

**UNLOCKING POSSIBILITIES WHILE PRESERVING PERFORMANCE:
PUTTING THE "INTERFACE" BACK IN BRAIN-COMPUTER INTERFACE**

by

David Edward Thompson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in The University of Michigan
2012

Doctoral Committee:

Research Assistant Professor Jane E. Huggins, Co-Chair
Professor Daniel P. Ferris, Co-Chair
Professor Douglas C. Noll
Professor Seth A. Warschausky
Assistant Professor Zeeshan Syed

To my Grandmother Fern, who I hope is still watching and cheering me on.

To my family, those still here and those who have passed on.

To my wife and daughter.

Acknowledgements

This dissertation represents the culmination of several years of research and design, and over those years I have had the fortune to meet many wonderful people both in the lab and outside of it. Meanwhile, many of those that I knew before I began this work have continued to be an important part of my life during it, though I've seen less of most of them than I might have wished to.

My wife, Amity, has been a steadfast source of support and inspiration, both in writing the dissertation and in the years proceeding. It is no exaggeration to say that without her, this document would never have existed. This last month in particular she's shouldered an awful lot of the load to give me time to write and edit. I'll pay you back sometime, love.

My parents, brother and three sisters, assorted nieces and nephews... all have been there for me when I've needed them, and are a big part of who I am today. Here's hoping I find somewhere closer after graduation.

My daughter, Keziah, was born during the writing phase of my dissertation. While in perfect honesty I've spent fewer hours working since she came (the last few weeks excepted), I've found myself more productive during each working hour. While she hasn't had much to say, yet, I like to think she's encouraging me every time she says "Ah-goo."

My principal investigator, Jane Huggins, deserves mention here, not only for mentoring in research, writing, and grantsmanship, but also for being a pleasure to talk to about out-of-lab topics. Sorry about the parking tickets...

The post-doc in our lab, Stefanie Blain-Moraes, has been an incredible help to me both in the lab and a good friend outside of it. She's taught me a lot about writing for publication, and how to overlook the frustrating parts of peer review and focus on the value instead. A lot of my best memories in Ann Arbor involve her, our respective spouses, and food. Also, Dominion.

Michael McCann, an undergraduate in our lab who rudely went off to other graduate programs, was also a pleasure to work with. I spent more than a few hours outside of work with him as well.

A lot of other people have played a role as well – the Cynnabar fencers, my role-playing group: Paul & Andria, Kelly & DeForest, Joe & Mindy, the lunch-hour basketball crew, and probably a lot of other people that don't deserve me forgetting to list them (Sephra and Lee, for instance). All of them have made Ann Arbor a little more of a home-away-from-home for me, and that much harder to leave.

I'd also like to thank Lyle McWilliams, who taught me the value of simulation and double-checking before building. I give him the credit for my first prototype of the device in this dissertation working correctly.

Last but not least, the NSF, who have their own language they prefer to use:

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 0718128.

Funding sources for the experiments are listed in the individual chapters.

Preface

With the exception of the first and last chapters, this dissertation is comprised of manuscripts in various stages of preparation, submission, and publication. I have attempted to fit these manuscripts into the overall story of my research, preserving the integrity of each work while minimizing overlapping or repeated text and content, but some degree of repeated text was unavoidable.

The ordering of the chapters was one of the most difficult parts of this work, due to a high degree of interdependence between the manuscripts. Regardless of the order chosen, earlier chapters need to reference concepts from later chapters. In the end, I ordered the chapters by chronology; this order minimizes the number of forward references and has the advantage of simplicity, though other orderings might have increased the appearance of the work's coherence.

Table of Contents

Dedication	ii
Acknowledgements	iii
Preface	v
List of Figures	ix
List of Tables	xii
Chapter 1 Motivation	1
Introduction	1
Background	3
Significance	5
References	8
Chapter 2 Design of a Plug-and-Play BCI	12
Introduction	12
Methods	15
Results	22
Discussion	25
Conclusion	28
Acknowledgements	28
References	29
Chapter 3 Evaluation of the Plug-and-Play BCI for Communication	31
Introduction	31
Methodology	34
Results	39
Discussion	43
Study Limitations	46

Conclusion	47
Acknowledgements.....	47
References.....	48
Chapter 4 Evaluation of the Plug-and-Play BCI for Control of a Wheelchair Tilt-In-Space System.....	50
Introduction.....	50
Methods.....	52
Results.....	55
Discussion.....	58
Study Limitations.....	60
Conclusion	62
Acknowledgements.....	62
References.....	63
Chapter 5 Measuring Performance	65
Introduction.....	65
Level 1 Performance Metrics.....	71
Level 1 Performance Metric Recommendations.....	77
Level 2 Performance Metrics.....	79
Level 2 Performance Metric Recommendations.....	85
Discussion.....	87
Limitations	89
Conclusion	91
Acknowledgements.....	91
References.....	92
Chapter 6 Estimating and Predicting BCI Accuracy	98
Introduction.....	99
Methods.....	102
Results.....	106
Discussion.....	109
Study Limitations.....	111
Conclusion	112

