were as follows.

Participants Participants were 21 BCI-naive subjects. They reported no current or prior neurological disorder and normal hearing. The latter was not formally tested. Subjects were financially compensated for their participation with eight euro per hour. Age ranged from 20 to 57 (mean: 34.1; SD: 11.4). Procedures were approved by the Ethics Committee of the Charité University Hospital. All subjects provided verbal and written informed consent and subsequent analysis and presentation of the data was anonymized.

Stimulation A six source setup was used (see Figure 3.1-C). Sources were evenly distributed in a circle, with 60° azimuth angle between them to stay well over the resolution of human spatial hearing (see Section 2.1). The pseudo-random stimulus sequence was generated with two constraints: (1) there were at least two other stimuli between two stimuli from one speaker, and (2)

Table 4.1.: The stimulus for each source was a complex of a tone (Base tone) with harmonics and band-pass filtered white noise (Noise range). The noise component extended into the HFR for most components. The source numbers refer to the speaker locations in Figure 3.1-C.

Source	Base tone [Hz]	Noise range [KHz]
1	762	3.6 - 9.2
2	528	3.2 - 8.0
3	1099	4.0 - 10.5
4	635	3.4 - 8.6
5	915	3.8 - 9.8
6	440	3.0 - 7.5

two successive stimuli never came from neighboring speakers. The SOA was set to 175 ms, which emerged as the best setting in Experiment 3. The circle radius was reduced to ~65 cm, based on subjective experience.

Stimuli consisted of a low frequency tone with seven harmonics and a high frequent, band-pass filtered noise overlay (see Table 4.1) similar to those of Experiments 3 and 4. The noise component of most sources now extended into the HFR (Langendijk et al., 2002). Though the six source setup reintroduces the front-back confusions, it is counteracted by double cueing with the additional tone.

Protocol Subjects were first familiarized with the sounds, after which about 30 minutes of calibration data were recorded. Subjects performed 48 trials, eight for each source. At the start of a trial, the target source was indicated both visually (one of six on-screen dots highlighted) and auditorily (by playing the source-specific stimulus from the target source). After that, stimulation was purely auditory and subjects were asked to mentally count and report the number of target appearances. Each calibration trial consisted of 15 iterations. However, in order to have a varying number of targets, and thus a more challenging task, a prequel of varying length (1-3 iterations) was added.

After calibration, subjects were asked to write one of two sentences word for word: “FRANZ JAGT IM TAXI DURCH BERLIN” or “SYLVIA WAGT QUICK DEN JUX BEI PFORZHEIM”. Sentences were randomly assigned to the subjects. Trial length was fixed to 15 iterations and no prequel was included. The stimulation sequence and the overhead of verbal label- and result presentation collectively took around 34 seconds per trial. With a two step letter selection process, the maximum theoretical writing speed is thus .89 char/min.

Data Acquisition EEG was recorded using a fixed set of 61 Ag/AgCl electrodes and BrainAmp amplifiers (Brain Products, Munich, Germany). All impedances were kept below 15 kΩ. Only for subject *VPfce* did the impedance exceeded this threshold at the end of the session.

The set of intervals $\mathcal{T}$ that was used for classification was found using the algorithm described in Section 2.4.3, and manually adapted if necessary. Typically, a set of two to four intervals with high discriminative information content were chosen, such that both early and late components were represented. This resulted in an overall dimensionality of 112 to 224 features.

4.3.1.2. Results

Neurophysiology Online data of all subjects were pooled and the grand average ERP responses were calculated, as shown in Figure 4.3 for all three components. Similar to Experiment 3, the

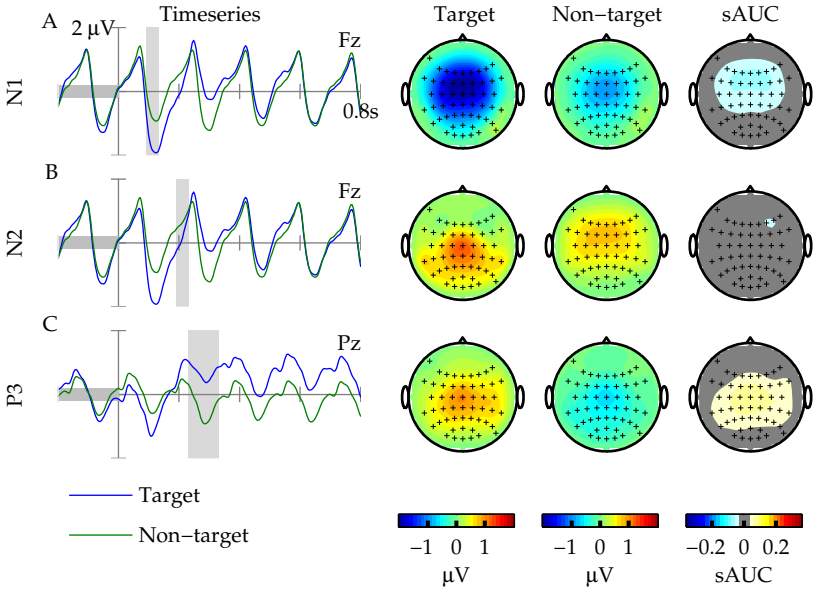


Figure 4.3.: The grand average ERP response is plotted for Experiment 5. Each row represents a different ERP component. Only the N1 and the P3 carry discriminative information. See Figure 2.4 for a full explanation of the figure elements.

P3 response is superimposed on the rhythmic perturbations caused by the early components. Discriminative information is centered around the centro-parietal region, and of similar strength as in Experiment 3. Like for Experiment 3, the discriminative informations for the N1 is located frontotemporal, and is similar in power to the P3 for the selected intervals. The N2 peak has again lost most of its predictive power.

BCI Performance All performance results are summarized in Table 4.2. BCI selection accuracy did neither significantly correlate linearly with age ($r(19) = -0.073$; $p = 0.76$), nor did it significantly differ with gender, as confirmed by a t-test ($t(19) = 0.93$; $p = 0.36$).

Selection accuracy and writing speed are visualized in Figure 4.4. Average selection accuracy was 77.4 %, as indicated by the black line in Figure 4.4-A. Of the 21 subjects, 16 subjects (or 76 %) had a selection accuracy higher than 70 % and were able to write a full German sentence at first try. The other five subjects got at most one word correct before their experiment was stopped. Figure 4.4-B shows the writing proficiency in terms of char/min, for those subjects that were able to write. The black line represents their average performance at .59 char/min.

Figure 4.5 shows the selection accuracy as a function of the word number for all subjects that were able to write. Word number is taken as an approximation of time, with word one indicating the first word that was written online. The slope of a linear least-squares fit between selection accuracy and time (blue line) does not differ significantly from zero ($r(110) = 0.096$; $p = 0.31$), indicating that the performance did not increase over time.

In terms of ITR, the average performance over all 21 subjects was 2.72 bits/min (maximum 4.82 bits/min), which includes also subjects that did not reach a performance level appropriate for writing (see Table 4.2).

Table 4.2.: Performance summary for Experiment 5. The performance measures selection accuracy (Sel. Acc.), char/min and ITR are given for all subjects and all sentences, along with the age and gender of the subject. Subjects are sorted according to their selection accuracy. Missing values for the char/min (-) indicate unfinished sessions.

User	Gender	Age	Sel. Acc. [%]	char/min	ITR [bits/min]
VPfaz	F	27	98.4	.83	4.79
VPfcc	M	28	98.4	.84	4.82
VPkw	F	57	97.0	.81	4.57
VPfca	M	23	94.7	.77	4.18
VPfcd	M	25	92.9	.77	4.02
VPfaw	M	40	91.7	.73	3.82
VPfar	M	31	85.6	.58	3.12
VPfav	M	40	84.9	.59	3.04
VPfcb	M	21	84.7	.49	3.01
VPfck	M	43	83.2	.50	2.91
VPfau	F	25	81.9	.50	2.78
VPfcj	F	23	80.6	.41	2.65
VPfcg	F	33	79.5	.40	2.57
VPfch	M	55	79.3	.46	2.57
VPfcm	M	51	78.3	.47	2.55
VPfax	M	28	72.1	.32	1.97
VPfci	M	28	58.6	-	1.21
VPfat	F	25	54.1	-	.96
VPfcl	F	44	51.7	-	.83
VPfce	M	47	40.0	-	.42
VPfas	F	29	37.5	-	.33
Means	F:8/21	34.1	77.4	.59	2.72

Directional Preference Figure 4.6 displays the matrix of pair-wise F-scores (see Section 2.6.1). Directions without any degrading factor had a median pairwise F-score of .84 (IQR: .023, N: 4). F-scores for spatial neighbors (median: .88, IQR: .055, N: 12) and pitch neighbors (median: .83, IQR: .022, N: 10) had similar F-scores, but the front-back confusion (median: .71, IQR: .083, N: 4) appeared to have a degrading effect on the pairwise discriminability. Significance testing using Bonferroni corrected Kruskal-Wallis H tests ($\alpha = 0.002$) did not reveal significant differences.

From the pitch neighbors, the highest and lowest pairs were confused most often. Subjective reporting of the subjects was in line with this result.

Questionnaires Subjects were in good mood (VAS-mood, median: 8.2, IQR: 2.8) and were highly motivated (VAS-motivation, median: 9.0, IQR: 1.23). Median overall workload was rated at 49.67 (IQR: 29.0), with the main contributions coming from *mental demand* (median: 20.0, IQR: 17.0) and *effort* (median: 10.0, IQR: 14.67). Subjects rated the four motivation factors of the QCMBCI questionnaire as follows: *mastery confidence* (median: 5.5, IQR: 2.06), *incompetence fear* (median: 1.8, IQR: 1.2), *interest* (median: 5.6, IQR: 2.2), and *challenge* (median: 5.0, IQR: 1.19).

Subjects were split in two groups, based on their ability to write, and all factors were compared independently using Kruskal-Wallis H tests. Figure 4.7 shows the result of this split on the questionnaire scores. Overall workload was rated significantly higher ($\chi^2(1) = 6.14$, $p = .01$) by the subjects that could not write (median: 75.33, IQR: 15.66) than by those that could write (median:

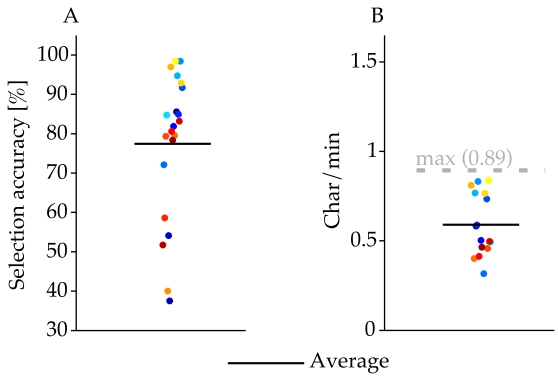


Figure 4.4.: The online writing performance, as expressed in selection accuracy (A) and char/min (B). The latter only includes subjects for which the char/min could be calculated. The black lines indicate the average performance. For the char/min, the theoretical limit is given, based on the SOA, overhead and number of iterations.

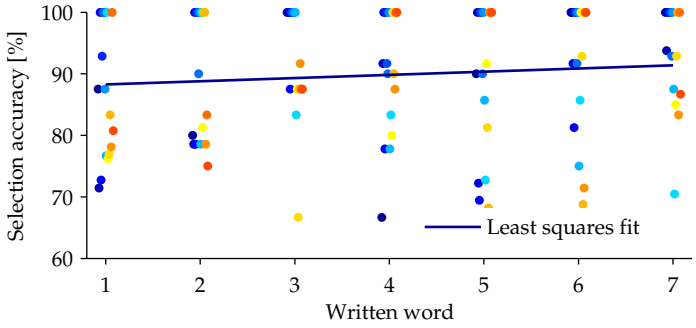


Figure 4.5.: The selection accuracy is given for all written words for all subjects that could write. Word number is taken as an approximation of time, with word one referring to the first word that was written online. The blue line indicates the least-squares fit over these points. The non-significant slope indicates that no learning effect could be found within the first session.

45.84, IQR: 20.17). Several factors contributed significantly to this increase in overall workload; *Performance* (Write median: 4.0, IQR: 4.00; No write median: 16.0, IQR: 12.25; $\chi^2(1) = 3.97$, $p = 0.046$), *Effort* (Write median: 9.33, IQR: 9.16; No write median: 24, IQR: 7.5; $\chi^2(1) = 4.11$, $p = 0.043$), and *Frustration* (Write median: 1.33, IQR: 4.67; No write median: 8, IQR: 7.5; $\chi^2(1) = 6.02$, $p = .014$).

Neither the VAS nor the QCM-BCI questionnaire showed any significant difference that could explain the (in)ability to write.

Approximating the Online Char/min The relation between the selection accuracy, char/min, and symb/min metrics was investigated offline. In Figure 4.8-A, the online char/min scores are plotted as a function of the selection accuracy for each subject (dots). Additionally, the theoretical upper boundary (only level one errors, striped line), and the theoretical lower boundary (only level two errors, striped-dotted line) on the symb/min are plotted, along with the case of equal error distribution on both levels (solid line). The empirical char/min falls within the lower range of the symb/min. The black line in Figure 4.8-A is the least-squares fit over the non-zero char/min

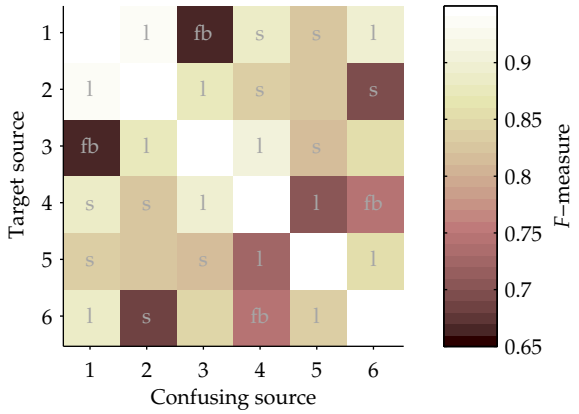


Figure 4.6.: The grand average source discrimination performance is depicted as a confusion matrix containing F-scores (see Section 2.6.1). Lower F-scores indicate worse performance. Again, sources with a front/back confusion (fb) are clearly more difficult to separate. Note that these data represent the results from EEG classification, and not behavioral data.

performances as a function of their selection accuracy. Clearly, a selection accuracy that is at least approximately 60 % is needed to write with the spelling interface.

In Figure 4.8-B, the online char/min is plotted against the symb/min, as calculated post-hoc on the same data. Though the symb/min overestimates the char/min performance, the fitted line is parallel to the diagonal line, indicating a uniform bias over all values of char/min. Thus, though slightly biased, the symb/min is a good estimator for the online char/min.

4.3.1.3. Summary

Using the AMUSE paradigm in an online setting resulted in similar, though less discriminative ERP components as those found in Chapter 3. The speller based on this setup allowed 16 out of 21 subjects to write a complete German sentence. Subjects that were unable to write reported a significantly higher perceived workload, in particular for the factors *Performance*, *Effort*, and *Frustration*.

4.4. Discussion

In this chapter, the principles of AMUSE were coupled to a spelling interface. It allowed the majority of healthy subjects to write a full sentence with an average speed of .59 char/min, thereby confirming the validity of AMUSE as a BCI paradigm. This speed is competitive with contemporary auditory BCI spellers (see bold printed studies in Table 1.1). For the sake of comparison, it is important to once more emphasize the conservative, but realistic nature of the protocol; all mistakes had to be corrected by the user, at the expense of time.

4.4.1. Directional Preference

Learning from Experiments 1 to 4, the stimuli were optimized by adding a unique tone to each source. Additionally, the noise spectrum for most sources extended into the HFR to further

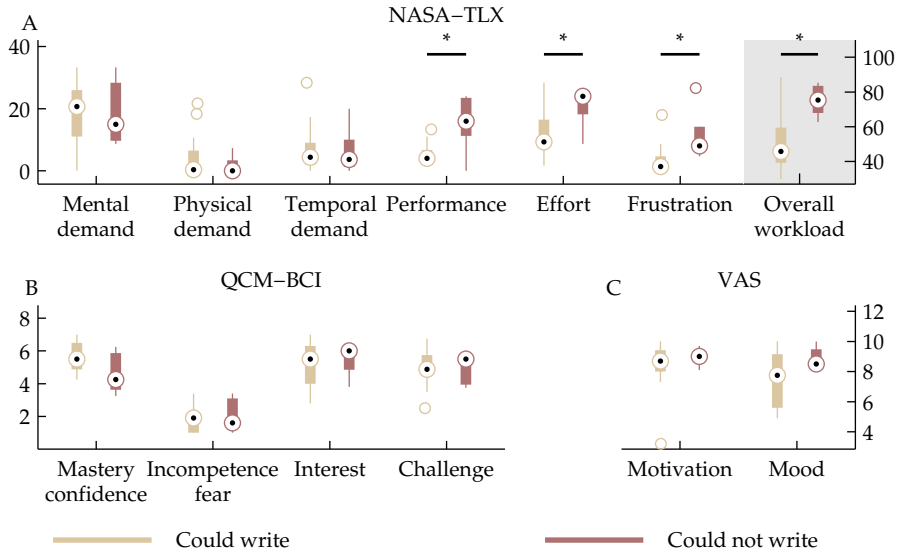


Figure 4.7.: Results for all factors of the questionnaires are compared between healthy subjects that could write and those that could not write. Lower is better for all factors of the NASA-TLX questionnaire and factor *incompetence fear* of the QCM-BCI questionnaire. Higher is better for all other factors. Significant differences are indicated with an asterix (* $p < 0.05$, ** $p < 0.01$).

facilitate separation (see Table 4.2). Unfortunately, these changes could not completely prevent the cone of confusion effect; errors due to the front-back confusions were still most frequent. A direct comparison between Figures 3.3 and 4.6 is difficult, as the first is based on behavioral data and the second on the classifier scores based on EEG data. The intrinsically noisy EEG channel adds to the confusion, resulting in lower overall F-scores. However, the reduction in F-scores for front-back pairs compared to those of pairs without detrimental factors for Experiment 1 (20 % reduction) and Experiment 5 (15.5 % reduction) indicates an improvement in resolving front-back pairs. Unfortunately, this can not be attributed to the improved stimuli alone, as several confounds exist (number of classes, offline versus online, and increased workload due to the spelling task).

It is known that for posterior sources, the power in the high frequencies is lower than for the same sound from frontal sources (Blauert, 1996). However, the frontal sources (1 and 6) both had lower frequency ranges in the stimulus complex than their respective confusers in the back (sources 3 and 4), which may have contributed to the remaining confusion. Changing the order of sounds to prevent this could lead to better separation. Furthermore, the equidistant source distribution that was arbitrarily chosen for Experiments 1 to 5 led to sources on the cone of confusion. By changing the angle between source pairs, a layout could be achieved that uses the entire space around the subject, without putting sources on the cone of confusion.

4.4.2. Neurophysiology

Much like in Experiment 3, AMUSE elicited discriminative N1 and P3 components in the current experiment. The N2 component is again unrecoverable. Given the high similarity in stimulation between condition C175 of Experiment 3 and the current experiment, the similarity in ERP

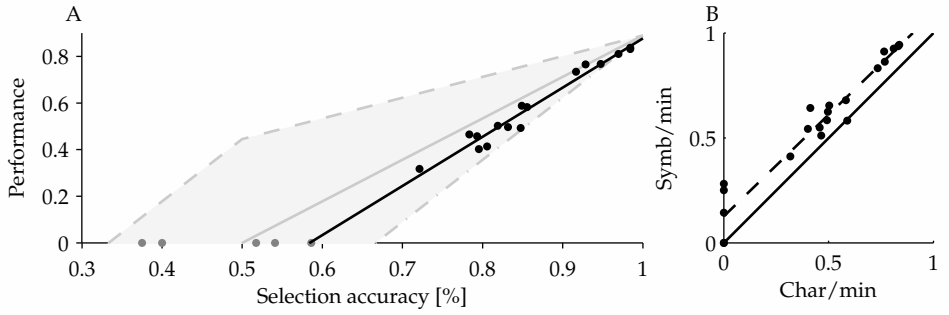


Figure 4.8.: Investigation of the quality of symb/min as an approximation of char/min. A) Both the online char/min and the range of theoretically possible symb/min values are plotted as a function of the selection accuracy. The symb/min is given as a range, since errors can be made mostly on level 1 (striped line), level 2 (striped-dotted line) or equally on both (solid line). B) The symb/min is estimated for each subject and plotted as a function of the true char/min.

components may not be surprising. Nevertheless, Figure 4.3 is the grand average based on data taken from online writing sessions. The task in Experiment 3 was simply to attend to a given target, whereas the task in Experiment 5 included identification of the next target and keeping track of the position in the sentence (though largely helped in both by the verbal and visual labels). It was shown that workload can hamper the elicitation of ERP components (Ullsperger et al., 2001). However, in the current experiment most of the added workload falls in between trials; the central task of detecting the target (once identified) remains the same, explaining the high similarity in ERP components between this experiment and Experiment 3.

4.4.3. Performance

By adding a spelling interface on top of AMUSE, 16 out of 21 subjects were able to write a full German sentence. Only subjects that had a selection accuracy larger than 70 % were able to write. Interestingly, this threshold for communication was previously suggested by Kübler et al. (2001), albeit for a very different setup (a binary tree speller based on SCP). Nevertheless, this threshold is widely used in BCI literature for all sorts of paradigms. It appears to be roughly correct for the AMUSE-based speller (though a linear regression suggests that selection accuracies higher than 60 % would suffice).

Those subjects that could write, did so on average with 0.59 char/min, which is about 66 % of the theoretical maximum speed. To put it differently, they could write approximately one symbol every 1:40 minute. Furthermore, the visual presentation of the speller interface, as done in Furdea et al. (2009) and Klobassa et al. (2009), was replaced by auditory labels before each trials. Though this increased the time needed per trial, the result is a speller that is almost completely independent of the visual modality. The latter was confirmed by the successful participation of a single blind person for demonstration purposes (she was not formally part of Experiment 5). The high time overhead of verbal label presentation let to an average ITR that, with 2.72 bits/min, is below the results obtained offline in condition C175 of Experiment 3. Nevertheless, individual subjects obtain scores that come close to the average predicted ITR of 5.4 bits/min.

With 0.59 char/min, the online results for AMUSE represented an increase in performance over most contemporary non-visual, covert attention-based BCIs (see Table 1.1). Nevertheless, the tactile speller reported by van der Waal et al. (2012) and the T9-based extension of AMUSE reported

by Höhne et al. (2011b) both resulted in higher writing performances. Both show the potential for improvement on the current results, for instance by using application side intelligence (see Chapter 6).

According to Guger et al. (2009), over 90 % of participants are able to control their ERP-based BCIs. While that may be true for the overt attention-based, visual paradigms they tested, it may be a too optimistic estimate for ERP-based BCIs in general. Those BCI paradigms are still most prevalent in literature, but they may not match fully with the needs and abilities of all end-users (Ramos Murguialday et al., 2011; Treder et al., 2010; Brunner et al., 2010). Furthermore, the learning effect is generally small for such overt attention-based ERP BCI paradigms, and average performance is high and consistent over subsequent trials. However, Klobassa et al. (2009) showed that the performance could increase significantly over sessions for their auditory ERP-based paradigm. Such a learning curve was neither observed within Experiment 5 (see Figure 4.5), nor between Experiments 5 and 8 (the latter included those subjects from Experiment 5 that could write). Nevertheless, it is possible that the percentage of people able to use AMUSE would increase with more training.

4.4.4. Questionnaires

The result of the NASA-TLX is useful in comparing different systems, as its outcome can only be interpreted in relative terms. Unfortunately, it has only recently been introduced in BCI research as a tool to measure workload, and has been employed in a limited number of studies (amongst which Aloise et al., 2012; Carlson et al., 2013; Combaz et al., 2013; Felton et al., 2012; Gürkök et al., 2011; Holz et al., 2013a; Käthner et al., 2013; Legeny et al., 2013; Riccio et al., 2011; Rohm et al., 2013; Zickler et al., 2013; Zickler et al., 2011). Though single subject / single session overall workload as low as five has been reported (Zickler et al., 2013), the typical range of reported workload scores does include the workload score for AMUSE.

To investigate the difference in the users' state and perceived workload between subjects able and unable to write, those groups were compared on each factor of the applied questionnaires individually. The overall workload and three individual factors of the NASA-TLX questionnaire differed significantly between these groups. Subjects that were unable to write reported a significantly higher overall workload (median: 75.33) than those that were able to write (median: 45.84). The large increase in workload was mainly due to the factors *Performance*, *Effort* and *Frustration*. Käthner et al. (2013) found a negative correlation between accuracy and overall workload for their auditory BCI. The current results indicate the same effect. However tempting, assuming a causal relationship in either direction would be unsubstantiated without further experiments. Such experiments would be of high value for the future design of BCIs.

5. AMUSE End-user Tests

The concept of AMUSE was designed to increase the usefulness of covert attention-based BCI. Results from the first two stages of the BCI design chain have confirmed the beneficial effect that adding spatial information to the stimuli in auditory BCIs has on the performance. As covert attention-based BCIs specifically target end-users with severe limitations in their daily living, the next step in the BCI design chain is testing the paradigm with this end-user population. Such a test should not only assess AMUSE in its current form with the target population, but it should also generate valuable end-user data for future designs following the user-centered design principles (see Section 1.6).

In this chapter, AMUSE is tested with five end-users over several sessions, using a spelling interface similar to that of Chapter 4. The obtained results of a single end-user are subsequently put into context with data taken from a different study in which she was enrolled (Riccio et al., 2012b; Schreuder et al., 2012a). The latter investigated an overt attention-based visual BCI, and gives valuable complementary insights into her performance.

This chapter is based on previous publications, in particular Quek et al. (2012), Riccio et al. (2012b), Schreuder et al. (2011b), Schreuder et al. (2012a), Schreuder et al. (2013b), and Tangermann et al. (2011)

5.1. Prior Work

BCI has been heralded as the ultimate technology for locked-in patients, and the vast majority of BCI literature states to serve this ultimate goal. Despite this, only a minority of published work is actually tested with those end-users. Kübler et al. (2013) estimated that less than 10 % of BCI related publications involve end-users (for reviews on end-user studies, see Kübler et al. (2008) and Mak et al. (2011)). The majority of proposed BCI algorithms and paradigms thus do not pass beyond the first two stages of the BCI design chain (see Section 1.6). This is likely due to the limited access to such end-users that exists for many BCI research groups. It has lead to a translational gap where the end-user relevance of proposed AT oriented BCIs remains unclear (Kübler et al., 2013). But the importance of such tests goes beyond the validation of a novel method; end-user tests produce invaluable data on the end-users' needs and abilities that can be used for future design decisions in BCI research.

Nevertheless, considerable evidence has been provided that BCIs can work for these end-users. Using SCPs, Birbaumer et al. (1999; 2000) were able to provide severely paralyzed end-users with a slow, but functional communication setup called the "*Though Translation Device*". Several end-users could independently use the system for writing with a speed of approximately 0.5 char/min. Because of the inherent slowness of this signal, researchers subsequently turned to other brain signals to increase the writing speed.

Different degrees of successful control by end-users have been reported for oscillatory features (ERD, Cincotti et al., 2008b; Holz et al., 2013a; Kübler et al., 2005; Neuper et al., 2003; Pfurtscheller et al., 2003; Rohm et al., 2013; Tonin et al., 2011; Wolpaw et al., 2004), (mainly visual) ERP (Donchin et al., 2000; Escolano et al., 2010; Hoffmann et al., 2008; Kaufmann et al., 2013a; Marchetti et al., 2013; Mugler et al., 2010; Münßinger et al., 2010; Nijboer et al., 2008a; Piccione et al., 2006; Sellers et al., 2010; Silvoni et al., 2009; Townsend et al., 2010; Zickler et al., 2011) and SSVEP (Combaz

et al., 2013). Furthermore, BCI technology is slowly moving out of the lab and into the end-users' home (Münßinger et al., 2010; Sellers et al., 2010; Vaughan et al., 2006).

Still, those studies that directly compared healthy users and end-users, consistently found a drop in performance for end-users (Cincotti et al., 2008b; Donchin et al., 2000; Mugler et al., 2010; Münßinger et al., 2010; Piccione et al., 2006; Sellers et al., 2006a). In some studies, the achieved control level could not sustain the intended task (Combaz et al., 2013; Hill et al., 2006; Kübler et al., 2009; Ortner et al., 2013).

With some exceptions, the studies included end-users that had only low or moderate impairment in daily life. Success in studies that tested end-users in CLIS has been limited (Kübler et al., 2008). The needs and abilities of this group of end-users that require a BCI the dearest may differ from end-users with LIS or even less impairment (Ramos Murguialday et al., 2011). Covert attention ERP-based BCIs were intended specifically for this end-user group, but only a handful has thus far been tested with any group of end-users, LIS or CLIS.

Marchetti et al. (2013) recently showed that their covert attention-based visual BCI could be operated by all ten end-users diagnosed with ALS. However, the level of impairment was minor for eight end-users and moderate for the remaining two. The four choice tactile BCI presented in Kaufmann et al. (2013a) resulted in discriminative features that could be classified offline with an average selection accuracy of 90 % on two conditions. Unfortunately, results in a preliminary online test were poor. The single near-LIS end-user that was included had only transient thumb and vertical eye movement remaining after brainstem stroke. Similarly, Sellers et al. (2006a) reported a four class auditory BCI tested with three end-users diagnosed with ALS that suffered from a varying degree of disability (though not in LIS). End-users achieved an average selection accuracy of 65.9 % for the auditory condition, relying on spoken stimuli. Finally, Kübler et al. (2009) reported on an auditory speller that was tested with four end-users diagnosed with ALS. Two end-users were nearly in LIS whilst the other two had residual motor control. Average selection accuracy for the first two end-users was 0 %, whereas the latter two had an average selection accuracy of 25 % and 23.52 %. Stimuli consisted of spoken numbers.

The above results indicate the importance of testing a paradigm with the end-user, in particular with end-users that are in a state that requires a BCI-based AT, to identify and mitigate the weaknesses of the paradigm.

5.2. AMUSE End-user Validation

Two experiments constitute the end-user tests for AMUSE. First, in Experiment 6, a speller similar to the one introduced in Chapter 4 is tested with five end-users with a high degree of paralysis. In Experiment 7, the AMUSE data is compared to a visual paradigm for a single end-user, along with a set of neuropsychological assessments, creating a rich account of the needs of such end-users for future BCI designs.

5.2.1. Experiment 6: End-user Validation of Online Writing

End-user tests for AMUSE were done in collaboration with the group of Dr. Donatella Mattia, head of Clinical Neurophysiology at the Fondazione Santa Lucia rehabilitation clinic in Rome, Italy.

5.2.1.1. Setting

For reasons given in Section 1.3, the patient trials included five end-users with different levels of advanced physical impairment. See Table 5.1 for a summary of the end-users demographics. Inabilities were caused by stroke (N: 2) and TBI (N: 3). At the time of the trials, four end-users were

Table 5.1.: Demographic and Clinical Data for End-users

ID	Syndrome	Age	Months post incident	Speech impaired	Movement impaired
VPro20ma	TBI	34	54	Able	Hemiplegic
VPgl29ma	TBI	34	192	Mute	Spastic tetraplegic
VPan09ma	TBI	21	30	Anarthric	Spastic tetraplegic
VPeg02ma	Ischemic stroke	46	24	Mute	Tetraplegic
VPda07ma	Hemorrhagic stroke	20	8	Mute	Paraplegic

in the chronic state and end-user *VPda07ma* was in the acute state. Given the level of disability, three out of five end-users could directly benefit from a robust BCI system in their daily activities. In other words, they are a true representation of the potential end-user of BCI technology.

The AMUSE speller setup was kept mostly as used in Experiment 5, with minor adaptations. First, the speaker positioning was slightly adjusted to compensate for the use of a headrest by several end-users and in an attempt to further reduce the front-back confusions (see Figure 3.1-D). Second, the SOA was increased to 300 ms, both at the request of the end-users and to allow the P3 component to develop beyond that of the faster condition (as was found in Experiment 3).

Furthermore, the complexity of the feedback was slowly increased within and over sessions. That way, the end-users could get familiar with the interface, without being faced with too many errors in the beginning; a modification that was done to prevent perceived failure and demotivation (Kleih et al., 2010). Three online modes were introduced. First, a *no feedback* mode was used. The online BCI was running in the background, but was not providing the user with any feedback. It allowed the experimenter to have an objective online performance estimate. Then, a *mask feedback* mode was used. It resembled the normal feedback from Experiment 5, but handles misinterpretations differently. Instead of displaying an incorrectly written letter, it displayed 1) the correct letter in white if both selections were correct, 2) the correct letter in light gray if only one of the two selections was correct, and 3) only a dot if both selections were wrong. The feedback thus focussed on the positive results, masking any errors from explicit display. Lastly, a *free feedback* mode was added, where all letters are displayed as interpreted, thus showing misspelled letters. It operated in copy-spelling mode and misspelled letters did not have to be corrected. The feedback modes, in order of complexity, were thus *no feedback*, *mask feedback* and *free feedback*. The point of transition to a more complex feedback was decided by the experimenter, and was done in any case if the user reached a selection accuracy >70 %.

A total of five sessions were intended with each end-user. Sessions were at a one week interval unless the end-user was prevented, for instance due to illness. The first session was used for introducing the AMUSE paradigm to the user and recording initial calibration data. In each of the following sessions, additional calibration data could be acquired. However, since all online tasks were essentially a version of the copy-spelling task, labels were available for the online data as well. Given the limited time that was available per user and session, these online data were typically used to recalibrate the classifier. A shrinkage regularized classifier was trained on features found by the algorithmic method (see Section 2.4).

A variable number of online runs was performed per session; each run consisted of copy-spelling a single word. The exact number depended on the available time. In addition to the questionnaires (see Section 2.5.2), the end-users were subjected to several neurophysiological and

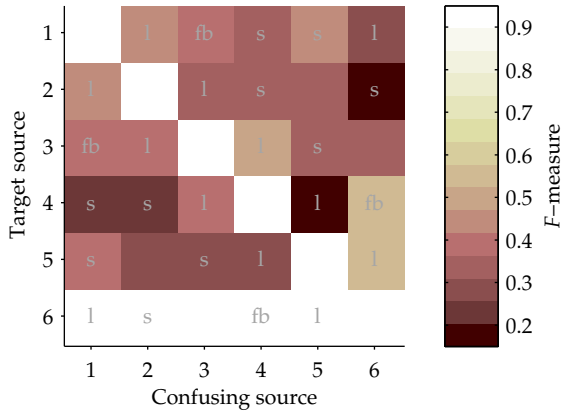


Figure 5.1.: The grand average end-user source discrimination performance is depicted as a confusion matrix containing F-scores (see Section 2.6.1). Lower F-scores indicate worse performance. The general darkness on this scale reflects the poor recognition performance. End-users did not need to correct committed mistakes. As source six was therefore almost never the target, it was excluded from further analysis.

-psychological tests (see Experiment 7 for details). BAERs were recorded once for the assessment of the brainstem function and the evaluation of the hearing ability; results were in the normal range.

Standard Auditory Oddball During the first and last session, each user additionally performed a standard, non-spatial auditory oddball task. It consisted of a two-tone discrimination task (target: 1000 Hz, 20 % probability; non-target: 500 Hz, 80 % probability). A total of 400 tones were recorded, with an SOA of 1000 ms. The end-user was instructed to focus on and count the high pitched targets. Both target and non-target tones came from source six (front-left, see Figure 3.1-D). These settings follow the recommended clinical protocol (Duncan et al., 2009).

5.2.1.2. Results

Most of the included end-users participated in parallel to their normal therapy, which greatly limited the amount of time that was available per session. On average, around 40 minutes of recording time was available per session (range: 5-75 minutes). Furthermore, some end-users were very tired during the experiments, due to the extensive physiotherapy they received during the rest of the day. For this reason, end-user *VPan09ma* discontinued his participation after four sessions. All other end-users performed five or more sessions.

Directional Preference Though formally no sources lied on a cone of confusion, subjects indicated the presence of a front-back confusion between source pairs 1 & 3 and 4 & 6. For comparison with previous experiments, the matrix with pairwise F-scores is given in Figure 5.1. Clearly, performance on all speaker pairs is heavily impaired, as indicated by the overall 'darkness' of the plot on this scale. Median F-scores ranged from 0.32 to 0.38. Group-wise testing using Bonferroni corrected Kruskal-Wallis H tests did not reveal significant differences. As no errors had to be corrected, the sixth direction was virtually never the target. The sixth row in the matrix in Figure 5.1 is empty due to this lack of data.

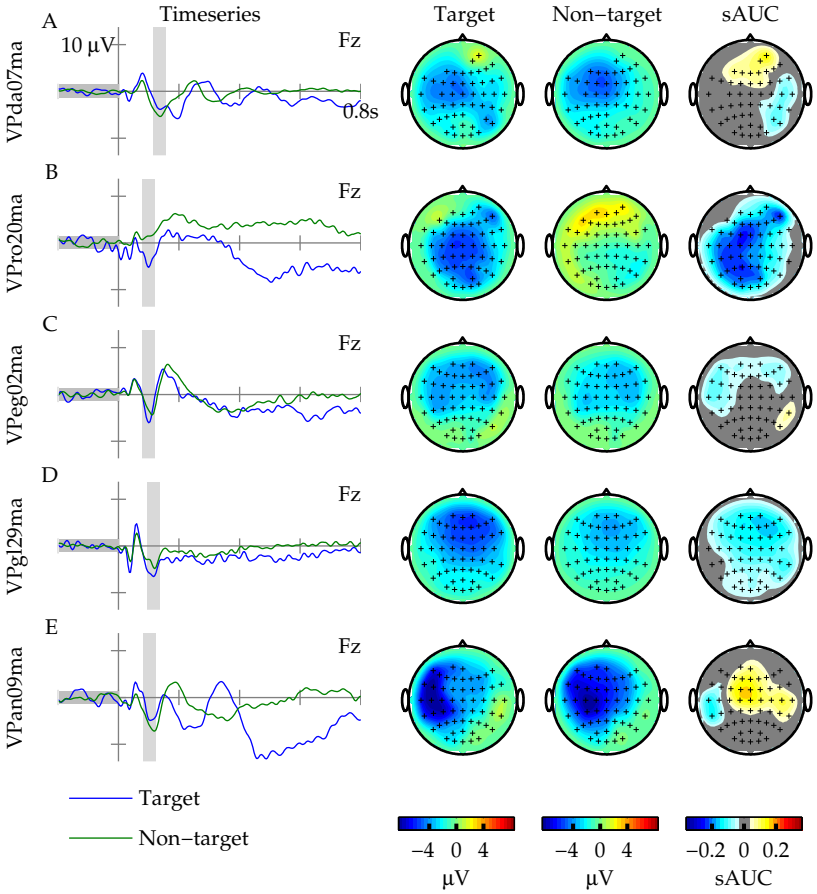


Figure 5.2.: Shown is the standard oddball N1 response for all end-users. All available datasets are collapsed for each end-user. Given the heterogeneity of the group, no grand average is presented. See Figure 2.4 for a full explanation of the figure elements.

Physiology Figures 5.2 and 5.3 show the ERP responses to the standard oddball task for the N1 and P3 components, respectively. Given the heterogeneity of the end-users' clinical profiles, no grand averages were calculated. All available standard oddball data were collapsed per end-user to increase the SNR, even if those data were recorded on different sessions.

The N1 component was visible for all subjects, but the distribution and power of the discriminative information varied greatly between end-users. For subject *VPro20ma* it was strongly binary-class discriminative, with a diffuse distribution. But also end-users *VPeg02ma* and VPgl29ma had a binary-class discriminative N1. Additionally, the N2 component was highly discriminative for subjects *VPda07ma*, *VPro20ma*, *VPan09ma* and (positively) subject *VPeg02ma* (not shown). Only end-users *VPeg02ma* and *VPan09ma* showed a component that, based on timing, polarity and spatial distribution, resembled a P3 component. For these end-users, the P3 was also highly binary-class discriminative. Target and non-target ERP responses for end-user *VPgl29ma* were highly overlapping for all components. All end-users were enrolled in the full AMUSE protocol.