Acknowledgements.....	112
References.....	113
Chapter 7 Conclusion.....	115
Contributions.....	115
Study Limitations.....	117
Future Work.....	118
References.....	121

List of Figures

Figure 1: A P300 BCI matrix used for in some of our experiments. The desired text is shown in the first line, with the current target letter shown in parentheses. Text produced by the BCI appears on the second line. In this picture, the third row is being presented as a stimulus.....	3
Figure 2: The role of Interface Technology connecting BCI Input Devices to any Device Controller, within the framework of input device emulation.	13
Figure 3: Overview of the MBOD hardware	16
Figure 4: Example sentence, actual letters selected, target text including error correction, and correctness of each selection (1 indicates correct, 0 indicates incorrect). Note that if no incorrect selections were made, Sentence and Target would be a perfect match and the backspace (◀) would not appear. In this example, 26 of 29 selections were correct, producing an accuracy of 89.7%.	37
Figure 5: Average accuracy by device for each participant. Top) participants without ALS; Bottom) Participants with ALS. Error bars shows standard deviation across sessions.	40
Figure 6: Average BCI accuracy vs. ALS-FRS.	42
Figure 7: Mean BCI Accuracy vs. # of Sequences in the three tilt cases.	55
Figure 8: Architecture of a BCI-based AAC system that is comprised of two modules: (1) a BCI Control module that translates brain signals into logical control outputs and (2) a Selection Enhancement module that translates logical control to semantic control. Performance of BCI-based AAC systems can be measured at three levels (labeled Level 1, Level 2, Level 3) within this architecture; each level of measurement is currently assessed by a variety of often incommensurable performance metrics.....	67

Figure 9: Distribution of the many metrics combinations that have been used in the literature to report performance of a BCI used by human participants for communication from January 2005 – October 2011.....	69
Figure 10: a) Comparison of each scalar Level 2 metric on data from a P300 copy-spelling task with correction, 75 sentences from 22 users, sorted by OCM. All metrics were converted into characters per minute. Eff_{SYS} and Eff_{SYS}' were calculated on the assumption that the accuracy for untested outputs was equal to the mean of the accuracies of all tested outputs. b) The ECM for the first datapoint presented in a). For space reasons, the 73 ECMs corresponding to the other datapoints are not presented. This illustrates the lack of accessibility of this metric.	86
Figure 11: Example of augmented Level 1 performance metric for a P300-speller BCI. Both the ITR and the accuracy are reported with respect to time, enabling comparison with other BCI Control Modules. Note that ITR was calculated including the time between selections.	88
Figure 12: Simulated P300 responses demonstrating the effects of amplitude variation and latency jitter on the average response. The amplitude and latency of the response were varied by identical proportions. The distortion in the average is clearly visible when latency jitter is present.	101
Figure 13: Accuracy plotted against classifier-based latency jitter estimates (CBLE) using two classifiers. Left - results from a classifier using least-squares (LS) regression; right - the popular step-wise linear discriminant analysis (SWLDA) classifier.	106
Figure 14: Predicting first-day accuracy based on a 5-character dataset, by classifier type and estimation method. Left - least-squares classification (LS); right - classification using step-wise linear discriminant analysis (SWLDA). Top - prediction based on 5-character accuracy; bottom – prediction based on classifier-based latency jitter estimates (CBLE).....	107
Figure 15: Predicting first-day accuracy based on training data, by classifier type and estimation method. Left - least-squares classification (LS); right - classification using step-wise linear discriminant analysis (SWLDA). Top - prediction based on	

training accuracy; middle – prediction based on leave-one-out cross-validation
accuracy on training data; bottom – prediction based on classifier-based latency
jitter estimates (CBLE). 108

Figure 16: Mean classifier score as a function of latency shift..... 120

List of Tables

Table 1: Design specification for an Interface Technology to connect a BCI Input Device to a Device Controller	14
Table 2: Supported inputs of common assistive technology devices	17
Table 3: Testing results for the MBOD with different output devices, using BCI control. 'Y' indicates successful operation, 'N' indicates unsuccessful operation, and 'N/A' indicates that the input mode is not supported by the device. Asterisks indicate notes in the following subsections. Non-Windows operating systems were tested with simulated brain activity	22
Table 4: Means and 95% confidence intervals for effects of environment and diagnosis.	41
Table 5: Number of sequences used in the offline analysis, with accuracy and throughput at that number of sequences. All numbers were calculated from data in the no tilt case, and ideal sequences and throughput were calculated using BCI-Utility.	56
Table 6: Estimates and confidence intervals for the differences between tilt environments. Lower and upper bounds are based on 95% confidence intervals.	56
Table 7: Estimates and confidence intervals for the differences between characters interrupted by movements and those that were not. Lower and upper bounds are based on 95% confidence intervals	57
Table 8: Comparison of common Level 1 BCI-based AAC performance metrics. Check marks indicate that the metrics fulfill the evaluation criterion.....	78
Table 9: Comparison of common Level 2 BCI-based AAC performance metrics. Check marks indicate that the metric fulfills the evaluation criterion.	84
Table 10: Confidence Intervals for Correlation Coefficients.....	108

Chapter 1

Motivation

Introduction

User-centered design dictates that the needs and wants of the users should be considered at all parts of the design, rather than expecting the user to adapt to a predetermined use model [1]. This requires careful consideration of the living conditions and other life factors of potential users, and preferably involvement of potential users in the design process itself [2]. This dissertation includes the application of these principles to the design of a Brain-Computer Interface (BCI).

According to the Wolpaw definition, a BCI is a "non-muscular channel for sending messages and commands to the external world" [3]. From a clinical perspective, a BCI represents the only extant hope of communication for people with total locked-in syndrome, a condition characterized by a complete lack of voluntary control of movement. BCIs may also be of use to people who are locked-in or nearly locked-in, but still retain some limited control of movement. However, in this setting, BCIs must compete with existing assistive technology (AT) such as eye-gaze systems and switch-scanning systems controlled by microswitches or sip-puff tubes. BCIs are often slower than existing AT solutions, though in at least one case a BCI has proven to be the preferred means of communication for a locked-in individual [4]. For people with normal motion, BCIs are typically only of novelty interest due to low information throughput.

The first focal point of this dissertation, that of unlocking possibilities, is driven by understanding of the user populations. First, surveys of individuals with ALS [5] as well as a focus group of individuals with ALS and their caregivers [6] indicate that these people are interested in BCI control of existing devices. Second, several of the conditions that can lead to the locked-in or total locked-in state are progressive, or

worsening with time (see Significance, below, for a discussion of user populations). Amyotrophic lateral sclerosis (ALS) in particular can last many years, while the patient experiences gradual loss of voluntary movement control. In the early course of the disease, individuals with ALS, particularly those with bulbar onset, may use augmentative and alternative communication (AAC) devices such as communication boards or text-to-speech converters to enable communication with loved ones and caregivers. As the disease progresses, the ability to use those AAC devices will fade, though the AAC devices may provide several years of service. These AAC devices, which may be standalone devices or computer software, are often highly configured to a user's needs, and through years of use are extremely familiar to the user.

While BCI researchers have begun to incorporate AAC techniques such as word prediction into their BCI systems [7], true AAC integration is rare. Instead, a BCI is treated as a replacement for existing AAC systems. In this model, users are expected to adapt to the BCI's AAC capabilities, discarding their familiar software or devices. By contrast, a plug-and-play BCI could work with existing AAC solutions, enabling users to continue working with their familiar, customized software and hardware. Through the principle of input emulation, such a plug-and-play BCI could be used as an input device to control existing AAC solutions. For users who already have an emotional or resource investment in a particular AAC technology, this approach could substantially reduce the adoption barrier to BCI technology.

The second focal point of this work is that of preserving performance. The number of individuals in a totally locked-in state is small; if BCIs were useful to individuals with less severe impairments, the commercial viability of BCI systems would be improved. As mentioned previously, unless the user is totally locked-in, existing AT solutions often outperform BCIs. If BCIs are to succeed in this broader user population, performance must be maintained, if not improved, as the possible uses are expanded.

The question of preserving performance is complicated by the fact that there is no broadly accepted metric of performance in the BCI field. The last two chapters of this work focus on the issue of measuring performance, utilizing data from the experiments in earlier chapters.

Background

BCI defined

A Brain-Computer Interface (BCI) is a "non-muscular channel for sending messages and commands to the external world [3]". Since BCIs do not require muscular activation, they have the potential to help individuals with the most severe and profound movement disabilities. Many different BCIs have been proposed, based on various physiological phenomena including the P300 response [8], sensorimotor rhythms [9], slow cortical potentials [10], and visual evoked potentials [11]. For current reviews, see [12–16].

P300 BCI defined

The P300-based BCI, first proposed by Farwell and Donchin in [8], is one of the easiest to learn and most effective BCIs. It has specifically been suggested for clinical use [17], and has been used for several years by a person unable to communicate through other means [4]. The P300 BCI can be used both for communication and environmental control (e.g. changing television channels), e.g. [3], [8], [18–25].