Figure 5.4 is similar to Figure 5.3, but displays the AMUSE calibration data for each subject, collapsed over sessions. As with the results from the standard oddball trials, responses from

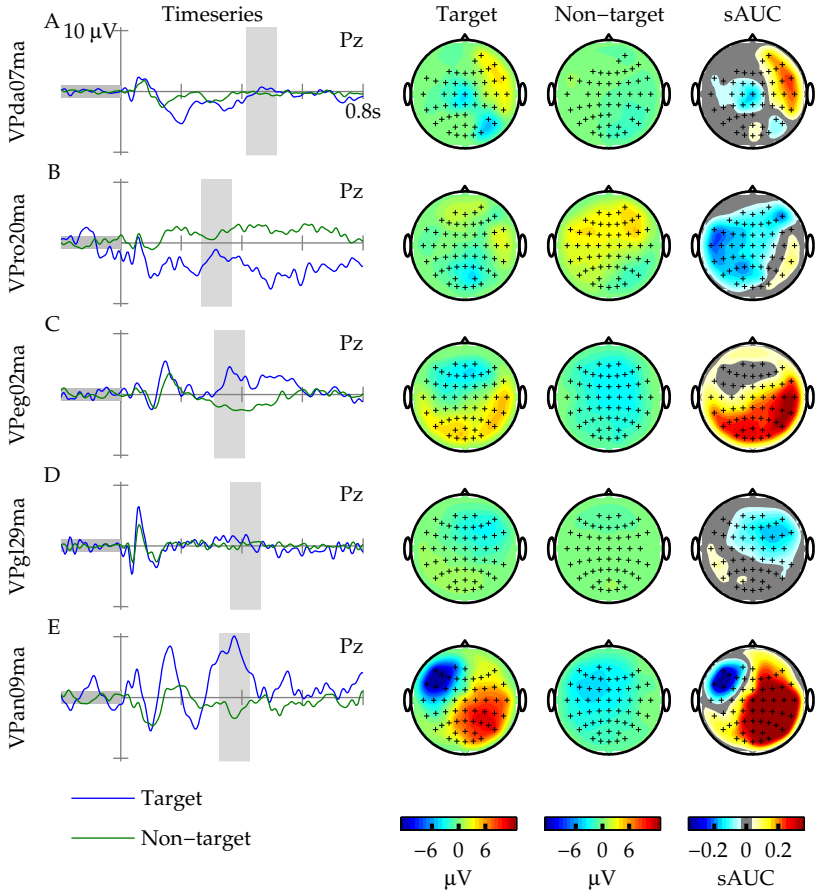


Figure 5.3.: Shown is the standard oddball P3 response for all end-users. All available datasets are collapsed for each end-user. Given the heterogeneity of the group, no grand average is presented. See Figure 2.4 for a full explanation of the figure elements.

VPeg02ma and *VPan09ma* contained a component that temporally and spatially resembled a P3, though it lost most of its discriminative power. Additionally, for end-user *VPda07ma* a diffusely distributed positive target response is found. The strong complex of early components found for subject *VPgl29ma* contained no binary-class discriminative information. In fact, the N2 again lacked binary-class discriminative information for all subjects, and the N1 contained low and diffuse binary-class discriminative information for subject *VPro20ma* only (both not shown).

BCI Performance Given the six class BCI, chance level for the selection accuracy is at 16.7 % (see Section 2.6.2). Figure 5.5 shows the selection accuracy per session for all end-users. Though most end-users consistently performed above chance level, the performance was poor and only reached a score larger than 60 % on one occasion (a score that is strictly required for writing with the AMUSE paradigm, according to Figure 4.8).

An increase over sessions is visible on the average performance line, but no conclusions can be drawn due to the low statistical power of five end-users. However, at least two end-users followed this trend (*VPan09ma* and *VPeg02ma*). They were the same end-users that have a relatively typical

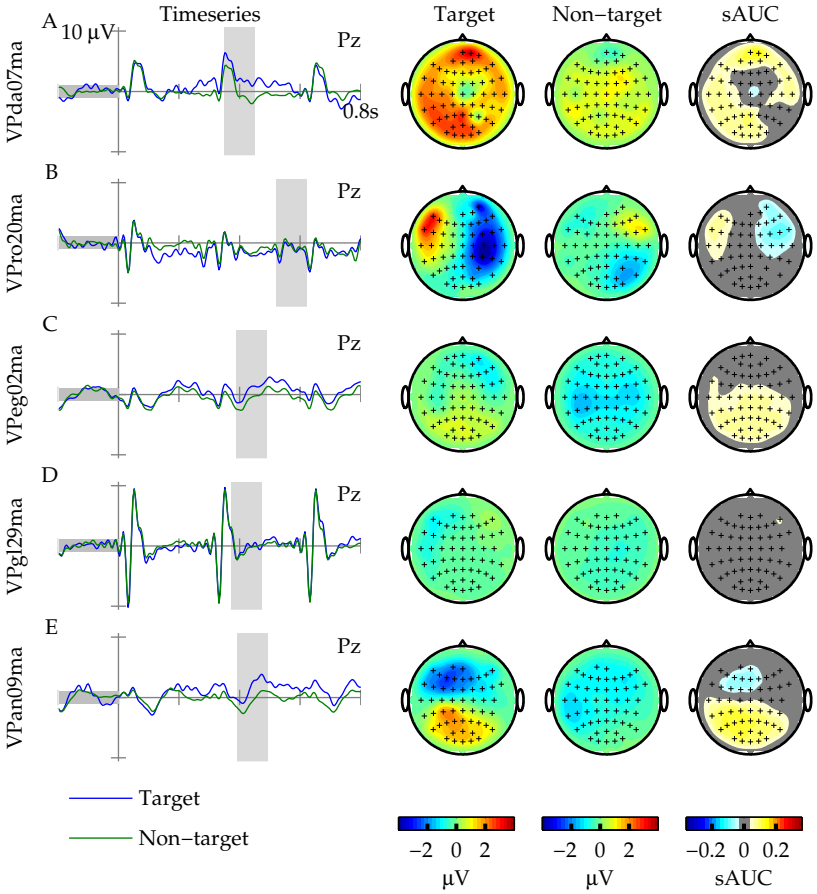


Figure 5.4.: Physiological responses for the AMUSE training trials are shown for all end-users. All available datasets are collapsed for each end-user. Given the heterogeneity of the group, no grand average is presented. See Figure 2.4 for a full explanation of the figure elements.

ERP response to AMUSE. Note that end-user *VPan09ma* only had three BCI sessions, and the average for the fourth BCI session was thus calculated over the remaining subjects. Any additional sessions by other end-users were not considered.

End-users were advanced to the feedback runs without reaching the 70 % limit, such that they could experience online feedback. End-users mainly used the *no feedback* and *mask feedback* mode. Figure 5.5 shows the data collapsed over all modes. Performances were not statistically different between the *no feedback* mode (mean: 33.7 %; SD: 13.9 %) and the *mask feedback* mode (mean: 42.6 %; SD: 18.6 %), as confirmed by a paired t-test ($t(10) = -1.76$; $p = 0.11$).

Further offline analysis of the data did not increase the performance substantially.

Questionnaires The following questionnaire results were taken from the first session where the end-user performed a trial with online feedback. End-users indicated to have average mood (VAS-mood, median: 6.90, IQR: 5.82) and motivation (VAS-motivation, median: 5.80, IQR: 6.08). Median overall workload was rated at 42.0 (IQR: 48.91), with the main contributions coming from

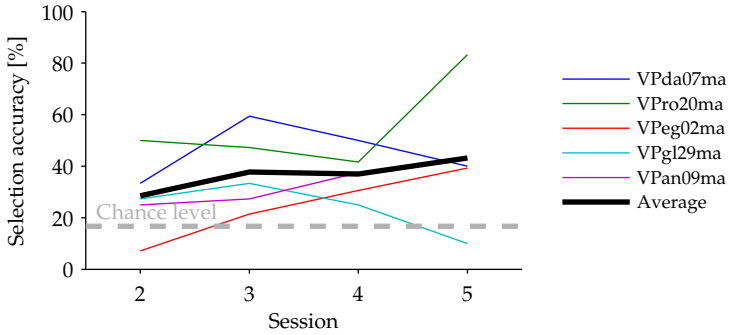


Figure 5.5.: Average online selection accuracy is given per EEG session for all end-users, along with the session average. Three out of five end-users performed above chance consistently; the remaining two end-users performed below chance on at most one session.

mental demand (median: 23.3, IQR: 20.0) and *performance* (median: 10.0, IQR: 6.0). End-users rated the four motivation factors of the QCM-BCI questionnaire as follows: *mastery confidence* (median: 5.5, IQR: 1.5), *incompetence fear* (median: 2.0, IQR: 3.3), *interest* (median: 4.8, IQR: 1.25), and *challenge* (median: 6.5, IQR: 0.81).

Figure 5.6 shows the end-user questionnaire results, together with those of the healthy subjects from Experiment 5. Factor-wise comparisons revealed a significant change in the *challenge* factor of the QCM-BCI questionnaire ($\chi^2(1) = 6.99$, $p < 0.01$), with end-users rating the motivation by *challenge* significantly higher than healthy users (median: 5.0, IQR: 1.19). Similar results were found if questionnaire data from the very first available session were taken.

None of the factors showed a significant correlation with time (session number) for any end-user.

5.2.1.3. Summary

Despite good results with healthy users, none of the five end-users was able to write with the AMUSE paradigm combined with the spelling interface. Physiological results showed atypical patterns for most end-users. Given the low performance, no real directional preference could be established, though subjective reports indicated the persistence of a front-back confusion. End-users drew more motivation from the challenge of participating than did healthy users.

5.2.2. Experiment 7: Single Subject Case-Study

The following experiment was published in an article jointly authored with Angela Riccio (Schreuder et al., 2013b), and this section would not have existed without her expert input.

A second round of end-user experiments was conducted about one year after Experiment 6, again in collaboration with Dr. Mattia's Lab in Rome, Italy. This time, a novel visual application was tested, which was based on the visual matrix speller (Farwell et al., 1988). Though the full results of that study are beyond the scope of this thesis, the fact that end-user *VPeg02ma* participated in both studies gives valuable insight into her results in Experiment 6. The current experiment thus combines data from both Experiment 6 and the visual study, in order to give an account of the needs and abilities of such an end-user (in accordance with Section 1.6).

For a better understanding of these results, a more detailed profile of end-user *VPeg02ma* is required. She is a 48 year old Italian woman. At age 44 she suffered from an ischemic stroke in the

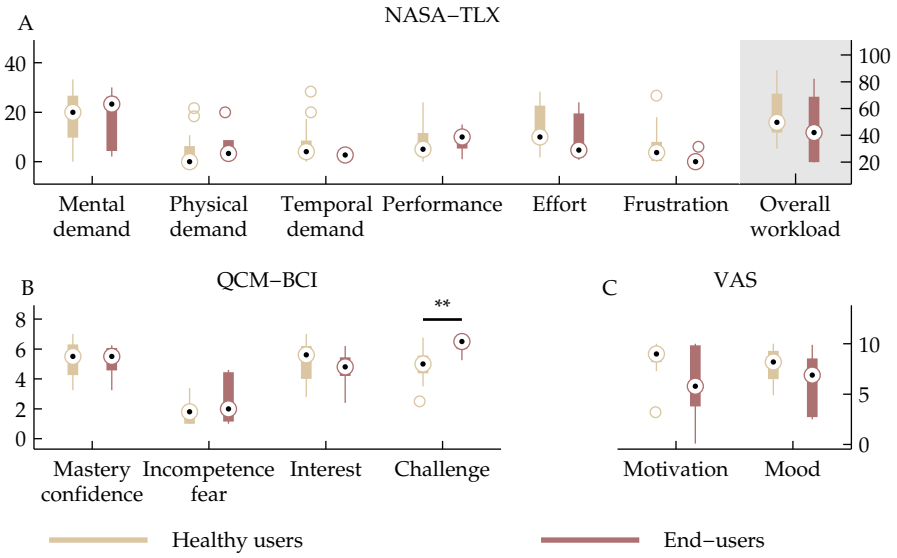


Figure 5.6.: Results for all factors of the questionnaires are compared between end-users and healthy subjects from Experiment 5. Lower is better for all factors of the NASA-TLX questionnaire and factor *incompetence fear* of the QCM-BCI questionnaire. Higher is better for all other factors. Significant differences are indicated with an asterix (* $p < 0.05$, ** $p < 0.01$).

area of the basilar artery, after which she showed a clinical picture characterized by tetraplegia and severe dysarthria. Her impairment thus led to a major lack in communication. An MRI scan, acquired ten days after the ischemic event, showed an altered signal intensity in the infero-posterior area of the left cerebellar hemisphere, in the upper area of the right cerebellar hemisphere, in the cerebellar vermis, and in the midbrain with greater extension on the left. It also showed a small hemorrhagic rift within the left cerebellar hemisphere lesion and in the central pontine area. Twenty days after the ischemic stroke, *VPeg02ma* was alert and able to localize sound stimuli by turning her eyes towards the sound source. Reasonable changes in facial expression were present. Her motor disability was characterized by motor tetraplegia with hypotonia and symmetrical generalized hyperreflexia. She was thus diagnosed with LIS. Immediately after the diagnosis, a low-tech binary model of communication was set up, exploiting eye gaze. *VPeg02ma* was trained to communicate by focusing her gaze on an item on an alphanumeric communication board, which she still used at the time of the experiments.

Before her stroke, *VPeg02ma* worked in the field of graphic arts and played the drums in a band. She was confident with using computers and other technology. At the time of testing, *VPeg02ma* had the ability to perform inaccurate movements with the right arm and the head, had preserved facial expressions and had precise eye movements. The communication of primary needs was only possible with the support of her communication board. In addition, she could acknowledge requests by a button press. *VPeg02ma* was curious and motivated to participate in BCI studies. During her participation in Experiment 6 she was 46 years old; one year later she joined the visual experiment.

5.2.2.1. Setting

For comparison with the results obtained in Experiment 6, the Photobrowser setup was used (Quek et al., 2012; Riccio et al., 2012b; Schreuder et al., 2012a; Tangermann et al., 2011). The Photobrowser is an extension of the visual matrix speller (Farwell et al., 1988). Here, the principles of the matrix speller were used for driving an application that allows the user to receive, view and share photos with friends via the internet. It is a social application that puts the end-user back in the center of his own group of significant persons. For this purpose, the characters were exchanged for objects such as images, folders and function buttons. The social aspect of the Photobrowser application was inspired by recent findings that about 20 % of end-users rank 'independent participation in social life' amongst the three most important things that would increase their quality of life (see Section 1.2, or Zickler et al., 2009).

On top of the simple brightness highlighting from Farwell et al. (1988), more complex stimuli were available that were found to be particular salient in prior experiments (Tangermann et al., 2011). *VPeg02ma* chose to use the complex stimulation and an SOA of 220 ms. She was able and allowed to direct her gaze at the target image.

Two preparation sessions and six online EEG sessions were performed. Informed consent was given during the first preparation session, and aspects of her quality of life and her use of AT were assessed by questionnaires. The second preparation session was used to introduce her to the paradigm and to collect initial calibration data. Data from this session were recorded with the same equipment as that in Experiment 6. Based on this initial recording, an optimal channel set was used for all further sessions (*Fz*, *Cz*, *CP5,1,z,2,6*, *P5,1,2,6*, *PO7,z,8*, *O1,2*), which was recorded with a g.USBamp EEG amplifier (g.Tec, Graz, Austria).

VEPs were recorded in order to detect possible lesions or inflammations of the optical nerve; the results were within the normal range.

Standard Visual Oddball A standard visual oddball assessment was applied before the first, and after the last session, using eight channels from the g.USBamp EEG amplifier (channels: *Fz*, *Cz*, *P3,z,4*, *Oz*, *PO7,8*). The standard visual oddball consisted of a set of symbols (X and O) that were serially presented in the middle of a computer screen (white on black), with an SOA of 2100 ms. The end-user was told to focus on one particular symbol (X) and count the number of occurrences, whilst ignoring the other symbol (O). Target probability was the same as for the auditory oddball (20 % target, 80 % non-target). A total of 500 stimuli were recorded.

Neuropsychological Assessment In the context of a standard clinical diagnosis process, end-user *VPeg02ma* was subjected twice to a neuropsychological assessment of general cognitive impairment, attention, memory, working memory and executive functions. The first assessment took place about four months before the start of Experiment 6 and the second assessment around six months after the last Photobrowser session. The test battery was administered in a quiet room and over several days to prevent fatigue. During both assessments, the end-user was motivated and cooperative.

Her general cognitive level was tested by means of the mini mental state examination (MMSE) (Magni et al., 1996). The two subtests of the scale that were not applicable due to *VPeg02ma*'s physical condition (spontaneous writing subtest and constructive praxis ability subtest) were not presented to her. Missing tests were scored with the maximum score. The attention capacities were evaluated using tests of divided attention and selective attention, taken from a computerized test battery (Zimmermann et al., 1995). In the latter tests, *VPeg02ma* was asked to provide an answer by pressing a key. The forward and backward digit-span tasks (Orsini, 2003) were utilized to assess verbal short-term memory and working-memory, respectively, and Corsi block-tapping test (CBTT) and Corsi supraspan block-tapping test (CSBTT) (Corsi, 1972; Spinnler et al., 1987) were

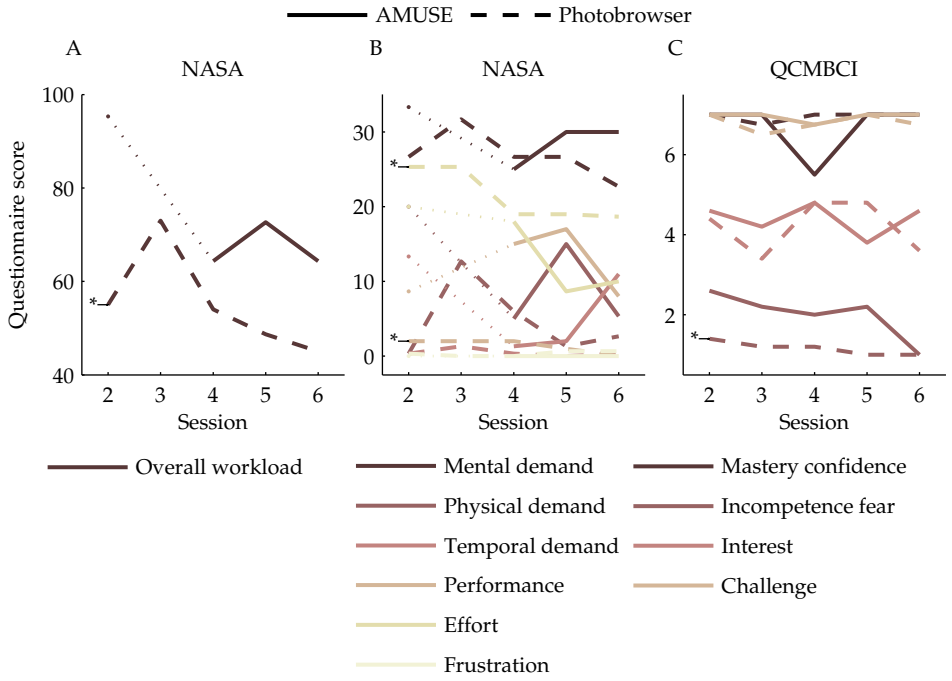


Figure 5.7.: VPeg02ma's questionnaire results for both the AMUSE and the Photobrowser sessions, over all EEG sessions. Factors that change significantly over sessions are indicated with an asterix (* $p < 0.05$). The NASA-TLX score for session three of the AMUSE experiment is missing due to time restrictions. The dotted line interpolates between the scores of sessions one and three.

used to assess visuo-spatial working memory. *VPeg02ma* was alert and able to pay attention even for prolonged periods of time. Also administered were the Wisconsin card sorting test (WCST) (Grant et al., 1948; Heaton et al., 1993; Heaton et al., 2000) to assess executive functions, and the N-back test (Zimmermann et al., 1995) to assess the control of information flow and the updating of information in working memory.

The computerized tests provide normative values for reaction times as well as for accuracies. However, only scores of accuracy were considered in drawing the neuropsychological profile, due to the slowness of *VPeg02ma's* motor response. As the end-user was unable to give a verbal response, the backward and forward digit span tests were administered by asking her to point to a series of numbers on her communication board.

5.2.2.2. Results

Questionnaires Figure 5.7-C shows that *VPeg02ma* was comparably motivation over the 5 sessions for both paradigms, as measured by the four factors of the QCM-BCI questionnaire. The difference between the scores never exceeded 1.2 points (out of 7). Though *VPeg02ma* found both paradigms to be highly *challenging*, her confidence in *mastering* them was equally high and *fear of incompetence* was minimal for both. She was moderately *interested* in both paradigms. These results thus give no systematical explanation for any performance difference. Figure 5.7-A shows that *VPeg02ma* scored the perceived overall workload, as measured by the NASA-TLX, higher for the AMUSE-based speller than for the Photobrowser for all sessions.

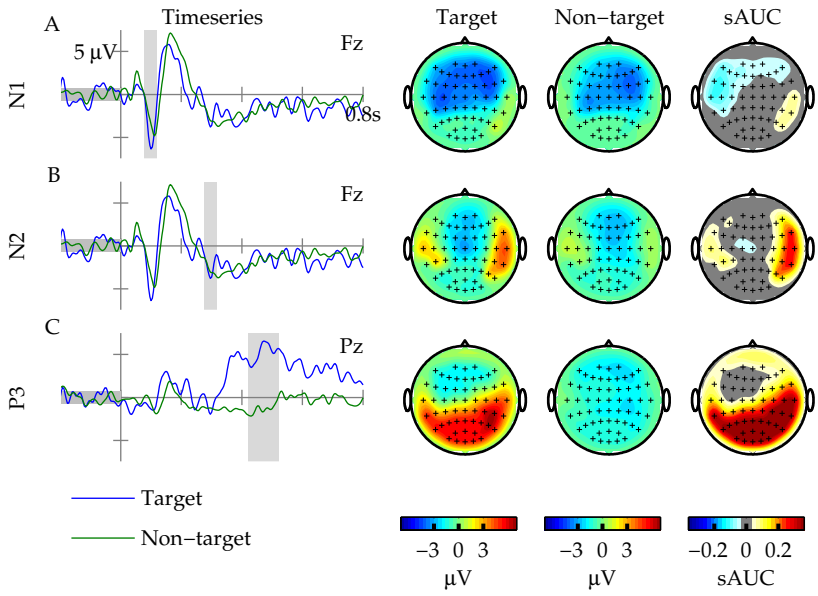


Figure 5.8.: The grand average ERP response to the standard auditory oddball is plotted for end-user *VPeg02ma*. Each row represents a different ERP component. The P3 is highly binary-class discriminative. See Figure 2.4 for a full explanation of the figure elements.

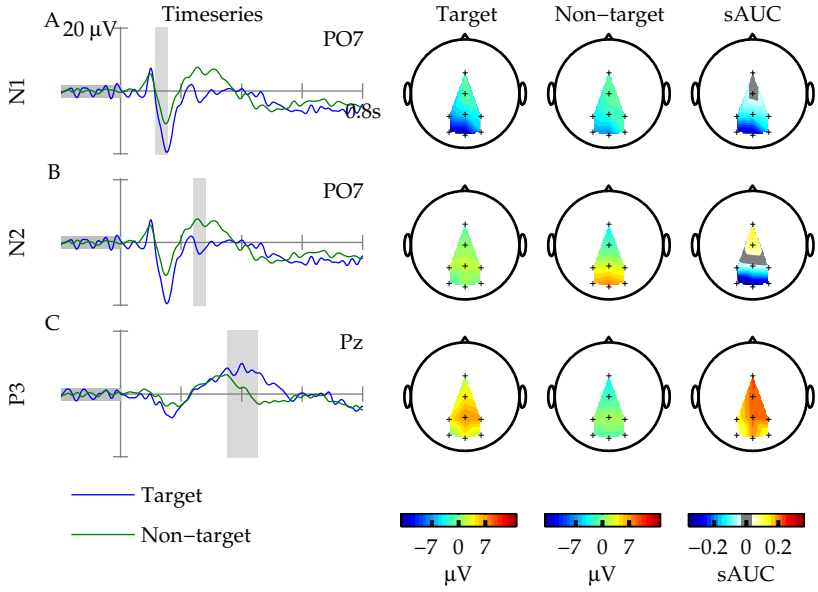


Figure 5.9.: The grand average ERP response to the standard visual oddball is plotted for end-user *VPeg02ma*. Each row represents a different ERP component. The early components are particularly binary-class discriminative. See Figure 2.4 for a full explanation of the figure elements.

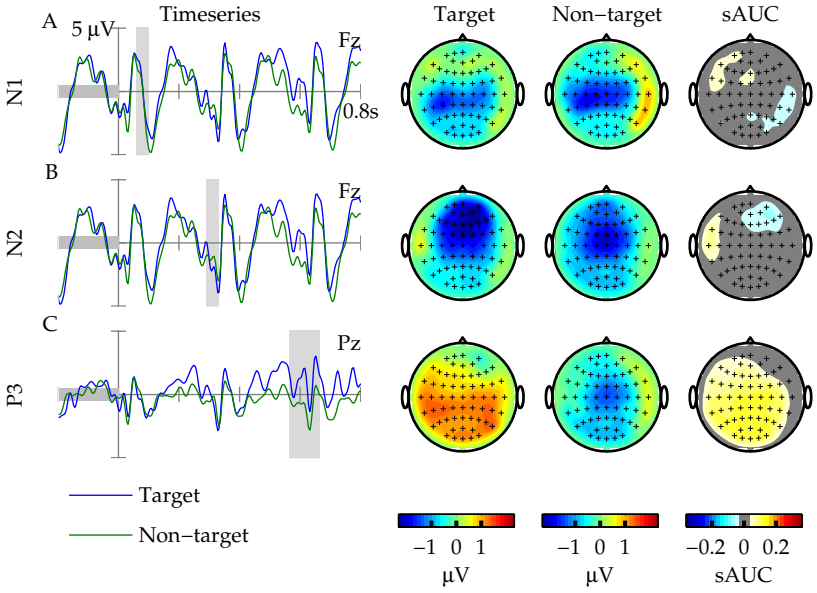


Figure 5.10.: The grand average ERP response to AMUSE is plotted for the calibration data of end-user *VPeg02ma*. Each row represents a different ERP component. The P3 lost most of its binary-class discriminative information. See Figure 2.4 for a full explanation of the figure elements.

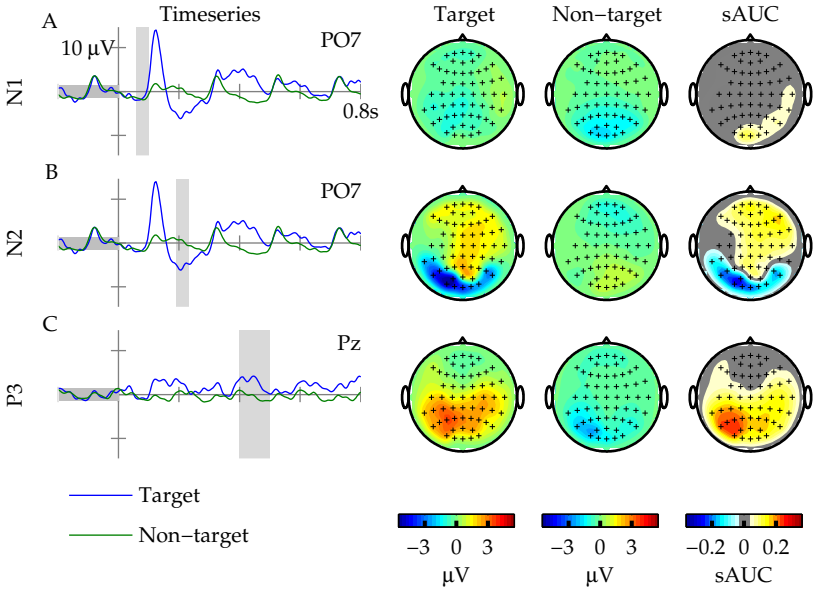


Figure 5.11.: The grand average ERP response to the Photobrowser is plotted for the calibration data of end-user *VPeg02ma*. Each row represents a different ERP component. The P3 and N2 remain binary-class discriminative, as is the additional P2 compenent. See Figure 2.4 for a full explanation of the figure elements.

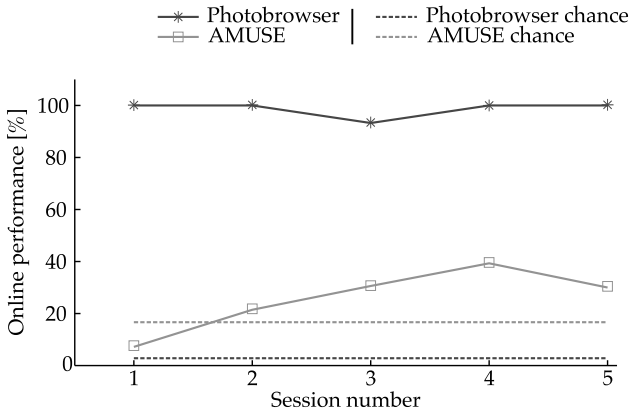


Figure 5.12.: *VPeg02ma*'s average online selection accuracy is given for five sessions for both the Photobrowser and the AMUSE-based speller. Additionally, the chance levels are given for both BCI systems. She had near perfect performance with the Photobrowser, much in contrast to AMUSE.

A non-parametric Spearman's rank correlation was calculated to address the relationship between time (session number) and 1) motivational factors rated with the QCM-BCI questionnaire and 2) overall workload and its factors obtained through the NASA-TLX questionnaire, as was done in (Nijboer et al., 2010). Significant results are indicated with a asterix (*) in Figure 5.7. Only for the Photobrowser did the results show a significant negative correlation between the session number and 1) factor *incompetence fear* ($rho = -0.95$; $p = 0.01$) from the QCM-BCI questionnaire, and factors 2) *overall workload* ($rho = -0.90$; $p = 0.04$), 3) *performance* ($rho = -.90$; $p = 0.04$), and *effort* ($rho = -0.95$; $p = 0.014$) from the NASA-TLX questionnaire.

Physiology Figures 5.8 and 5.9 show *VPeg02ma's* physiological response to the standard auditory- and visual oddball, respectively. The data are taken from the recording closest to the BCI sessions. Highlighted are the N1, N2 and P3 interval. For both modalities, typical early negative (N1 and N2) and late positive (P3) components could be evoked. For the visual oddball, the early N1 and N2 components over the occipital channels were particularly pronounced and loaded with discriminative information. In contrast, the auditory oddball resulted in a strong parietal P3 component and very little discriminative power in the early components.

Similarly, Figures 5.10 and 5.11 represent the BCI calibration data of the first EEG session for AMUSE and the Photobrowser, respectively. Discriminative information for the AMUSE paradigm was highly reduced, and low binary-class discriminative information between the target and the non-target ERPs was only found for the P3. In contrast, the visual ERP detected for the Photobrowser paradigm was characterized by strong differences between target and non-target epochs, especially in the early components. In addition to the strong N2, an occipital P2 component was clearly visible and highly discriminative.

BCI Performance *VPeg02ma*'s online performance for both experiments can be found in Figure 5.12. The performance line for AMUSE is the same as that in Figure 5.5, with one additional session. The difference in performance between both experiments was remarkable. For the Photobrowser, online selection accuracy was 100 % in four out of five sessions, with the remaining session reaching 93.2 % (chance level below 3 %). *VPeg02ma* thus had near-perfect control over the Photobrowser application. The performance for the AMUSE paradigm is above chance level in four out of five sessions, but never allowed *VPeg02ma* to write.

Table 5.2.: Results of the neuropsychological assessment for end-user *VPeg02ma*, and their respective cut-off values below which a score is marked as pathological with an asterix (*).

	Experiment 6	Visual Experiment	Cut-off
MMSE	30	30	22
Attention			
Selective attention	45	45	5
Divided attention	3*	3*	5
Memory Digit span			
Forward	8	6	5
Backward	10	10	5
CBTT	2	0*	0
CSBTT	0*	0*	0
N back (2 back)	1*	1*	5
WCST	+	+	

Neuropsychological Assessment The scores obtained by *VPeg02ma* in both neuropsychological evaluations, and the thresholds for all tests, are reported in Table 5.2. *VPeg02ma* was alert, fully cooperative, and well oriented in time and space during both evaluations. She performed normally in the MMSE, showing a normal general cognitive level. With regard to attention abilities, her level of performance was comparable between the two assessments when considering the accuracy scores. The scores obtained in the tests were reported in percentiles, and to detect the presence of an attention impairment, a fifth percentile cut-off point was used (Pero et al., 2006; Zoccolotti et al., 2000).

VPeg02ma performed above the fifth percentile for the selective attention task, showing retained capacity to select target stimuli and to inhibit the response to non-target stimuli. On the contrary, she performed below the cut-off on the tasks for divided attention, which is required to process a visual and an auditory task in parallel.

With regard to the memory abilities, both her verbal short term span (forward digit-span task; (Orsini, 2003)) and her verbal working memory span (backward digit-span tasks; (Orsini, 2003)) were within the normal range, though the first showed a decrease from the first to the second assessment. However, even if the scores obtained by *VPeg02ma* in both the forward and the backward digit span task did not fall in the pathological range, they were in the lower limit of the normal range (between the fifth and tenth percentile).

VPeg02ma's performance on the visuo-spatial CBTT task was normal during the first assessment. During the second assessment, a decrease of performance was observed, leading to a pathological score. Furthermore, *VPeg02ma* showed an impairment on the CSBTT task and in the 2-back test. Scores obtained in the WCST task were in the normal range.

Summary Contrary to her performance in Experiment 6, end-user *VPeg02ma* had near perfect control over a visual paradigm. Her motivation on both paradigms was similar, but the perceived workload was consistently higher for the AMUSE paradigm. *VPeg02ma*'s cognitive profile was characterized by deficits in divided attention, the visuo-spatial learning ability and in the capability to control the information flow and to update information in working memory. In the second assessment, a deficit of visuo-spatial memory span was also found.

5.3. Discussion

Unfortunately, AMUSE's success in the first two steps of the BCI design chain (Experiments 1 to 5) could not be transfer directly to the end-users tests. Despite several improvements to the paradigm, none of the five end-users was able to use the AMUSE driven spelling interface with the current settings. As discussed in Section 1.6, and in particular because of the success with healthy users, such an initial attempt is the start of a revision process to further adapt AMUSE to the needs and abilities of the end-users. Accordingly, the needs and abilities of a single end-user are exposed in Experiment 7 for future designs.

The included end-users differ from the cases most often described in BCI literature, where end-users suffering from the effects of ALS are typically tested. The latter is a difficult user group in itself, as the impact of ALS extends beyond motor impairment; a large number of patients also shows cognitive impairment (Woolley et al., 2010). The same is true for patients that suffer from the effects of a stroke or TBI (Schretlen et al., 2003; Tatemichi et al., 1994). Both groups may thus have limited cognitive resources available for controlling a BCI, which was the very reason for drawing on the ability of human spatial hearing (as also suggested in Kübler et al., 2009). The reasons why AMUSE did not succeed in this particular experimental setting are discussed in the following sections, and suggestions for future modifications are provided.

5.3.1. Directional Preference

Despite additional improvements with respect to Experiment 5, the discrimination of all source pairs was heavily reduced. Since no behavioral tests for spatial localization were performed for the end-users, it remains unclear if this is due to an actual inability to separate the sound sources, or the lack of discriminative ERPs in the EEG. As the standard auditory oddball task also resulted in a-typical ERP components for most subjects, the latter reason seems to explain the poor performance at least in part. Under this overall poor performance, no differences are found between the types of source pairs.

The SOA was set to 300 ms in the current experiment by the end-users' request and to allow for the P3 to develop completely. The larger P3 that was found for condition C300 than for condition C175 in Experiment 3 supported this decision. However, the reduced speed may have prevented the end-users from segregating the stimulation sequence into separate target and non-target sequences (Bregman et al., 2000, and Section 3.4).

5.3.2. Neurophysiology

In the standard oddball task, the target and non-target stimuli differed in pitch and not spatial location. The pitch difference that was used was recommended for clinical application (Duncan et al., 2009) and the task can be considered easy for healthy subjects. Nevertheless, the standard auditory oddball ERP responses for end-users in Experiment 6 were characterized by visible N1 components with low binary-class information, and a clear and discriminative P3 component for only two out of five end-users.

For BCI applications, any kind of discriminative information in the EEG is useful for binary-class separation, whether in early- or late components. Unfortunately, most of this information is lost in the faster AMUSE calibration data. Though early components are still clearly visible, even in channel *Pz*, the target and non-target response overlap and do not aid the classification. The positive P3 response is visible for three subjects, but also with limited binary-class discriminative information. As the data is taken from calibration trials, there is no additional workload such as discussed in Section 4.4. The reduction of discriminative information is thus due to 1) the increased number of classes, 2) the faster SOA, or 3) the spatial stimulus placement. Furthermore,

the end-users' engagement on the task may have been different. All are plausible causes for a reduction in binary-class discriminative information. Unfortunately it is impossible to disentangle these effects.

From the single-subject comparison, it is clear that the visual paradigm relies heavily on early components. In the visual standard oddball, all components have a similar sAUC; they contribute approximately equally to the class separation task. For the Photobrowser paradigm, the early N2 and an additional P2 component are particularly discriminative, with the P3 losing in power. The reliance of overt attention-based, visual BCIs on early components, of which the Photobrowser is a variation, is well known (Allison et al., 2003; Bianchi et al., 2010).

Early components where shown to be much less powerful binary-class predictors for covert attention-based paradigms (Brunner et al., 2010; Treder et al., 2010), of which the auditory standard oddball and AMUSE are instances. They thus rely more on the contribution of later, endogenous components such as the P3. Accordingly, the early components carry very limited information for end-user *VPeg02ma* in the standard oddball task, and virtually none for AMUSE. Any classification would thus be based on the P3, which is highly discriminative for the standard oddball but heavily reduced for AMUSE. As will be discussed in Section 5.3.5, the end-users (*VPeg02ma* in particular) could possibly not afford the cognitive resources that are needed for producing a P3 under the higher workload of AMUSE (Riccio et al., 2012a).

5.3.3. Questionnaires

End-user questionnaires, taken from the first session with an online run, were compared to those of the healthy users in Experiment 5 (see Figure 5.7). Despite the poor performance, end-users reported a similar (tending to lower) perceived overall workload on the NASA-TLX questionnaire. For a one dimensional MI-based BCI, Felton et al. (2012) too found no significant difference between healthy subjects and end-users. No further BCI studies comparing healthy subjects and end-users have thus far assessed workload through the NASA-TLX. Based on Felton et al. (2012) and the comparison of Experiments 5 and 6, it can tentatively be derived that both groups perceive the workload of BCI similarly, despite the difference in performance. Whether the cognitive resources that are needed to deal with such workload are available is another question entirely, and a matter of current investigation (see Section 5.3.5).

Where healthy subjects consistently showed high motivation and mood on the VAS questionnaires, the end-users showed considerably more variability in their response. Though this did not lead to a significant reduction on either scale, it did reflect the end-users' state well; they were often exhausted from prior therapy which noticeably effected their attitude towards the session. Kleih et al. (2010) found a significant positive correlation between the VAS motivation score and amplitude of the P3 at electrode *Cz*, on 33 healthy subjects. Nijboer et al. (2010) also suggested that "*motivational factors may be related to performance ... in some participants and should be monitored in standard BCI protocols*". As discussed earlier, the P3 is an important component for classification in covert attention-based BCI. The reduced motivation may thus partially account for the poor performance, but only on some sessions and for some subjects.

The only significant difference between healthy subjects and end-users was found for the *challenge* factor of the QCM-BCI questionnaire. Since the QCM-BCI questionnaire was always applied before each session, healthy users were BCI naïve upon rating. End-user data were taken from the first session with an online run. They thus had experienced several BCI sessions already, which may influence their feedback. However, comparing the healthy user data with the end-user data taken from the very first available recording session gave a similar result; only the *challenge* factor differed significantly. Thus, the challenge of controlling a BCI was a stronger motivational factor for end-users than for healthy subjects.