The P300 BCI gets its name from the P300 or "oddball" response, first reported in [26], which occurs when a stimulus of interest is presented among a sea of distracters. This response is well studied in the field of neuroscience (for reviews see [27–29]), so much so that it is currently used to study population differences (e.g. [30–33]) and other neurological responses (e.g. [34–36]). The stimuli may be of various forms, including audible [37–39] and tactile [40], but this work will focus on the visual mode, which is

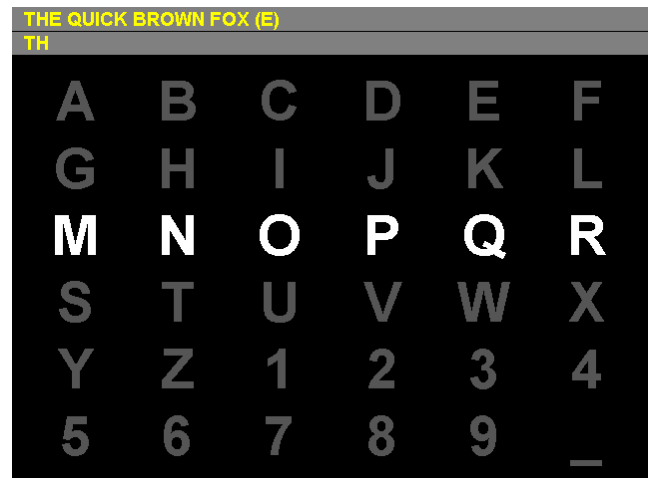


Figure 1: A P300 BCI matrix used for in some of our experiments. The desired text is shown in the first line, with the current target letter shown in parentheses. Text produced by the BCI appears on the second line. In this picture, the third row is being presented as a stimulus.

more common in BCI research. The P300 is evident in electroencephalogram (EEG) recordings, and while EEG has the lowest signal-to-noise ratio (SNR) of almost all BCI options, it is non-invasive, comparatively inexpensive, and useful for evaluating the possible efficacy of more invasive BCIs.

P300 BCI in action

The P300 BCI is a system for making selections from a large number of options. Typically, these options are letters, numbers, and special keys such as backspace, but may include symbolic concepts or environmental controls. The options are presented to the user, who is asked to attend to the choice he or she wishes to make. Often, choices (stimuli) are presented by flashing rows and columns on the computer screen in a setup similar to that of **Figure 1**, but they may be presented through more advanced methods [20], [41–43].

Repeated observations required

Because of the low SNR of EEG, however, detecting the presence or absence of a P300 response is not a simple task. Most algorithms rely on signal averaging to reduce the noise. Therefore every choice must be presented to the user multiple times before a selection can be made, dramatically slowing information throughput. In almost all algorithms, there is a tradeoff between better accuracy and faster selections which can be described by various metrics (see Chapter 5 for details).

In some situations, accuracy may be far more important than speed. An example of such a situation is environmental control, where a single mistake could have consequences that are difficult to quantify (e.g. accidentally changing the channel instead of the volume during an important play of a sports game) or potentially very serious (e.g. wheelchair control). Due to the difficulty of quantifying costs and benefits in environmental control tasks, the comparison in Chapter 5 focuses on the communication task. It is worth noting, however, that one of the metrics that performs the best in that comparison can in fact be used for those control tasks with quantifiable costs.

Significance

Potential user population

The relatively slow throughput of BCIs relative to other assistive technology limits its use to those populations with severe motor impairments. In the BCI literature, several diseases or conditions have been mentioned (e.g. [3]) as capable of affecting motion to a degree that a BCI might be an appropriate option. Brief summaries of the symptoms and incidence rates appear below.

- Amyotrophic lateral sclerosis (ALS), called motor neuron disease (NMD) in Europe and commonly known as Lou Gherig's Disease. Incidence is about 2 per 100,000 people [44]. Prevalence in the U.S is unknown [45] but estimated to be 20,000 to 30,000 people [45],[46]. ALS is classically characterized by a progressive loss of motor control with preservation of sensory and cognitive function, though recently a link between ALS and fronto-temporal dementia has been of considerable research interest [47]. ALS is of primary research interest because most people with the disease could progress to a point where a BCI might be of use, though few do at present (see the following section on end-of-life decisions).
- Muscular dystrophy (MD), a condition with many distinct symptomologies and etiologies. Many types of MD exist, [48] lists 6 categories and 50 distinct types, including Duchenne MD (DMD), Becker MD, Limb-Girdle Dystrophies (LGMD), and Facioscapulohumeral MD (FSHD). As the number of forms might suggest, the symptoms of MD are quite heterogeneous and describing them here is beyond the scope of this work. Instead, I will simply note that most forms of MD are associated with muscle weakness of some degree, though only some subset of the population will experience symptoms severe enough to warrant a BCI. The most common form is Duchenne Muscular Dystrophy (DMD), which affects about 1 in 3500 male children [48]. DMD is a progressive disease eventually affecting all skeletal muscle tissue; most people with DMD could progress to a point where a BCI is a useful option, though the end-of-life considerations in the following section apply in this population as well.