The single-subject comparison showed a significant negative correlation between session number and *overall workload*, *performance* and *effort* of the NASA-TLX questionnaire, and *incompetence fear* for the QCM-BCI questionnaire, but only for the Photobrowser. *VPeg02ma* thus clearly got more comfortable with operating the Photobrowser over time, perhaps driven by her consistent (near-)perfect performance. The same was not true for the AMUSE paradigm, where no significant correlation was found. Unfortunately, the NASA-TLX questionnaire scores for session three were missing due to time restrictions, which likely influenced the statistical power. Nevertheless, visual inspection of the remaining four samples did not identify a linear trend. Her gradual increase in performance over sessions (see Figure 5.5) thus did not coincide with her level of comfort in operating the AMUSE-based speller.

5.3.4. Performance

Selection accuracy for AMUSE was consistently above chance level for three subjects. For the remaining two, at most one session was below chance level (see Figure 5.5). However, in contrast to the healthy users in Experiment 5, the end-users were unable to consistently control the spelling interface through AMUSE. For this reason, no ITR or char/min were calculated.

The difficulty in (directly) transferring a paradigm from healthy subjects to the end-user is unfortunately not new in BCI literature. Furdea et al. (2009) introduced an auditory speller, using spoken numbers to represent the rows and columns of the visual spelling matrix (Farwell et al., 1988), only to report on an unsuccessful attempt to apply it to end-users that same year (Kübler et al., 2009). In fact, they attributed the poor performance to the excessive cognitive demand of their auditory BCI, suggesting that improvements such as adding spatial features to the stimuli might alleviate part of this demand. Münßinger et al. (2010) included both healthy users and end-users in their first test of the Brain Painting application. End-user control was considerably less accurate, but enough to sustain the application. Similarly, end-users and healthy subjects trained to control a binary BCI differed considerably in their performance in Cincotti et al. (2008b). They are only several examples of the translational issues that exist with current BCI. However, as suggested in Section 1.6, disappointing performance on an initial attempt should be used to drive future modifications.

Covert attention, ERP-based BCIs such as AMUSE have the potential to increase performance over MI-based, binary BCIs. They typically require little to no training, making them a convenient alternative. However, to date they have hardly been validated with the end-users for which they are intended. Unfortunately, for the few cases where this is done, the result showed much room for improvement (Kaufmann et al., 2013a; Kübler et al., 2009; Sellers et al., 2006a). This is often attributed to the lack of explicit mapping of stimulus to class (Brunner et al., 2013; Riccio et al., 2012a). The lack of explicit mapping in auditory BCI has been addressed in particular by using spoken words that directly represent the class instead of artificial or natural sounds that are not directly mapped (Hill et al., 2014; Höhne et al., 2014; Sellers et al., 2006a). Unfortunately, unlike the visual matrix speller (Farwell et al., 1988), it is virtually impossible to present multiple letters at once, making the selection a lengthy process (but see Höhne et al. (2014) for how spatial cues can help here).

5.3.5. Comparison to a Visual Paradigm

Using data from a visual study that was performed in close temporal proximity to Experiment 6, and included one overlapping end-user, allowed for the direct comparison of a single subject.

VPeg02ma is affected by a brainstem ischemic stroke, causing tetraplegia with severe dysarthria which lead to a serious communication deficit. BCI technology could provide her with a realistic communication AT. In parallel to both studies, she was screened for attention, memory and

executive capabilities. Her cognitive profile was characterized by a deficit of divided attention, visuo-spatial learning ability, the ability to control the information flow and to update information in working memory. Her scores for short-term memory and verbal working memory also fell in the lower limit of the normal range. She showed highly binary-class discriminative components for the standard visual- and auditory oddball task. Unfortunately, *VPeg02ma* did not gain sufficient control over the AMUSE spelling application, much in contrast to the Photobrowser application where she had near perfect control. This result is not uncommon, as previous literature already described reduced performance for end-users with auditory BCIs when compared to visual BCIs (Kaufmann et al., 2013a; Kübler et al., 2009). However, in this case it can partially be explained by the clinical profile of *VPeg02ma*.

As found in Experiment 5 and underlined in Riccio et al. (2012a), a multiclass paradigm like AMUSE can feature high ITRs in comparison to binary auditory paradigms (Halder et al., 2010; Hill et al., 2005). This higher ITR comes with a trade-off, as an increased demand is put on the subject's attention and working memory to accomplish the task. Since AMUSE by design relies exclusively on the auditory modality, the user has to keep a mapping of the target symbol to the attended direction in mind. The mapping is presented to the user verbally in a sequential fashion before each trial, and the user has to memorize the tone and location of his target stimulus. This requires a greater involvement of working memory, the capacity to temporarily maintain and to manipulate the information necessary for solving such cognitive tasks. During an ongoing trial, the user is required to identify the target when it is presented in a random sequence together with five competing tones. Though the spatial feature of AMUSE reduces the difficulty of this task (see Chapter 3), it is still cognitively more demanding than a binary BCI. *VPeg02ma*'s clinical profile shows deficits in exactly those key abilities. She is thus not able to afford a more cognitively demanding task like the AMUSE paradigm.

The Photobrowser application, on the other hand, makes use of visual, overt attention: the user is allowed to gaze at the target directly. Though this may not always be accessible for end-users (Ramos Murguialday et al., 2011; Treder et al., 2010; Brunner et al., 2010), it works well for those in possession of gaze- and eyelid control (Münßinger et al., 2010; Nijboer et al., 2008b; Silvoni et al., 2009; Zickler et al., 2013). For such overt visual spellers, the class to stimulus mapping is explicit, i.e. the class label itself is the stimulus. Allowing the subject to gaze directly at his target reduces the need for keeping the target in memory; once a target is picked, the user simply keeps it locked in gaze. This relies less on those cognitive functions where *VPeg02ma* shows clinical deficits, and allowed her to have near perfect control over the visual application.

5.3.6. Future Directions for AMUSE

AMUSE has proven to increase performance for healthy subjects in a direct comparison with a non-spatial setup in Experiments 3 and 4, and improved the state of the art in covert attention-based BCI. Though the initial attempt at transferring this success to end-user tests was unsatisfactory, it revealed several key targets for improvement of AMUSE to better fit the needs and abilities of the end-users. In particular the clinical profile of end-user *VPeg02ma* provided invaluable input for future revisions of the AMUSE paradigm for end-user application. To mitigate the performance drop for end-users such as *VPeg02ma*, the following aspects of AMUSE should at least be revised to reduce the draw on the limited cognitive resources.

Direct mapping - So far, the stimuli consisted of spatially distributed, artificial tones. While this provides some benefits in terms of signal processing, it forces the user to maintain an implicit mapping of stimulus to class. In order to reduce the cognitive demands of auditory BCI, a range of other stimulus defining features has been tested in BCI literature. Many studies use natural sounds such as speech (Furdea et al., 2009; Gao et al., 2011; Guo et al., 2010; Guo et al., 2009; Höhne et al., 2012a; Kübler et al., 2009; Lopez-Gordo et al., 2012; Sellers et al., 2006a; Yoshimoto

et al., 2012) and environmental sounds (Heo et al., 2013; Klobassa et al., 2009; Ruf et al., 2013), which may have several benefits. First, human listeners are tuned to process these sounds, which may reduce the mental strain (Höhne et al., 2012a). Second, the mapping from stimulus to class can be made explicit (Höhne et al., 2014). In combination with the spatial aspect of AMUSE, such direct mapping could lower the cognitive demand.

Reduced number of classes - Taking six sources and two levels (6×2 design) is a design decision that worked out well for healthy subjects. With an undo action on the second level it results in 30 classes, enough to encompass the entire alphabet and some punctuation symbols. Alternatively, reducing the number of sources to four and increasing the number of levels to three (4×3 design) would give a total of 36 classes, assuming an undo action on the second and third level. Though this adds an additional selection to a character, each selection itself is less complex which too could reduce the cognitive demand.

Removing front-back confusions - In particular in combination with the above adaptation, the front-back confusions should be removed entirely by only using frontal sources. Experiment 2 showed that this is possible for five sources; higher densities of frontal sources were not tested. Therefore, at least the 4×3 design should be feasible in the frontal half only.

Changing the SOA - As discussed above, the greater SOA of 300 msec that was used for the end-users may have hampered their ability to segregate the target- from the non-target stream. Lowering the SOA again to 175 msec may engage the end-users' stream segregation abilities much like the healthy users reported it did for them.

It is my firm believe that through these and further modifications, AMUSE will proof useful for end-users.

6. Data-Driven Performance Optimization

The development of new paradigms, based on new end-user understanding, can lead to an increase of both performance and usability. The AMUSE paradigm is one of several of such recent advances toward covert attention-based BCI (Acqualagna et al., 2013; Höhne et al., 2011b; Kaufmann et al., 2011; Treder et al., 2011). Improvement can, however, also be gained on other aspects such as better recording equipment, smarter artifact removal tools, informed feature extraction, or state of the art classifiers. In particular for ERP-based BCIs there is a further opportunity for optimization.

As described in Section 2.3.1, a trial typically consists of several iterations. This repeated stimulation is done in order to improve the relatively low SNR of the individual epochs, and to thereby increase the accuracy of the decision. However, as each additional iteration prolongs the length of a trial, it introduces a trade-off between accuracy and speed, which allows for data-driven optimization. Longer stimulation duration typically means higher accuracy but lower selection speed, whereas a shorter stimulation duration may result in lower accuracy but higher speed. The optimum in this trade-off of course depends on the SNR of the ERPs, but is to a large part also determined by application factors such as the number of trials needed for correcting a wrong decision, the number of trials needed to write a character and automatic error correction features of the application. All these should be taken into account when fine-tuning this trade-off.

Typically, the above trade-off is not dealt with during the online use of a BCI, and an arbitrary number of iterations is fixed beforehand. Subsequently, offline post-hoc analyses on the online data are used to generate performance curves over the number of iterations, given some metric (for an example, see Figure 3.9). The peak performance is then taken as a measure of the possible performance of the system. This simple strategy of determining an "optimal" number of iterations is widely used in BCI literature, but generalization might be troublesome due to the post-hoc nature. The determined number of iterations would have been optimal in retrospect, but this does not necessarily hold for future data, as EEG is prone to changes reflecting all sorts of user states. Such post-hoc analyses thus tend to over-fit the true performance and may give an overly optimistic estimate of the true performance.

Ultimately, a BCI should be able to determine on the fly if enough evidence is gathered. The length of any trial would not be set beforehand, but depends on the current data. This optimization problem is the focus of this chapter. A framework is introduced that encompasses the definition of previously proposed stopping methods. Using this framework, a new method is proposed and validated. In Experiment 8 it is applied in a spelling application of AMUSE similar to Experiment 5, and in Experiments 9 to 11 it is benchmarked offline on artificial and real BCI datasets, comparing it against several previously proposed methods.

Parts of this chapter are based on prior publications Schreuder et al. (2013a), Schreuder et al. (2011a), and Schreuder et al. (2011c).

6.1. Prior Work

The simplest way of dealing with the aforementioned trade-off is to learn the optimal number of iterations from the calibration data and fix this for online use. This will be referred to as *fixed* stopping. Once trained, it does not adapt to changes in the online data. However, EEG signals are non-stationary and the covariance structure changes over time. The distribution of the online data may thus differ from the calibration data, and this may be increasingly so as time passes.

For this reason, dynamic methods have been proposed in literature (Höhne et al., 2010b; Jin et al., 2011; Kindermans et al., 2014b; Kindermans et al., 2013; Lenhardt et al., 2008; Liu et al., 2010; Throckmorton et al., 2013; Verschore et al., 2012). For successful application, such methods are trained on a limited amount of calibration data and they should be robust against distortions that are commonly observed in EEG data. They are mostly trained on the classifier scores that stem from the calibration data, and do not need the raw EEG data. In online mode, a trial is then stopped at any point if enough evidence (in the form of classifier scores) has been accumulated for a correct decision, i.e. if the trained threshold is surpassed.

Possibly the simplest approach to this problem is by Jin et al. (2011). Their threshold simply requires the online trial to come to the same decision in a number of consecutive iterations; a number that can be optimized on the calibration data. Contrary, Höhne et al. (2010b) run a number of one-sided t-tests after each iteration (one class versus the rest, for each class), and the trial is stopped if a pre-trained α level is reached. Liu et al. (2010) rely on the sample distance to the separating hyperplane; by summing these distances over iterations, the first class to reach the threshold is considered the target and the trial is terminated. Lastly, Lenhardt et al. (2008) rely on the fact that the difference between target and non-target classifier scores should be maximized; after normalizing the scores, thereby pushing target scores to one and all others to zero, the sum of the scores should be below a pre-trained threshold.

Obviously, Bayesian methods could deal with this problem efficiently (Throckmorton et al., 2013). One such method was not explicitly designed for the purpose of dynamically stopping a trial, but rather to distinguish between control and non-control states (Zhang et al., 2008). With slight adaptations, it can however be used for the early stopping task. Recently, another method was introduced that combines both a dynamic stopping method and an unsupervised classification scheme, thus adapting to the ongoing data even further (Kindermans et al., 2014a; Kindermans et al., 2013; Verschore et al., 2012).

Technical details on some of the above methods are described in Appendix A.1, where they are reformalized in the proposed framework (see below). Full details on each method can be found in the original manuscripts.

Box 6.1: Dynamic Stopping Framework Notations

Γ: Paradigm parameters (SOA, overhead, C)

Φ: Hyperparameter of the stopping method that should be optimized

Θ: Set of trained parameters of a stopping method

L: Training function which returns parameters Θ

F: Online decision function

M: Sub-function describing the data

6.2. A Dynamic Stopping Framework

Dynamic stopping is a rather new line of investigation in BCI literature. Nevertheless, the above methods share certain principles, that lead to the formulation of a framework for describing them. The terminology for all methods was unified for easier comparison.

On top of the notations already given in Box 2.4 on page 25, the framework requires some additional notation that is listed in Box 6.1 on page 88. It is important to note that the framework assumes that classifiers are trained to assign (more) negative scores to target epochs.

By definition, all methods have access to $\boldsymbol{D}^{train}$, $\boldsymbol{l}$ and Γ for calibration purposes; for testing purposes only $\boldsymbol{D}^{test}$, Γ and Θ are available. $\boldsymbol{D}^{train}$ and $\boldsymbol{D}^{test}$ contain the classification scores for disjoint sets of trials, sorted according to class and iteration (see Section 2.4.5). Γ is a set of parameters describing the application that the data come from, such as SOA, the number of classes, and the number of selections necessary for writing a single character, $\boldsymbol{l}$ contains the true labels of the data and Θ is the set of hyper-parameters as returned by L and trained on $\boldsymbol{D}^{train}$. As there is no access to raw EEG data, the methods are implemented as a module between the BCI and the application. This ensures maximum compatibility with existing processing pipelines.

Given the training data $\boldsymbol{D}^{train}$, all dynamic stopping methods are trained with a function L, such that

$$\Theta = L\left(\boldsymbol{D}^{train}, \boldsymbol{l}, \Gamma\right) \tag{6.1}$$

The type and number of parameters contained in Θ differ for each method. Given Θ, function F is called after each iteration of an online trial. It is defined for the current trial t at iteration j by

$$F\left(\boldsymbol{D}^{test}, \Theta, t, j\right) = \begin{cases} c^w_{j,t} & \text{if stop possible} \\ \emptyset, & \text{otherwise} \end{cases} \tag{6.2}$$

where $c^w_{j,t}$ indicates the winning class, and the empty set $\emptyset$ indicates that no early stop should be taken. In case a trial reaches the maximum number of J iterations without triggering an early stop, the winning class is calculated and returned. Unless defined directly by the method, the winning class $c^w_{j,t}$ is calculated according to Equation 2.4. In most cases, function F relies on a subfunction M, which summarizes the data. These functions are at the core of each dynamic stopping method.

6.3. Rank diff - A Dynamic Stopping Method

Several methods for dealing with the above optimization problem were already introduced. They all have their own set of merits and characteristics. However, as was shown in literature, the P3 response is not always stable over trials, or even within a trial, as habituation can occur (see Section 2.2.2.2). As the P3 contributes considerably to the classification task, the resulting classifier scores can be assumed to vary systematically over the course of a trial. This is not explicitly modeled in the methods mentioned previously. For some of those it may not pose a big problem, due to the nature of the method (Jin et al., 2011). However, a too large fluctuation of the distributions within a trial could have detrimental effects on others (Höhne et al., 2010b; Lenhardt et al., 2008; Liu et al., 2010).

Therefore, rank diff was designed specifically with this effect in mind. By using several thresholds, one for each iteration, the goal was to create a method that deals well with habituation.

Using the above framework, rank diff is based on $M\left(\boldsymbol{D}, t, j\right)$, defined for iteration j and trial t as

$$M(\boldsymbol{D}, t, j) = \min_{c^* \neq c^w_{j,t}} \left(\text{median}\left(\boldsymbol{D}_{c^w_{j,t},1\ldots j,t}\right) - \text{median}\left(\boldsymbol{D}_{c^*,1\ldots j,t}\right)\right) \tag{6.3}$$

where $c^{w}_{j,t}$ is found using Equation 2.4 and c^{*} runs over all other classes. A larger value for M means a better separation of the class medians, and thus more confidence in a decision. $\Theta = \{\Delta_1, \ldots, \Delta_J\}$ contains J iteration dependent thresholds Δ. To determine Δ_j (the threshold for iteration j), $L\left(\boldsymbol{D}^{train}, \boldsymbol{l}, \Gamma, j\right)$ calculates $M\left(\boldsymbol{D}^{train}, t, j\right)$ at iteration j for each training trial. These are assigned to one of two vectors, depending on whether the estimated class label was correct ($\vec{+}_j$) or incorrect ($\vec{-}_j$) according to Equation 2.4. To implement a security margin that prevents too optimistic early stops, Δ_j is then defined as

$$\Delta_j = \max\left\{\max\left(\vec{-}_j\right), \Phi \cdot \text{median}\left(\vec{+}_j\right)\right\} \tag{6.4}$$

For online use, decision function F is then defined as

$$F\left(\boldsymbol{D}^{test}, \Theta, t, j\right) = \begin{cases} c^{w}_{j,t} & \text{if } M\left(\boldsymbol{D}^{test}, t, j\right) > \Delta_j \\ \emptyset, & \text{otherwise} \end{cases} \tag{6.5}$$

Hyperparameter Φ is set globally over all iterations, and could be set using prior knowledge. For example, a threshold with $\Phi = 1$ is rather conservative, as it is biased away from false positives (incorrect early stops). Such a conservative threshold would be useful when a false positive is costly. For instance, in the AMUSE-based speller described in Experiment 5, correcting an error takes up to four additional selections. For applications where a false selection is less costly, the threshold can be set less conservative by taking a smaller value for Φ.

6.4. Online Validation

As discussed in Section 2.5.3, the most conclusive way of validating a method is to apply it in an online BCI session. To validate the rank diff method, such an online experiment was performed, using a spelling interface that was otherwise similar to Experiment 5.

6.4.1. Experiment 8: Online Validation of Rank Diff

For a within subject comparison, a subset of subjects from Experiment 5 were tested in a slightly modified setup, including the rank diff method for dynamic stopping.

6.4.1.1. Setting

Subjects who were able to successfully write the full sentence with the AMUSE system in Experiment 5 were asked to come back for a second session. This session consisted of a calibration phase, equal to that of Experiment 5, and the online writing of two German sentences. The first sentence was the remaining sentence from Experiment 5, whereas the second sentence could be chosen by the subject but should be at least five words long.

The protocol was adjusted in several ways, the main change being the introduction of rank diff for dynamic stopping of trials. Rank diff was set to terminate a trial after at least 4 and up to 15 iterations when enough evidence for a correct selection was found, i.e. when $M\left(\boldsymbol{D}^{test}, t, j\right) > \Delta_j$. As an additional change, subjects were requested to learn the selections that are needed for any character by heart before the session. They were no longer provided with auditory labels prior to a trial, which reduced the trial time by about 6.2 seconds. The maximum theoretical writing speed with 15 iterations was thus 1.10 char/min.

During the first sentence, subjects got visual information on the labels to familiarize them again with the interface. During the second sentence this was turned off, thus relying exclusively on the auditory sense. During both sentences, subjects still received auditory information after a selection to know their current location in the spelling tree.

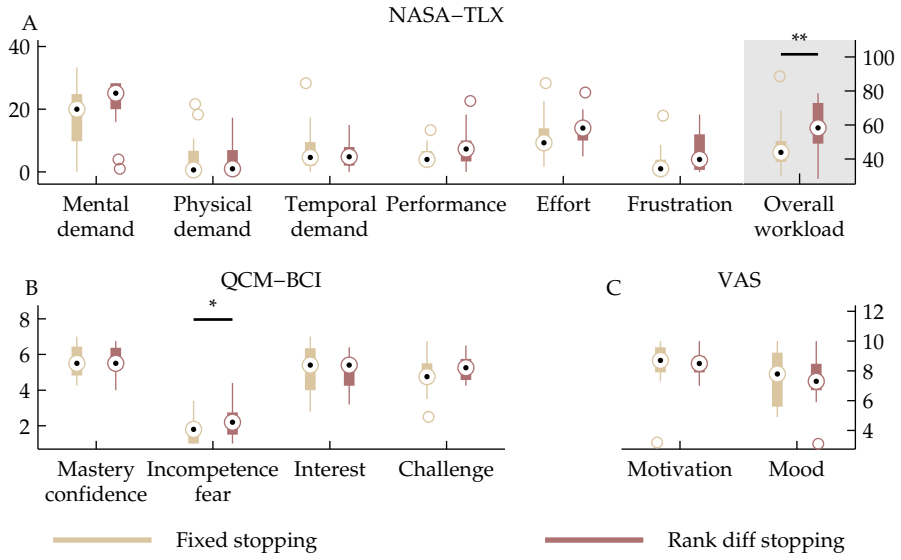


Figure 6.1.: Results for all factors of the questionnaires are compared between healthy subjects using the AMUSE-based speller with rank diff or without (Experiment 5). Lower is better for all factors of the NASA-TLX questionnaire and factor *incompetence fear* of the QCM-BCI questionnaire. Higher is better for all other factors. Significant differences are indicated with an asterix (* $p < 0.05$, ** $p < 0.01$). Note that this is a within subject comparison.

Note on Rank diff in Experiment 8 During the online validation, a preliminary version of rank diff was used, which differed slightly from the final method that was introduced in Section 6.3 and used in Experiments 9 to 11. Here, Φ was fixed to 1 and a third-order polynomial was fitted to the maximum values of $\vec{=}$ for smoothing, such that $f = \text{fit3}\left(\left[\max(\vec{=}_1), \ldots, \max(\vec{=}_j)\right]\right)$, with $f\left(j\right)$ being the smoothed value at iteration j. Equation 6.4 then reads as follows:

$$\Delta_j = \max\left\{f\left(j\right), \text{median}\left(\vec{+}_j\right)\right\} \tag{6.6}$$

Setting Φ to 1 effectively allows early stops only for trials where $M\left(\boldsymbol{D}, t, j\right)$ is at least larger than the median M score at iteration j of all training trials that resulted in a correct decision at that iteration.

Furthermore, for finding the decision thresholds Δ, the trained classifier was reapplied to the complete calibration data on which it was trained, instead of using cross-validation. Although this can be considered overfitting and might lead to an overestimation of the class distances, it is not consider as a caveat when searching for a conservative threshold.

6.4.1.2. Results

Of 15 returning subjects, 14 wrote both full German sentences. Subject *VPfch* was able to write the first sentence, but took so long that the session had to be aborted after this. Subject *VPfax* was unable to return for this experiment, although he was successful in the first session. Unless stated otherwise, the statistical testing was done using two-sided paired t-tests.

Questionnaires Questionnaire results can be found in Figure 6.1. The results for the subjects that were able to write in Experiment 5 is given for comparison. Subjects were in good mood

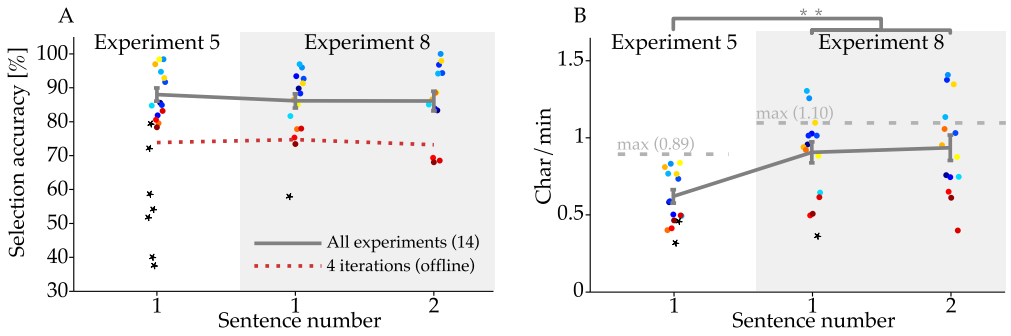


Figure 6.2.: BCI performance in terms of selection accuracy (A) and char/min (B) is presented for both sentences. Data from Experiment 5 is included for comparison. Average performance is represented by the gray line and is calculated only for those subjects that finished all sentences. Excluded subjects are marked with a star. (A) The red, dotted line shows the selection accuracy when only four iterations are considered. (B) The increase in the theoretical maximum is due to the omitting of spoken labels.

(VAS-mood, median: 7.3; IQR: 1.78) and motivation (VAS-motivation, median: 8.5; IQR: 1.1). According to the NASA-TLX questionnaire, median overall workload was rated at 58.34 (IQR; 24.0), with the main contributions from *mental demand* (median: 25.16; IQR: 8.33) and *effort* (median: 14.0; IQR: 6.0). Subjects rated the four motivation factors of the QCM-BCI questionnaire as follows: *mastery confidence* (median: 5.5; IQR: 1.12), *incompetence fear* (median: 2.2; IQR: 1.25), *interest* (median: 5.4; IQR: 1.35) and *challenge* (median: 5.25; IQR: 1.19). In comparison with data from Experiment 5, a paired Wilcoxon signed-ranks test confirmed a marginally significant effect for factor *incompetence fear* of the QCM-BCI questionnaire (Z=-2.05, p = .04) and the *overall workload* of the NASA-TLX questionnaire (Z=-2.67, p = 0.05).

Accuracy Individual performance measures for each participating subject can be found in Figure 6.2 and Table 6.1. Average selection accuracy per sentence was 84.3 % and 86.1 % for sentence 1 and 2, respectively. The increase in accuracy when compared to Experiment 5 is due to the drop out of less-performing subjects. When recalculating the mean for Experiment 5 and sentence 1 over only those subjects that finished all three sentences (N=14), this difference disappears (represented by the gray line in Figure 6.2-A). For further analysis, only those 14 subjects are regarded.

No significant difference existed ($t(13) = 0.041$; $p = 0.97$) between selection accuracy in sentence 1 (mean: 86.1 %; SD: 7.8) and 2 (m=86.1 %; SD: 11). The difference between these conditions is that for sentence 2 the visually presented support labels were switched off. This shows that the spelling tree can easily be learned and thus does not need visual presentation.

As the number of iterations varied in the current experiment, an offline analysis was performed to objectively assess any learning effect that might remain from Experiment 5. For each sentence, the online classification scores of only the first four iterations were used to calculate the selection accuracy. The red, dotted line in Figure 6.2-A shows the result. Clearly, no systematic learning effect was present. As there was also no significant difference in accuracy ($t(13) = 0.88$; $p = .40$) between Experiment 5 (mean: 88.0 %; SD: 7.3) and sentence 2 (gray line), it can thus be concluded that rank diff did not introduce additional errors.

Writing Proficiency Figure 6.2-B presents the writing proficiency for each subject in terms of char/min for the current experiment, and for Experiment 5 for comparison. The gray line

Table 6.1.: Performance summary for Experiment 8. The performance measures selection accuracy [%], char/min and ITR [bits/min] are given for all subjects and both sentences, along with age and gender. Subjects are sorted according to their selection accuracy on Experiment 5 for comparison. Missing values indicate that the sentence was not written by the user (x).

			Sentence 1			Sentence 2		
User	Gender	Age	Sel. Acc.	char/min	ITR	Sel. Acc.	char/min	ITR
VPfaz	F	27	95.9	1.26	6.93	100.0	1.41	8.19
VPfcc	M	28	85.1	.88	4.41	88.6	.88	5.01
VPkw	F	57	86.5	.94	4.98	86.5	.95	5.09
VPfca	M	23	97.0	1.30	7.36	94.2	1.14	6.38
VPfcd	M	25	91.3	1.10	5.66	97.9	1.35	7.56
VPfaw	M	40	92.7	1.01	5.58	94.3	1.03	5.94
VPfar	M	31	89.8	.96	5.16	83.3	.76	4.20
VPfav	M	40	88.3	1.01	5.57	96.8	1.38	7.44
VPfcb	M	21	81.7	.65	4.02	85.1	.75	4.84
VPfck	M	43	78.0	.61	3.27	68.5	.40	2.39
VPfau	F	25	93.4	1.03	5.92	83.9	.74	4.33
VPfcj	F	23	75.3	.50	3.07	69.3	.65	2.50
VPfcg	F	33	77.8	.92	3.80	88.5	1.06	5.57
VPfch	M	55	57.9	.36	1.68	x	x	x
VPfcm	M	51	73.4	.51	2.85	68.1	.61	2.43
Means	F:5/15	34.8	84.3	.87	4.68	86.1	.94	5.13

represents the sentence-wise average for only those subjects that finished all sentences (N=14), which will be considered for further analyses.

As expected from the selection accuracy, there is no significant difference in char/min ($t(13) = 0.6$; $p = 0.57$) between sentence 1 (mean: 0.91 char/min; SD: 0.25) and sentence 2 (mean: 0.94 char/min; SD: 0.31). However, when comparing sentence 2 with Experiment 5 (mean: 0.62 char/min; SD: 0.16), a significant increase of 52 % can be observed ($t(13) = -4.7$; $p < 0.01$). Given the similar selection accuracy, the increase must be due to improvements in the interface's efficiency. These were twofold: 1) dynamic stopping was introduced and 2) the verbal labels were omitted. To assess the influence of the dynamic stopping method independently, the 6.2 seconds time gain from the latter modification was re-added to sentence 2 in an offline simulation. A significant increase of 37 % remains ($t(13) = 4.05$; $p < 0.01$) from Experiment 5 to the simulated sentence 2 (mean: 0.85 char/min; SD: 0.26).

The average ITR in Experiment 5 was 2.72 bits/min (maximum 4.82 bits/min). When considering only those subjects that were able to write in Experiment 5 and this experiment, the average ITR over Experiment 5 was 3.49 bits/min. With the two improvements, this increased by 46 % to an average of 5.13 bits/min, with a maximum of 8.19 bits/min.

Robustness Over all subjects, a total of 3297 online trials were recorded in this experiment. Rank diff produced early stops on 41.3 % of these trials. The maximum number of 15 iterations (full trial) was reached in the remaining 58.7 % of the trials. This was in cases where the threshold was not exceeded in iterations four to 14.

About 12.6 % of early stops resulted in an incorrect decisions, which is 5.2 % of the overall number of trials. This reflects the conservative threshold policy as described in Section 6.3. At

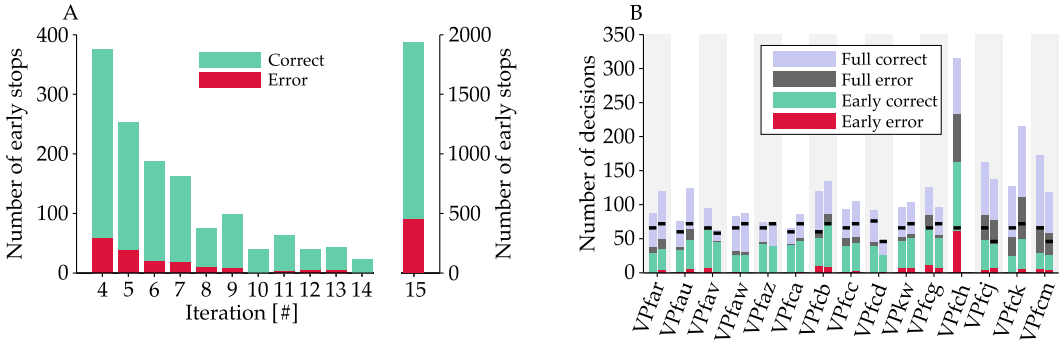


Figure 6.3.: Frequency of errors on early stops and full trials. (A) Data is collapsed over all subjects that participated in Experiment 8. The number of correctly and incorrectly classified trials are plotted as a function of the iteration in which a trial was stopped. (B) The same data is plotted separately for each subject and both sentences. The black markers indicate the minimum number of selections necessary for writing a sentence. Subject *VPfch* did not perform sentence 2.

23.5 %, the percentage of errors was almost twice as high for full trials. As depicted in Figure 6.3-A, the majority of early stops is performed directly at the fourth iteration, with the number of stops decreasing until finally all remaining trials are stopped at the 15th iteration. At each individual iteration the percentage of incorrect decisions is lower than for full trials.

Figure 6.3-B depicts the frequency of the four possible decision types for each subject. It again shows that the error rate for early stops is relatively low, and similar for all subjects. Only subject *VPfch* had such a high error rate – both for early stops and full trials – that there was no time left for writing the second sentence. The black lines indicate the minimum number of decisions necessary for writing the particular sentence.

An obvious explanation for the good performance of rank diff is that mostly the "*difficult-to-decide-trials*" reach the full number of iterations, which would in fact be by design. Figure 6.4-A displays the separability of data taken from Experiment 5, as expressed in AUC calculated at iteration four and 15. As no dynamic stopping was performed in Experiment 5, all trials reached the full 15 iterations. Correctly classified trials mostly reside in the top left corner, at high AUCs, whereas erroneous trials have lower AUC both at four and 15 iterations. More importantly, a correlation exists between the AUC at iteration four and 15, as indicated by the gray line ($r(362) = 0.639$; $p < 0.01$). The AUC at iteration four can thus serve as an approximation of the final quality of a trial.

Figure 6.4-B displays the same data for Experiment 8, which includes rank diff for dynamic stopping; Figure 6.4-C summarizes B over the x-axis. Trials are marked as correct or erroneous on stopped or full trials. Indeed, the trials that were stopped correctly are those with high AUC at iteration four; trials with average AUC typically reach the full number of 15 trials. Erroneously stops occur on trials with the lowest AUC. The latter effect is due to a wrong class triggering rank diff and leading to an AUC below 0.5. Note that the class labels are required for calculating the AUC, and that these are not available online. For the sake of clarity, the total number of trials was reduced randomly by a factor four for Figure 6.4-A and B, keeping the relative group size in tact.

6.4.1.3. Summary

Rank diff was validated in an online experiment. The performance increased significantly due to rank diff when compared to the fixed number of iterations that was used in Experiment 5. Rank

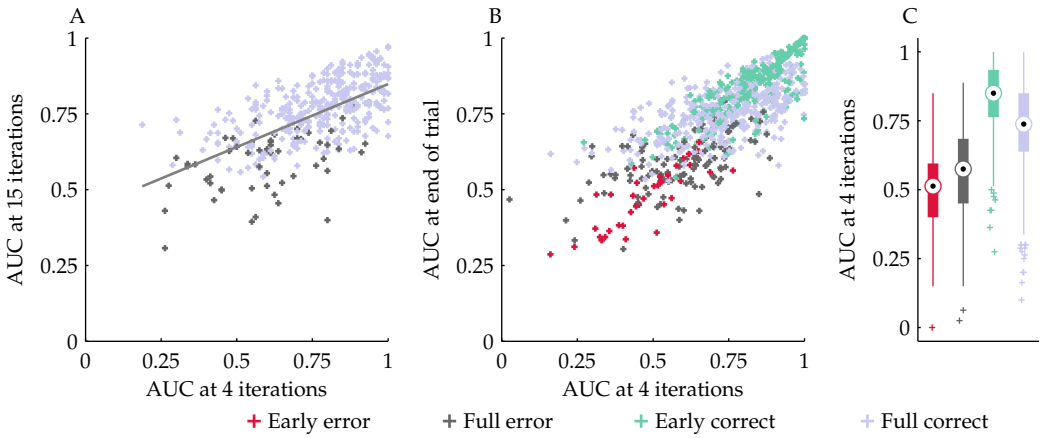


Figure 6.4.: The quality of a trial determines its length. (A) The quality of trials from Experiment 5 as expressed in terms of the AUC at iterations four and 15. (B) Similar to (A) but for data of Experiment 8 with dynamic stopping. Correctly stopped trials have the highest AUC at iteration four. (C) Summary of the x-axis of (B). The number of trials for (A) and (B) was randomly reduced by a factor four, maintaining relative group sizes.

diff did not reduce the accuracy with the conservative settings that were tested, and correctly stopped trials with highly separable target and non-target binary-classes.

6.5. Benchmark Comparison

To test and compare rank diff against some of the methods previously proposed in literature, a selection of these methods was implemented. Without marginalizing the contributions of all the authors on the original articles, they are referred to with the name of the lead author as *Liu* (Liu et al., 2010), *Jin* (Jin et al., 2011), *Lenhardt* (Lenhardt et al., 2008), *Höhne* (Höhne et al., 2010b), and *Zhang* (Zhang et al., 2008). Condensed details on each method can be found in Appendix A.1. All methods were validated both on artificial data, checking for common data distortions (Experiment 9), and on real EEG data of several paradigms (Experiments 10 and 11). All stopping methods were provided with the output of a linear classifier for training and testing purposes. Although more sophisticated classifiers are sometimes used (outputting for instance probabilities), the most widely used classifier in ERP-based BCI remains LDA. For this reason, all methods were restricted to work with data obtained from a regularized LDA. Preprocessing was similar to the pipeline described in Section 2.4. Each stimulus was classified independently, and the accumulation of the evidence was left to the stopping methods. This ensures the applicability of the results, as the methods can be realized as a simple module between the existing BCI and the application.

All analyses were performed offline, but maximum generalization was ensured by leaving out all online data for validation (see Section 2.5.3 for a discussion). The experimental protocol is depicted in Figure 6.5 and followed these four steps:

1. EEG data are split into training and test epochs (see conditions below).
2. Training epochs are transformed into classifier scores through two-fold cross-validation. Additionally, a classifier is trained on all training epochs and applied to the test epochs.

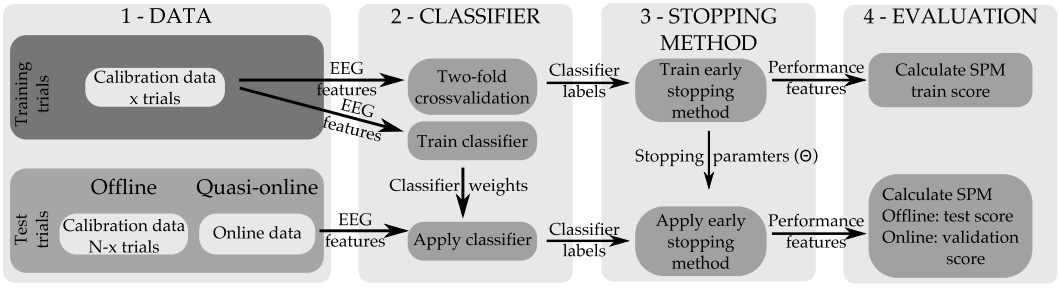


Figure 6.5.: The validation procedure is the same for both the offline and the quasi-online conditions, and for EEG- and artificial data. The number of training trials (x) was changed to evaluate the robustness of the methods. For all other analyses 21 training trials were used.

3. The classifier scores from the cross-validation are used to obtain Θ for each stopping method. The trained stopping methods are then applied to the classifier scores from the test epochs, to determine the stopping point for each trial.
4. Selection accuracy and stimulation duration are transformed into symb/min. The symb/min obtained from the validation data is used throughout this manuscript, unless stated otherwise.

Two conditions need to be distinguished. **Offline**: the above procedure takes place inside an eight-fold cross-validation. All data come from the original calibration data only, and at each fold 21 trials are used for training and three trials are used for testing. All parameters are estimated within the cross-validation procedure. **Quasi-online**: this condition was performed once, after the optimization procedures were fully debugged and produced plausible results on the offline data. Here, epochs from 21 trials of the calibration data were used as training data, and the test trials consisted of all online data.

The procedures were the same for the artificially generated data, with the addition that the experiments were run 40 times per condition. All reported performance *gains* are calculated with the *no stopping* condition as a baseline condition, by subtracting the *no stopping* performance. Negative performance gains thus mean that the stopping method results in a performance that is worse than when *no stopping* was applied.

Under the framework, all methods have one or more hyperparameters Φ, which allows the user to influence the trade-off between speed and robustness. It requires some form of optimization within the training function $\boldsymbol{L}$. Interestingly, most of the original publications do not provide a way to properly estimate the value of this hyperparameter.

Therefore, hyperparameter Φ is optimized using a resampling procedure on the classifier scores obtained from a two-fold cross-validation on the training epochs (see Figure 6.5). Two thirds of the classifier scores are randomly selected to estimate the necessary data statistics (for instance the intensities in *Lenhardt* or the priors, means, and variances in *Zhang*). Given these statistics, the remaining training data is used to optimize hyperparameter Φ for symb/min, using a simple linear search. This procedure is repeated 15 times and the median hyperparameter is subsequently applied to the test trials.

Some methods – *(Fixed) optimal*, *Höhne*, *Jin* – do not require such a compound optimization, and could be calculated directly. However, they too are estimated using the resampling procedure for extra robustness and better comparability. In these cases the hyperparameter is optimized on the larger two-thirds data split directly. In the case of two mutually dependent hyperparameters

(*Jin* and *Lenhardt*), these are estimated simultaneously by extending the linear search to a two dimensional grid search.

6.5.1. Experiment 9: Artificial Data

The most important requirement for the methods is that their performance should at worst drop to the baseline performance of the *no stopping* strategy whenever data become difficult to separate. In other words, they should gracefully degrade. Several artificial datasets were constructed to test the robustness of the stopping methods for data distortions that are typically observed in classifier outputs during BCI online sessions (see Figure 6.6).

6.5.1.1. Setting

Unless stated otherwise, data are sampled from two normal distributions (target and non-target) with unit variance. They represent the classifier outputs directly, and not the EEG data or features thereof. A training set and a validation set were generated independently. For the calculation of the symb/min metric, all of the following datasets were assumed to have an SOA of 200 ms, six classes, a two-level interface, and an overhead of ten seconds per step. The influence of the following distortions types were investigated (but not interactions between them):

Separation To mimic data that is generally difficult to classify, the distance between the target and non-target binary-class mean is varied, effectively changing the AUC from 0.5 for no separability, to one for perfect separability. This is done equally for both training and validation data. Based on the results, all other datasets were constructed with an AUC value of 0.85.

Outliers Artifacts in the EEG create outliers in the classifier outputs. These can generally be removed from the calibration data before training the classifier. Particularly during longer use, the number of artifacts in the online phase will be larger and removing them is not straightforward. Thus, between 0 % (no outliers) and 65 % (many outliers) are introduced to both binary-classes, but in the validation data only. Outliers are sampled randomly from a uniform distribution on [-3, 3].

Confusions A typical phenomenon in some BCIs is a systematical confusion, where one particular non-target is mistaken for a target more often than other non-targets (Höhne et al., 2012a). This leads to a partial overlap of non-target and target classifier scores. To simulate this, one non-target class was shifted from the non-target distribution (normal separation) towards the target distribution (no separation) in several steps. This is a special case of the separation dataset, where only one non-target class is affected by bad separability. It is applied to both training and validation data.

Datasets with changes in terms of drift, scale, distribution and number of classes were also tested. Since their influence on the performance was minimal, they are not described fully, but only depicted in Figure 6.6.

6.5.1.2. Results

All methods were robust to most types of data distortion. The largest performance deteriorations were caused by the binary-class separation, outliers and systematical confusions, which are shown in Figure 6.7-B. Results are summarized by their mean performance gain over the *no stopping* condition (see Figure 6.7-A) over the given parameter space. The gain of a specific setting was considered significant if at least 90 % of the 40 simulations yielded an output on the same side of the baseline (black bars). Bars are gray where no such significance was found.

Separation - Most methods behave well for data that is difficult to separate, meaning that they do not decrease performance below the *no stopping* baseline when data separation goes to zero.

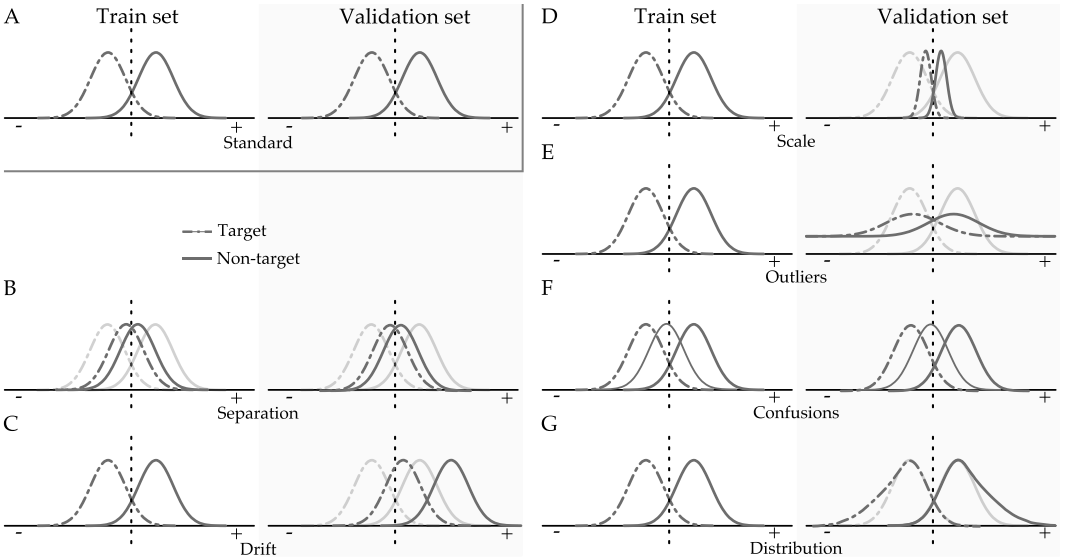


Figure 6.6.: Sample distributions of classifier scores for artificial datasets are shown. The "Train set" columns represent the distribution of the classifier outputs of training data on arbitrary scales; the "Validation set" columns display the classifier outputs for online data on the same scales. (A) shows the standard sample without distortions between training and validation data. B-G depict specific forms of data distortions that were tested in Experiment 9. Whenever the data are distorted in B-G, the standard distribution is indicated by the faded lines. The 7th dataset only changes the number of non-target classes (not shown here).

Liu clearly struggles with data as it becomes difficult to separate, as it severely undershoots the baseline performance. This is reflected in the relatively low average gain. To a lesser extend, *Jin* suffers from the same effect for the most difficult data. For difficult data, *Rank diff* shows a constant tiny, but significant reduction. *Zhang* is exceptionally resilient to inseparable data, with a significant performance gain for all but the most inseparable data.

Outlier - All methods have severe problems in dealing with outliers. None of the methods reached a positive average gain over the given parameter space. Though *Liu* performs worst again, here the picture is not that clear. In fact, all methods severely deteriorate performance compared to baseline when many outliers are present, though the level of tolerance changes slightly between methods. It is important to note that the number of outliers went as high as 65 %, and that these outliers were not observed in the calibration data.

Confusion - All methods are influenced by systematic confusions, but they degraded gracefully to the baseline. Interestingly, also *Liu* does not show any significant performance reduction from baseline, even though *confusion* is a special case of *separation*.

Rank diff was amongst the top three in terms of average gain for the above three datasets. The results on the other datasets can be found in the Appendix, Figure A.1. It is worth noting that *Liu* is the only method that is clearly not robust against drifts, showing a performance peak around the no-drift parameter and falling off to both sides with positive and negative drifts. Furthermore, *Liu*, *Rank diff*, and *Zhang* show reduced performance gain when the data is downscaled. For all other methods, and for upscaling only marginal influences were found. An increased number of classes reduced the gain for all methods, and a more skewed distribution showed higher gains for all methods.

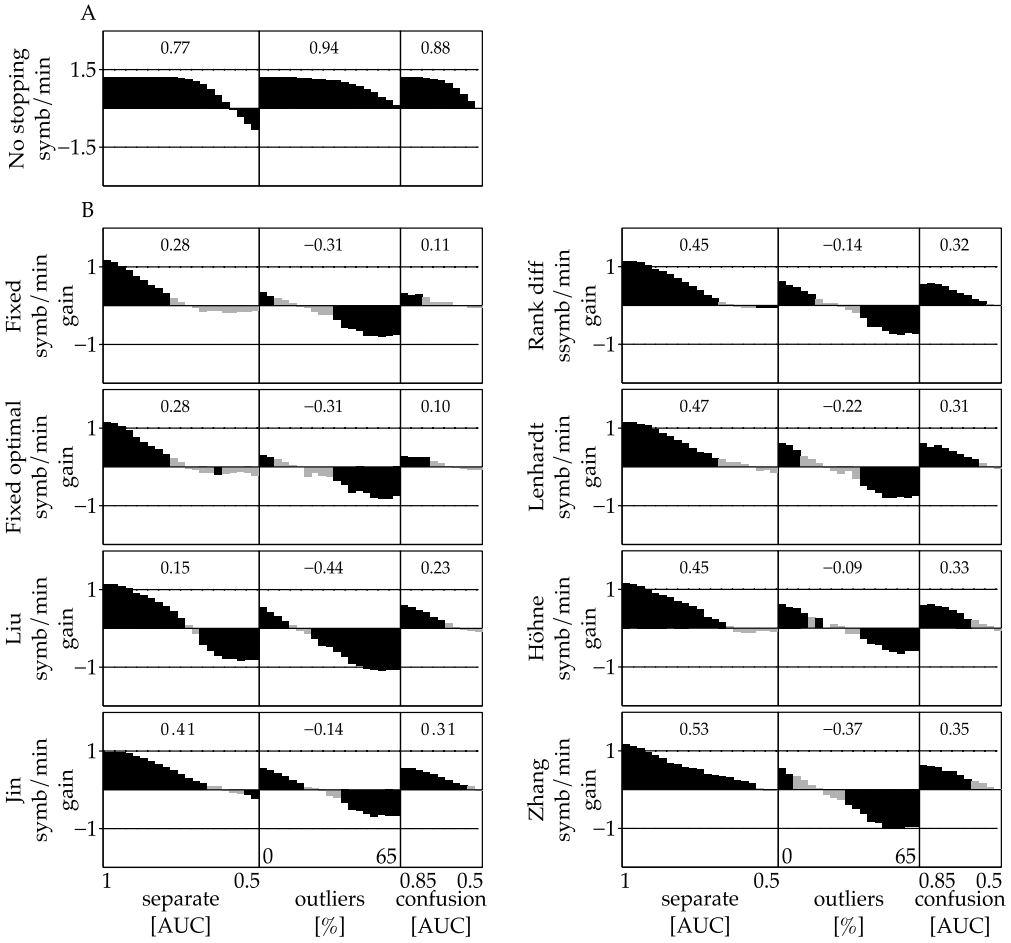


Figure 6.7.: Results on artificial data sets. The three artificial datasets that had particular influence on the early-stopping methods are displayed. A) To indicate the effect size, the baseline performance of *no stopping* is shown. B) The bars indicate the absolute performance gain over the *no stopping* condition, and are averaged over 40 simulations. Bars are black if 90 % of the 40 simulations result in the same sign, and gray otherwise. The plotted numbers indicate the average gain over the depicted parameter space of each distortion method and stopping strategy, with larger positive numbers indicating a more robust behavior of the stopping strategy.

6.5.1.3. Summary

Using artificial data, the robustness of different dynamic stopping methods was compared. Most methods behave well in the sense that they do not reduce performance beyond the baseline of *no stopping*. Only for a high number of outliers did all methods degrade beyond this baseline. Method *Liu* was prone to almost all data distortions. Rank diff was consistently amongst the top three in performance.

Table 6.2.: Dataset parameters. Given are the number of subjects (N), number of channels, modality (A = auditory, V = visual), number of classes, levels (number of trials per character), the maximum number of iterations in the original data, the SOA [ms] and the overhead [s]. AMUSE data is taken from Experiment 5 where the labels were read out to the user, explaining the large overhead.

Dataset	N	Channels	Modality	Class	Levels	Max. It.	SOA	Overhead
Experiment 5	16	61	A	6	2	15	175	18.25
Hex-o-spell	13	63	V	6	2	10	217	6
Center speller	13	63	V	6	2	10	217	8.25
Cake speller	13	63	V	6	2	10	217	6
RSVP speller	12	55	V	30	1	10	83	12
MVEP speller	16	57	V	6	2	10	266	11.7

6.5.2. Experiment 10: BCI Data Analyses

As artificial data can only go so far in estimating the usefulness of any method, all methods were additionally validated on pre-recorded EEG data from BCI experiments. The datasets represented a variety of covert attention-based BCI paradigms, as they were available within the BBCI group. However, as will be discussed in Section 6.6.5, any improvement in the AUC of a classifier results in better gains for most dynamic stopping methods. With overt attention-based BCIs typically resulting in data with better binary-class separation, the results on the datasets used here can be considered a lower estimate of the performance gain on overt attention-based BCIs.

6.5.2.1. Setting

EEG data from 83 sessions were taken, originating from Experiment 5 and five visual ERP paradigms: the rapid serial visual presentation (RSVP) speller (Acqualagna et al., 2011), the MVEP speller (Schaeff et al., 2012), and the ERP-based Hex-o-spell, center speller and cake speller (Treder et al., 2011). Datasets were previously recorded for the respective publications. All datasets consisted of an offline calibration part and an online spelling part; in all the original studies the users' task was to write text. Details on the datasets can be found in Table 6.2. All calibration datasets were reduced to 24 trials. Due to the varying number of classes and iterations per trial, there was still a discrepancy between datasets. Since it is not the primary goal here to compare datasets, this was not considered a problem. All online results are based on 21 training trials for calibration, to resemble the amount of data available in one cross-validation fold. Only results explicitly testing the influence of training data had a varying number of training trials.

All data were high-pass filtered above 0.1 Hz during recording. For the current analyses they were further low-pass filtered below 30 Hz (order eight Chebyshev II filter, 30 Hz pass-band, 42 Hz stop-band, 50 dB damping) before being downsampled to 100 Hz and segmented according to the stimulus onset. Single epochs were baselined to the interval 150 ms prior to the stimulus onset. For feature extraction, data were further downsampled using the unevenly spaced intervals based on prior knowledge (see Section 2.4.3). The linear classifier described in Section 2.4.4 was used.

6.5.2.2. Results

While the results on the artificial data give an intuition of the properties of the various methods, the truly relevant question is their behavior on real BCI data, which are summarized in Figure 6.8. On average over all datasets, all methods improved the performance. However, this did not hold when observing the paradigms individually.

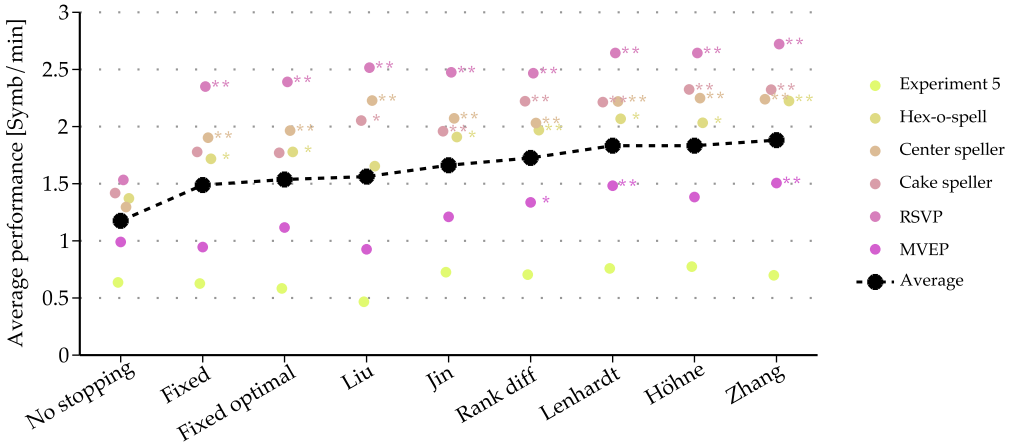


Figure 6.8.: Quasi-online performance. Averaged symb/min performance is given for all stopping methods (columns) and experimental paradigms (dots). The connected black dots represent the averages over experimental paradigms. It was calculated over all subjects directly and accounts for the different number of subjects per paradigm. An asterix marks the significance level of a performance increase over no-stopping (* $p < 0.05$; ** $p < 0.01$), as found by a Bonferroni corrected paired t-test.

All methods significantly increased the performance on two datasets (center speller and RSVP speller), and all methods but *fixed* and *fixed optimal* managed to increase the performance on the cake speller dataset (for significance levels see Figure 6.8). Performance on the hex-o-spell speller dataset was significantly increased by all but *Liu*. Significant performance increase for the MVEP speller dataset was found for rank diff, *Zhang*, and *Lenhardt*. None of the methods significantly increased (or decreased) performance on the AMUSE speller dataset.

In summary, no method increased performance significantly on all datasets. Rank diff, *Zhang*, and *Lenhardt* did so for five datasets, *Höhne* and *Jin* for four datasets, and *fixed*, *fixed optimal*, and *Liu* for three datasets.

Figure 6.9 shows the average gain over *no stopping* for each stopping method, accumulated over datasets. Over all datasets, *Höhne*, *Lenhardt*, and *Zhang* performed particularly well, as indicated by the high cumulative average gain (4.17, 4.14, and 4.47 respectively). Directly after these three methods, *Jin* and rank diff followed with cumulative gains of 3.11, and 3.49. The cumulative gain for *Liu* was competitive (2.83), in particular for the "easier" RSVP- and center speller datasets. When subtracting the negative gain for the AMUSE- and MVEP spellers it reduced to 2.60.

Both of the fixed methods (*fixed* and *fixed optimal*) showed a cumulative gain over *no stopping* (2.08 and 2.36 after subtraction negative paradigms). However, their gains were clearly smaller when compared to most dynamic methods. A full image depicting individual gains for all subjects and all stopping methods can be found in the Appendix, Figure A.2. In absolute terms, the fixed methods and *Liu* reduced performance for about 26.5 % to 30 % of the subjects, against 8.4 % to 15.7 % for the dynamic methods.

Required Training Data Figure 6.10 shows the effect of the training set size on the estimated online performance, as averaged over all datasets. Shown is the additional *upper bound* method, along with the previously introduced methods. The *upper bound* method uses the true labels, and is therefore only shown to indicate the highest achievable performance (see Appendix A.1). With the minimum number of three training trials, all methods but *Jin* and *Lenhardt* decreased

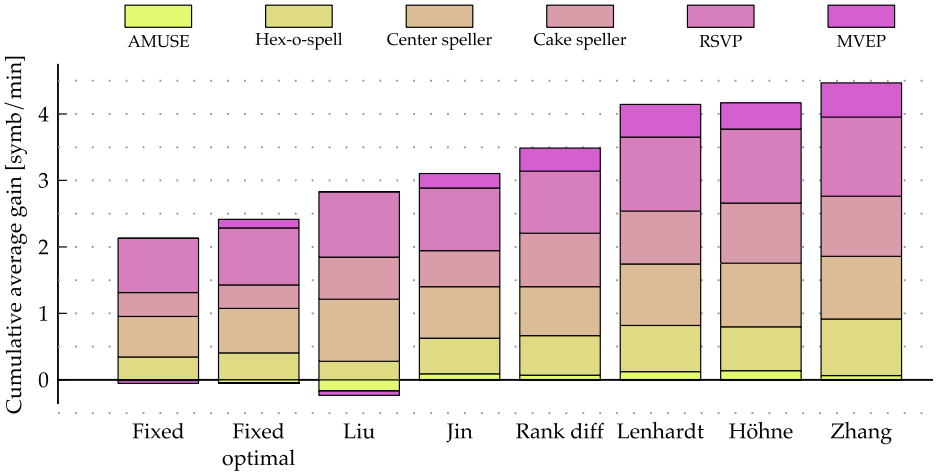


Figure 6.9.: Cumulative gain. The y-axis shows the cumulative average gain over *no-stopping* for each of the stopping methods. The height of a segment indicates the average performance gain which was realized for a particular dataset. Negative average performance gains are displayed below the zero line. It is interesting to note that the performance metric is symb/min, and thus directly interpretable.

the performance with respect to *no stopping*, on average. With 12 training trials or more, all methods increased performance over the *no stopping* condition, on average. Performance for the *no stopping* condition converged after about 12 trials, as the accuracy saturated to 100 % for several paradigms. For all stopping methods, the relative gain increased substantially during these first four conditions (three to 12 training trials). After 12 training trials, most dynamic methods kept increasing in performance until trial 21. Indeed, the *upper bound* also kept increasing until 21 trials, showing a performance gain that would be impossible without such stopping methods.

Importantly, on average all dynamic methods, except for *Liu*, consistently outperformed the *fixed* and *fixed optimal* method from six training trials upwards. Most dynamic approaches could thus exploit the data better than simply fixing a number of iterations, even with small amounts of training data. *Liu* needed on average 15 trials or more to outperform *fixed (optimal)*. Though the effect size varied, the average findings largely transfer to the individual datasets.

6.5.2.3. Summary

All tested methods increased the spelling performance over *no stopping* on average. As could be expected from Experiment 9, *Liu* struggles on two datasets, for which it in fact decreased the average performance. The *fixed (optimal)* methods also improved performance, but they are clearly outperformed by the dynamic stopping methods. With as little as 12 training trials, all methods showed positive performance gains compared to *no stopping*.

6.5.3. Experiment 11: Approximating the Hyperparameters

The proposed scheme for estimating hyperparameter Φ workse well, but it may pose a hurdle for some. However, approximations of this optimal Φ may exist, which would simplify the estimation process and potentially increase adoption of these methods in BCI literature.

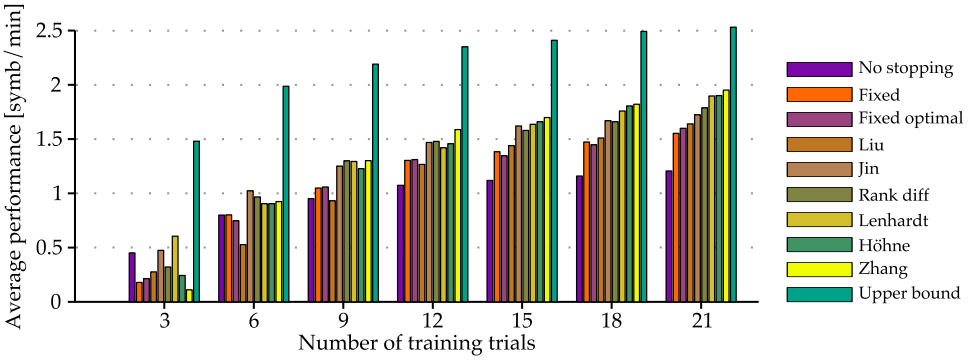


Figure 6.10.: Influence of the number of training trials on the performance of stopping methods. With low number of training trials, most dynamic methods deteriorate the performance with respect to *no stopping*. When increasing the number of training trials, the dynamic methods start outperforming *no stopping*. Their performance increase converges at larger number of trials, indicating a gain which is impossible to achieve without using dynamic methods.

6.5.3.1. Setting

The size and heterogeneity of the datasets in Experiment 10 can be used to find a proxy for the optimization of hyperparameter Φ. For this purpose, the optimized values of Φ for all training sets were investigated for dependencies on data features. Any resulting independent variable should be chosen such, that it can easily be derived from a calibration dataset. As the AUC of the training data highly correlated with Φ (see Appendix A.2, Figure A.3), and it is independent of the sample size, it was chosen as the independent variable. Dependent variable Φ was then linearly fitted to the AUC of the training data. Fitting was done using iteratively re-weighted least squares analysis (Holland et al., 1977) to reduce the influence of outliers. This gives two coefficients for each method, which transform the AUC of a dataset into a corresponding hyperparameter $\Phi^\dagger$, which approximates Φ. All EEG datasets were reanalyzed using this approximation by $\Phi^\dagger$, and the performance was compared to the optimized variant. In the case of multiple hyperparameters, the hyperparameter with the highest correlation coefficient was approximated whilst fixing the other. For *Jin*, j^{mean} was fixed to the minimum of two iterations, and for *Lenhardt*, S_1 was fixed to 1.

6.5.3.2. Results

Even across different paradigms, with all the different settings, the optimized Φ was a rather robust function of the AUC (see Appendix, Figure A.3). Settings such as the SOA and overhead thus appear to play a minor role. The black line in Figure A.3 shows the fit of all the data; the gray lines fit the data each with one dataset left out. Only for method *Jin* did these fits not clearly overlap, which is likely due to the fixing of j^{mean} in the online case, the inherent discrete nature of the hyperparameters and the small parameter range. All other methods had a robust fit, regardless of the aforementioned factors and the paradigm fitted. The resulting coefficients can be found in Table 6.3 and can be used to approximate Φ according to the equation

$$\Phi^\dagger = a \cdot \mathrm{AUC} + b \tag{6.7}$$

The result of reanalyzing all data, using Equation 6.7 to determine $\Phi^\dagger$ from the training data, can be found in Figure 6.11. Results are summarized by their median, 25 and 75 percentiles

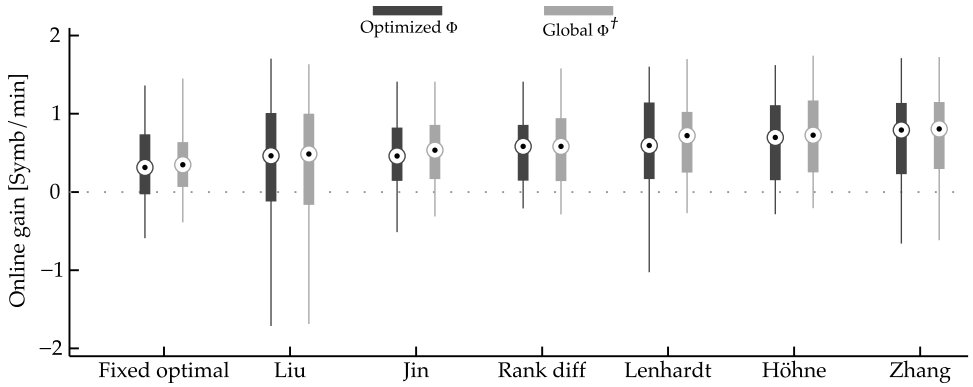


Figure 6.11.: Performance comparison between optimized hyperparameter Φ and a global $\Phi^\dagger$, the latter of which was estimated by linear regression. The optimized Φ boxplots are the quasi-online performance results of each method, when trained using the subject specific optimization (Φ). The global $\Phi^\dagger$ boxplots indicate the results, when the methods are rerun and $\Phi^\dagger$ is directly estimated from the AUC of the training data, using the coefficients in Table 6.3. Performance differences are not significant, as confirmed by a Bonferroni corrected paired t-test.

and minimum and maximum gain. Though the use of $\Phi^\dagger$ significantly decreased the training performance for several methods when compared to the optimal Φ (not shown), it did not significantly change the online performance for any method. Significance was tested using a Bonferroni corrected, paired t-test.

6.5.3.3. Summary

Using a simplifying approximation of the optimization scheme, similar performance gains could be achieved. The provided coefficients transform the AUC of a training dataset into approximations of hyperparameter Φ for each method in a straightforward manner.

6.6. Discussion

Increase in performance due to any dynamic stopping method is promising, but has been largely neglected in literature. Though several methods have been introduced before, they seldomly find their way into online studies. Rank diff was validated online, and benchmarked against other methods on a large body of EEG data.

6.6.1. Online Validation of Rank diff

By reducing the overhead, introducing rank diff and inclusion of only those subjects that were able to write, the ITR jumped from 2.72 bits/min in Experiment 5 to 5.13 bits/min in Experiment 8. When considering only those subjects from Experiment 5 that completed sentence 2 of Experiment 8, the ITR is adjusted to 3.49 bits/min. The reduced overhead and rank diff still resulted in an increase of approximately 46 %, to bring the ITR to the level that was predicted offline in Experiment 3.

Importantly, both the average ITR and the char/min represent an improvement in performance for auditory BCI even to date. In fact, it is similar to the tactile BCI reported in van der Waal et al.

Table 6.3.: Optimization approximation. Given the tested datasets, hyperparameter Φ was found to correlate with the AUC of the training data. Coefficients of the linear fit are given here for each method. The approximated $\Phi^{\dagger}$ can be recovered from the training data using Equation 6.7, which can form a starting point for new experiments or further optimization.

Method	a	b
Fixed optimal	-21.96	22.03
Liu	-4.58	5.46
Jin	-7.90	8.86
Rank diff	-1.97	1.77
Zhang	0.80	-0.30
Höhne	0.10	-0.04
Lenhardt	-1.07	1.07

(2012) and approaches the writing performance of some covert attention-based visual spellers (Acqualagna et al., 2013; Liu et al., 2011).

Performance in traditional visual ERP applications is generally good even at a low number of iterations. However, things get more difficult when overt attention is impossible; the performance breaks down and a quality check as provided by a dynamic stopping method may be desirable. The same can be said for auditory paradigms, where covert attention is the standard and the resulting ERPs are less strong. Here, a decision is enforced after 15 iterations, but an additional benefit of dynamic stopping methods could be to refrain from any decision if no threshold is surpassed. It might prove useful to allow for trials were no decisions is made, to implement a no-control state or to simply refrain from taking uncertain decisions. This could be particularly useful in long-term use of a system, where at times the user may not want to interact with the system (Huggins et al., 2011).

Interestingly, questionnaire results show that the returning subjects reported a slight increase in *incompetence fear* on the QCM-BCI questionnaire, despite the fact that they all mastered the task in Experiment 5. It is likely that the expectation they expressed in Experiment 5, as BCI novices, was adjusted in the current experiment after having had first hand experience with BCI.

Subjects furthermore reported a higher perceived *overall workload* on the NASA-TLX questionnaire for the AMUSE driven speller with rank diff. As this questionnaire was in both cases assessed at the end of the session, and the effect size is considerable, it likely reflects a real increase in workload. To formally assess the source of this increase, a control group would have been necessary, to control for the session factor; subjects could simply have expected to improve more than they did. Nevertheless, it is possible that the changes resulted in additional workload. Since the verbal labels before a trial were omitted, the subject was required to mentally prepare for the next trial by recollecting the next target from memory. Even though subjects were perfectly able to do so, it may require more attentional resources. Rank diff too may have caused additional workload; according to subjective reports, an erroneous early stop was particularly frustrating. Though objectively rank diff did not introduce more errors, such events could cause the subject to get distracted and shortly loose focus, adding to the perceived workload.

With an early decision in 41 % of all trials and a false positive rate of 5 %, rank diff can be considered a conservative yet effective method for dynamic stopping. When compared to Experiment 5 where no dynamic stopping was applied, early stops by rank diff did not lead to a difference in accuracy. Consequently, rank diff alone boosted the average writing speed by 37 %. It therefore validates the claim that rank diff can produce valuable performance gains.

6.6.2. Benchmarking with Existing Datasets

Several methods for dynamic stopping have been proposed in literature, but a direct performance comparison has so far been difficult due the variety of paradigms, preprocessing steps, timings and evaluation metrics. Here, several of such methods were systematically tested within the same theoretical framework, on the same datasets and under the same conditions. Artificial data were generated and evaluated to test the methods' performance under strictly controlled conditions. Subsequently, their performance was evaluated on pre-recorded real EEG data from a variety of different BCI experiments. The goal was not only to compare rank diff to other methods; this evaluation strategy allows to draw conclusions on the benefits and drawbacks of these methods in general.

It can be said that early stopping, be it dynamic or fixed, increased the performance in the majority of cases. This was particularly true when a paradigm resulted in highly discriminative classifier scores, with average performance gains of up to 1.19 symb/min for the RSVP speller data with the stopping method of *Zhang*, representing a 78 % performance increase over *no stopping*. Note that, due to the use of the symb/min metric, this can be interpreted in a straightforward manner: on average 1.19 more characters can be written each minute. The covert attention-based visual spellers that already perform well can thus be further optimized using dynamic stopping.

Most methods could be applied even when the classifier scores were less discriminative; though they might no longer increase the performance, they reduced to the baseline performance at worst. Artificial data (Figure 6.7) showed that this was the case for all methods but *Liu* and – in case of very difficult data – also for *Jin*. Thinking about patient application of stopping methods, this is an important feature, as patient data is typically less clean and more difficult to classify. Most methods can thus be applied without worrying about performances degradation.

Adding a robust early stopping module to the ERP-based BCI is a good idea, provided that enough data is available. Training an early stopping method on top of the classifier effectively means estimating more parameters from the same amount of data. This can pose a problem, as data is often limited. However, over all datasets, a number of nine to twelve training trials sufficed for the methods to increase the online performance on average. Given the difference in epochs per trial between datasets, the actual number may vary for different paradigms. Though nine trials is an amount that is normally available for training, the more difficult the data are, the more trials may be needed. For difficult datasets (such as the MVEP- or AMUSE speller datasets), dynamic stopping methods need more data to perform reliably.

Before additional modules are added to a complex BCI system, another aspect needs to be taken into consideration: the tested methods have a different implementation complexity. The fixed methods are obviously the simplest. Taking the best performing methods *Lenhardt*, *rank diff*, *Höhne* and *Zhang*, arguably the simplest method is *Höhne*, given that the needed statistical tests are part of most data analysis toolboxes. The Bayesian statistics behind *Zhang* could be considered more complex and requires a custom implementation. *Rank diff* and *Lenhardt* are rather simple to implement, and fall somewhere between them. All methods showed runtime performances that are well in the range of what would be applicable online (typically between 3 and 27 ms per decision). As the methods were not optimized for runtime performance, these values can most likely be reduced.

6.6.3. Dynamic Versus Fixed Methods

Early stopping methods allow the BCI to adapt to the current state of the user. In doing so, they can reduce the time that is necessary for a selection, and thus increase the performance of the system. This seems to be particularly true for dynamic methods, since they can adapt to changes in the data. Though *fixed* and *fixed optimal* are the simplest methods and require very little adaptation

of the BCI system, they are not dynamic; during online application the number of iterations is fixed regardless of the data coming in.

On average the *fixed* methods do increase the performance with respect to *no stopping*. Unfortunately, their performance gains leave much room for improvement, which is observed both in artificial data and real EEG data.

6.6.4. User-optimized Φ Versus Global $\Phi^{\dagger}$

As most of the original manuscripts do not specify how to find an appropriate value for hyperparameter Φ, a general optimization scheme is proposed. With the symb/min, it uses a realistic metric as objective function and leads to considerable performance gains. To optimize Φ, realistic simulations for the online performance for different values of Φ are required, which can be time consuming and error-prone to implement. For the BCI practitioner, a simpler way of estimating Φ would be convenient. For instance, a regression function would help to determine an at least *suitable* hyperparameter for a new BCI user, given some collected training data. The empirical observation that Φ is highly correlated with the AUC of the training data, led to a set of coefficients that translate a given AUC into an approximation of Φ, called $\Phi^{\dagger}$. Using $\Phi^{\dagger}$, the online performance gains were similar to the performance gains using Φ.

Apart from an interesting basic finding, this result has practical relevance. The AUC can be estimated reliably with a relatively low amount of data, say a few trials. Using the regression coefficients, the problem with training on a low number of training trials may thus be overcome. But even with an appropriate amount of training data, they alleviate the practitioner of the task of running simulations to optimize S. They thereby reduce the barrier for using stopping methods in similar paradigms.

6.6.5. General Remarks

As discussed in Section 2.5.3, the danger of over-fitting the data is always luring in offline analyses, and online validations are considered to be the last conclusive evidence. Such evidence was provided for rank diff in Experiment 8, but is not feasible for large scale comparisons such as those in Experiments 9 to 11. In this case, the protocol was kept strict to mimic such a validation, leaving the online data untouched until the very end. Such a quasi-online protocol bypasses the need for online validation, whilst keeping the validity of the test in tact. However, one clear difference is that the human is not in the loop. Potentially, and as observed in Experiment 8, shortening the trials can impact the perceived workload. This study showed the principle validity of the tested methods. Influences of such methods on the human in the loop remain an open topic, though at least for *Jin* (Jin et al., 2011), *Lenhardt* (Lenhardt et al., 2008), and of course *rank diff* (Experiment 8) there is online evidence.

Method *Liu* differed slightly from the original implementation, to comply with the testing framework. The linear kernel support vector machine (SVM) originally employed is very similar to the LDA used here. This design difference is not likely to be the cause for the poor performance. The effect of temporal smoothing the features, as proposed by *Liu*, could result in cleaner data. At the same time it introduces strong dependencies between samples, which violates the assumptions of most classifiers. The influence of this on the current results is unclear.

For the sake of comparability, the current analyses were done using an LDA, the most prominent classifier in ERP-based BCI literature. However, given that for all methods the gained performance depends on the AUC of the classifier data, it could be expected that any improvement made to the classifier (and thus the AUC) could pay off double: it improves the classification performance in itself and additionally benefits more from the stopping methods. The influence of different, potentially non-linear methods remains an open question.

7. Summary and Conclusions

The term BCI was introduced by Vidal (1973) and has been the topic of extensive investigation ever since. Visual ERP-based BCIs have been particularly successful, both with patients and with healthy users. However, they rely to a large extend on overt attention, i.e. eye gaze. Therefore, they only serve a subset of the end-users that are in need of BCI as an AT. Covert attention-based BCIs could remedy this shortcoming.

Work on covert attention-based BCIs arguably started with a series of papers by Hill et al. (2004; 2005). However, the urgency of such covert attention-based BCIs was emphasized recently by several studies showing the detrimental effect of visual impairment on the traditional visual BCIs (Ramos Murguialday et al., 2011; Treder et al., 2010; Brunner et al., 2010). This sparked additional interest in covert attention-based BCIs. Unfortunately, successful visual paradigms can not simply be transferred to the tactile or auditory domains. Special paradigms are needed to account for the lack of direct mapping and lower number of classes. In addition, ERP based BCIs, covert and overt the like, have an intrinsic optimization parameter that influences the accuracy and speed of the system: the number of iterations. This parameter is almost never optimized, and the same fixed number of iterations is used for all subjects.

The work presented in this thesis contributed to both these challenges, thereby significantly increasing the applicability of ERP-based BCIs.

In Chapter 3 an alternative auditory paradigm was introduced and successfully tested offline with healthy subjects. By using spatially distributed stimuli, a multiclass covert attention-based BCI was realized, called AMUSE. Subjects were able to focus their attention to a single target direction in a stream of target and non-target stimuli differing only in their spatial location. When asked to respond to target stimuli with a key press, they did so with high accuracy, thus confirming hypothesis *i – subjects can focus on one amongst several spatial stimuli.* The resulting ERP responses were typical and ERP components contained discriminative information for separating the target stimuli from the non-target stimuli for up to eight classes. Components as early as the N1 were attention modulated for spatial cues but not for pitch cues, clearly showing the vital information that is added by the spatial location. Hypothesis *ii – spatial stimuli alone are sufficient to elicit binary-class discriminative ERPs* could thus be confirmed. The performances achieved with AMUSE, as estimated from offline analysis, represented a major jump forward compared to contemporary alternatives. This performance brakes down almost entirely without the spatial properties of the stimuli. Care is to be taken with sources that are placed on the cone of confusion. The loss of information in the binaural cues for spatial localization (ITD and ILD) cause a reduction in the recognition, as shown behaviorally.