- Cerebral palsy (CP), a congenital condition with a variety of motor symptoms, usually including spasticity. Incidence is about 2.4 per 1000 births [49], though only a small fraction of these individuals will present symptoms severe enough to require a BCI for communication. Also, spastic movement presents serious challenges to neural recordings necessary for BCI, and as such is rarely studied; the only example I can find in the literature is [50].
- Brainstem stroke, traumatic brain injury, spinal muscular atrophy, and high spinal cord injury have also been mentioned, but I have yet to see studies including people from these populations in the literature.

As BCI performance increases, the technology may become attractive to user populations with less severe impairments.

Progressive diseases and end-of-life decisions:

An important point is that while the above diseases or conditions can produce a locked-in state, most of the progressive diseases rarely do so at present. ALS, for example, is on average fatal within two to five years from symptom onset [51], [52], generally due to respiratory failure. The totally locked-in state from ALS was first observed in patients who underwent mechanical ventilation [53]. The use of mechanical ventilation has been shown to increase survival significantly, sometimes by ten or more years [54]. However, most patients do not choose to undergo mechanical ventilation, partially because of perceived quality of life issues related to communication [54]. Similarly, mechanical ventilation is known to prolong the life of those with DMD for about 10 years [48], but relatively few patients choose to undergo the procedures.

BCIs could enable the most affected of this population to communicate more effectively. As the inability to communicate is related to both loss of decisional control for the patient and anxiety and frustration for the caregiver [55], this change could be expected to improve the quality of life of both the patients and their caregivers. The resulting difference in quality of life may also affect the number of people who choose to undergo mechanical ventilation, substantially impacting the survival rate of those with the disease. Thus, not only could BCIs improve the quality of life of their potential user

population, the widespread clinical availability of BCIs could increase the size of the potential user population by affecting end-of-life decisions.

User desires and preferences

Our laboratory was the first to perform structured inquiry into the desires and preferences of potential BCI users (see [5], [6]). While these investigations revealed many things, this dissertation will focus on the following facts:

- Users are very interested in using a BCI for a variety of tasks [5], including interfacing with existing hardware devices [6]
- BCI speed is not yet up to user expectations [5], though perhaps this could be offset by existing AT solutions such as word prediction [6]

These two facts form the central goal of this dissertation: Unlocking possibilities while preserving performance.

References

- [1] D. A. Norman, *The Design of Everyday Things*. Basic Books, 2002.
- [2] C. Abras, D. Maloney-Krichmar, and J. Preece, "User-centered design," Bainbridge, W. *Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications, 2004.
- [3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, Jun. 2002.
- [4] E. W. Sellers, T. M. Vaughan, and J. R. Wolpaw, "A brain-computer interface for long-term independent home use," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 5, pp. 449–455, Oct. 2010.
- [5] J. E. Huggins, P. A. Wren, and K. L. Gruis, "What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis*, vol. 12, no. 5, pp. 318–324, Sep. 2011.
- [6] S. Blain-Moraes, R. Schaff, K. L. Gruis, J. E. Huggins, and P. A. Wren, "Barriers to and mediators of brain-computer interface user acceptance: focus group findings," *Ergonomics*, vol. 55, no. 5, pp. 516–525, 2012.
- [7] D. B. Ryan, G. E. Frye, G. Townsend, D. R. Berry, S. Mesa-G, N. A. Gates, and E. W. Sellers, "Predictive spelling with a P300-based brain-computer interface: increasing the rate of communication," *International Journal of Human-Computer Interaction*, vol. 27, no. 1, pp. 69–84, 2011.
- [8] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, Dec. 1988.
- [9] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalography and Clinical Neurophysiology*, vol. 78, no. 3, pp. 252–259, Mar. 1991.
- [10] A. Kubler, B. Kotchoubey, N. Ghanayim, T. Hinterberger, J. Perelmouter, M. Schauer, C. Fritsch, and N. Birbaumer, "A thought translation device for brain computer communication," *Studia Psychologica*, vol. 40, no. 1–2, pp. 17–31, 1998.
- [11] E. E. Sutter, "The brain response interface: communication through visually-induced electrical brain responses," *Journal of Microcomputer Applications*, vol. 15, no. 1, pp. 31–45, Jan. 1992.
- [12] L. Fernando Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, Feb. 2012.
- [13] H. Cecotti, "Spelling with non-invasive brain-computer interfaces - Current and future trends," *Journal of Physiology-Paris*, vol. 105, no. 1–3, pp. 106–114, Jun. 2011.
- [14] E. Pasqualotto, S. Federici, and M. O. Belardinelli, "Toward functioning and usable brain-computer interfaces (BCIs): a literature review," *Disability and Rehabilitation: Assistive Technology*, vol. 7, no. 2, pp. 89–103, Mar. 2012.
- [15] S. C. Kleih, T. Kaufmann, C. Zickler, S. Halder, F. Leotta, F. Cincotti, F. Aloise, A. Riccio, C. Herbert, D. Mattia, and A. Kübler, "Out of the frying pan into the fire--the P300-based BCI faces real-world challenges," *Progress in Brain Research*, vol. 194, pp. 27–46, 2011.
- [16] J. N. Mak, Y. Arbel, J. W. Minett, L. M. McCane, B. Yuksel, D. Ryan, D. Thompson, L. Bianchi, and D. Erdogmus, "Optimizing the P300-based brain-computer interface: Current status, limitations and future directions," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025003, Apr. 2011.