In BCI research, a new paradigm is often considered validated if it works in an online setting. This evidence was provided in Chapter 4. AMUSE was used to drive a two-leveled spelling application using six sound sources. The vast majority of healthy subjects was able to write a full German sentence at first try, thus confirming hypothesis *iii – with spatial stimuli, a covert attention-based, auditory multiclass BCI can be realized.* Contrary to many other auditory BCIs, the AMUSE-based speller did not rely on visual instructions, as they were given verbally using text synthesis software. It could thus be operated with minimal reliance on the visual domain. Using the practical char/min metric for writing speed, AMUSE set a new performance standard for non-visual covert attention-based BCI, and rivaled the performance of the covert attention-based visual paradigms at the time of introduction. Thus, hypothesis *iv – spatial stimuli allow for*

useful BCI performance in a strictly auditory setting for healthy subjects and end-users – could be confirmed for healthy subjects.

As the focus of covert attention-based BCIs, and thus of AMUSE, is on end-user AT applications, the next step was testing AMUSE with a population of end-users. This was done with five end-users with advanced paralysis over five sessions in Chapter 5. End-users perceived the workload induced by the speller in a similar way as did healthy subjects. Nevertheless, none of the end-users was able to successfully operate the same spelling setup that was successfully controlled by healthy subjects. Hypothesis *iv – spatial stimuli allow for useful BCI performance in a strictly auditory setting for healthy subjects and end-users* – could thus not be confirmed for end-users. As most other covert attention-based BCI, AMUSE lacks the explicit class mapping from the visual matrix speller (Farwell et al., 1988) and its derivates. It could be shown for at least one end-user, that the additional cognitive resources that this requires may not be available. Suggestions for future modifications of AMUSE are given that should make it less cognitively demanding. Reducing the draw on cognitive resources poses a real challenge for the future design of covert attention-based BCIs.

Performance improvements can not only be achieved through better paradigm design. BCI research relies heavily on state of the art methods from statistics and machine learning for artifact rejection, feature extraction and classification. Only recently has there been more focus on evidence accumulation and dynamic decision making, despite the tremendous impact they can have. Chapter 6 introduces rank diff, a new method for addressing this last step in the processing chain. It successfully increased the online writing performance with AMUSE for healthy subjects by almost 40 %, confirming hypothesis *v – properly dealing with the speed / accuracy trade-off increases performance significantly.* Due to rank diff, AMUSE approaches current visual solutions for covert attention-based BCI both in terms of ITR and char/min. In a direct comparison with several other (non-)dynamic methods, rank diff performed competitively amongst the top three for artificial data. For real data, rank diff is often conservative (by design), resulting in a low number of subjects for whom the performance was reduced, at the expense of less gain than other methods. However, the key result is that most methods (including rank diff) do increase the performance significantly and they do so with relatively little training data. They also consistently outperform a fixed number of iterations, even if this number was optimized on the calibration data, confirming hypothesis *vi – smart methods outperform simply picking an "optimal" number of stimuli.*

In the presented work, confirmation was successfully found for all hypotheses for healthy subjects. One key hypothesis – *spatial stimuli allow for useful BCI performance in a strictly auditory setting for healthy subjects and end-users* (*iv*) – could unfortunately not be confirmed for end-users. In fact, this hypothesis is quite bold, as it represents the holy grail of the majority of BCI research: successful end-user application of a BCI paradigm. This difficulty in transferring a successful paradigm to end-users is unfortunately not new in BCI research. Nevertheless, it adds valuable information about end-user translation to the relatively small body of end-user literature. Furthermore, a partial explanation for the poor performance was formulated, aiding the future design of covert attention-based BCI.

The fundamental concept of AMUSE has triggered several further studies, both internally in the BBCI group (Höhne et al., 2010b; Höhne et al., 2011b; Rost et al., 2010; Schreuder et al., 2012b) and in other labs (Belitski et al., 2011; Cai et al., 2013; Ferracuti et al., 2013; Gao et al., 2011; Käthner et al., 2013; Nambu et al., 2013; Ruf et al., 2013; Webb, 2013).

References

Acqualagna, L. and B. Blankertz (2011). "A Gaze Independent Spelling Based on Rapid Serial Visual Presentation". In: *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Boston, USA: IEEE EMBS, pp. 4560–4563 (cit. on pp. 8, 100).
DOI: 10.1109/IEMBS.2011.6091129

– (2013). "Gaze-Independent BCI-Spelling Using Rapid Serial Visual Presentation (RSVP)". In: *Clinical Neurophysiology* 124.5, pp. 901–908 (cit. on pp. 6, 8, 55, 87, 105).
DOI: 10.1016/j.clinph.2012.12.050

Acqualagna, L., M. S. Treder, M. Schreuder, and B. Blankertz (2010). "A Novel Brain-Computer Interface Based on the Rapid Serial Visual Presentation Paradigm". In: *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Buenos Aires, Argentina: IEEE EMBS, pp. 2686–2689 (cit. on pp. 6, 53).
DOI: 10.1109/IEMBS.2010.5626548

Allison, B. Z. and J. A. Pineda (2003). "ERPs Eevoked by Different Matrix Sizes: Implications for a Brain Computer Interface (BCI) System". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11.2, pp. 110–113 (cit. on pp. 22, 24, 83).
DOI: 10.1109/TNSRE.2003.814448

Allison, B. Z. and J. a. Pineda (2006). "Effects of SOA and Flash Pattern Manipulations on ERPs, Performance, and Preference: Implications for a BCI System". In: *International Journal of Psychophysiology* 59.2, pp. 127–140 (cit. on p. 51).
DOI: 10.1016/j.ijpsycho.2005.02.007

Aloise, F., P. Aricò, F. Schettini, A. Riccio, S. Salinari, D. Mattia, F. Babiloni, and F. Cincotti (2012). "A Covert Attention P300-Based Brain-Computer Interface: Geospell". In: *Ergonomics* 55.5, pp. 538–551 (cit. on pp. 6, 7, 55, 65).
DOI: 10.1080/00140139.2012.661084

Aloise, F., P. Aricò, F. Schettini, S. Salinari, D. Mattia, and F. Cincotti (2013). "Asynchronous Gaze-Independent Event-Related Potential-Based Brain-Computer Interface". In: *Artificial Intelligence in Medicine* 59.2, pp. 61–69 (cit. on p. 7).
DOI: 10.1016/j.artmed.2013.07.006

Aloise, F., F. Schettini, P. Aricò, F. Leotta, S. Salinari, D. Mattia, F. Babiloni, and F. Cincotti (2011). "P300-Based Brain-Computer Interface for Environmental Control: an Asynchronous Approach". In: *Journal of Neural Engineering* 8.2, p. 025025 (cit. on p. 3).
DOI: 10.1088/1741-2560/8/2/025025

Ang, K. K., C. Guan, K. S. G. Chua, B. T. Ang, C. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang (2010). "Clinical Study of Neurorehabilitation in Stroke using EEG-Based Motor Imagery Brain-Computer Interface with Robotic Feedback". In: *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Buenos Aires, Argentina: IEEE EMBS, pp. 5549–5552 (cit. on p. 3).
DOI: 10.1109/IEMBS.2010.5626782

Arnott, S. R. and C. Alain (2002). "Effects of Perceptual Context on Event-Related Brain Potentials During Auditory Spatial Attention". In: *Psychophysiology* 39.5, pp. 625–632 (cit. on p. 35).
DOI: 10.1111/1469-8986.3950625

Baillet, S., J. C. Mosher, and R. M. Leahy (2001). "Electromagnetic Brain Mapping". In: *IEEE Signal Processing Magazine* 18.6, pp. 14–30 (cit. on p. 19).
DOI: 10.1109/79.962275

Bartz, D. and K.-R. Müller (2013). "Generalizing Analytic Shrinkage for Arbitrary Covariance Structures". In: *Advances in Neural Information Processing Systems 26*. Ed. by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger. Lake Tahoe, Nevada: MIT Press, pp. 1869–1877 (cit. on p. 28).

Batteau, D. W. (1967). "The Role of the Pinna in Human Localization". In: *Proceedings of the Royal Society of London. Series B* 168.1011, pp. 158–180 (cit. on p. 16).
DOI: 10.1098/rspb.1967.0058

Bauer, G., F. Gerstenbrand, and E. Rumpl (1979). "Varieties of the Locked-In Syndrome". In: *Journal of Neurology* 221.2, pp. 77–91 (cit. on p. 3).
DOI: 10.1007/BF00313105

Belitski, A., J. D. R. Farquhar, and P. Desain (2011). "P300 Audio-Visual Speller". In: *Journal of Neural Engineering* 8.2, p. 025022 (cit. on pp. 5, 51, 52, 110).
DOI: 10.1088/1741-2560/8/2/025022

Bennemann, J., C. Freigang, E. Schröger, R. Rübsamen, and N. Richter (2013). "Resolution of Lateral Acoustic Space Assessed by Electroencephalography and Psychoacoustics". In: *Frontiers in Psychology* 4.338 (cit. on pp. 15, 50).
DOI: 10.3389/fpsyg.2013.00338

Bensch, M., A. a. Karim, J. Mellinger, T. Hinterberger, M. Tangermann, M. Bogdan, W. Rosenstiel, and N.-P. Birbaumer (2007). "Nessi: an EEG-Controlled Web Browser for Severely Paralyzed Patients". In: *Computational Intelligence and Neuroscience* 2007.71863, p. 5 (cit. on p. 3).
DOI: 10.1155/2007/71863

Bentley, A. S. J., C. M. Andrew, and L. R. John (2008). "An Offline Auditory P300 Brain-Computer Interface Using Principal and Independent Component Analysis Techniques for Functional Electrical Stimulation Application". In: *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Vancouver, Canada: IEEE EMBS, pp. 4660–4663 (cit. on p. 50).
DOI: 10.1109/IEMBS.2008.4650252

Berger, H. (1929). "Über das Elektrenkephalogramm des Menschen". In: *Archiv für Psychiatrie und Nervenkrankheiten* 87.1, pp. 527–570 (cit. on p. 19).
DOI: 10.1007/BF01797193

Bianchi, L., S. Sami, A. Hillebrand, I. P. Fawcett, L. R. Quitadamo, and S. Seri (2010). "Which Physiological Components are More Suitable for Visual ERP Based Brain-Computer Interface? A Preliminary MEG/EEG Study". In: *Brain Topography* 23.2, pp. 180–185 (cit. on pp. 22, 24, 83).
DOI: 10.1007/s10548-010-0143-0

Bießmann, F. (2007). "Error Correcting Codes for the P300 Visual Speller". MA thesis. Eberhard-Karls-University Tübingen (cit. on p. 5).

Bin, G., X. Gao, Y. Wang, Y. Li, B. Hong, and S. Gao (2011). "A High-Speed BCI Based on Code Modulation VEP". In: *Journal of Neural Engineering* 8.2, p. 025015 (cit. on p. 5).
DOI: 10.1088/1741-2560/8/2/025015

Birbaumer, N.-P. and L. G. Cohen (2007). "Brain-Computer Interfaces: Communication and Restoration of Movement in Paralysis". In: *Journal of Physiology* 579.3, pp. 621–636 (cit. on p. 3).
DOI: 10.1113/jphysiol.2006.125633

Birbaumer, N.-P., N. Ghanayim, T. Hinterberger, I. Iversen, B. I. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor (1999). "A Spelling Device for the Paralysed". In: *Nature* 398.6725, pp. 297–298 (cit. on pp. 2–4, 67).
DOI: 10.1038/18581

Birbaumer, N.-P., A. Kübler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. I. Kotchoubey, N. Neumann, and H. Flor (2000). "The Thought Translation Device (TTD) for Completely Paralyzed Patients". In: *IEEE Transactions on Rehabilitation Engineering* 8.2, pp. 190–193 (cit. on p. 67).
DOI: 10.1109/86.847812

Blankertz, B., G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio (2007). "The Non-Invasive Berlin Brain-Computer Interface: Fast Acquisition of Effective Performance in Untrained Subjects". In: *NeuroImage* 37.2, pp. 539–550 (cit. on p. 5).
DOI: 10.1016/j.neuroimage.2007.01.051

Blankertz, B., G. Dornhege, M. Krauledat, M. Schröder, J. Williamson, R. Murray-Smith, and K.-R. Müller (2006). "The Berlin Brain-Computer Interface Presents the Novel Mental Typewriter Hex-o-Spell". In: *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*. Ed. by G. R. Müller-Putz, C. Brunner, R. Leeb, R. Scherer, A. Schlögl, G. Wriessnegger, and G. Pfurtscheller. Graz, Austria: Verlag der Technischen Universität Graz, pp. 108–109 (cit. on pp. 3, 5, 56).

Blankertz, B., G. Dornhege, C. Schäfer, R. Krepki, J. Kohlmorgen, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio (2003). "Boosting Bit Rates and Error Detection for the Classification of Fast-Paced Motor Ccommands Based on Single-Trial EEG Analysis". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11.2, pp. 127–131 (cit. on p. 3).
DOI: 10.1109/TNSRE.2003.814456

Blankertz, B., S. Lemm, M. S. Treder, S. Haufe, and K.-R. Müller (2011). "Single-Trial Analysis and Classification of ERP Components – a Tutorial". In: *NeuroImage* 56.2, pp. 814–825 (cit. on pp. 21, 25, 27, 28, 32).
DOI: 10.1016/j.neuroimage.2010.06.048

Blankertz, B., M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. Ramsey, I. Sturm, G. Curio, and K.-R. Müller (2010). "The Berlin Brain–Computer Interface: Non-Medical Uses of BCI Technology". In: *Frontiers in Neuroscience* 4.17 (cit. on p. 3).
DOI: 10.3389/fnins.2010.00198

Blankertz, B., R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller (2008). "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis". In: *IEEE Signal Processing Magazine* 25.1, pp. 41–56 (cit. on pp. 1, 4).
DOI: 10.1109/MSP.2008.4408441

Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization*. Rev. Ed. Cambridge, Massachusetts: MIT Press, p. 508 (cit. on pp. 15, 50, 63).
ISBN: 9780262024136

Brainard, D. H. (1997). "The Psychophysics Toolbox". In: *Spatial Vision* 10.4, pp. 433–436 (cit. on p. 37).
DOI: 10.1163/156856897X00357

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Vol. 19. Massachusetts, USA: MIT Press, p. 773 (cit. on p. 52).
ISBN: 0262022974

Bregman, A. S., P. a. Ahad, P. A. C. Crum, and J. O'Reilly (2000). "Effects of Time Intervals and Tone Durations on Auditory Stream Segregation". In: *Perception & Psychophysics* 62.3, pp. 626–636 (cit. on pp. 52, 82).
DOI: 10.3758/BF03212114

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde: in Ihren Prinzipien Dargestellt auf Grund des Zellenbaues*. Leipzig: Barth, p. 324 (cit. on p. 18).

Broetz, D., C. Braun, C. Weber, S. R. Soekadar, A. Caria, and N.-P. Birbaumer (2010). "Combination of Brain-Computer Interface Training and Goal-Directed Physical Therapy in Chronic Stroke: a

Case Report". In: *Neurorehabilitation and Neural Repair* 24.7, pp. 674–679 (cit. on p. 3).
DOI: 10.1177/1545968310368683

Brouwer, A.-M. and J. B. F. van Erp (2010). "A Tactile P300 Brain-Computer Interface". In: *Frontiers in Neuroscience* 4.19 (cit. on pp. 6, 8).
DOI: 10.3389/fnins.2010.00019

Brown, A. W., C. L. Leibson, J. F. Malec, P. K. Perkins, N. N. Diehl, and D. R. Larson (2004). "Long-Term Survival after Traumatic Brain Injury: A Population-Based Analysis". In: *NeuroRehabilitation* 19.1, pp. 37–43 (cit. on p. 4).
DOI: 10.1097/HTR.0b013e318280d3e6

Brungart, D. S., N. I. Durlach, and W. M. Rabinowitz (1999). "Auditory Localization of Nearby Sources. II. Localization of a Broadband Source". In: *The Journal of the Acoustical Society of America* 106.4, pp. 1956–1968 (cit. on p. 50).
DOI: 10.1121/1.427943

Brunner, P., S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk (2010). "Does the 'P300' Speller Depend on Eye Gaze?" In: *Journal of Neural Engineering* 7.5, p. 056013 (cit. on pp. 7, 65, 83, 85, 109).
DOI: 10.1088/1741-2560/7/5/056013

Brunner, P. and G. Schalk (2013). "Toward Gaze-Independent Brain-Computer Interfaces". In: *Clinical Neurophysiology* 124.5, pp. 831–833 (cit. on pp. 7, 84).
DOI: 10.1016/j.clinph.2013.01.017

Buch, E., C. Weber, L. G. Cohen, C. Braun, M. a. Dimyan, T. Ard, J. Mellinger, A. Caria, S. R. Soekadar, A. Fourkas, and N.-P. Birbaumer (2008). "Think to Move: a Neuromagnetic Brain-Computer Interface (BCI) System for Chronic Stroke". In: *Stroke* 39.3, pp. 910–917 (cit. on p. 4).
DOI: 10.1161/STROKEAHA.107.505313

Cai, Z., S. Makino, and T. M. Rutkowski (2013). "Brain Evoked Potential Latencies Optimization for Spatial Auditory Brain–Computer Interface". In: *Cognitive Computation* (cit. on pp. 6, 9, 52, 110).
DOI: 10.1007/s12559-013-9228-x

Canton, R. (1875). "The Electric Currents of the Brain". In: *43rd Annual Meeting of the British Medical Association / British Medical Journal*. Vol. 2. 765, p. 278 (cit. on p. 19).
DOI: 10.1136/bmj.2.765.257

Carlile, S., P. Leong, and S. Hyams (1997). "The Nature and Distribution of Errors in Sound Localization by Human Listeners". In: *Hearing Research* 114.1–2, pp. 179–196 (cit. on pp. 16, 51).
DOI: 10.1016/S0378-5955(97)00161-5

Carlson, T., L. Tonin, S. Perdikis, R. Leeb, and J. d. R. Millán (2013). "A Hybrid BCI for Enhanced Control of a Telepresence Robot". In: *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Osaka, Japan: IEEE EMBS, pp. 3097–3100 (cit. on p. 65).
DOI: 10.1109/EMBC.2013.6610196

Carmena, J. M., M. A. Lebedev, R. E. Crist, J. E. O'Doherty, D. M. Santucci, D. F. Dimitrov, P. G. Patil, C. S. Henriquez, and M. A. L. Nicolelis (2003). "Learning to Control a Brain-Machine Interface for Reaching and Grasping by Primates". In: *PLoS Biology* 1.2, E42 (cit. on pp. 2, 4).
DOI: 10.1371/journal.pbio.0000042

Caroscio, J. T., M. N. Mulvihill, R. Sterling, and B. Abrams (1987). "Amyotrophic Lateral Sclerosis. Its Natural History". In: *Neurologic Clinics* 5.1, pp. 1–8 (cit. on p. 4).

Chatelle, C., S. Chennu, Q. Noirhomme, D. Cruse, A. M. Owen, and S. Laureys (2012). "Brain-Computer Interfacing in Disorders of Consciousness". In: *Brain Injury* 26.12, pp. 1510–1522 (cit. on p. 3).
DOI: 10.3109/02699052.2012.698362

Chatterjee, A., V. Aggarwal, A. Ramos, S. Acharya, and N. V. Thakor (2007). "A Brain-Computer Interface with Vibrotactile Biofeedback for Haptic Information". In: *Journal of NeuroEngineering and Rehabilitation* 4.40, p. 12 (cit. on p. 8).
DOI: 10.1186/1743-0003-4-40

Cincotti, F., L. Kauhanen, F. Aloise, T. Palomäki, N. Caporusso, P. Jylänki, D. Mattia, F. Babiloni, G. Vanacker, M. Nuttin, M. G. Marciani, and J. d. R. Millán (2007). "Vibrotactile Feedback for Brain-Computer Interface Operation". In: *Computational Intelligence and Neuroscience* 2007, p. 48937 (cit. on p. 8).
DOI: 10.1155/2007/48937

Cincotti, F., D. Mattia, F. Aloise, S. Bufalari, L. Astolfi, F. De Vico Fallani, A. Tocci, L. Bianchi, M. G. Marciani, S. Gao, J. d. R. Millán, and F. Babiloni (2008a). "High-Resolution EEG Techniques for Brain-Computer Interface Applications". In: *Journal of Neuroscience Methods* 167.1, pp. 31–42 (cit. on p. 4).
DOI: 10.1016/j.jneumeth.2007.06.031

Cincotti, F., D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A. Cherubini, M. G. Marciani, and F. Babiloni (2008b). "Non-Invasive Brain-Computer Interface System: Towards its Application as Assistive Technology". In: *Brain Research Bulletin* 75.6, pp. 796–803 (cit. on pp. 2, 3, 5, 67, 68, 84).
DOI: 10.1016/j.brainresbull.2008.01.007

Collinger, J. L., B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. C. McMorland, M. Velliste, M. L. Boninger, and A. B. Schwartz (2013). "High-Performance Neuroprosthetic Control by an Individual with Tetraplegia". In: *Lancet* 381.9866, pp. 557–564 (cit. on pp. 3, 4).
DOI: 10.1016/S0140-6736(12)61816-9

Combaz, A., C. Chatelle, A. Robben, G. Vanhoof, A. Goeleven, V. Thijs, M. M. van Hulle, and S. Laureys (2013). "A Comparison of Two Spelling Brain-Computer Interfaces Based on Visual P3 and SSVEP in Locked-In Syndrome". In: *PLoS ONE* 8.9, e73691 (cit. on pp. 3, 65, 67, 68).
DOI: 10.1371/journal.pone.0073691

Conroy, M. a. and J. Polich (2007). "Normative Variation of P3a and P3b from a Large Sample". In: *Journal of Psychophysiology* 21.1, pp. 22–32 (cit. on p. 22).
DOI: 10.1027/0269-8803.21.1.22

Cook, A. M. and J. M. Polgar (2008). *Cook and Hussey's Assistive Technologies: Principles and Practice*. 3rd Ed. St. Louis, Missouri: Mosby Elsevier, p. 592 (cit. on p. 2).
ISBN: 9780323039079

Corsi, P. M. (1972). "Human Memory and the Medial Temporal Region of the Brain". PhD thesis. Montreal, Canada: McGill University, p. 78 (cit. on pp. 76, 144).

Coull, J. T. (1998). "Neural Correlates of Attention and Arousal: Insights from Electrophysiology, Functional Neuroimaging and Psychopharmacology". In: *Progress in Neurobiology* 55.4, pp. 343–361 (cit. on p. 24).
DOI: 10.1016/S0301-0082(98)00011-2

Courchesne, E., R. Y. Courchesne, and S. A. Hillyard (1978). "The Effect of Stimulus Deviation on P3 Waves to Easily Recognized Stimuli". In: *Neuropsychologia* 16.2, pp. 189–199 (cit. on p. 23).
DOI: 10.1016/0028-3932(78)90106-9

Coyle, S., T. Ward, C. Markham, and G. McDarby (2004). "On the Suitability of Near-Infrared (NIR) Systems for Next-Generation Brain–Computer Interfaces". In: *Physiological Measurement* 25.4, pp. 815–822 (cit. on p. 4).
DOI: 10.1088/0967-3334/25/4/003

Cruse, D., S. Chennu, C. Chatelle, T. a. Bekinschtein, D. Fernández-Espejo, J. D. Pickard, S. Laureys, and A. M. Owen (2011). "Bedside Detection of Awareness in the Vegetative State: a Cohort Study". In: *Lancet* 378.9809, pp. 2088–2094 (cit. on pp. 3, 5).
DOI: 10.1016/S0140-6736(11)61224-5

Daly, J. J. and J. R. Wolpaw (2008). "Brain-Computer Interfaces in Neurological Rehabilitation". In: *Lancet Neurology* 7.11, pp. 1032–1043 (cit. on p. 3).
DOI: 10.1016/S1474-4422(08)70223-0

Davis, P. A. (1939). "Effects of Acoustic Stimuli on the Waking Human Brain". In: *Journal of Neurophysiology* 2.6, pp. 494–499 (cit. on pp. 24, 51).

Dennis, M. S., J. P. Burn, P. A. Sandercock, J. M. Bamford, D. T. Wade, and C. P. Warlow (1993). "Long-Term Survival After First-Ever Stroke: the Oxfordshire Community Stroke Project". In: *Stroke* 24.6, pp. 796–800 (cit. on p. 4).
DOI: 10.1161/01.STR.24.6.796

Deouell, L. Y., A. Parnes, N. Pickard, and R. T. Knight (2006). "Spatial Location is Accurately Tracked by Human Auditory Sensory Memory: Evidence from the Mismatch Negativity". In: *The European Journal of Neuroscience* 24.5, pp. 1488–1494 (cit. on p. 36).
DOI: 10.1111/j.1460-9568.2006.05025.x

Desain, P., A. M. G. Hupse, M. G. J. Kallenberg, B. J. de Kruif, and R. S. Schaefer (2006). "Brain-Computer Interfacing Using Selective Attention and Frequency-Tagged Stimuli". In: *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*. Ed. by G. R. Müller-Putz, C. Brunner, R. Leeb, R. Scherer, A. Schlögl, G. Wriessnegger, and G. Pfurtscheller. Graz, Austria: Verlag der Technischen Universität Graz, pp. 98–99 (cit. on p. 8).

Dobkin, B. H. (2007). "Brain-Computer Interface Technology as a Tool to Augment Plasticity and Outcomes for Neurological Rehabilitation". In: *Journal of Physiology* 579.3, pp. 637–642 (cit. on p. 3).
DOI: 10.1113/jphysiol.2006.123067

Donchin, E., K. Spencer, and R. Wijesinghe (2000). "The Mental Prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface". In: *IEEE Transactions on Rehabilitation Engineering* 8.2, pp. 174–179 (cit. on pp. 3, 67, 68).
DOI: 10.1109/86.847808

Donkelaar, H. J. ten and K. Kaga (2011). "The Auditory System". In: *Clinical Neuroanatomy - Brain Circuitry and Its Disorders*. Ed. by H. J. ten Donkelaar. 1st ed. Berlin / Heidelberg: Springer. Chap. 7, pp. 305–330 (cit. on p. 17).
DOI: 10.1007/978-3-642-19134-3_7

Dornhege, G., J. d. R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller (2007). *Toward Brain-Computer Interfacing*. Ed. by G. Dornhege, J. d. R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller. 1st Ed. Cambridge, Massachusetts: MIT Press, p. 520 (cit. on p. 1).
ISBN: 9780262042444

Duda, R. O., P. E. Hart, and D. G. Stork (2000). *Pattern Classification*. 2nd Ed. Wiley & Sons, p. 680 (cit. on p. 32).
ISBN: 9780471056690

Duncan, C. C., R. J. Barry, J. F. Connolly, C. Fischer, P. T. Michie, R. Näätänen, J. Polich, I. Reinvang, and C. van Petten (2009). "Event-Related Potentials in Clinical Research: Guidelines for Eliciting, Recording, and Quantifying Mismatch Negativity, P300, and N400". In: *Clinical Neurophysiology* 120.11, pp. 1883–1908 (cit. on pp. 21, 28, 41, 70, 82).
DOI: 10.1016/j.clinph.2009.07.045

Escolano, C., J. M. Antelis, and J. Minguez (2012). "A Telepresence Mobile Robot Controlled With a Noninvasive Brain-Computer Interface". In: *IEEE Transactions on Systems, Man, and Cybernetics* 42.3, pp. 793–804 (cit. on p. 2).
DOI: 10.1109/TSMCB.2011.2177968

Escolano, C., A. Ramos Murguialday, T. Matuz, N.-P. Birbaumer, and J. Minguez (2010). "A Telepresence Robotic System Operated with a P300-Based Brain-Computer Interface: Initial Tests with ALS Patients". In: *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Buenos Aires, Argentina: IEEE EMBS,

pp. 4476–4480 (cit. on pp. 2, 67).
DOI: 10.1109/IEMBS.2010.5626045

Farquhar, J. D. R., J. Blankespoor, R. J. Vlek, and P. Desain (2008). "Towards a Noise-Tagging Auditory BCI-Paradigm". In: *Proceedings of the 4th International BCI Conference, Graz*. Ed. by G. R. Müller-Putz, C. Brunner, R. Leeb, G. Pfurtscheller, and C. Neuper. Graz, Austria: Verlag der Technischen Universität Graz, pp. 50–55 (cit. on pp. 6, 8, 50, 52).

Farwell, L. A. and E. Donchin (1988). "Talking off the Top of your Head: Toward a Mental Prosthesis Utilizing Event-Related Brain Potentials". In: *Electroencephalography and Clinical Neurophysiology* 70.6, pp. 510–523 (cit. on pp. 3, 5, 9, 22, 25, 26, 33, 74, 76, 84, 110).
DOI: 10.1016/0013-4694(88)90149-6

Fawcett, T. (2006). "An Introduction to ROC Analysis". In: *Pattern Recognition Letters* 27.8, pp. 861–874 (cit. on p. 31).
DOI: 10.1016/j.patrec.2005.10.010

Fazli, S., J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K.-R. Müller, and B. Blankertz (2012). "Enhanced Performance by a Hybrid NIRS-EEG Brain Computer Interface". In: *NeuroImage* 59.1, pp. 519–529 (cit. on p. 4).
DOI: 10.1016/j.neuroimage.2011.07.084

Felton, E. A., J. C. Williams, G. C. Vanderheiden, and R. G. Radwin (2012). "Mental Workload during Brain-Computer Interface Training". In: *Ergonomics* 55.5, pp. 526–537 (cit. on pp. 65, 83).
DOI: 10.1080/00140139.2012.662526

Ferracuti, F., A. Freddi, S. Iarlori, S. Longhi, and P. Peretti (2013). "Auditory Paradigm for a P300 BCI System Using Spatial Hearing". In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Tokyo, Japan: IEEE, pp. 871–876 (cit. on p. 110).
DOI: 10.1109/IROS.2013.6696453

Finke, A., A. Lenhardt, and H. J. Ritter (2009). "The MindGame: a P300-Based Brain-Computer Interface Game". In: *Neural Networks* 22.9, pp. 1329–1333 (cit. on p. 3).
DOI: 10.1016/j.neunet.2009.07.003

Friedrich, E. V. C., C. Neuper, and R. Scherer (2013). "Whatever Works: a Systematic User-Centered Training Protocol to Optimize Brain-Computer Interfacing Individually". In: *PLoS ONE* 8.9, e76214 (cit. on p. 10).
DOI: 10.1371/journal.pone.0076214

Furdea, A., S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N.-P. Birbaumer, and A. Kübler (2009). "An Auditory Oddball (P300) Spelling System for Brain-Computer Interfaces". In: *Psychophysiology* 46.3, pp. 617–625 (cit. on pp. 5, 6, 9, 35, 50, 51, 53, 55, 64, 84, 85).
DOI: 10.1111/j.1469-8986.2008.00783.x

Galán, F., M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán (2008). "A Brain-Actuated Wheelchair: Asynchronous and Non-Invasive Brain-Computer Interfaces for Continuous Control of Robots". In: *Clinical Neurophysiology* 119.9, pp. 2159–2169 (cit. on pp. 2, 5).
DOI: 10.1016/j.clinph.2008.06.001

Gao, H., M. Ouyang, D. Zhang, and B. Hong (2011). "An Auditory Brain-Computer Interface Using Virtual Sound Field". In: *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Boston, USA: IEEE EMBS, pp. 4568–4571 (cit. on pp. 6, 9, 51, 52, 85, 110).
DOI: 10.1109/IEMBS.2011.6091131

Gerson, A. D., L. C. Parra, and P. Sajda (2006). "Cortically Coupled Computer Vision for Rapid Image Search". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14.2, pp. 174–179 (cit. on p. 3).
DOI: 10.1109/TNSRE.2006.875550
ISBN: 0001401106

Gerven, M. van, J. D. R. Farquhar, R. S. Schaefer, R. J. Vlek, J. Geuze, A. Nijholt, N. Ramsey, P. Haselager, L. Vuurpijl, S. Gielen, and P. Desain (2009). "The Brain-Computer Interface Cycle". In: *Journal of Neural Engineering* 6.4, p. 041001 (cit. on p. 1).
DOI: 10.1088/1741-2560/6/4/041001

Giguère, C. and S. M. Abel (1993). "Sound Localization: Effects of Reverberation Time, Speaker Array, Stimulus Frequency, and Stimulus Rise/Decay". In: *The Journal of the Acoustical Society of America* 94.2, pp. 769–776 (cit. on p. 50).
DOI: 10.1121/1.408206

Gloor, P. (1985). "Neuronal Generators and the Problem of Localization in Electroencephalography: Application of Volume Conductor Theory to Electroencephalography". In: *Journal of Clinical Neurophysiology* 2.4, pp. 327–354 (cit. on p. 19).

Gonsalvez, C. J., R. J. Barry, J. a. Rushby, and J. Polich (2007). "Target-to-Target Interval, Intensity, and P300 from an Auditory Single-Stimulus Task". In: *Psychophysiology* 44.2, pp. 245–250 (cit. on pp. 22, 51).
DOI: 10.1111/j.1469-8986.2007.00495.x

Gonsalvez, C. J., E. Gordon, S. Grayson, R. J. Barry, I. Lazzaro, and H. Bahramali (1999). "Is the Target-to-Target Interval a Critical Determinant of P3 Amplitude?" In: *Psychophysiology* 36.5, pp. 643–654 (cit. on p. 51).
DOI: 10.1111/1469-8986.3650643

Good, M. D. and R. H. Gilkey (1996). "Sound Localization in Noise: the Effect of Signal-to-Noise Ratio". In: *The Journal of the Acoustical Society of America* 99.2, pp. 1108–1117 (cit. on p. 50).
DOI: 10.1121/1.415233

Grant, D. A. and E. Berg (1948). "A Behavioral Analysis of Degree of Reinforcement and Ease of Shifting to New Responses in a Weigl-Type Card-Sorting Problem". In: *Journal of experimental psychology* 38.4, pp. 404–411 (cit. on p. 77).
DOI: 10.1037/h0059831

Green, D. M. and J. A. Swets (1966). *Signal Detection Theory and Psychophysics*. Reprint. Vol. 1. New York City: John Wiley & Sons, Inc., p. 521 (cit. on p. 32).
ISBN: 0932146236

Grozea, C., C. D. Voinescu, and S. Fazli (2011). "Bristle-Sensors–Low-Cost Flexible Passive Dry EEG Electrodes for Neurofeedback and BCI Applications". In: *Journal of Neural Engineering* 8.2, p. 025008 (cit. on pp. 5, 20).
DOI: 10.1088/1741-2560/8/2/025008

Guger, C., S. Daban, E. W. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger (2009). "How Many People are Able to Control a P300-Based Brain-Computer Interface (BCI)?" In: *Neuroscience Letters* 462.1, pp. 94–98 (cit. on pp. 5, 65).
DOI: 10.1016/j.neulet.2009.06.045

Guo, J., S. Gao, and B. Hong (2010). "An Auditory Brain-Computer Interface Using Active Mental Response". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.3, pp. 230–235 (cit. on pp. 9, 51, 85).
DOI: 10.1109/TNSRE.2010.2047604

Guo, J., B. Hong, F. Guo, X. Gao, and S. Gao (2009). "An Auditory BCI Using Voluntary Mental Response". In: *2009 4th International IEEE/EMBS Conference on Neural Engineering*. Antalya, Turkey: IEEE, pp. 455–458 (cit. on pp. 6, 9, 35, 50, 51, 53, 85).
DOI: 10.1109/NER.2009.5109331

Gürkök, H., G. Hakvoort, and M. Poel (2011). "Evaluating User Experience in a Selection Based Brain-Computer Interface Game - A Comparative Study". In: *Entertainment Computing - ICEC 2011*. Ed. by J. C. Anacleto, S. Fels, N. Graham, B. Kapralos, M. S. El-Nasr, and K. Stanley. Vol. 6972. Vancouver, Canada: Springer, pp. 77–88 (cit. on p. 65).
DOI: 10.1007/978-3-642-24500-8_9

Halder, S., E. M. Hammer, S. C. Kleih, M. Bogdan, W. Rosenstiel, N.-P. Birbaumer, and A. Kübler (2013). "Prediction of Auditory and Visual P300 Brain-Computer Interface Aptitude". In: *PLoS ONE* 8.2, e53513 (cit. on pp. 24, 51).
DOI: 10.1371/journal.pone.0053513

Halder, S., M. Rea, R. Andreoni, F. Nijboer, E. M. Hammer, S. C. Kleih, N.-P. Birbaumer, and A. Kübler (2010). "An Auditory Oddball Brain-Computer Interface for Binary Choices". In: *Clinical Neurophysiology* 121.4, pp. 516–523 (cit. on pp. 6, 9, 55, 85).
DOI: 10.1016/j.clinph.2009.11.087

Hart, S. G. and L. E. Staveland (1988). "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research". In: *Advances in Psychology* 52.3, pp. 139–183 (cit. on pp. 29, 147).
DOI: 10.1016/S0166-4115(08)62386-9

Hartmann, W. M. (1983). "Localization of Sound in Rooms". In: *The Journal of the Acoustical Society of America* 74.5, pp. 1380–1391 (cit. on pp. 16, 50).
DOI: 10.1121/1.390163

Haufe, S. (2011). "Towards EEG Source Connectivity Analysis". PhD thesis. Technische Universität Berlin (cit. on pp. 4, 19).

Haufe, S., M. S. Treder, M. F. Gugler, M. Sagebaum, G. Curio, and B. Blankertz (2011). "EEG Potentials Predict Upcoming Emergency Brakings during Simulated Driving". In: *Journal of Neural Engineering* 8.5, p. 056001 (cit. on p. 3).
DOI: 10.1088/1741-2560/8/5/056001

Heaton, R. K., G. J. Chelune, J. L. Talley, G. G. Kay, and G. Curtiss (1993). *Wisconsin Card Sorting Test Manual – Revised and Expanded*. Odessa, FL: Psychological Assessment Resources Inc (cit. on p. 77).

– (2000). *WCST: Wisconsin Card Sorting Test (Italian)*. Firenze: Giunti-Organizzazioni Speciali (cit. on p. 77).

Heo, J., H. J. Baek, S. H. Hong, and K. S. Park (2013). "Auditory Brain Computer Interface Using Natural Sounds". In: *Proceedings of the 5th International BCI Meeting*. Ed. by J. d. R. Millán, S. Gao, G. R. Müller-Putz, J. R. Wolpaw, and J. E. Huggins. Monterey, USA: Verlag der Technischen Universität Graz (cit. on p. 86).
DOI: 10.3217/978-3-85125-260-6-158

Hild ll, K. E., M. Kurimo, and V. D. Calhoun (2010). "The Sixth Annual MLSP Competition, 2010". In: *IEEE International Workshop on Machine Learning for Signal Processing*. Vol. 6. Mlsp. Kittila, Finland: IEEE, pp. 107–111 (cit. on p. 3).
DOI: 10.1109/MLSP.2010.5589227

Hill, N. J., T. N. Lal, K. Bierig, N.-P. Birbaumer, and B. Schölkopf (2004). "Attentional Modulation of Auditory Event-Related Potentials in a Brain-Computer Interface". In: *IEEE International Workshop on Biomedical Circuits and Systems, 2004*. IEEE, S3/17–20 (cit. on pp. 6, 8, 35, 52, 53, 109).
DOI: 10.1109/BIOCAS.2004.1454156

– (2005). "An Auditory Paradigm for Brain-Computer Interfaces". In: *Advances in Neural Information Processing Systems 17*. Ed. by L. K. Saul, Y. Weiss, and L. Bottou. Cambridge, MA: MIT Press, pp. 569–576 (cit. on pp. 8, 50, 85, 109).

Hill, N. J., T. N. Lal, M. Schröder, T. Hinterberger, B. Wilhelm, F. Nijboer, U. Mochty, G. Widman, C. Elger, B. Schölkopf, A. Kübler, and N.-P. Birbaumer (2006). "Classifying EEG and ECoG Signals Without Subject Training for Fast BCI Implementation: Comparison of Nonparalyzed and Completely Paralyzed Subjects". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14.2, pp. 183–186 (cit. on pp. 3, 68).
DOI: 10.1109/TNSRE.2006.875548

Hill, N. J., A. Moinuddin, A.-K. Häuser, S. Kienzle, and G. Schalk (2012a). "Communication and Control by Listening: Toward Optimal Design of a Two-Class Auditory Streaming Brain-Computer Interface". In: *Frontiers in Neuroscience* 6.181 (cit. on pp. 6, 8, 52).
DOI: 10.3389/fnins.2012.00181

Hill, N. J., E. Ricci, S. Haider, L. M. McCane, S. Heckman, J. R. Wolpaw, and T. M. Vaughan (2014). "A Practical, Intuitive Brain-Computer Interface for Communicating "Yes" or "No" by Listening". In: *Journal of Neural Engineering* 11.3, p. 035003 (cit. on p. 84).
DOI: 10.1088/1741-2560/11/3/035003

Hill, N. J. and B. Schölkopf (2012b). "An Online Brain-Computer Interface based on Shifting Attention to Concurrent Streams of Auditory Stimuli". In: *Journal of Neural Engineering* 9.2, p. 026011 (cit. on pp. 6, 8, 35, 52, 55).
DOI: 10.1088/1741-2560/9/2/026011

Hillyard, S. A., R. Hink, V. L. Schwent, and T. W. Picton (1973). "Electrical Signs of Selective Attention in the Human Brain". In: *Science* 182.4108, pp. 177–182 (cit. on p. 24).
DOI: 10.1126/science.182.4108.177

Hinterberger, T., N. Neumann, M. Pham, A. Kübler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor, and N.-P. Birbaumer (2004). "A Multimodal Brain-Based Feedback and Communication System". In: *Experimental Brain Research* 154.4, pp. 521–526 (cit. on p. 8).
DOI: 10.1007/s00221-003-1690-3

Hochberg, L. R., M. D. Serruya, G. M. Friehs, J. a. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue (2006). "Neuronal Ensemble Control of Prosthetic Devices by a Human with Tetraplegia". In: *Nature* 442.7099, pp. 164–171 (cit. on p. 3).
DOI: 10.1038/nature04970

Hoffmann, U., J.-M. Vesin, T. Ebrahimi, and K. Diserens (2008). "An Efficient P300-Based Brain-Computer Interface for Disabled Subjects". In: *Journal of Neuroscience Methods* 167.1, pp. 115–125 (cit. on p. 67).
DOI: 10.1016/j.jneumeth.2007.03.005

Hofman, M. and J. van Opstal (2003). "Binaural Weighting of Pinna Cues in Human Sound Localization". In: *Experimental Brain Research* 148.4, pp. 458–470 (cit. on p. 16).
DOI: 10.1007/s00221-002-1320-5

Höhne, J., K. Krenzlin, S. Dähne, and M. Tangermann (2012a). "Natural Stimuli improve Auditory BCIs with Respect to Ergonomics and Performance". In: *Journal of Neural Engineering* 9.4, p. 045003 (cit. on pp. 51, 85, 86, 97).
DOI: 10.1088/1741-2560/9/4/045003

Höhne, J., M. Schreuder, B. Blankertz, K.-R. Müller, and M. Tangermann (2010a). "Novel Paradigms for Auditory P300 Spellers with Spatial Hearing: Two Online Studies". In: *Frontiers in Computational Neuroscience - Bernstein Conference on Computational Neuroscience*. 44. Berlin, Germany: Frontiers Research Foundation (cit. on p. 55).
DOI: 10.3389/conf.fncom.2010.51.00044

– (2011a). "Novel Paradigms for Auditory P300 Spellers with Spatial Hearing: Two Online Studies". In: *Proceedings of the 2nd TOBI Workshop / International Journal of Bioelectromagnetism*. Vol. 13. 2, pp. 96–97 (cit. on p. 55).

Höhne, J., M. Schreuder, B. Blankertz, and M. Tangermann (2010b). "Two-Dimensional Auditory P300 Speller with Predictive Text System". In: *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Buenos Aires, Argentina: IEEE EMBS, pp. 4185–4188 (cit. on pp. 9, 52, 88, 89, 95, 110, 153).
DOI: 10.1109/IEMBS.2010.5627379

– (2011b). "A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System". In: *Frontiers in Neuroscience* 5.99 (cit. on pp. 6, 9, 52, 55, 65, 87, 110).
DOI: 10.3389/fnins.2011.00099

Höhne, J., M. Schreuder, and M. Tangermann (2011c). "Auditory ERP Speller Applications as a Tool for BCI End-Users". In: *Frontiers in Human Neuroscience - XI International Conference on Cognitive Neuroscience (ICON XI)*. Vol. 108. Palma de Mallorca, Spain: Frontiers Research Foundation (cit. on p. 55).
DOI: 10.3389/conf.fnhum.2011.207.00108

Höhne, J. and M. Tangermann (2012b). "How Stimulation Speed Affects Event-Related Potentials and BCI Performance". In: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. San Diego, USA: IEEE EMBS, pp. 1802–1805 (cit. on p. 52).
DOI: 10.1109/EMBC.2012.6346300

– (2014). "Towards User-friendly Spelling with an Auditory Brain-Computer Interface: The Char-Streamer Paradigm". In: *PLoS ONE* 9.6, e98322 (cit. on pp. 84, 86).
DOI: 10.1371/journal.pone.0098322

Holland, P. W. and R. E. Welsch (1977). "Robust Regression Using Iteratively Reweighted Least-Squares". In: *Communications in Statistics Theory and Methods* 6.9, pp. 813–827 (cit. on p. 103).
DOI: 10.1080/03610927708827533

Holz, E. M., J. Höhne, P. Staiger-Sälzer, M. Tangermann, and A. Kübler (2013a). "Brain-Computer Interface Controlled Gaming: Evaluation of Usability by Severely Motor Restricted End-Users". In: *Artificial Intelligence in Medicine* 59.2, pp. 111–120 (cit. on pp. 3, 65, 67).
DOI: 10.1016/j.artmed.2013.08.001

Holz, E. M., T. Kaufmann, L. Desideri, M. Malavasi, E.-j. Hoogerwerf, and A. Kübler (2013b). "User Centered Design in BCI Development". In: *Towards Practical Brain-Computer Interfaces*. Ed. by B. Z. Allison, S. Dunne, R. Leeb, J. Del R. Millán, and A. Nijholt. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 155–172 (cit. on pp. 10, 29).
DOI: 10.1007/978-3-642-29746-5_8

Hong, B., B. Lou, J. Guo, and S. Gao (2009). "Adaptive Active Auditory Brain Computer Interface". In: *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Minneapolis, USA: IEEE EMBS, pp. 4531–4534 (cit. on p. 50).
DOI: 10.1109/IEMBS.2009.5334133

Huggins, J. E., P. A. Wren, and K. L. Gruis (2011). "What Would Brain-Computer Interface Users want? Opinions and Priorities of Potential Users with Amyotrophic Lateral Sclerosis". In: *Amyotrophic Lateral Sclerosis* 12.5, pp. 318–324 (cit. on pp. 7, 35, 105).
DOI: 10.3109/17482968.2011.572978

ISO 9241-210 (2010). *Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. Tech. rep., p. 32 (cit. on pp. 10, 150).

Jin, J., B. Z. Allison, E. W. Sellers, C. Brunner, P. Horki, X. Wang, and C. Neuper (2011). "An Adaptive P300-Based Control System". In: *Journal of Neural Engineering* 8.3, p. 36006 (cit. on pp. 88, 89, 95, 107, 152).
DOI: 10.1088/1741-2560/8/3/036006

Johnson, R. J. (1993). "On the Neural Generators of the P300 Component of the Event-Related Potential". In: *Psychophysiology* 30.1, pp. 90–97 (cit. on p. 23).
DOI: 10.1111/j.1469-8986.1993.tb03208.x

Kandel, E. R., J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth (2012). *Principles of Neural Science*. 5th Ed. New York City: McGraw-Hill Professional, p. 1760 (cit. on p. 17).
ISBN: 0071390111

Kanoh, S., K.-i. Miyamoto, and T. Yoshinobu (2008). "A Brain-Computer Interface (BCI) System Based on Auditory Stream Segregation". In: *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 1. Vancouver, Canada: IEEE EMBS, pp. 642–645 (cit. on pp. 6, 8, 50, 52).
DOI: 10.1109/IEMBS.2008.4649234

Karim, A. A., T. Hinterberger, J. Richter, J. Mellinger, N. Neumann, H. Flor, A. Kübler, and N.-P. Birbaumer (2006). "Neural Internet: Web Surfing with Brain Potentials for the Completely Paralyzed". In: *Neurorehabilitation and Neural Repair* 20.4, pp. 508–515 (cit. on p. 3).
DOI: 10.1177/1545968306290661

Karmali, F., M. Polak, and A. Kostov (2000). "Environmental Control by a Brain-Computer Interface". In: *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Chicago, USA: IEEE EMBS, pp. 2990–2992 (cit. on p. 3).
DOI: 10.1109/IEMBS.2000.901508

Käthner, I., C. A. Ruf, E. Pasqualotto, C. Braun, N.-P. Birbaumer, and S. Halder (2013). "A Portable Auditory P300 Brain-Computer Interface with Directional Cues". In: *Clinical Neurophysiology* 124.2, pp. 327–338 (cit. on pp. 6, 9, 52, 55, 65, 110).
DOI: 10.1016/j.clinph.2012.08.006

Kaufmann, T., E. M. Holz, and A. Kübler (2013a). "Comparison of Tactile, Auditory, and Visual Modality for Brain-Computer Interface Use: a Case Study with a Patient in the Locked-In State". In: *Frontiers in Neuroscience* 7.129 (cit. on pp. 6, 8, 35, 67, 68, 84, 85).
DOI: 10.3389/fnins.2013.00129

Kaufmann, T., S. M. Schulz, C. Grünzinger, and A. Kübler (2011). "Flashing Characters with Famous Faces Improves ERP-Based Brain-Computer Interface Performance". In: *Journal of Neural Engineering* 8.5, p. 056016 (cit. on pp. 5, 87).
DOI: 10.1088/1741-2560/8/5/056016

Kaufmann, T., S. M. Schulz, A. Köblitz, G. Renner, C. Wessig, and A. Kübler (2013b). "Face Stimuli Effectively Prevent Brain-Computer Interface Inefficiency in Patients with Neurodegenerative Disease". In: *Clinical Neurophysiology* 124.5, pp. 893–900 (cit. on p. 5).
DOI: 10.1016/j.clinph.2012.11.006

Kelly, S. P., E. C. Lalor, R. B. Reilly, and J. J. Foxe (2005). "Visual Spatial Attention Tracking Using High-Density SSVEP Data for Independent Brain-Computer Communication". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 13.2, pp. 172–178 (cit. on pp. 6, 8).
DOI: 10.1109/TNSRE.2005.847369

Kim, D.-W., J.-H. Cho, H.-J. Hwang, J.-H. Lim, and C.-H. Im (2011a). "A Vision-Free Brain-Computer Interface (BCI) Paradigm Based on Auditory Selective Attention". In: *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Boston, USA: IEEE EMBS, pp. 3684–3687 (cit. on p. 8).
DOI: 10.1109/IEMBS.2011.6090623

Kim, D.-W., H.-J. Hwang, J.-H. Lim, Y.-H. Lee, K.-Y. Jung, and C.-H. Im (2011b). "Classification of Selective Attention to Auditory Stimuli: Toward Vision-Free Brain-Computer Interfacing". In: *Journal of Neuroscience Methods* 197.1, pp. 180–185 (cit. on pp. 6, 8, 52).
DOI: 10.1016/j.jneumeth.2011.02.007

Kindermans, P.-J., M. Schreuder, B. Schrauwen, K.-R. Müller, and M. Tangermann (2014a). "True Zero-Training Brain-Computer Interfacing – An Online Study". In: *PLoS ONE* 9.7, e102504 (cit. on p. 88).
DOI: 10.1371/journal.pone.0102504

Kindermans, P.-J., M. Tangermann, K.-R. Müller, and B. Schrauwen (2014b). "Integrating Dynamic Stopping, Transfer Learning and Language Models in an Adaptive Zero-Training ERP Speller". In: *Journal of Neural Engineering* 11.3, p. 035005 (cit. on p. 88).

Kindermans, P.-J., H. Verschore, and B. Schrauwen (2013). "A Unified Probabilistic Approach to Improve Spelling in an Event-Related Potential-Based Brain-Computer Interface". In: *IEEE Transactions on Biomedical Engineering* 60.10, pp. 2696–2705 (cit. on p. 88).
DOI: 10.1109/TBME.2013.2262524

Kleih, S. C., F. Nijboer, S. Halder, and A. Kübler (2010). "Motivation Modulates the P300 Amplitude during Brain-Computer Interface Use". In: *Clinical Neurophysiology* 121.7, pp. 1023–1031 (cit.

on pp. 69, 83).
DOI: 10.1016/j.clinph.2010.01.034

Klobassa, D. S., T. M. Vaughan, P. Brunner, N. E. Schwartz, J. R. Wolpaw, C. Neuper, and E. W. Sellers (2009). "Toward a High-Throughput Auditory P300-Based Brain-Computer Interface". In: *Clinical Neurophysiology* 120.7, pp. 1252–1261 (cit. on pp. 5, 6, 9, 35, 50, 51, 55, 64, 65, 86).
DOI: 10.1016/j.clinph.2009.04.019

Kohlmorgen, J., G. Dornhege, M. L. Braun, B. Blankertz, K.-R. Müller, G. Curio, K. Hagemann, A. Bruns, M. Schrauf, and W. E. Kincses (2007). "Improving Human Performance in a Real Operating Environment Through Real-Time Mental Workload Detection". In: *Toward Brain-Computer Interfacing*. Ed. by G. Dornhege, J. d. R. Millán, T. Hinterberger, D. J. Mcfarland, and K.-R. Müller. Cambridge, Massachusetts: MIT Press. Chap. 24, pp. 409–422 (cit. on p. 3).
ISBN: 9780262042444

Kotchoubey, B. I., J. S. Jordan, B. Grözinger, K. P. Westphal, and H. H. Kornhuber (1996). "Event-Related Brain Potentials in a Varied-Set Memory Search Task: A Reconsideration". In: *Psychophysiology* 33.5, pp. 530–540 (cit. on p. 50).
DOI: 10.1111/j.1469-8986.1996.tb02429.x

Krepki, R., B. Blankertz, G. Curio, and K.-R. Müller (2007). "The Berlin Brain-Computer Interface (BBCI): Towards a New Communication Channel for Online Control in Gaming Applications". In: *Multimedia Tools and Applications* 33.1, pp. 73–90 (cit. on p. 3).
DOI: 10.1007/s11042-006-0094-3

Kübler, A. and N.-P. Birbaumer (2008). "Brain–Computer Interfaces and Communication in Paralysis: Extinction of Goal Directed Thinking in Completely Paralysed Patients?" In: *Clinical Neurophysiology* 119.11, pp. 2658–2666 (cit. on pp. 67, 68).
DOI: 10.1016/j.clinph.2008.06.019

Kübler, A., A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. I. Kotchoubey (2009). "A Brain-Computer Interface Controlled Auditory Event-Related Potential (P300) Spelling System for Locked-In Patients". In: *Annals of the New York Academy of Sciences* 1157.1. Ed. by L. S. Schiff N D, pp. 90–100 (cit. on pp. 3, 5, 6, 9, 35, 50, 51, 68, 82, 84, 85).
DOI: 10.1111/j.1749-6632.2008.04122.x

Kübler, A., E. M. Holz, T. Kaufmann, and C. Zickler (2013). "A User Centred Approach for Bringing BCI Controlled Applications to End-Users". In: *Brain-Computer Interface Systems – Recent Progress and Future Prospects*. Ed. by R. Fazel-Rezai. Intechopen, pp. 1–20 (cit. on pp. 29, 67).
DOI: 10.5772/55802

Kübler, A., N. Neumann, J. Kaiser, B. I. Kotchoubey, T. Hinterberger, and N.-P. Birbaumer (2001). "Brain-Computer Communication: Self-Regulation of Slow Cortical Potentials for Verbal Communication". In: *Archives of Physical Medicine and Rehabilitation* 82.11, pp. 1533–1539 (cit. on pp. 52, 64).
DOI: 10.1053/apmr.2001.26621

Kübler, A., F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N.-P. Birbaumer, and J. R. Wolpaw (2005). "Patients with ALS Can Use Sensorimotor Rhythms to Operate a Brain-Computer Interface". In: *Neurology* 64.10, pp. 1775–1777 (cit. on pp. 3, 5, 67).
DOI: 10.1212/01.WNL.0000158616.43002.6D

Kutas, M., G. McCarthy, and E. Donchin (1977). "Augmenting Mental Chronometry : The P300 as a Measure of Stimulus Evaluation Time". In: *Science* 197.4305, pp. 792–795 (cit. on pp. 22, 23).
DOI: 10.1126/science.887923

Laar, B. van de, B. Reuderink, D. Plass-Oude Bos, and D. Heylen (2010). "Evaluating User Experience of Actual and Imagined Movement in BCI Gaming". In: *International Journal of Gaming and Computer-Mediated Simulations* 2.4, pp. 33–47 (cit. on p. 3).
DOI: 10.4018/jgcms.2010100103

LaFleur, K., K. Cassady, A. Doud, K. Shades, E. Rogin, and B. He (2013). "Quadcopter Control in Three-Dimensional Space Using a Noninvasive Motor Imagery-Based Brain-Computer Interface". In: *Journal of Neural Engineering* 10.4, p. 046003 (cit. on pp. 2, 5).
DOI: 10.1088/1741-2560/10/4/046003

Lalor, E. C., S. P. Kelly, C. Finucane, R. Burke, R. Smith, R. B. Reilly, and G. McDarby (2005). "Steady-State VEP-Based Brain-Computer Interface Control in an Immersive 3D Gaming Environment". In: *EURASIP Journal on Advances in Signal Processing* 2005.19, pp. 3156–3164 (cit. on p. 3).
DOI: 10.1155/ASP.2005.3156

Lammers, W. J. and P. Badia (1989). "Habituation of P300 to Target Stimuli". In: *Physiology & Behavior* 45.3, pp. 595–601 (cit. on p. 23).
DOI: 10.1016/0031-9384(89)90079-6

Langendijk, E. H. and A. W. Bronkhorst (2002). "The Contribution of Spectral Cues to Human Sound Localization". In: *The Journal of the Acoustical Society of America* 112.4, p. 1583 (cit. on pp. 16, 50, 58).
DOI: 10.1121/1.1501901

Laureys, S., F. Pellas, P. Van Eeckhout, S. Ghorbel, C. Schnakers, F. Perrin, J. Berré, M.-E. Faymonville, K.-H. Pantke, F. Damas, M. Lamy, G. Moonen, and S. Goldman (2005). "The Locked-In Syndrome : What is it Like to be Conscious but Paralyzed and Voiceless?" In: *Progress in Brain Research*. Ed. by S. Laureys. Vol. 150. Amsterdam, the Netherlands: Elsevier B.V., pp. 495–511 (cit. on p. 4).
DOI: 10.1016/S0079-6123(05)50034-7

Ledoit, O. and M. Wolf (2004). "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices". In: *Journal of Multivariate Analysis* 88.2, pp. 365–411 (cit. on p. 28).
DOI: 10.1016/S0047-259X(03)00096-4

Legeny, J., R. Viciana-Abad, and A. Lecuyer (2013). "Toward Contextual SSVEP-Based BCI Controller: Smart Activation of Stimuli and Control Weighting". In: *IEEE Transactions on Computational Intelligence and AI in Games* 5.2, pp. 111–116 (cit. on p. 65).
DOI: 10.1109/TCIAIG.2013.2252348

Lemm, S., B. Blankertz, T. Dickhaus, and K.-R. Müller (2011). "Introduction to Machine Learning for Brain Imaging". In: *NeuroImage* 56.2, pp. 387–399 (cit. on p. 29).
DOI: 10.1016/j.neuroimage.2010.11.004

Lenhardt, A., M. Kaper, and H. J. Ritter (2008). "An Adaptive P300-Based Online Brain-Computer Interface". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 16.2, pp. 121–130 (cit. on pp. 88, 89, 95, 107, 152).
DOI: 10.1109/TNSRE.2007.912816

León-Carrión, J., P. van Eeckhout, and M. D. R. Domínguez-Morales (2002). "The Locked-In Syndrome: A Syndrome Looking for a Therapy". In: *Brain Injury* 16.7, pp. 555–569 (cit. on p. 4).
DOI: 10.1080/02699050110119466

Leuthardt, E. C., G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran (2004). "A Brain-Computer Interface Using Electrocorticographic Signals in Humans". In: *Journal of Neural Engineering* 1.2, pp. 63–71 (cit. on p. 4).
DOI: 10.1088/1741-2560/1/2/001

Liu, T., L. Goldberg, S. Gao, and B. Hong (2010). "An Online Brain-Computer Interface Using Non-Flashing Visual Evoked Potentials". In: *Journal of Neural Engineering* 7.3, p. 036003 (cit. on pp. 88, 89, 95, 151).
DOI: 10.1088/1741-2560/7/3/036003

Liu, Y., Z. Zhou, and D. Hu (2011). "Gaze Independent Brain-Computer Speller with Covert Visual Search Tasks". In: *Clinical Neurophysiology* 122.6, pp. 1127–1136 (cit. on pp. 6, 7, 55, 105).
DOI: 10.1016/j.clinph.2010.10.049

Lopez, M. A., H. Pomares, A. Prieto, and F. Pelayo (2008). "Signal Processing and Perceptrons in an Auditory Based Brain-Computer Interface". In: *2008 8th International Conference on Hybrid*

Intelligent Systems. Barcelona, Spain: IEEE, pp. 781–786 (cit. on pp. 8, 50).
DOI: 10.1109/HIS.2008.77

Lopez-Gordo, M. A., E. Fernandez, S. Romero, F. Pelayo, and A. Prieto (2012). "An Auditory Brain–Computer Interface Evoked by Natural Speech". In: *Journal of Neural Engineering* 9.3, p. 036013 (cit. on pp. 6, 8, 52, 85).
DOI: 10.1088/1741-2560/9/3/036013

Lopez-Gordo, M. A., R. Ron-Angevin, and F. P. Valle (2011). "Auditory Brain-Computer Interfaces for Complete Locked-In Patients". In: *Advances in Computational Intelligence / Proceedings of the 11th International Work-Conference on Artificial Neural Networks, IWANN*. Ed. by J. Cabestany, I. Rojas, and G. Joya. Vol. 6691. Torremolinos-Málaga, Spain: Springer Berlin Heidelberg, pp. 378–385 (cit. on p. 9).
DOI: 10.1007/978-3-642-21501-8_47

Lulé, D., Q. Noirhomme, S. C. Kleih, C. Chatelle, S. Halder, A. Demertzi, M.-A. Bruno, O. Gosseries, A. Vanhaudenhuyse, C. Schnakers, M. Thonnard, A. Soddu, A. Kübler, and S. Laureys (2013). "Probing Command Following in Patients with Disorders of Consciousness Using a Brain-Computer Interface". In: *Clinical Neurophysiology* 124.1, pp. 101–106 (cit. on p. 3).
DOI: 10.1016/j.clinph.2012.04.030

Magni, E., G. Binetti, A. Bianchetti, R. Rozzini, and M. Trabucchi (1996). "Mini-Mental State Examination: a Normative Study in Italian Elderly Population". In: *European Journal of Neurology* 3.3, pp. 198–202 (cit. on pp. 76, 147).
DOI: 10.1111/j.1468-1331.1996.tb00423.x

Mak, J. N., Y. Arbel, J. W. Minett, L. M. Mccane, B. Yuksel, D. B. Ryan, D. Thompson, L. Bianchi, and D. Erdogmus (2011). "Optimizing the P300-Based Brain – Computer Interface : Current Status, Limitations and Future Directions". In: *Journal of Neural Engineering* 8.2, p. 025003 (cit. on p. 67).
DOI: 10.1088/1741-2560/8/2/025003

Makous, J. C. and J. C. Middlebrooks (1990). "Two-Dimensional Sound Localization by Human Listeners". In: *The Journal of the Acoustical Society of America* 87.5, pp. 2188–2200 (cit. on pp. 15–17, 51).
DOI: 10.1121/1.399186

Mandel, C., T. Luth, T. Laue, T. Rofer, A. Graser, and B. Krieg-Bruckner (2009). "Navigating a Smart Wheelchair with a Brain-Computer Interface Interpreting Steady-State Visual Evoked Potentials". In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. St. Louis, USA: IEEE, pp. 1118–1125 (cit. on p. 2).
DOI: 10.1109/IROS.2009.5354534

Marchetti, M., F. Piccione, S. Silvoni, L. Gamberini, and K. Priftis (2013). "Covert Visuospatial Attention Orienting in a Brain-Computer Interface for Amyotrophic Lateral Sclerosis Patients". In: *Neurorehabilitation and Neural Repair* 27.5, pp. 430–438 (cit. on pp. 4, 6, 7, 67, 68).
DOI: 10.1177/1545968312471903

Mason, S. G., A. Bashashati, M. Fatourechi, K. F. Navarro, and G. E. Birch (2007). "A Comprehensive Survey of Brain Interface Technology Designs". In: *Annals of Biomedical Engineering* 35.2, pp. 137–169 (cit. on pp. 1, 4).
DOI: 10.1007/s10439-006-9170-0

Mattia, D., F. Pichiorri, P. Aricò, F. Aloise, and F. Cincotti (2013a). "Hybrid Brain-Computer Interaction for Functional Motor Recovery after Stroke". In: *Converging Clinical and Engineering Research on Neurorehabilitation*. Ed. by J. L. Pons, D. Torricelli, and M. Pajaro. Vol. 1. Berlin, Heidelberg: Springer, pp. 1275–1279 (cit. on p. 3).
DOI: 10.1007/978-3-642-34546-3_213

Mattia, D., F. Pichiorri, M. Molinari, and R. Rupp (2013b). "Brain Computer Interface for Hand Motor Function Restoration and Rehabilitation". In: *Towards Practical Brain-Computer Inter-*

faces. Ed. by B. Z. Allison, S. Dunne, R. Leeb, J. Del R. Millán, and A. Nijholt. Berlin, Heidelberg: Springer, pp. 131–153 (cit. on p. 3).
DOI: 10.1007/978-3-642-29746-5_7

Mellinger, J., G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N.-P. Birbaumer, and A. Kübler (2007). "An MEG-Based Brain-Computer Interface (BCI)". In: *NeuroImage* 36.3, pp. 581–593 (cit. on p. 4).
DOI: 10.1016/j.neuroimage.2007.03.019

Mertens, R. and J. Polich (1997). "P300 from a Single-Stimulus Paradigm: Passive Versus Active Tasks and Stimulus Modality". In: *Electroencephalography and Clinical Neurophysiology* 104.6, pp. 488–497 (cit. on p. 23).
DOI: 10.1016/S0168-5597(97)00041-5

Middlebrooks, J. C. and D. M. Green (1991). "Sound Localization by Human Listeners". In: *Annual Review of Psychology* 42.1, pp. 135–159 (cit. on p. 15).
DOI: 10.1146/annurev.ps.42.020191.001031

Millán, J. d. R., R. Rupp, G. R. 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 (2010). "Combining Brain-Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges". In: *Frontiers in Neuroscience* 4.161 (cit. on p. 2).
DOI: 10.3389/fnins.2010.00161

Mills, A. W. (1958). "On the Minimum Audible Angle". In: *The Journal of the Acoustical Society of America* 30.4, pp. 237–246 (cit. on pp. 15, 17, 50).
DOI: 10.1121/1.1909553

Mondor, T. A. and R. J. Zatorre (1995). "Shifting and Focusing Auditory Spatial Attention". In: *Journal of Experimental Psychology* 21.2, pp. 387–409 (cit. on p. 15).
DOI: 10.1037/0096-1523.21.2.387

Mugler, E. M., C. a. Ruf, S. Halder, M. Bensch, and A. Kübler (2010). "Design and Implementation of a P300-Based Brain-Computer Interface for Controlling an Internet Browser". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.6, pp. 599–609 (cit. on pp. 3, 67, 68).
DOI: 10.1109/TNSRE.2010.2068059

Müller, K.-R., C. W. Anderson, and G. E. Birch (2003). "Linear and Nonlinear Methods for Brain-Computer Interfaces". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11.2, pp. 165–169 (cit. on p. 28).
DOI: 10.1109/TNSRE.2003.814484

Müller, K.-R. and B. Blankertz (2006). "Toward Noninvasive Brain-Computer Interfaces". In: *Signal Processing Magazine* 23.5, pp. 125–126, 128 (cit. on pp. 5, 56).
DOI: 10.1109/MSP.2006.1708426

Müller, K.-R., M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz (2004). "Machine Learning Techniques for Brain-Computer Interfaces". In: *Biomedical Engineering / Biomedizinische Technik* 49.1, pp. 11–22 (cit. on p. 32).

Müller, K.-R., S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf (2001). "An Introduction to Kernel-Based Learning Algorithms". In: *IEEE Transactions on Neural Networks* 12.2, pp. 181–201 (cit. on p. 29).
DOI: 10.1109/72.914517

Müller, K.-R., M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz (2008). "Machine Learning for Real-Time Single-Trial EEG-Analysis: From Brain-Computer Interfacing to Mental State Monitoring". In: *Journal of Neuroscience Methods* 167.1, pp. 82–90 (cit. on pp. 3, 27).
DOI: 10.1016/j.jneumeth.2007.09.022

Müller-Putz, G. R., D. S. Klobassa, C. Pokorny, G. Pichler, H. Erlbeck, R. G. L. Real, A. Kübler, M. Risetti, and D. Mattia (2012). "The Auditory P300-Based SSBCI: A Door to Minimally Conscious Patients?" In: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. San Diego, USA: IEEE EMBS, pp. 4672–4675 (cit. on p. 3).
DOI: 10.1109/EMBC.2012.6347009

Müller-Putz, G. R., R. Scherer, C. Neuper, and G. Pfurtscheller (2006). "Steady-State Somatosensory Evoked Potentials: Suitable Brain Signals for Brain-Computer Interfaces?" In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14.1, pp. 30–37 (cit. on pp. 6, 8).
DOI: 10.1109/TNSRE.2005.863842

Müller-Putz, G. R., R. Scherer, G. Pfurtscheller, and R. Rupp (2005). "EEG-Based Neuroprosthesis Control: a Step Towards Clinical Practice". In: *Neuroscience Letters* 382.1–2, pp. 169–174 (cit. on p. 5).
DOI: 10.1016/j.neulet.2005.03.021

Münßinger, J. I., S. Halder, S. C. Kleih, A. Furdea, V. Raco, A. Hösle, and A. Kübler (2010). "Brain Painting: First Evaluation of a New Brain-Computer Interface Application with ALS-Patients and Healthy Volunteers". In: *Frontiers in Neuroscience* 4.182 (cit. on pp. 3, 5, 67, 68, 84, 85).
DOI: 10.3389/fnins.2010.00182

Murphy, T. I. and S. J. Segalowitz (2004). "Eliminating the P300 Rebound in Short Oddball Paradigms". In: *International Journal of Psychophysiology* 53.3, pp. 233–238 (cit. on p. 23).
DOI: 10.1016/j.ijpsycho.2004.05.001

Näätänen, R., A. W. K. Gaillard, and S. Mäntysalo (1978). "Early Selective-Attention Effect on Evoked Potential Reinterpreted". In: *Acta psychologica* 42.4, pp. 313–329 (cit. on p. 24).
DOI: 10.1016/0001-6918(78)90006-9

Näätänen, R., P. Paavilainen, T. Rinne, and K. Alho (2007). "The Mismatch Negativity (MMN) in Basic Research of Central Auditory Processing: A Review". In: *Clinical Neurophysiology* 118.12, pp. 2544–2590 (cit. on p. 24).
DOI: 10.1016/j.clinph.2007.04.026

Näätänen, R. and T. W. Picton (1987). "The N1 Wave of the Human Electric and Magnetic Response to Sound: a Review and an Analysis of the Component Structure". In: *Psychophysiology* 24.4, pp. 375–425 (cit. on pp. 24, 52).
DOI: 10.1111/j.1469-8986.1987.tb00311.x

Näätänen, R., M. Simpson, and N. E. Loveless (1982). "Stimulus Deviance and Evoked Potentials". In: *Biological Psychology* 14.1–2, pp. 53–98 (cit. on p. 24).
DOI: 10.1016/0301-0511(82)90017-5

Nambu, I., M. Ebisawa, M. Kogure, S. Yano, H. Hokari, and Y. Wada (2013). "Estimating the Intended Sound Direction of the User: Toward an Auditory Brain-Computer Interface Using Out-of-Head Sound Localization". In: *PLoS ONE* 8.2, e57174 (cit. on pp. 6, 9, 52, 110).
DOI: 10.1371/journal.pone.0057174

Neuper, C., G. R. Müller, A. Kübler, N.-P. Birbaumer, and G. Pfurtscheller (2003). "Clinical Application of an EEG-Based Brain–Computer Interface: a Case Study in a Patient with Severe Motor Impairment". In: *Clinical Neurophysiology* 114.3, pp. 399–409 (cit. on pp. 3, 67).
DOI: 10.1016/S1388-2457(02)00387-5

Nicolelis, M. A. L. (2003). "Brain-Machine Interfaces to Restore Motor Function and Probe Neural Circuits". In: *Nature Reviews Neuroscience* 4.5, pp. 417–422 (cit. on pp. 2, 4).
DOI: 10.1038/nrn1105

Nijboer, F., N.-P. Birbaumer, and A. Kübler (2010). "The Influence of Psychological State and Motivation on Brain-Computer Interface Performance in Patients with Amyotrophic Lateral Sclerosis - a Longitudinal Study". In: *Frontiers in Neuroscience* 4.55 (cit. on pp. 4, 29, 80, 83, 148).
DOI: 10.3389/fnins.2010.00055

Nijboer, F., A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N.-P. Birbaumer, and A. Kübler (2008a). "An Auditory Brain-Computer Interface (BCI)". In: *Journal of Neuroscience Methods* 167.1, pp. 43–50 (cit. on pp. 8, 67).
DOI: 10.1016/j.jneumeth.2007.02.009

Nijboer, F., E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N.-P. Birbaumer, and A. Kübler (2008b). "A P300-Based Brain-Computer Interface for People with Amyotrophic Lateral Sclerosis". In: *Clinical Neurophysiology* 119.8, pp. 1909–1916 (cit. on pp. 3, 4, 85).
DOI: 10.1016/j.clinph.2008.03.034

Nijholt, A., D. Plass-Oude Bos, and B. Reuderink (2009). "Turning Shortcomings into Challenges: Brain–Computer Interfaces for Games". In: *Entertainment Computing* 1.2, pp. 85–94 (cit. on p. 3).
DOI: 10.1016/j.entcom.2009.09.007

Nykopp, T. (2001). "Statistical Modelling Issues for the Adaptive Brain Interface". MA thesis. Teknillinen Korkeakoulu (cit. on p. 32).

Oldfield, S. R. and S. P. A. Parker (1984). "Acuity of Sound Localisation: a Topography of Auditory Space. I. Normal Hearing Conditions". In: *Perception* 13.5, pp. 581–600 (cit. on p. 17).
DOI: 10.1068/p130581

Orsini, A. (2003). "La Memoria Diretta e la Memoria Inversa di Cifre in Soggetti dai 16 ai 64 anni". In: *Bollettino di Psicologia Applicata* 239, pp. 73–77 (cit. on pp. 76, 81).

Ortner, R., Z. Lugo, R. Pruckl, C. Hintermuller, Q. Noirhomme, and C. Guger (2013). "Performance of a Tactile P300 Speller for Healthy People and Severely Disabled Patients". In: *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Osaka, Japan: IEEE EMBS, pp. 2259–2262 (cit. on pp. 8, 68).
DOI: 10.1109/EMBC.2013.6609987

Owen, A. M., M. R. Coleman, M. Boly, M. H. Davis, S. Laureys, and J. D. Pickard (2006). "Detecting Awareness in the Vegetative State". In: *Science* 313.5792, p. 1402 (cit. on p. 3).
DOI: 10.1126/science.1130197

Pasqualotto, E., A. Simonetta, V. Gnisci, and S. Federici (2011). "Toward a Usability Evaluation of BCIs". In: *Proceedings of the 2nd TOBI Workshop / International Journal of Bioelectromagnetism* 13.3, pp. 121–122 (cit. on p. 29).

Pero, S., C. Incoccia, B. Caracciolo, P. Zoccolotti, and R. Formisano (2006). "Rehabilitation of Attention in Two Patients with Traumatic Brain Injury by Means of 'Attention Process Training'". In: *Brain Injury* 20.11, pp. 1207–1219 (cit. on p. 81).
DOI: 10.1080/02699050600983271

Perrott, D. R. (1984). "Concurrent Minimum Audible Angle: A Re-Examination of the Concept of Auditory Spatial Acuity". In: *The Journal of the Acoustical Society of America* 75.4, pp. 1201–1206 (cit. on p. 17).
DOI: 10.1121/1.390771

Perrott, D. R. and K. Saberi (1990). "Minimum Audible Angle Thresholds for Sources Varying in Both Elevation and Azimuth". In: *The Journal of the Acoustical Society of America* 87.4, pp. 1728–1731 (cit. on p. 17).
DOI: 10.1121/1.399421

Pfurtscheller, G. and F. H. Lopes da Silva (1999). "Event-Related EEG/MEG Synchronization and Desynchronization: Basic Principles". In: *Clinical Neurophysiology* 110.11, pp. 1842–1857 (cit. on pp. 1, 5).
DOI: 10.1016/S1388-2457(99)00141-8

Pfurtscheller, G., G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp (2003). "'Thought' – Control of Functional Electrical Stimulation to Restore Hand Grasp in a Patient with Tetraplegia".

In: *Neuroscience Letters* 351.1, pp. 33–36 (cit. on pp. 3, 5, 67).
DOI: 10.1016/S0304-3940(03)00947-9

Pham, M., T. Hinterberger, N. Neumann, A. Kübler, N. Hofmayer, A. Grether, B. Wilhelm, J.-J. Vatine, and N.-P. Birbaumer (2005). "An Auditory Brain-Computer Interface Based on the Self-Regulation of Slow Cortical Potentials". In: *Neurorehabilitation and Neural Repair* 19.3, pp. 206–218 (cit. on p. 8).
DOI: 10.1177/1545968305277628

Phukan, J., M. Elamin, P. Bede, N. Jordan, L. Gallagher, S. Byrne, C. Lynch, N. Pender, and O. Hardiman (2012). "The Syndrome of Cognitive Impairment in Amyotrophic Lateral Sclerosis: a Population-Based Study". In: *Journal of Neurology, Neurosurgery, and Psychiatry* 83.1, pp. 102–108 (cit. on p. 4).
DOI: 10.1136/jnnp-2011-300188

Piccione, F., F. Giorgi, P. Tonin, K. Priftis, S. Giove, S. Silvoni, G. Palmas, and F. Beverina (2006). "P300-Based Brain Computer Interface: Reliability and Performance in Healthy and Paralysed Participants". In: *Clinical Neurophysiology* 117.3, pp. 531–537 (cit. on pp. 3, 67, 68).
DOI: 10.1016/j.clinph.2005.07.024

Picton, T. W. (1992). "The P300 Wave of the Human Event-Related Potential". In: *Journal of Clinical Neurophysiology* 9.4, pp. 456–479 (cit. on p. 22).

Picton, T. W. and S. A. Hillyard (1974a). "Human Auditory Evoked Potentials. I: Evaluation of Components". In: *Electroencephalography and Clinical Neurophysiology* 36, pp. 179–190 (cit. on p. 21).
DOI: 10.1016/0013-4694(74)90155-2

– (1974b). "Human Auditory Evoked Potentials. II: Effects of Attention". In: *Electroencephalography and Clinical Neurophysiology* 36, pp. 191–199 (cit. on p. 21).
DOI: 10.1016/0013-4694(74)90156-4

Pokorny, C., D. S. Klobassa, G. Pichler, H. Erlbeck, R. G. L. Real, A. Kübler, D. Lesenfants, D. Habbal, Q. Noirhomme, M. Risetti, D. Mattia, and G. R. Müller-Putz (2013). "The Auditory P300-Based Single-Switch Brain-Computer Interface: Paradigm Transition from Healthy Subjects to Minimally Conscious Patients". In: *Artificial Intelligence in Medicine* 59.2, pp. 81–90 (cit. on pp. 3, 9).
DOI: 10.1016/j.artmed.2013.07.003

Polich, J. (1987). "Task Difficulty, Probability, and Inter-Stimulus Interval as Determinants of P300 from Auditory Stimuli". In: *Electroencephalography and Clinical Neurophysiology* 68.4, pp. 311–320 (cit. on p. 51).
DOI: 10.1016/0168-5597(87)90052-9

– (1989). "Frequency, Intensity, and Duration as Determinants of P300 from Auditory Stimuli". In: *Journal of Clinical Neurophysiology* 6.3, pp. 277–286 (cit. on p. 22).