- [17] F. Nijboer, N. Birbaumer, and A. Kübler, "The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis - a longitudinal study," *Frontiers in Neuroscience*, vol. 4, 2010.
- [18] F. Piccione, F. Giorgi, P. Tonin, K. Priftis, S. Giove, S. Silvoni, G. Palmas, and F. Beverina, "P300-based brain computer interface: Reliability and performance in healthy and paralysed participants," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 531–537, Mar. 2006.
- [19] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 15–21, Jan. 2008.
- [20] G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. E. Frye, C. K. Hauser, N. E. Schwartz, T. M. Vaughan, J. R. Wolpaw, and E. W. Sellers, "A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109–1120, Jul. 2010.
- [21] C. Wang, C. Guan, and H. Zhang, "P300 brain-computer interface design for communication and control applications," in *Proceedings of IEEE EMBS*, vol. 5, pp. 5400–5403, 2005.
- [22] B. Rebsamen, E. Burdet, Cuntai Guan, Chee Leong Teo, Qiang Zeng, M. Ang, and C. Laugier, "Controlling a wheelchair using a BCI with low information transfer rate," in *2007 10th International IEEE Conference on Rehabilitation Robotics*, 2007, pp. 1003–1008.
- [23] B. Rebsamen, C. Guan, H. Zhang, C. Wang, C. Teo, M. H. Ang Jr., and E. Burdet, "A brain controlled wheelchair to navigate in familiar environments," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 99, p. 1, 2010.
- [24] D. E. Thompson, J. J. Baker, W. A. Sarnacki, and J. E. Huggins, "Plug-and-play brain-computer interface keyboard performance," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, New York, 2009, pp. 426–428.
- [25] D. E. Thompson, M. T. McCann, and J. E. Huggins, "Controlling wheelchair tilt with a brain-computer interface," presented at the *4th International BCI Meeting*, Asilomar, California, 2010.
- [26] S. Sutton, M. Braren, J. Zubin, and E. R. John, "Evoked-potential correlates of stimulus uncertainty," *Science*, vol. 150, no. 700, pp. 1187–1188, Nov. 1965.
- [27] G. Hajcak, A. MacNamara, and D. M. Olvet, "Event-related potentials, emotion, and emotion regulation: an integrative review," *Developmental Neuropsychology*, vol. 35, no. 2, pp. 129–155, Feb. 2010.
- [28] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007.
- [29] T. W. Picton, "The P300 wave of the human event-related potential," *Journal of Clinical Neurophysiology*, vol. 9, no. 4, pp. 456–479, Oct. 1992.
- [30] C. G. Matas, S. L. de A. Matas, C. R. S. de Oliveira, and I. C. Gonçalves, "Auditory evoked potentials and multiple sclerosis," *Arq Neuropsiquiatr*, vol. 68, no. 4, pp. 528–534, Aug. 2010.
- [31] Y. Anjana, F. Khaliq, and N. Vaney, "Event-related potentials study in attention deficit hyperactivity disorder," *Functional Neurology*, vol. 25, no. 2, pp. 87–92, Jun. 2010.
- [32] M. J. S. Cóser, P. L. Cóser, F. S. Pedrosa, R. Rigon, and E. Cioqueta, "P300 auditory evoked potential latency in elderly," *Brazilian Journal of Otorhinolaryngology*, vol. 76, no. 3, pp. 287–293, Jun. 2010.
- [33] J. M. Ford and A. Pfefferbaum, "Event-related potentials and eyeblink responses in automatic and controlled processing: effects of age," *Electroencephalography and Clinical Neurophysiology*, vol. 78, no. 5, pp. 361–377, May 1991.