– (2007). "Updating P300: an Integrative Theory of P3a and P3b". In: *Clinical Neurophysiology* 118.10, pp. 2128–2148 (cit. on pp. 22, 23).
DOI: 10.1016/j.clinph.2007.04.019

Popescu, F., S. Fazli, Y. Badower, B. Blankertz, and K.-R. Müller (2007). "Single Trial Classification of Motor Imagination Using 6 Dry EEG Electrodes". In: *PLoS ONE* 2.7, e637 (cit. on pp. 5, 20).
DOI: 10.1371/journal.pone.0000637

Posner, M. I., M. J. Nissen, and R. M. Klein (1976). "Visual Dominance: an Information-Processing Account of its Origins and Significance". In: *Psychological Review* 83.2, pp. 157–171 (cit. on p. 24).
DOI: 10.1037/0033-295X.83.2.157

Pritchard, W. S., S. A. Shappell, and M. E. Brandt (1991). "Psychophysiology of N200/N400: A Review and Classification Scheme". In: *Advances in Psychophysiology* 4, pp. 43–106 (cit. on

pp. 24, 51).
ISBN: 1853020796

Quek, M., A. Ramsay, A. Crossan, A. Riccio, M. Schreuder, J. Höhne, S. Dähne, D. Mattia, M. Tangermann, and R. Murray-Smith (2012). "A BCI-Controlled Photo Browser for Social Integration: Perspectives of Friends and Family". In: *Proceedings of the 3rd TOBI Workshop*. Ed. by S. Kleih, T. Kaufmann, B. Hörning, and A. Kübler. Würzburg, Germany: Universität Würzburg, pp. 87–88 (cit. on pp. 67, 76).

Rader, S. K., J. L. Holmes, and E. J. Golob (2008). "Auditory Event-Related Potentials during a Spatial Working Memory Task". In: *Clinical Neurophysiology* 119.5, pp. 1176–1189 (cit. on p. 36).
DOI: 10.1016/j.clinph.2008.01.014

Rakerd, B. and W. M. Hartmann (2005). "Localization of Noise in a Reverberant Environment". In: *Auditory Signal Processing*. Ed. by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet. New York City: Springer New York, pp. 413–421 (cit. on p. 16).
DOI: 10.1007/0-387-27045-0_51

Ramos Murguialday, A., N. J. Hill, M. Bensch, S. Martens, S. Halder, F. Nijboer, B. Schölkopf, N.-P. Birbaumer, and A. Gharabaghi (2011). "Transition from the Locked In to the Completely Locked-In State: a Physiological Analysis". In: *Clinical Neurophysiology* 122.5, pp. 925–933 (cit. on pp. 7, 65, 68, 85, 109).
DOI: 10.1016/j.clinph.2010.08.019

Ravden, D. and J. Polich (1998). "Habituation of P300 from Visual Stimuli". In: *International Journal of Psychophysiology* 30.3, pp. 359–365 (cit. on p. 23).
DOI: 10.1016/S0167-8760(98)00039-7

Rayleigh, L. (1907). "On Our Perception of Sound Direction". In: *Philosophical Magazine Series 6* 13.74, pp. 214–232 (cit. on p. 16).
DOI: 10.1080/14786440709463595

Rheinberg, F., R. Vollmeyer, and B. D. Burns (2001). "FAM: Ein Fragebogen zur Erfassung aktuller Motivation in Lern- und Leistungssituationen". In: *Diagnostica* 47.2, pp. 57–66 (cit. on pp. 29, 148).
DOI: 10.1026//0012-1924.47.2.57

Riccio, A., F. Leotta, L. Bianchi, F. Aloise, C. Zickler, E.-J. Hoogerwerf, A. Kübler, D. Mattia, and F. Cincotti (2011). "Workload Measurement in a Communication Application Operated through a P300-Based Brain-Computer Interface". In: *Journal of Neural Engineering* 8.2, p. 025028 (cit. on pp. 29, 65).
DOI: 10.1088/1741-2560/8/2/025028

Riccio, A., D. Mattia, L. Simione, M. Olivetti, and F. Cincotti (2012a). "Eye-Gaze Independent EEG-Based Brain-Computer Interfaces for Communication". In: *Journal of Neural Engineering* 9.4, p. 045001 (cit. on pp. 7, 8, 83–85).
DOI: 10.1088/1741-2560/9/4/045001

Riccio, A., M. Schreuder, S. Dähne, A. Ramsay, M. Quek, A. Crossan, J. Höhne, S. C. Kleih, A. Kübler, R. Murray-Smith, M. Tangermann, and D. Mattia (2012b). "Bridging the Gap between BCI Technology Design and Severely Disabled End-Users: Experience from a Photo Browser Validation". In: *Proceedings of the 3rd TOBI Workshop*. Ed. by S. Kleih, T. Kaufmann, B. Hörning, and A. Kübler. Würzburg, Germany: Universität Würzburg, pp. 105–106 (cit. on pp. 67, 76).

Riccio, A., L. Simione, F. Schettini, A. Pizzimenti, M. Inghilleri, M. O. Belardinelli, D. Mattia, and F. Cincotti (2013). "Attention and P300-Based BCI Performance in People with Amyotrophic Lateral Sclerosis". In: *Frontiers in Human Neuroscience* 7.732 (cit. on p. 4).
DOI: 10.3389/fnhum.2013.00732

Ringholz, G. M., S. H. Appel, and M. Bradshaw (2005). "Prevalence and Patterns of Cognitive Impairment in Sporadic ALS". In: *Neurology* 65.4, pp. 586–590 (cit. on p. 4).
DOI: 10.1212/01.wnl.0000172911.39167.b6

Ritter, W., R. Simson, H. G. Vaughan, and D. Friedman (1979). "A Brain Event Related to the Making of a Sensory Discrimination". In: *Science* 203.4387, pp. 1358–1361 (cit. on p. 24).
DOI: 10.1126/science.424760

Roffler, S. K. and R. A. Butler (1968). "Factors That Influence the Localization of Sound in the Vertical Plane". In: *The Journal of the Acoustical Society of America* 43.6, pp. 1255–1259 (cit. on p. 16).
DOI: 10.1121/1.1910976

Rohm, M., M. Schneiders, C. Müller, A. Kreilinger, V. Kaiser, G. R. Müller-Putz, and R. Rupp (2013). "Hybrid Brain-Computer Interfaces and Hybrid Neuroprostheses for Restoration of Upper Limb Functions in Individuals with High-Level Spinal Cord Injury". In: *Artificial Intelligence in Medicine* 59.2, pp. 133–142 (cit. on pp. 3, 5, 65, 67).
DOI: 10.1016/j.artmed.2013.07.004

Rost, T., M. Schreuder, and M. Tangermann (2010). "Optimized Spatial Auditory Stimuli for an ERP-Based BCI Paradigm". In: *Proceedings of the 1st TOBI Workshop*. Graz, Austria: Technische Universität Graz, p. 45 (cit. on p. 110).

Ruf, C. A., N. Simon, I. Käthner, E. Pasqualotto, N.-P. Birbaumer, and S. Halder (2013). "Listen to the Frog! An Auditory P300 Brain-Computer Interface With Directional Cues and Natural Sounds". In: *Proceedings of the 5th International BCI Meeting*. Ed. by J. d. R. Millán, S. Gao, G. R. Müller-Putz, J. R. Wolpaw, and J. E. Huggins. Monterey, USA: Verlag der Technischen Universität Graz (cit. on pp. 51, 52, 86, 110).
DOI: 10.3217/978-3-85125-260-6-157

Ruffini, G., S. Dunne, E. Farres, I. Cester, P. C. P. Watts, S. P. Silva, C. Grau, L. Fuentemilla, J. Marco-Pallares, and B. Vandecasteele (2007). "ENOBIO Dry Electrophysiology Electrode; First Human Trial Plus Wireless Electrode System". In: *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Lyon, France: IEEE EMBS, pp. 6690–6694 (cit. on pp. 5, 20).
DOI: 10.1109/IEMBS.2007.4353895

Rutkowski, T. M., A. S. Cichocki, and D. P. Mandic (2009). "Spatial Auditory Paradigm for Brain Computer/Machine Interfacing". In: *International Workshop on the Principles and Applications of Spatial Hearing*. Ed. by Y. Suzuki, D. S. Brungart, Y. Iwaya, K. Iida, D. Cabrera, and H. Kato. Vol. 2009. Miyagi, Japan: World Scientific, pp. 1–2 (cit. on p. 36).

Rutkowski, T. M., F. Vialatte, A. S. Cichocki, D. P. Mandic, and A. K. Barros (2006). "Auditory Feedback for Brain Computer Interface Management – An EEG Data Sonification Approach". In: *Knowledge-Based Intelligent Information and Engineering Systems*. Ed. by B. Gabrys, R. J. Howlett, and L. C. Jain. Vol. 4253. Bournemouth, UK: Springer Berlin Heidelberg, pp. 1232–1239 (cit. on p. 8).
DOI: 10.1007/11893011_156

Sannelli, C. (2012). "Optimizing Spatial Filters to Reduce BCI Inefficiency". PhD thesis. Technische Universität (cit. on p. 4).

Sarro, L., F. Agosta, E. Canu, N. Riva, A. Prelle, M. Copetti, G. Riccitelli, G. Comi, and M. Filippi (2011). "Cognitive Functions and White Matter Tract Damage in Amyotrophic Lateral Sclerosis: a Diffusion Tensor Tractography Study". In: *American Journal of Neuroradiology* 32.10, pp. 1866–1872 (cit. on p. 4).
DOI: 10.3174/ajnr.A2658

Schaeff, S., M. S. Treder, B. Venthur, and B. Blankertz (2012). "Exploring Motion VEPs for Gaze-Independent Communication". In: *Journal of Neural Engineering* 9.4, p. 045006 (cit. on pp. 6, 8, 35, 53, 55, 100).
DOI: 10.1088/1741-2560/9/4/045006

Schalk, G., K. J. Miller, N. R. Anderson, J. A. Wilson, M. D. Smyth, J. G. Ojemann, D. W. Moran, J. R. Wolpaw, and E. C. Leuthardt (2008). "Two-Dimensional Movement Control Using Electro-

corticographic Signals in Humans". In: *Journal of Neural Engineering* 5.1, pp. 75–84 (cit. on p. 4).
DOI: 10.1088/1741-2560/5/1/008

Schlögl, A., J. Kronegg, J. E. Huggins, and S. G. Mason (2007). "Evaluation Criteria for BCI Research". In: *Toward Brain-Computer Interfacing*. Ed. by G. Dornhege, J. d. R. Millán, T. Hinterberger, D. J. Mcfarland, and K.-R. Müller. Cambridge, MA: MIT Press. Chap. 19, pp. 297–312 (cit. on p. 32).
ISBN: 9780262042444

Schretlen, D. J. and A. M. Shapiro (2003). "A Quantitative Review of the Effects of Traumatic Brain Injury on Cognitive Functioning". In: *International Review of Psychiatry* 15.4, pp. 341–349 (cit. on p. 82).
DOI: 10.1080/09540260310001606728

Schreuder, M. (2008a). *Classification of the P300 Response to Auditory Cues Differing in Spatial Location*. Master Report, Vrije Universiteit Amsterdam, the Netherlands (cit. on p. 35).

– (2008b). "The Non-Invasive BCI Cycle Revised – An Overview of the Components Involved in BCI Setups". MA thesis. Vrije Universiteit, Amsterdam, p. 38 (cit. on p. 1).

Schreuder, M., B. Blankertz, and M. Tangermann (2010a). "A New Auditory Multi-Class Brain-Computer Interface Paradigm: Spatial Hearing as an Informative Cue". In: *PLoS ONE* 5.4. Ed. by J. Yan, e9813 (cit. on p. 35).
DOI: 10.1371/journal.pone.0009813

Schreuder, M., J. Höhne, B. Blankertz, S. Haufe, T. Dickhaus, and M. Tangermann (2013a). "Optimizing Event-Related Potential Based Brain-Computer Interfaces: a Systematic Evaluation of Dynamic Stopping Methods". In: *Journal of Neural Engineering* 10.3, p. 36025 (cit. on p. 87).
DOI: 10.1088/1741-2560/10/3/036025

Schreuder, M., J. Höhne, M. S. Treder, B. Blankertz, and M. Tangermann (2011a). "Performance Optimization of ERP-Based BCIs Using Dynamic Stopping". In: *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Boston, USA: IEEE EMBS, pp. 4580–4583 (cit. on p. 87).
DOI: 10.1109/IEMBS.2011.6091134

Schreuder, M., A. Riccio, F. Cincotti, M. Risetti, B. Blankertz, M. Tangermann, and D. Mattia (2011b). "Putting AMUSE to Work: an End-User Study". In: *Proceedings of the 2nd TOBI Workshop / International Journal of Bioelectromagnetism*. Vol. 13. 3, pp. 139–140 (cit. on p. 67).

Schreuder, M., A. Riccio, A. Ramsay, S. Dähne, J. Höhne, M. Quek, A. Crossan, D. Mattia, R. Murray-Smith, and M. Tangermann (2012a). "End User Performance in a Novel Social BCI Application: The Photobrowser". In: *Proceedings of the 3rd TOBI Workshop*. Ed. by S. Kleih, T. Kaufmann, B. Hörning, and A. Kübler. Würzburg, Germany: Universität Würzburg, pp. 16–17 (cit. on pp. 67, 76).

Schreuder, M., A. Riccio, M. Risetti, S. Dähne, A. Ramsay, J. Williamson, D. Mattia, and M. Tangermann (2013b). "User-Centered Design in Brain-Computer Interfaces – A Case Study". In: *Artificial Intelligence in Medicine* 59.2, pp. 71–80 (cit. on pp. 10, 67, 74).
DOI: 10.1016/j.artmed.2013.07.005

Schreuder, M., T. Rost, and M. Tangermann (2011c). "Listen, You are Writing! Speeding Up Online Spelling with a Dynamic Auditory BCI". In: *Frontiers in Neuroscience* 5.112 (cit. on pp. 55, 87).
DOI: 10.3389/fnins.2011.00112

Schreuder, M. and M. Tangermann (2010b). "Online Spelling Using the Brand New Spatial Auditory P300 Paradigm". In: *Proceedings of the 1st TOBI Workshop*. Graz, Austria: Technische Universität Graz, p. 22 (cit. on p. 55).

– (2010c). "Online Spelling Using the New Spatial Auditory BCI". In: *Proceedings of the 4th International BCI Meeting*. Monterey, USA (cit. on p. 55).

Schreuder, M., M. Tangermann, and B. Blankertz (2009). “Initial Results of a High-Speed Spatial Auditory BCI”. In: *International Journal of Bioelectromagnetism* 11.2, pp. 105–109 (cit. on pp. 7, 35, 36, 52).

Schreuder, M., M. E. Thurlings, A.-M. Brouwer, J. B. F. van Erp, and M. Tangermann (2012b). “Exploring the Use of Tactile Feedback in an ERP-Based Auditory BCI”. In: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. San Diego, USA: IEEE EMBS, pp. 6707–6710 (cit. on p. 110).
DOI: 10.1109/EMBC.2012.6347533

Schröder, M. and J. Trouvain (2003). “The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching”. In: *International Journal of Speech Technology* 6.4, pp. 365–377 (cit. on pp. 37, 56).
DOI: 10.1023/A:1025708916924

Schultze-Kraft, M., S. Dähne, G. Curio, and B. Blankertz (2013). “Temporal and Spatial Distribution of Workload-Induced Power Modulations of EEG Rhythms”. In: *Proceedings of the 5th International BCI Meeting*. Ed. by J. d. R. Millán, S. Gao, G. R. Müller-Putz, J. R. Wolpaw, and J. E. Huggins (cit. on p. 3).
DOI: 10.3217/978-3-85125-260-6-116

Schwent, V. L., S. A. Hillyard, and R. Galambos (1976). “Selective Attention and the Auditory Vertex Potential. I. Effects of Stimulus Delivery Rate”. In: *Electroencephalography and Clinical Neurophysiology* 40.6, pp. 604–614 (cit. on p. 52).
DOI: 10.1016/0013-4694(76)90135-8

Sellers, E. W. and E. Donchin (2006a). “A P300-Based Brain-Computer Interface: Initial Tests by ALS Patients”. In: *Clinical Neurophysiology* 117.3, pp. 538–548 (cit. on pp. 3, 4, 6, 50–53, 55, 68, 84, 85).
DOI: 10.1016/j.clinph.2005.06.027

Sellers, E. W., D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw (2006b). “A P300 Event-Related Potential Brain–Computer Interface (BCI): The Effects of Matrix Size and Inter Stimulus Interval on Performance”. In: *Biological Psychology* 73.3, pp. 242–252 (cit. on p. 46).
DOI: 10.1016/j.biopsycho.2006.04.007

Sellers, E. W., T. M. Vaughan, and J. R. Wolpaw (2010). “A Brain-Computer Interface for Long-Term Independent Home Use”. In: *Amyotrophic Lateral Sclerosis* 11.5, pp. 449–455 (cit. on pp. 3, 4, 67, 68).
DOI: 10.3109/17482961003777470

Shannon, C. E. and W. Weaver (1950). “The Mathematical Theory of Communication”. In: *The Mathematical Gazette* 34.310, pp. 312–313 (cit. on p. 32).
DOI: 10.2307/3611062

Silvoni, S., A. Ramos Murguialday, M. Cavinato, C. Volpato, G. Cisotto, A. Turolla, F. Piccione, and N.-P. Birbaumer (2011). “Brain-Computer Interface in Stroke: A Review of Progress”. In: *Clinical EEG and Neuroscience* 42.4, pp. 245–252 (cit. on p. 3).
DOI: 10.1177/155005941104200410

Silvoni, S., C. Volpato, M. Cavinato, M. Marchetti, K. Priftis, A. Merico, P. Tonin, K. Koutsikos, F. Beverina, and F. Piccione (2009). “P300-Based Brain-Computer Interface Communication: Evaluation and Follow-Up in Amyotrophic Lateral Sclerosis”. In: *Frontiers in Neuroscience* 3.60 (cit. on pp. 3, 67, 85).
DOI: 10.3389/neuro.20.001.2009

Sitaram, R., H. Zhang, C. Guan, M. Thulasidas, Y. Hoshi, A. Ishikawa, K. Shimizu, and N.-P. Birbaumer (2007). “Temporal Classification of Multichannel Near-Infrared Spectroscopy Signals of Motor Imagery for Developing a Brain-Computer Interface”. In: *NeuroImage* 34.4, pp. 1416–1427 (cit. on p. 4).
DOI: 10.1016/j.neuroimage.2006.11.005

Sonnadara, R. R., C. Alain, and L. J. Trainor (2006). "Effects of Spatial Separation and Stimulus Probability on the Event-Related Potentials Elicited by Occasional Changes in Sound Location". In: *Brain Research* 1071.1, pp. 175–185 (cit. on p. 36).
DOI: 10.1016/j.brainres.2005.11.088

Spinnler, H. and G. Tognoni (1987). *Standardizzazione e Taratura Italiana di Test Neuropsicologici: Gruppo Italiano per lo Studio Neuropsicologico dell'Invecchiamento*. Milan, Italy: Masson Italia Periodici (cit. on p. 76).

Squires, N. K., K. C. Squires, and S. A. Hillyard (1975). "Two Varieties of Long-Latency Positive Waves Evoked by Unpredictable Auditory Stimuli in Man". In: *Electroencephalography and Clinical Neurophysiology* 38.4, pp. 387–401 (cit. on pp. 21–23, 51).
DOI: 10.1016/0013-4694(75)90263-1

Starr, A. and J. Achor (1975). "Auditory Brain Stem Responses in Neurological Disease". In: *Archives of Neurology* 32.11, pp. 761–768 (cit. on p. 21).
DOI: 10.1001/archneur.1975.00490530083009

Student, B. (1908). "The Probable Error of a Mean". In: *Biometrika* 6.1, pp. 1–25 (cit. on p. 32).

Sutton, S., M. Braren, J. Zubin, and E. R. John (1965). "Evoked-Potential Correlates of Stimulus Uncertainty". In: *Science* 150.3700, pp. 1187–1188 (cit. on pp. 1, 22, 24).
DOI: 10.1126/science.150.3700.1187

Tangermann, M., J. Höhne, and M. Schreuder (2013). "Fixed-Sequence Stimulus Presentation in ERP-BCI". In: *Proceedings of the 4th TOBI Workshop*. Sion, Switzerland: École Polytechnique Fédéderal de Lausanne, pp. 141–142 (cit. on p. 22).

Tangermann, M., J. Höhne, H. Stecher, and M. Schreuder (2012). "No Surprise — Fixed Sequence Event-Related Potentials for Brain-Computer Interfaces". In: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. San Diego, USA: IEEE EMBS, pp. 2501–2504 (cit. on p. 22).
DOI: 10.1109/EMBC.2012.6346472

Tangermann, M., M. Krauledat, K. Grzeska, M. Sagebaum, B. Blankertz, C. Vidaurre, and K.-R. Müller (2009). "Playing Pinball with Non-Invasive BCI". In: *Advances in Neural Information Processing Systems 21*. Ed. by D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou. Cambridge, MA: MIT Press, pp. 1641–1648 (cit. on p. 3).

Tangermann, M., M. Schreuder, S. Dähne, J. Höhne, S. Regler, A. Ramsay, M. Quek, J. Williamson, and R. Murray-Smith (2011). "Optimized Stimulation Events for a Visual ERP BCI". In: *Proceedings of the 2nd TOBI Workshop / International Journal of Bioelectromagnetism*. Vol. 13. 3, pp. 119–120 (cit. on pp. 5, 67, 76).

Tatemichi, T. K., D. W. Desmond, Y. Stern, M. Paik, M. Sano, and E. Bagiella (1994). "Cognitive Impairment after Stroke: Frequency, Patterns, and Relationship to Functional Abilities". In: *Journal of Neurology, Neurosurgery, and Psychiatry* 57.2, pp. 202–207 (cit. on p. 82).
DOI: 10.1136/jnnp.57.2.202

Teder-Sälejärvi, W. A. and S. A. Hillyard (1998). "The Gradient of Spatial Auditory Attention in Free Field: An Event-Related Potential Study". In: *Perception & Psychophysics* 60.7, pp. 1228–1242 (cit. on pp. 15, 35).
DOI: 10.3758/BF03206172

Throckmorton, C. S., K. A. Colwell, D. B. Ryan, E. W. Sellers, and L. M. Collins (2013). "Bayesian Approach to Dynamically Controlling Data Collection in P300 Spellers". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 21.3, pp. 508–517 (cit. on p. 88).
DOI: 10.1109/TNSRE.2013.2253125

Thurlings, M. E., J. B. F. van Erp, A.-M. Brouwer, B. Blankertz, and P. Werkhoven (2012). "Control-Display Mapping in Brain-Computer Interfaces". In: *Ergonomics* 55.5, pp. 564–580 (cit. on p. 8).
DOI: 10.1080/00140139.2012.661085

Thurlow, W. R., J. W. Mangels, and P. S. Runge (1967). "Head Movements During Sound Localization". In: *The Journal of the Acoustical Society of America* 42.2, pp. 489–493 (cit. on p. 16).
DOI: 10.1121/1.1910605

Tomioka, R. and K.-R. Müller (2010). "A Regularized Discriminative Framework for EEG Analysis with Application to Brain-Computer Interface". In: *NeuroImage* 49.1, pp. 415–432 (cit. on pp. 1, 28).
DOI: 10.1016/j.neuroimage.2009.07.045

Tonet, O., M. Marinelli, L. Citi, P. M. Rossini, L. Rossini, G. Megali, and P. Dario (2008). "Defining Brain-Machine Interface Applications by Matching Interface Performance with Device Requirements". In: *Journal of Neuroscience Methods* 167.1, pp. 91–104 (cit. on p. 5).
DOI: 10.1016/j.jneumeth.2007.03.015

Tonin, L., T. Carlson, R. Leeb, and J. d. R. Millán (2011). "Brain-Controlled Telepresence Robot by Motor-Disabled People". In: *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Boston, USA: IEEE EMBS, pp. 4227–4230 (cit. on pp. 2, 67).
DOI: 10.1109/IEMBS.2011.6091049

Townsend, G., 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 (2010). "A Novel P300-Based Brain – Computer Interface Stimulus Presentation Paradigm: Moving Beyond Rows and Columns". In: *Clinical Neurophysiology* 121.7, pp. 1109–1120 (cit. on pp. 3–5, 67).
DOI: 10.1016/j.clinph.2010.01.030

Treder, M. S. (2012). "Special Section on Gaze-Independent Brain-Computer Interfaces". In: *Journal of Neural Engineering* 9.4, p. 040201 (cit. on p. 7).
DOI: 10.1088/1741-2560/9/4/040201

Treder, M. S. and B. Blankertz (2010). "(C)overt Attention and Visual Speller Design in an ERP-Based Brain-Computer Interface". In: *Behavioral and Brain Functions* 6.28 (cit. on pp. 6, 7, 53, 65, 83, 85, 109).
DOI: 10.1186/1744-9081-6-28

Treder, M. S., N. M. Schmidt, and B. Blankertz (2011). "Gaze-Independent Brain-Computer Interfaces Based on Covert Attention and Feature Attention". In: *Journal of Neural Engineering* 8.6, p. 066003 (cit. on pp. 5–8, 35, 55, 87, 100).
DOI: 10.1088/1741-2560/8/6/066003

Ullsperger, P., G. Freude, and U. Erdmann (2001). "Auditory Probe Sensitivity to Mental Workload Changes - an Event-Related Potential Study". In: *International Journal of Psychophysiology* 40.3, pp. 201–209 (cit. on p. 64).
DOI: 10.1016/S0167-8760(00)00188-4

Vaughan, T. M., D. J. McFarland, G. Schalk, W. A. Sarnacki, D. J. Krusienski, E. W. Sellers, and J. R. Wolpaw (2006). "The Wadsworth BCI Research and Development Program: At Home With BCI". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14.2, pp. 229–233 (cit. on pp. 4, 68).
DOI: 10.1109/TNSRE.2006.875577

Verschore, H., P.-J. Kindermans, D. Verstraeten, and B. Schrauwen (2012). "Dynamic Stopping Improves the Speed and Accuracy of a P300 Speller". In: *Artificial Neural Networks and Machine Learning – ICANN 2012*. Ed. by A. E. Villa, W. Duch, P. Érdi, F. Masulli, and G. Palm. Vol. 7552. Lausanne, Switzerland: Springer Berlin / Heidelberg, pp. 661–668 (cit. on p. 88).
DOI: 10.1007/978-3-642-33269-2_83

Vianello, A., G. Arcaro, A. Palmieri, M. Ermani, F. Braccioni, F. Gallan, G. Soraru', and E. Pegoraro (2011). "Survival and Quality of Life after Tracheostomy for Acute Respiratory Failure in Patients with Amyotrophic Lateral Sclerosis". In: *Journal of Critical Care* 26.3, 329.e7–14 (cit. on p. 4).
DOI: 10.1016/j.jcrc.2010.06.003

Vidal, J. J. (1973). "Toward Direct Brain-Computer Communication". In: *Annual Review of Biophysics and Bioengineering* 2.1, pp. 157–180 (cit. on pp. vi, 1, 2, 5, 109).
DOI: 10.1146/annurev.bb.02.060173.001105

Volosyak, I. (2011). "SSVEP-Based Bremen-BCI Interface–Boosting Information Transfer Rates". In: *Journal of Neural Engineering* 8.3, p. 036020 (cit. on p. 5).
DOI: 10.1088/1741-2560/8/3/036020

Vos, M. de, K. Gandras, and S. Debener (2014). "Towards a Truly Mobile Auditory Brain-Computer Interface: Exploring the P300 to Take Away". In: *International Journal of Psychophysiology* 91.1, pp. 46–53 (cit. on pp. 6, 9, 52).
DOI: 10.1016/j.ijpsycho.2013.08.010

Waal, M. van der, M. Severens, J. Geuze, and P. Desain (2012). "Introducing the Tactile Speller: an ERP-Based Brain-Computer Interface for Communication". In: *Journal of Neural Engineering* 9.4, p. 045002 (cit. on pp. 6, 8, 35, 55, 64, 104).
DOI: 10.1088/1741-2560/9/4/045002

Wallach, H. (1940). "The Role of Head Movements and Vestibular and Visual Cues in Sound Localization". In: *Journal of Experimental Psychology* 27.4, pp. 339–368 (cit. on p. 16).
DOI: 10.1037/h0054629

Wallach, H., E. Newman, and M. Rosenzweig (1949). "The Precedence Effect in Sound Localization". In: *The American Journal of Psychology* 62.3, pp. 315–336 (cit. on pp. 16, 50).

Webb, L. M. (2013). "Towards Navigation of Google Streetview Using a Spatial Auditory P300-Based Brain-Computer Interface". MA thesis. University of Cape Town (cit. on p. 110).

Weiskopf, N., K. Mathiak, S. W. Bock, F. Scharnowski, R. Veit, W. Grodd, R. Goebel, and N.-P. Birbaumer (2004). "Principles of a Brain-Computer Interface (BCI) Based on Real-Time Functional Magnetic Resonance Imaging (fMRI)". In: *IEEE Transactions on Biomedical Engineering* 51.6, pp. 966–970 (cit. on p. 4).
DOI: 10.1109/TBME.2004.827063

Westheimer, G. (1965). "Visual Acuity". In: *Annual Review of Psychology* 16, pp. 359–380 (cit. on p. 15).
DOI: 10.1146/annurev.ps.16.020165.002043

Wightman, F. L. and D. J. Kistler (1999). "Resolution of Front-Back Ambiguity in Spatial Hearing by Listener and Source Movement". In: *The Journal of the Acoustical Society of America* 105.5, pp. 2841–2853 (cit. on pp. 16, 50).
DOI: 10.1121/1.426899

Wills, S. a. and D. J. C. MacKay (2006). "DASHER – An Efficient Writing System for Brain-Computer Interfaces?" In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14.2, pp. 244–246 (cit. on p. 5).
DOI: 10.1109/TNSRE.2006.875573

Wolpaw, J. R., N.-P. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan (2000). "Brain-Computer Interface Technology: a Review of the First International Meeting". In: *IEEE Transactions on Rehabilitation Engineering* 8.2, pp. 164–173 (cit. on p. 32).
DOI: 10.1109/TRE.2000.847807

Wolpaw, J. R., N.-P. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan (2002). "Brain-Computer Interfaces for Communication and Control". In: *Clinical Neurophysiology* 113.6, pp. 767–791 (cit. on p. 1).
DOI: 10.1016/S1388-2457(02)00057-3

Wolpaw, J. R. and D. J. McFarland (2004). "Control of a Two-Dimensional Movement Signal by a Noninvasive Brain–Computer Interface in Humans". In: *Proceedings of the National Academy of Sciences* 101.51, pp. 17849–17854 (cit. on pp. 3, 5, 67).
DOI: 10.1073/pnas.0403504101

Wolters, C. and J. C. de Munck (2007). "Volume Conduction". In: *Scholarpedia* 2.3, p. 1738 (cit. on p. 19).
DOI: 10.4249/scholarpedia.1738

Woods, D. L. and R. Elmasian (1986). "The Habituation of Event-Related Potentials to Speech Ssounds and Tones". In: *Electroencephalography and Clinical Neurophysiology* 65.6, pp. 447–459 (cit. on p. 23).
DOI: 10.1016/0168-5597(86)90024-9

Woodworth, R. S. (1938). *Experimental Psychology*. 1st Ed. Oxford, England: Holt, p. 889 (cit. on p. 16).
ISBN: 0030028019

Woolley, S. C., D. H. Moore, and J. S. Katz (2010). "Insight in ALS: Awareness of Behavioral Change in Patients with and without FTD". In: *Amyotrophic Lateral Sclerosis* 11.1–2, pp. 52–56 (cit. on p. 82).
DOI: 10.3109/17482960903171110

Yoshimoto, S., Y. Washizawa, T. Tanaka, H. Higashi, and J. Tamura (2012). "Toward Multi-Command Auditory Brain Computer Interfacing Using Speech Stimuli". In: *Proceedings of the Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*. Hollywood, USA: IEEE (cit. on pp. 51, 85).

Yu, T., Y. Li, J. Long, and Z. Gu (2012). "Surfing the Internet With a BCI Mouse". In: *Journal of Neural Engineering* 9.3, p. 036012 (cit. on p. 3).
DOI: 10.1088/1741-2560/9/3/036012

Zander, T. O. and C. Kothe (2011a). "Towards Passive Brain-Computer Interfaces: Applying Brain-Computer Interface Technology to Human-Machine Systems in General". In: *Journal of Neural Engineering* 8.2, p. 025005 (cit. on p. 3).
DOI: 10.1088/1741-2560/8/2/025005

Zander, T. O., M. Lehne, K. Ihme, S. Jatzev, J. Correia, C. Kothe, B. Picht, and F. Nijboer (2011b). "A Dry EEG-System for Scientific Research and Brain-Computer Interfaces". In: *Frontiers in Neuroscience* 5.53 (cit. on pp. 5, 20).
DOI: 10.3389/fnins.2011.00053

Zhang, D., A. Maye, X. Gao, B. Hong, A. K. Engel, and S. Gao (2010). "An Independent Brain-Computer Interface Using Covert Non-Spatial Visual Selective Attention". In: *Journal of Neural Engineering* 7.1, p. 16010 (cit. on p. 8).
DOI: 10.1088/1741-2560/7/1/016010

Zhang, D., Y. Wang, A. Maye, A. K. Engel, X. Gao, B. Hong, and S. Gao (2007). "A Brain-Computer Interface Based on Multi-Modal Attention". In: *2007 3rd International IEEE/EMBS Conference on Neural Engineering*. Kohala Coast, HI: IEEE, pp. 414–417 (cit. on pp. 6, 8).
DOI: 10.1109/CNE.2007.369697

Zhang, H., C. Guan, and C. Wang (2008). "Asynchronous P300-Based Brain-Computer Interfaces: a Computational Approach with Statistical Models". In: *IEEE Transactions on Biomedical Engineering* 55.6, pp. 1754–1763 (cit. on pp. 88, 95, 153, 154).
DOI: 10.1109/TBME.2008.919128

Zickler, C., V. di Donna, V. Kaiser, A. Al-Khodairy, S. C. Kleih, A. Kübler, M. Malavasi, D. Mattia, S. Mongardi, C. Neuper, M. Rohm, R. Rupp, P. Staiger-Sälzer, and E.-J. Hoogerwerf (2009). "BCI Applications for People with Disabilities: Defining User Needs and User Requirements". In: *Assistive Technology from Adapted Equipment to Inclusive Environments, AAATE*. Ed. by P. L. Emiliani, L. Burzagli, A. Como, F. Gabbanini, and A.-L. Salminen. Vol. 25. Amsterdam, the Netherlands: IOS Press, pp. 185–189 (cit. on pp. 2, 76).
DOI: 10.3233/978-1-60750-042-1-185

Zickler, C., S. Halder, S. C. Kleih, C. Herbert, and A. Kübler (2013). "Brain Painting: Usability Testing According to the User-Centered Design in End Users with Severe Motor Paralysis". In: *Artificial*

Intelligence in Medicine 59.2, pp. 99–110 (cit. on pp. 10, 65, 85).
DOI: 10.1016/j.artmed.2013.08.003

Zickler, C., A. Riccio, F. Leotta, S. Hillian-Tress, S. Halder, E. M. Holz, P. Staiger-Sälzer, E.-J. Hoogerwerf, L. Desideri, D. Mattia, and A. Kübler (2011). "A Brain-Computer Interface as Input Channel for a Standard Assistive Technology Software". In: *Clinical EEG and Neuroscience* 42.4, pp. 236–244 (cit. on pp. 65, 67).
DOI: 10.1177/155005941104200409

Zimmermann, P. and B. Fimm (1995). "Test for Attentional Performance (TAP)". In: *PsyTest, Herzogenrath* (cit. on pp. 76, 77).

Zoccolotti, P., A. Matano, G. Deloche, A. Cantagallo, A. Passadori, M. Leclercq, L. Braga, N. Cremel, P. Pittau, and M. Renom (2000). "Patterns of Attentional Impairment Following Closed Head Injury: A Collaborative European study". In: *Cortex* 36.1, pp. 93–107 (cit. on p. 81).
DOI: 10.1016/S0010-9452(08)70839-6

List of Figures

1.1. Introduction: Vidal's Original Approach to BCI 2

2.1. Fundamentals: The Three Axes of Sound Localization 16
2.2. Fundamentals: View on a Model of the Left Brain Hemisphere 17
2.3. Fundamentals: Basics of EEG Generation 18
2.4. Fundamentals: Example of Typical ERP Components 20
2.5. Fundamentals: Two Alternative Feature Extraction Methods 27
2.6. FundamentalsL The ROC Curve Explained 31

3.1. Experiment 1 - 4: Schematic Representation of the Setups Used in This Thesis . . . 36
3.2. Experiment 3 - 4: Channels Used for Classification 38
3.3. Experiment 1: *F*-Score Matrix for Key Response With Healthy Subjects 40
3.4. Experiment 2: P3 Response to Slow Stimulation 42
3.5. Experiment 2: N2 Response to Slow Stimulation 43
3.6. Experiment 2: N1 Response to Slow Stimulation 44
3.7. Experiment 3 - 4: P3 Response to Fast Stimulation 47
3.8. Experiment 3 - 4: N1 Response to Fast Stimulation 47
3.9. Experiment 3 - 4: Evolution of Classification Performance Over Iterations 49

4.1. Experiment 5: Experimental Protocol for Online Writing 56
4.2. Experiment 5: Spelling Interface With Minimal Reliance on Visual Ability 57
4.3. Experiment 5: ERP Response to Online Spelling for Healthy Subjects 59
4.4. Experiment 5: Online BCI Performance for Healthy Subjects 61
4.5. Experiment 5: Online BCI Performance Per Word for Healthy Subjects 61
4.6. Experiment 5: *F*-Score Matrix for Healthy Subject Online Writing 62
4.7. Experiment 5: Questionnaire Result Online Writing Healthy Subjects 63
4.8. Experiment 5: Illustration of the Interdependence of symb/min and char/min . . . 64

5.1. Experiment 6: *F*-Score Matrix for End-user Online Writing 70
5.2. Experiment 6: N1 Response to the Standard Oddball Task for End-users 71
5.3. Experiment 6: P3 Response to the Standard Oddball Task for End-users 72
5.4. Experiment 6: ERP Response to AMUSE for End-users 73
5.5. Experiment 6: Online BCI Performance Over Sessions for End-users 74
5.6. Experiment 6: Questionnaire Result for End-users 75
5.7. Experiment 7: Detailed Questionnaire Result for VPeg02ma 77
5.8. Experiment 7: ERP Response to the Auditory Standard Oddball for VPeg02ma . . . 78
5.9. Experiment 7: ERP Response to the Visual Standard Oddball for VPeg02ma 78
5.10. Experiment 7: ERP Response to AMUSE for VPeg02ma 79
5.11. Experiment 7: ERP Response to the Photobrowser for VPeg02ma 79
5.12. Experiment 7: Online BCI Performance Over Sessions for VPeg02ma 80

6.1. Experiment 8: Questionnaire Result for Online Validation of rank diff 91
6.2. Experiment 8: Online BCI Performance With Rank diff 92

6.3. Experiment 8: Contribution of Errors by Rank diff 94
6.4. Experiment 8: Quality of Trials Stopped by Rank diff 95
6.5. Experiment 9-11: Visualization of the Validation Protocol 96
6.6. Experiment 9: Visual Representation of the Artificial Datasets 98
6.7. Experiment 9: Dynamic Stopping Results for the Artificial Datasets 99
6.8. Experiment 10: Dynamic Stopping Results for the EEG Datasets 101
6.9. Experiment 10: Dynamic Stopping Gain for the EEG Datasets 102
6.10. Experiment 10: Influence of Training Set Size on Dynamic Stopping Performance . 103
6.11. Experiment 11: Dynamic Stopping Results With Approximated Hyperparameter Φ 104

A.1. Appendix Experiment 9: Full Early Stopping Results for the Artificial Datasets . . . 155
A.2. Appendix Experiment 10: Full Early Stopping Results for the EEG Datasets 156
A.3. Appendix Experiment 11: Regression Fit of Φ to AUC 157

List of Tables

1.1. Introduction: State of the Art in Covert Attention-Based BCI 6

2.1. Fundamentals: Parameters for ERP Peak Picking 29

3.1. Experiment 1: Key Press Accuracy 39
3.2. Experiment 1: Key Press Result Per Source Location 41
3.3. Experiment 3 - 4: Stimulus Properties 45
3.4. Experiment 3 - 4: BCI Classification Performance Results 48

4.1. Experiment 5: Stimulus Properties 58
4.2. Experiment 5: Subject Demographics and Online Performance Summary 60

5.1. Experiment 6: Demographic and Clinical End-user Data 69
5.2. Experiment 7: Results of Neuropsychological Assessments for VPeg02ma 81

6.1. Experiment 8: Subject Demographics and Online Performance Summary 93
6.2. Experiment 10: EEG Dataset Parameters 100
6.3. Experiment 11: Regression Coefficients for Approximation of Hyperparameter Φ . 105

List of Terms and Abbreviations

ALS – *Amyotrophic Lateral Sclerosis*
Rapidly progressing neurodegenerative disease, causing muscle weakness in the initial phase and complete loss of motor control in the final state, resulting in LIS or CLIS. Most patients die of respiratory failure before entering CLIS, but the lifetime can be extended through tracheotomy. ALS has been the main drive of the AT application of BCI.
Detailed description: Section 1.3
See also: CLIS, LIS, TBI

AMUSE – *Auditory MUlticlass Spatial Event-elated potential*
The paradigm that is the central topic of this thesis. Core to AMUSE is the use of spatially separated auditory sources to allow for a multiclass, auditory BCI.
Detailed description: Section 3.2

ANOVA – *Analysis of Variance*
(Set of) parametric statistical test(s) for the analysis of differences between group means.