- [34] M. D. Bomba and A. Singhal, "ERP evidence of early cross-modal links between auditory selective attention and visuo-spatial memory," *Brain and Cognition*, vol. 74, no. 3, pp. 273–280, Dec. 2010.
- [35] M. A. Coates and K. B. Campbell, "Event-related potential measures of processing during an Implicit Association Test," *Neuroreport*, vol. 21, no. 16, pp. 1029–1033, Nov. 2010.
- [36] U. Martens, H. Leuthold, and S. R. Schweinberger, "On the temporal organization of facial identity and expression analysis: Inferences from event-related brain potentials," *Cognitive, Affective, and Behavioral Neuroscience*, vol. 10, no. 4, pp. 505–522, Dec. 2010.
- [37] E. W. Sellers and E. Donchin, "A P300-based brain-computer interface: Initial tests by ALS patients," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 538–548, 2006.
- [38] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, "An auditory oddball (P300) spelling system for brain-computer interfaces," *Psychophysiology*, vol. 46, no. 3, pp. 617–625, 2009.
- [39] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, "A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients," *Annals of the New York Academy of Sciences/Disorders of Consciousness*, vol. 1157, pp. 90–100, 2009.
- [40] A.-M. Brouwer and J. B. F. van Erp, "A tactile P300 brain-computer interface," *Frontiers in Neuroscience*, vol. 4.
- [41] J. Jin, B. Z. Allison, E. W. Sellers, C. Brunner, P. Horki, X. Wang, and C. Neuper, "Optimized stimulus presentation patterns for an event-related potential EEG-based brain-computer interface," *Medical and Biological Engineering and Computing*, Oct. 2010.
- [42] M. Salvaris and F. Sepulveda, "Visual modifications on the P300 speller BCI paradigm," *Journal of Neural Engineering*, vol. 6, no. 4, p. 046011, Aug. 2009.
- [43] G. Luo and W. Min, "Distance-constrained orthogonal latin squares for brain-computer interface," *Journal of Medical Systems*, Mar. 2010.
- [44] "CDC - Amyotrophic Lateral Sclerosis: FAQ." [Online]. Available: <https://wwwn.cdc.gov/ALS/ALSFAQ.aspx#b>. [Accessed: 19-Jun-2012].
- [45] "Who gets ALS? - The ALS Association." [Online]. Available: <http://www.alsa.org/about-als/who-gets-als.html>. [Accessed: 19-Jun-2012].
- [46] R. G. Miller, D. F. Gelinas, P. O'Connor, and I. Ebrary, *Amyotrophic lateral sclerosis*. New York: Demos Medical Pub., 2005.
- [47] C. Lomen-Hoerth, J. Murphy, S. Langmore, J. H. Kramer, R. K. Olney, and B. Miller, "Are amyotrophic lateral sclerosis patients cognitively normal?," *Neurology*, vol. 60, no. 7, pp. 1094–1097, Apr. 2003.
- [48] R. C. Griggs, A. A. Amato, and ScienceDirect (Online service), *Muscular dystrophies*. Amsterdam: Elsevier, 2011.
- [49] F. Miller and SpringerLink (Online service), *Cerebral palsy*. New York: Springer, 2005.
- [50] G. Pires, U. Nunes, and M. Castelo-Branco, "Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis," *Journal of Neuroscience Methods*, vol. 195, no. 2, pp. 270–281, Feb. 2011.
- [51] C. E. Jackson and W. W. Bryan, "Amyotrophic lateral sclerosis," *Seminars in Neurology*, vol. 18, no. 1, pp. 27–39, 1998.

- [52] C. L. Shoesmith and M. J. Strong, "Amyotrophic lateral sclerosis: update for family physicians," *Canadian Family Physician*, vol. 52, no. 12, pp. 1563–1569, Dec. 2006.
- [53] H. Hayashi and S. Kato, "Total manifestations of amyotrophic lateral sclerosis: ALS in the totally locked-in state," *Journal of the Neurological Sciences*, vol. 93, no. 1, pp. 19–35, Oct. 1989.
- [54] J. R. Bach, "Amyotrophic lateral sclerosis. Communication status and survival with ventilatory support," *American Journal of Physical Medicine and Rehabilitation*, vol. 72, no. 6, pp. 343–349, Dec. 1993.
- [55] M. B. Happ, "Interpretation of nonvocal behavior and the meaning of voicelessness in critical care," *Social Science and Medicine*, vol. 50, no. 9, pp. 1247–1255, May 2000.

Chapter 2

Design of a Plug-and-Play BCI

Copyright notice: The following text is reformatted from "A Multipurpose Brain-Computer Interface Output Device," as submitted to *Clinical EEG & Neuroscience*. The text and figures appear here with permission from SAGE, who owns all copyright to the work. See [1] for the final, definitive version of this work.

Introduction

Brain-computer interfaces (BCIs) are intended as a promising alternative to existing movement-controlled interfaces for individuals with severe motor impairments. BCIs have the potential to provide these individuals with direct-brain control of technologies such as computers, augmentative communication devices, environmental control systems and neural prostheses. Using this technology to maintain agency and their relationships with others, prospective users would potentially experience an increased level of independence and a higher quality of life [2].

Mason and Birch [3] proposed a framework for BCI design in which the user controls a device (e.g. a power wheelchair) through a series of functional components: (1) signal acquisition and amplification; (2) feature extraction; (3) feature translation; (4) control interfaces; and (5) device controllers. To date, the majority of BCI research has focused on the design and optimization of the first four components of the framework - for a thorough review of these elements of BCI design, the interested reader is referred to Mason's later paper [4].