AP – *Action Potential*
An action potential is the efferent signal of neurons passing through the axon and stimulating the release of neurotransmitters into the synaptic cleft.
Detailed description: Section 2.2
See also: PSP

AT – *Assistive Technology*
Name for the collection of technologies that support people with disabilities in regaining the ability to perform tasks that they are unable to do, or with greater than normal difficulty, due to their disability.
Detailed description: Section 1.2

AUC – *Area Under the ROC-Curve*
The area under the ROC curve summarizes the ROC curve in a single value.
Detailed description: Section 2.6.3
See also: ROC

Azimuth
Name for the angle that defines the location of a sound source on the horizontal plane.
Detailed description: Section 2.1

BAER – *Brainstem Auditory Evoked Response*
A complex of ERPs within the first ten milliseconds after an auditory stimulus, mainly originating in the brainstem.
Other names: Auditory Brainstem Response (ABR)
See also: ERP, VEP

BBCI – *Berlin Brain-Computer Interface*
BBCI refers to the BCI research group at the Berlin Institute of Technology, as well as the in-house developed toolbox of that group.

BCI – *Brain-Computer Interface*
A BCI establishes a direct connection between the brain and a computer or other device. It allows the user to convey his intent without using any of the brain's natural efferent pathways.
Detailed description: Section 1.1

bits/min
Unit of the ITR metric.
Detailed description: Section 2.6.4
See also: ITR

CBTT – *Corsi block-tapping test*
A psychological test for the assessment of visuo-spatial short term working memory (Corsi, 1972).
Detailed description: Section 5.2.2.1
See also: CSBTT

Char/Min – *Characters per Minute*
A metric for measuring the practical relevance of a BCI-based speller, indicating the actual number of characters that a user can write, or has written per minute.
Detailed description: Section 2.6.5
See also: Symb/min

Classifier
An algorithm for separating data into two or more classes, based on their statistical properties.
Detailed description: Section 2.4.4
See also: LDA, SVM, SWLDA

CLIS – *Completely Locked-In Syndrome*
A clinical condition where the patient is aware and awake, but unable to communicate due to complete paralysis. In contrast to LIS, a patient with CLIS has lost control over the eyes, which are often the last remaining communication possibility.
Other names: totally locked-in syndrome
Detailed description: Section 1.3
See also: LIS

Cone of confusion
The cone of confusion describes the effect that sound sources that have similar distances to each ear are difficult to separate due to the limited information contained in the binaural cues.
Detailed description: Section 2.1

Covert
Attention can be divided into covert and overt attention. Covert attention refers to the mental orienting of attentional resources. As an example, in visual attention it refers to the shift of attention to a particular part of the visual field, without redirecting the eye-gaze to it.
See also: Overt

CSBTT – *Corsi supraspan block-tapping test*
A psychological test for for the assessment of visuo-spatial short term working memory (Corsi, 1972).

Detailed description: Section 5.2.2.1
See also: CBTT

ECoG – *Electrocorticography*
Electrical neuroimaging method using recording electrodes that are inside the skull, but do not penetrate the cortical surface. ECoG is commonly used to determine the focus of intractable epilepsy and most ECoG-based BCI studies originate from this patient population.

EEG – *Electroencephalography*
Electrical neuroimaging method using recording electrodes on the scalp.
Detailed description: Section 2.2.1

EOG – *Electro Oculography*
Electrical recording of the eye movements, typically performed with two bipolar channels for vertical- and horizontal movements.

Epoch
The smallest time unit in an ERP-based BCI; the single stimulus. Additionally, epoch refers to the chunk of EEG data recorded for a single stimulus.
Detailed description: Section 2.3.1

ERD – *Event-Related Desynchronization*
The brain oscillates at rest, as is visible from the EEG. Directly after an event, the brain region responsible for initiating that event, or processing it, shows a reduction in synchronicity called ERD. ERDs are time- but not phaselocked.

ERP – *Event-Related Potential*
Brain response to external- or internal events, as measurable in the EEG. ERPs are both time- and phaselocked.
Detailed description: Section 2.2.2
See also: BAER, MMN, MVEP, N1, N2, P3, VEP

fMRI – *Functional Magnetic Resonance Imaging*
Magnetic resonance imaging (MRI) is a brain imaging technique that offers high spatial resolution and can target the subcortical structures. It is used to obtain anatomical images of the brain. With a particular modification, the oxygenation of the blood can be visualized, allowing for functional imaging (fMRI).

HFR – *High Frequency Range*
Upper range of the frequency spectrum that is audible to the human ear (roughly 8-20 Khz).

Hyperparameter
A free parameter of a method or model that is set by the user. Often, hyperparameters are set using cross-validation to find an optimal value, given the data.

ILD – *Interaural Level Differences*
One of the ears is partially masked by the head for sound sources that are off the median plane, and are thereby reduced in intensity. This is one of two binaural cues that the brain uses to localize sounds in space.
Detailed description: Section 2.1
See also: ITD

IQR – *Inter-Quartile Range*
Data descriptive summarizing the data in their distance between the 25% and the 75% percentiles. This measure is often used for data that is not Gaussian distributed, as it is a better estimator than the alternative standard deviation.
See also: SD

ITD – *Interaural Timing Differences*
Sounds have different path lengths to the ears for sound sources that are off the median plane, resulting in phase differences between the perceived sound in both ears. This is one of two binaural cues that the brain uses to localize sounds in space.
Detailed description: Section 2.1
See also: ILD

Iteration
An iteration refers to the consecutive stimulation of C unique stimuli, where C refers to the number of classes in the paradigm.
Detailed description: Section 2.3.1

ITR – *Information Transfer Rate*
Theoretical metric for measuring the quality of data transmission over a noisy channel, which is often used in BCI literature.
Detailed description: Section 2.6.4
See also: Bits/min

LDA – *Linear Discriminant Analysis*
Linear classifier algorithm often used in BCI literature.
Detailed description: Section 2.4.4
See also: Classifier

Likert scale
A discrete psychometric response scale. Typically, a statement is evaluated by the respondent with a fixed set of possible responses. For instance, the respondent could be asked to indicate how strongly he (dis)agrees with a statement, choosing from the following possible answers: strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree. This constitutes a 5-point Likert scale.
Detailed description: Section 2.5.2
See also: VAS

LIS – *Locked-In Syndrome*
A clinical condition where the patient is aware and awake, but unable to communicate due to near-complete paralysis. In contrast to CLIS, a patient with LIS has (residual) control over the eyes.
Detailed description: Section 1.3
See also: CLIS

MAA – *Minimum Audible Angle*
The minimum angle between two sources that allows a human listener to distinguish both sources uniquely.
Detailed description: Section 2.1

MEG – *Magnetoencephalograpy*
Magnetic neuroimaging method, using extremely sensitive magnetic sensors to measure

the electromagnetic fields produced by currents in the brain. The magnetic fields travel through the head without resistance, resulting in a better spatial resolution than EEG.

MI – *Motor Imagery*
A paradigm often used to elicit ERD, which relies on the fact that imagining a movement results in a desynchronization pattern that is similar to that when the movement is actually performed.
See also: ERD

MMN – *Mismatch Negativity*
A component of the ERP complex in the range of the N2. The MMN is an automatic response to changes in the stimulation pattern.
Detailed description: Section 2.2.2.3
See also: ERP, N2

MMSE – *Mini Mental State Examination*
A short questionnaire to assess the cognitive functioning that is often used in clinical settings (Magni et al., 1996)

MVEP – *Motion Visual Evoked Potential*
VEP evoked specifically by the moving of a stimulus.
See also: ERP, VEP

N1
Negative ERP component peaking around 100 ms after stimulus onset.
Detailed description: Section 2.2.2.4
See also: ERP

N2
Negative ERP component peaking around 200 ms after stimulus onset.
Detailed description: Section 2.2.2.3
See also: ERP, MMN

NASA-TLX – *NASA Task Load Index*
A questionnaire for assessing the overall perceived workload on six factors: *mental demand, physical demand, temporal demand, performance, effort,* and *frustration* (Hart et al., 1988).
Detailed description: Section 2.5.2

NIRS – *Near-Infrared Spectroscopy*
Neuroimaging method using the absorption pattern of near-infrared light by the brain to estimated the oxygenation of the brain area underlying the transmitters and receivers placed on the scalp.
See also: fMRI

Non-target
The stimuli that are to be ignored in an oddball paradigm.
Other names: standard
Detailed description: Section 2.2.2.1
See also: Target

Overt
Attention can be divided into covert and overt attention. Overt attention refers to the

orienting of attentional resources by physically shifting the sensor. As an example, in visual attention it refers to the shift of attention to a particular part of the visual field by redirecting the eye-gaze.
See also: Covert

P3
Positive ERP component peaking around 300 ms after stimulus onset.
Detailed description: Section 2.2.2.2
See also: ERP

PSP – *Postsynaptic Potential*
The transient potential deflection in a postsynaptic cell resulting from the release of neurotransmitters by a presynaptic cell. PSPs are the primary source of the EEG signal.
Detailed description: Section 2.2
See also: AP

QCM-BCI – *Questionnaire for Current Motivation in Brain-Computer Interfaces*
A questionnaire for assessing the subject's motivation on four factors: *mastery confidence*, *incompetence fear*, *interest*, and *challenge*. Introduced by Rheinberg et al. (2001) and adapted for use in BCI by Nijboer et al. (2010).
Detailed description: Section 2.5.2
See also: NASA-TLX

Rank diff
New method for optimizing the trade-off between accuracy and speed in ERP-based BCI, as introduced in this thesis.
Detailed description: Section 6.3

ROC – *Receiver-Operator Characteristic*
Graphical representation of the performance of a binary classifier.
Detailed description: Section 2.6.3
See also: AUC, Classifier, Signed r^2

RSVP – *Rapid Serial Visual Presentation*
Psychophysics paradigm in which images, symbols or characters are presented in rapid temporal succession. It is often used to study phenomena such as the attentional blink, but was more recently used as a BCI paradigm.

Run
A logical collection of trials within one session, sharing the same experimental condition. For online BCI spellers, a run typically refers to the writing of a single word or sentence.
Detailed description: Section 2.3.1

SCP – *Slow Cortical Potential*
Gradual changes in the EEG potential, lasting up to several seconds. After training, the SCPs can be initiated by the subject without external stimulus.

SD – *Standard deviation*
Statistical property of a Gaussian distribution, describing its width. 68.2% of the data in the distribution is contained within one SD above and below the sample mean.
See also: IQR

Session
A collection of experimental tasks that is performed on one day, and in one go. In BCI, it typically refers to the putting on- and removing of the cap and everything in between.
Detailed description: Section 2.3.1

Signed r^2
The pointwise biserial correlation coefficient (r^2) is a statistical method for describing the correlation between a continuous and a dichotomous variable. The signed version is often used in BCI literature to estimate the predictive power of individual features. It is obtained by multiplying the r^2 with the sign of r.
See also: ROC

SNR – *Signal-to-Noise Ratio*
A measure of the ratio of the power of a desired signal to that of the background noise.

SOA – *Stimulus Onset Asynchrony*
The time between the onsets of two consecutive stimuli, typically measured in milliseconds.

SSAEP – *Steady-State Auditory Evoked Potentials*
The phenomenon that a continuous auditory stimulus with a particular frequency resonates in the brain, causing a power increase in that frequency band in the brain region responsible for processing the stimulus.
Other names: Auditory Steady State Response (ASSR)
See also: SSSEP, SSVEP

SSSEP – *Steady-State Somatosensory Evoked Potentials*
The phenomenon that a continuous somatosensory stimulus with a particular frequency resonates in the brain, causing a power increase in that frequency band in the brain region responsible for processing the stimulus.
See also: SSAEP, SSVEP

SSVEP – *Steady-State Visually Evoked Potentials*
The phenomenon that a continuous visual stimulus with a particular frequency resonates in the brain, causing a power increase in that frequency band in the brain region responsible for processing the stimulus.
See also: SSAEP, SSSEP

SVM – *Support-Vector Machine*
A particular instance of a classifier, maximizing the margin between the samples of two classes. SVMs can be linear or non-linear, the latter of which is achieved by applying the kernel trick.
See also: Classifier

SWLDA – *Step-wise Linear Discriminant Analysis*
Extension of the LDA classifier. SWLDA includes an additional preprocessing step, selecting a subset of the features that significantly aid the classification task. All other features are set to zero, before applying standard LDA. SWLDA thus results in a sparse classifier.
See also: classifier, LDA

Symb/Min – *Symbols per Minute*
A metric measuring the performance of a communication BCI in terms of the number of correctly spelled symbols it produces per minute. It evaluates both the BCI and the control strategy as a whole. In contrast to the char/min metric, it approximates the performance

using the average accuracy and the time per trial.
Detailed description: Section 2.6.5
See also: Char/min, ITR

T9
Word prediction algorithm often used in older mobile phones for faster text input. Using a dictionary and word prediction, words are written using only nine keys and one keystroke per character.

Target
The stimulus that is to be attended amongst other stimuli.
Other names: deviant, oddball
Detailed description: Section 2.2.2.1
See also: Non-target

TBI – *Traumatic Brain Injury*
A clinical condition where the brain is traumatically damaged by an external force. TBI is a major cause of death and severe paralysis worldwide, causing LIS and CLIS.
Detailed description: Section 1.3
See also: ALS, CLIS, LIS

Trial
A series of stimuli, collectively leading to a single class decision. In two leveled spellers such as proposed in Experiment 5, two trials are required for writing a single character.
Detailed description: Section 2.3.1

User-centered design
ISO standard, formalizing the design process of technology around a user (ISO 9241-210, 2010). Key aspects are that the end-users has to be included at all stages of the design process and that all design decisions should be made with a deep understanding of the end-user.
Detailed description: Section 1.6

VAS – *Visual Analogue Scale*
A continuous psychometric response scale. Both ends of a line indicate an extreme response to a question, and respondents are requested to answer by making a mark on the line between the two extremes.
Detailed description: Section 2.5.2
See also: Likert scale

VEP – *Visually Evoked Potential*
Brain potentials that are evoked by visual stimuli. More specifically, VEP is typically used to refer to the early responses, associated with the physical processing of visual stimuli. VEPs can be used to assess the functioning of the optical nerve.
Detailed description: Section 5.2.2.1
See also: BAER, ERP

WCST – *Wisconsin card sorting test*
Neuropsychological test for the assessment of executive functions.
Detailed description: Section 5.2.2.1
See also: QCM-BCI

A. Appendix

A.1. Technical Details for Dynamic Stopping Methods

All early stopping methods are described here conceptually within the framework introduced in Section 6.2. A more detailed description is given if the implementation deviates from the original implementation. For a full account, the reader is referred to the original publications. All evaluations are done in the context of a text spelling application and thus by optimizing hyperparameter Φ of each stopping method for symb/min. The optimization procedure is the same as described in Section 6.5.

A.1.1. No Stopping

In this condition, no early stopping is performed. It uses all available iterations and is therefore the performance baseline. No parameters need to be estimated. Formally, this means that $\Theta = \{J\}$, the maximum number of iterations possible. $F\left(D^{test}, \Theta, t, j\right)$ returns $c^w_{j,t}$ (according to Equation 2.4) when $j = J$.

A.1.2. Fixed (Optimal)

The simplest way of optimizing the BCI is to estimate an optimal fixed number of iteration on the calibration data and to apply this for the online data. Because the number of iteration can be set in virtually any BCI system, this method requires no implementation effort. The number of iterations $j^\dagger$ which results in the highest symb/min on the training data is selected for online use. Θ is thus defined as $\Theta = \left\{j^\dagger\right\}$. As $j^\dagger$ is directly optimized it equals Φ. During online use, $F\left(D^{test}, \Theta, t, j\right)$ returns $c^w_{j,t}$ (Equation 2.4) when $j = j^\dagger$. For the *fixed* condition, the optimal number of iterations is estimated once from all training data; it serves as a evaluation of the post-hoc estimation of BCI performance which is generally given in BCI literature (see also Figure 3.9). For the *fixed optimal* condition, the resampling scheme is used to increase robustness.

A.1.3. Liu

This approach simply sums the outputs of a linear classifier per class and over iterations (Liu et al., 2010). M is defined by

$$M\left(\boldsymbol{D}, t, j\right) = \min_c \left(\sum_{j'=1}^{j} \boldsymbol{D}_{c,j',t} \right) \tag{A.1}$$

for trial t and iteration j, thus taking the most negative class. The average of all target classification scores in $\boldsymbol{D}^{train}$, called $\tilde{d}$, is taken as the initialization for optimizing $\Delta = \Phi\ \tilde{d}$. Θ then contains this threshold Δ for online use.

During online operation, the decision for an early stop is taken as follows:

$$F\left(\boldsymbol{D}^{test}, \Theta, t, j\right) = \begin{cases} c^w_{j,t} & \text{if } M\left(\boldsymbol{D}^{test}, t, j\right) < \Delta \\ \emptyset & \text{otherwise} \end{cases} \tag{A.2}$$

where $c^w_{j,t}$ is the class with the most negative summed score.

In the original work, EEG features from iteration i are smoothed together by averaging them with features from iterations $i-1$ and $i-2$ before classification. As here only access to the instantaneous classifier output is given by design, this preprocessing step is left out.

A.1.4. Jin

This method sets a minimum number of consecutive iterations (j^{min}) which must result in the same class prediction (Jin et al., 2011). To measure this, $M(\boldsymbol{D}, t, j)$ is defined as $1+n$, where n denotes the number of iterations directly prior to iteration j which all lead to the same decision as j does, according to Equation 2.4. The original paper performs an additional preprocessing step on $\boldsymbol{D}$ to give $\boldsymbol{D}^*$ where $\boldsymbol{D}^*_{c,j,t} = \frac{1}{j}\sum_{j'=1}^{j} \boldsymbol{D}_{c,j',t}$ for each $j \geq j^{mean}$. Since it only includes processing the classifier outputs, the same is done here. Both values of $\Theta = \{j^{min}, j^{mean}\}$ are themselves the hyperparameters ($\Phi_1 = j^{min}$ and $\Phi_2 = j^{mean}$); they are optimized directly.

In the online case, the decision function F looks as follows:

$$F(\boldsymbol{D}^{*,test}, \Theta, t, j) = \begin{cases} c^w_{j,t} & \text{if } M(\boldsymbol{D}^{*,test}, t, j) \geq j^{min} \\ & \text{and } j \geq j^{mean} \\ \emptyset, & \text{otherwise} \end{cases} \tag{A.3}$$

Here, both j^{min} and j^{mean} were bound between 2 and 5. As there are no decisions taken before $j < j^{mean}$, the earliest possible stop is at iteration $j^{mean} + j^{min}$.

A.1.5. Lenhardt

This method was originally proposed for a visual speller, and defines two thresholds (Lenhardt et al., 2008). For both thresholds, $\boldsymbol{D}$ is first transformed into $\boldsymbol{D}^*$ by $\boldsymbol{D}^*_{1\ldots C,j,t} = \text{norm}\left(\sum_{j'=1}^{j} -\boldsymbol{D}_{1\ldots C,j',t}\right)$, where $\text{norm}(X) = \frac{X-\min(X)}{\max(X-\min(X))}$, a vector normalization over the class dimension. *Lenhardt* requires the binary target class to have more positive labels. Since the binary target class was defined negative in this framework, a sign reversal is done within this transformation.

The first threshold is based on a measure called the "intensity", or M_1, defined for iteration j and trial t as

$$M_1(\boldsymbol{D}^*, t, j) = \sum_{c=1}^{C} \boldsymbol{D}^*_{c,t,j} \tag{A.4}$$

M_1 is thus simply the sum of all values in $\boldsymbol{D}^*_{1\ldots C,t,j}$. The second threshold is based on a slightly different measure M_2, defined as

$$M_2(\boldsymbol{D}^*, t, j) = 1 - \frac{\boldsymbol{D}^*_{k_2,t,j}}{\boldsymbol{D}^*_{k_1,t,j}} \tag{A.5}$$

where $k_1 = \arg\max_c\left(\boldsymbol{D}^*_{c,t,j}\right)$ and $k_2 = \arg\max_{c\neq k_1}\left(\boldsymbol{D}^*_{c,t,j}\right)$. M_2 is thus the ratio of the highest to the second highest value in $\boldsymbol{D}^*_{1\ldots C,t,j}$.

To find both thresholds, the training function $L(\boldsymbol{D}^{*,train}, \boldsymbol{l}, \Gamma)$ first calculates the intensity $M_1(\boldsymbol{D}^*, t, j)$ for the earliest possible stop which leads to a correct decision (according to Equation 2.4). This is done for each training trial, and the average of these intensities, $\tilde{i}$, serves as an initialization for optimizing $\Delta_1 = \Phi_1\ \tilde{i}$. The second threshold, Δ_2, is optimized on the training data directly ($\Delta_2 = \Phi_2$), which leads to $\Theta = \{\Delta_1, \Delta_2\}$.

Online decision for an early stop are taken by F as follows:

$$F\left(\boldsymbol{D}^{*,test},\Theta,t,j\right)=\begin{cases}c^{w}_{j,t} & \text{if } M_1\left(\boldsymbol{D}^{*,test},t,j\right)<\Delta_1\\ & \text{and } M_2\left(\boldsymbol{D}^{*,test},t,j\right)>\Delta_2\\ \emptyset, & \text{otherwise}\end{cases} \tag{A.6}$$

The original publication calculates the thresholds using data of several subjects. Here, it is restricted to the more common case of only having the calibration data of a single subject.

A.1.6. Höhne

This method was proposed for an auditory speller paradigm (Höhne et al., 2010b). A t-test with unequal variances (also referred to as *Welch's t-test*) is calculated on $\boldsymbol{D}_{1\ldots j}$ for each class against all other classes, according to

$$t\left(X_1,X_2\right)=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}}} \tag{A.7}$$

where $\overline{X}_i, s_i^2, N_i$ are sample mean, sample variance and sample size, respectively. In order to account for a varying number of data points, the t-test statistic $t\left(\boldsymbol{D}_{c,1\ldots j,t},\boldsymbol{D}_{\bar{c},1\ldots j,t}\right)$ is transformed into a p-value. This transformation is part of any statistics toolbox. $M\left(\boldsymbol{D},t,j\right)$ is then defined as this p-value.

The natural threshold on such a p-value is the significance level, or $\Theta=\{\alpha\}$. It is optimized for symb/min on the training data directly, with $\alpha=\Phi$. During online operation, the decision for an early stop is taken according to the following

$$F\left(\boldsymbol{D}^{test},\Theta,t,j\right)=\begin{cases}c^{w}_{j,t} & \text{if } M\left(\boldsymbol{D}^{test},t,j\right)<\alpha\\ \emptyset, & \text{otherwise}\end{cases} \tag{A.8}$$

where $c^{w}_{j,t}$ is the class with the smallest value of M.

A.1.7. Zhang

This method proposes a probabilistic approach to continuously tracking whether a BCI user is in control or non-control state (Zhang et al., 2008). The core of the algorithm is the formulation of a joint likelihood function, which accumulates the evidence for being attended for each stimulus over iterations. Here, this evidence is used as a criterion for early stopping.

The method requires that the distributions of the classifier outputs for target and non-target stimuli, f_{a} and f_{u}, are known. In practice, they are estimated from the training data. Here, univariate normal distributions for the classifier outputs are assumed, which are fully characterized by their means and variances $\mu_{\mathrm{a/u}}$ and $\sigma^2_{\mathrm{a/u}}$. Note that this assumption is compatible with the assumption of multivariate normal distributed *data* made by the LDA classifier.

Making the further simplifying assumption that the EEG data are independent across iterations, as well as across stimulations within one iteration, the joint *likelihood* of observing the classifier outputs of the first j iterations given that stimulus c is attended is

$$f\left(\vec{D}^{test}_{1\ldots C,1\ldots j,t}\,|\,c\right)=\prod_{j^*=1}^{j} f_{\mathrm{a}}\left(\vec{D}^{test}_{c,j^*,t}\right)\prod_{c^*\neq c} f_{\mathrm{u}}\left(\vec{D}^{test}_{c^*,j^*,t}\right) \tag{A.9}$$

According to Bayes' theorem, the *posterior probability* of stimulus c being attended after the first j iterations is given by

$$P\left(c \mid \vec{D}^{test}_{1\ldots C,1\ldots j,t}\right) = \frac{f\left(\vec{D}^{test}_{1\ldots C,1\ldots j,t} \mid c\right) P(c)}{\sum_{\tilde{c}=1}^{C} f\left(\vec{D}^{test}_{1\ldots c,1\ldots j,t} \mid \tilde{c}\right) P(\tilde{c})} \tag{A.10}$$

where $P(c)$, $c = 1,\ldots,C$ denote the *prior probabilities* with which the C classes are selected by the user. These probabilities may be estimated in the training phase or, e. g., derived from a language model in spelling applications. If all classes are assumed to be equally likely, uniform priors $P(c) = 1/C$, $c = 1,\ldots,C$ are used. Note that in practice, numerical inaccuracies may occur when evaluating Equations A.9 and A.10 due to excessively small values of the joint likelihood function. Therefore, all computations are carried out on log probabilities, as in the original publication (Zhang et al., 2008).

The joint likelihoods and posterior probabilities of each class are updated after each iteration. The function $M\left(\boldsymbol{D}^{test}, t, j\right)$ is defined as the maximum of the posterior probabilities of all classes,

$$M\left(\boldsymbol{D}^{test}, t, j\right) = \max_{\check{c}=1,\ldots,C} P\left(\check{c} \mid \vec{D}^{test}_{1\ldots C,1\ldots j,t}\right) \tag{A.11}$$

A natural stopping condition is attained, if Equation A.11 exceeds $1-\alpha$, where $\alpha = \Phi$ is a hyper-parameter to be optimized based on the training data. Thus, $\Theta = \left\{\{P(c)\}_{c=1}^{C}, \mu_{\text{a/u}}, \sigma^2_{\text{a/u}}, \alpha\right\}$. The decision at iteration j then reads

$$F\left(\boldsymbol{D}^{test}, \Theta, t, j\right) = \begin{cases} c^{w}_{j,t} & \text{if } M\left(\boldsymbol{D}^{test}, t, j\right) \geq 1-\alpha \\ \emptyset, & \text{otherwise} \end{cases} \tag{A.12}$$

where $c^{w}_{j,t} = \arg\max_{\check{c}=1,\ldots,C} P(\check{c} \mid \vec{D}^{test}_{1\ldots C,1\ldots j,t})$.

A.1.8. Upper bound

The *upper bound* method uses knowledge about the class labels and is not applicable for normal online use. It is applied for comparison reasons and to deliver an upper bound on the performance of any of the other stopping methods. It simply stops a trial at the first iteration which leads to a correct decision. If a trial does not result in the correct decision at any iteration, the full number of iterations is applied.

A.2. Additional Results Experiments 9 and 10

On the following pages, Figures A.1 to A.3 provide the full results for Experiments 9 to 11.

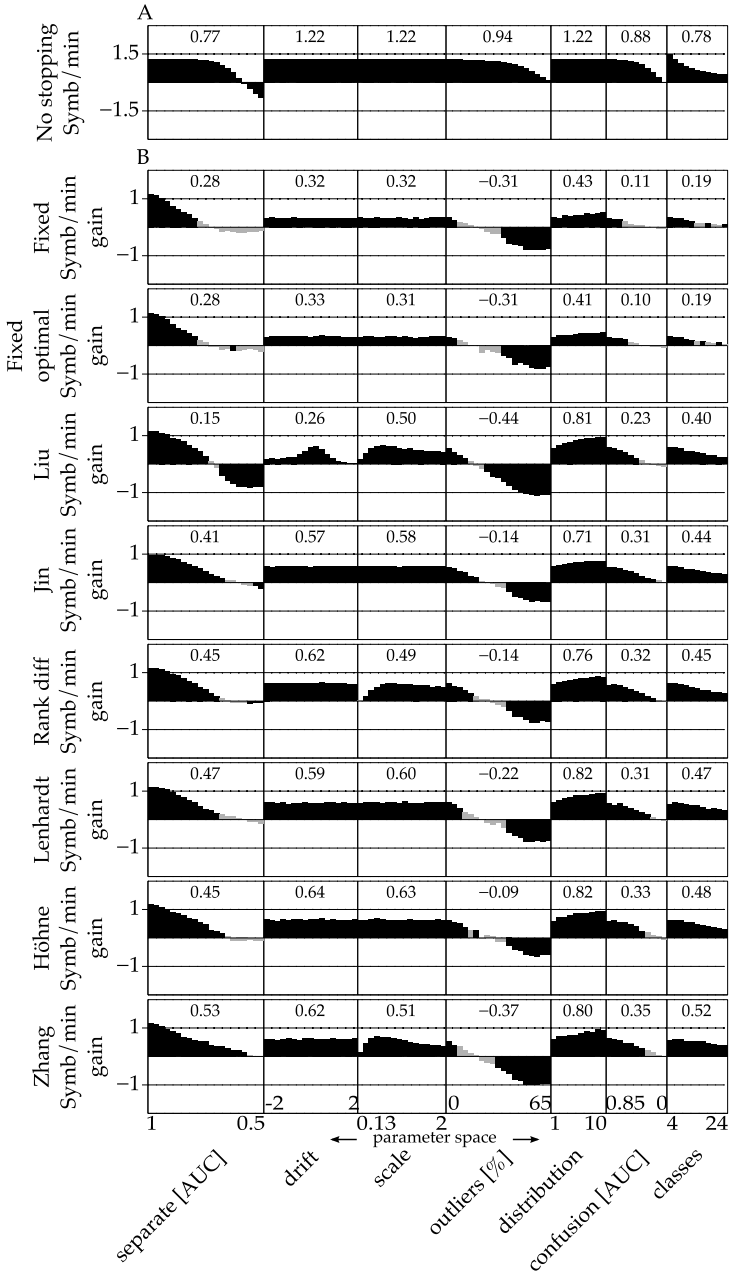


Figure A.1.: Full early stopping results on artificial datasets. A) To indicate the effect size, the baseline performance of *no stopping* is shown. B) Shown is the performance gain over the *no stopping* condition for all methods, on several artificially created datasets. With many outliers in the online data, all methods reduce the performance. All methods show a similar dependency on data separation, and apart from *Liu* and, to a much lesser extend, *Jin* and *Rank diff*, all methods reduce to baseline for inseparable data. *Liu* as the only method is susceptible to all sorts of data distortion, including drifts and scaling.

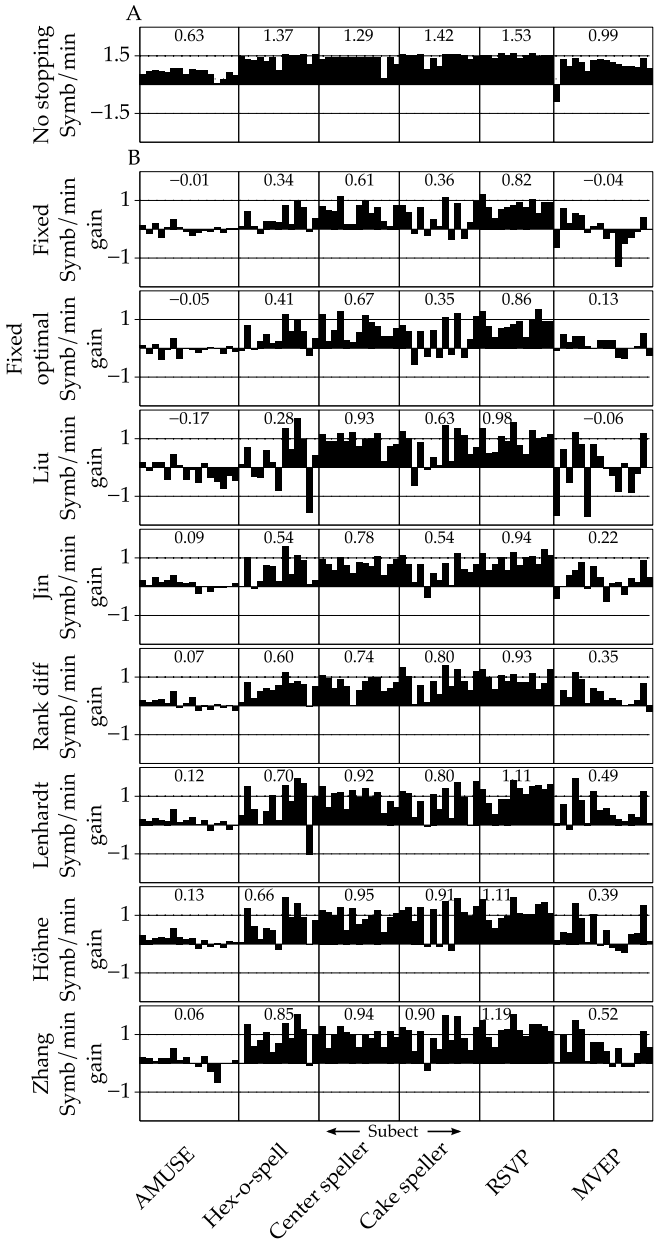


Figure A.2.: Full early stopping results on EEG datasets. A) To indicate the effect size, the baseline performance of *no stopping* is shown. B) Shown are for each subject and each stopping method the achieved gains over the *no stopping* condition. Numbers indicate the average gain per method per paradigm. Negative gains mean that a method reduced the performance for a subject (negative bar) or on average over an entire paradigm (negative number).

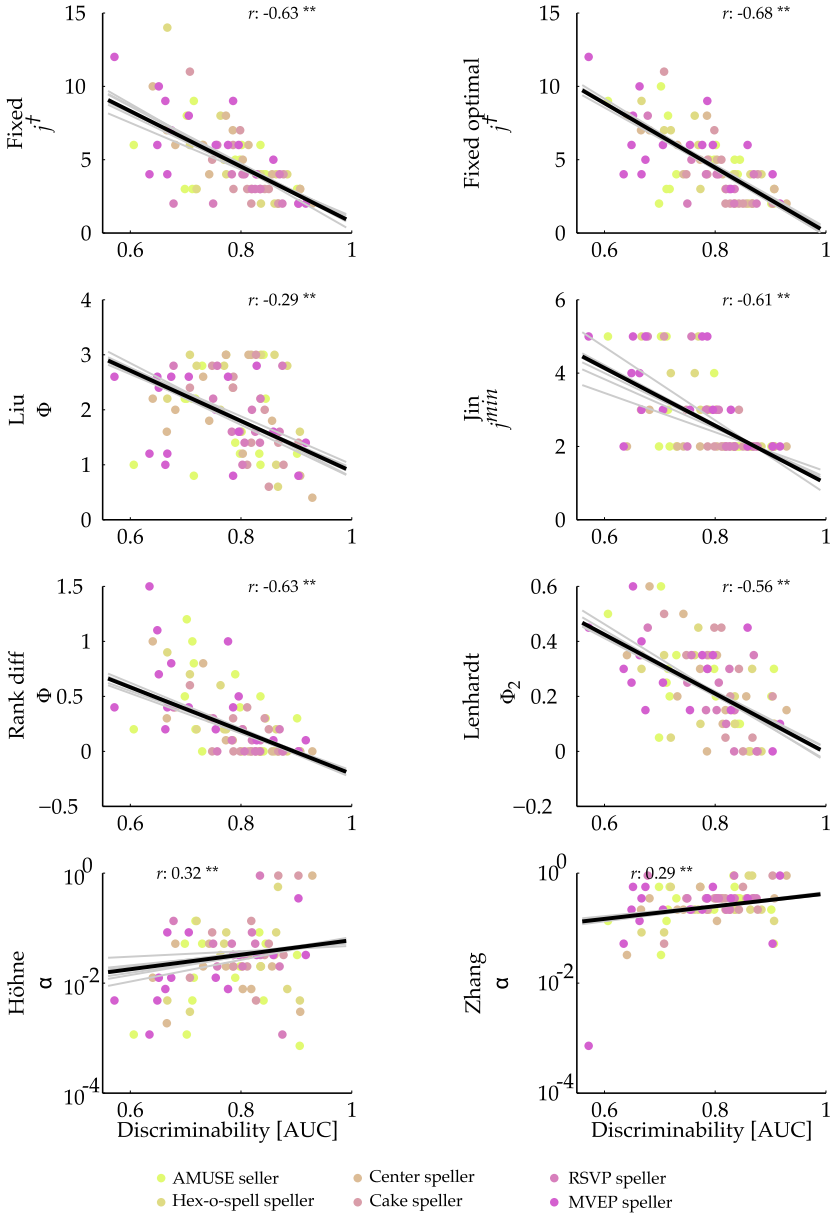


Figure A.3.: Regression results for $\Phi^{\dagger}$. The individually optimized scaling parameter Φ is plotted against the AUC of the training data. High linear correlations are found between these, which leads to a set of coefficients for each methods that allow the direct estimate of $\Phi^{\dagger}$ from the training data. The black line indicates the linear fit for all data, using an iteratively reweighted least squares fit. The gray lines indicate the fit of all but one dataset, showing high resemblance with the overall fit. Note that the α values for *Höhne* and *Zhang* are plotted on a log scale, given the nature of the parameter.

To date, the brain-computer interface (BCI) based on visual stimulation is by far the most investigated. This makes a lot of sense given the excellent visual capabilities of humans, and it has been a most successful approach. Nevertheless, a non-negligible part of the BCI end-user population with advanced paralysis is incapable of directing their eye gaze or of seeing at all. Traditional visual BCI will rarely be a solution for these end-users and alternative paradigms are required that rely on covert attention. Furthermore, though BCI already benefit greatly from statistical methods coming from the field of machine learning, a faster and robust performance is required for adoption by the clinical community.

This thesis contributes in two key facets in an attempt to take a step towards full inclusion of all end-users.

ISBN 978-3-7375-0485-0