As BCI design matures from a theoretical laboratory technology to a practical system that can be used in real-world situations by individuals with disabilities, there is growing interest in seamlessly interfacing the first four framework elements with existing device controllers. This approach has been taken by one group with a commercial robot and a commercial environmental control device [5], as opposed to the more common

strategy of recreating existing AT device controllers, e.g. a programmable IR controller [6]. In light of the goal of seamlessly interfacing with existing device controllers, BCIs can be modeled under the principles of input device emulation. In this model, the BCI control (e.g. the electrodes, amplifier, feature extractor and feature translator) and the Control Interface are combined into a single input device. This BCI input device, via an interface technology, would emulate standard input devices and be able to interface with various Device Controllers, as depicted in **Figure 2**. The importance of input device emulation as a design criterion has been emphasized in assistive technology design, maximizing the number of people who can connect special input devices [7], and providing a strategy that can accommodate the users' changing needs [8]. This latter issue is of particular importance to individuals with degenerative conditions such as amyotrophic lateral sclerosis (ALS) who experience frequent changes in their physical ability to interact with their environment; the ability to connect various input devices into the same Device Controller would minimize cost and adaptation for the user and his/her caregivers.

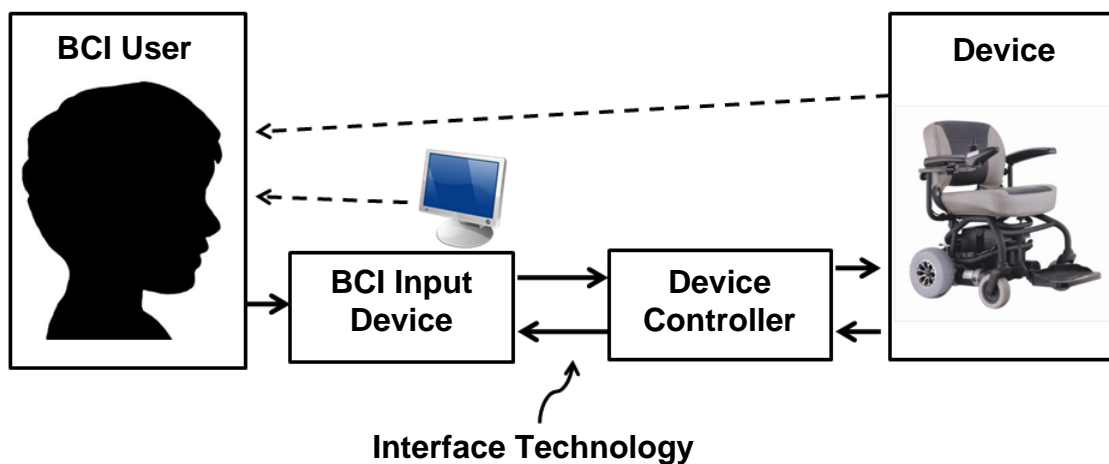


Figure 2: The role of Interface Technology connecting BCI Input Devices to any Device Controller, within the framework of input device emulation.

In order to realize this model, an effective interface technology must be developed that can connect the BCI Input Device to the Device Controller. Such an interface technology should be compatible with a variety of Device Controllers, able to interface with any BCI output, be convenient to use, and provide intuitive control of the Device

Controller to the BCI Input Device so that the resulting system can easily be configured to user needs and preferences. The design specifications for such an interface technology are presented in **Table 1**.

Table 1: Design specification for an Interface Technology to connect a BCI Input Device to a Device Controller

<i>Specification</i>	<i>Explanation</i>
1. Device Controller compatibility	Compatible with various assistive technologies, operating systems, etc.
2. BCI Input Device compatibility	Compatible with BCI2000 and custom BCI software
3. Convenience	Plug-and-play, no need for external power, configurable
4. Intuitive command structure	Complete and intuitive control of all output options

This paper presents the design and implementation of one example of an interface technology that connects BCI Input Devices to Device Controllers: the Multi-Purpose BCI Output Device (MBOD). The MBOD is capable of translating plain-text User Datagram Protocol (UDP) outputs such as those produced by BCI2000¹ (an open-source BCI implementation) into physical switch closures, Universal Serial Bus (USB) keyboard key-presses, or USB mouse movements and clicks. The design of the MBOD according to the specifications presented in **Table 1** will be described in the methods section, while the results section will present the results of testing the MBOD with various assistive technologies, operating systems and mobility devices. The discussion section will include the merits and limitations of the MBOD, and the paper will conclude with a note on the usefulness of the MBOD as a research tool to facilitate the study of clinical BCI use and the expectation of the eventual incorporation of this interface functionality into commercial BCIs.

¹ bci2000.org

Methods

Design Overview

The MBOD is designed to provide various output options to a BCI, enabling the combination of the BCI with existing AT devices and software. In order to interface with AT devices or programs with device controllers that support a USB keyboard or mouse, the MBOD is capable of acting as an intervening layer of hardware between two USB hosts (e.g. computers). An intervening hardware layer is necessary because USB hosts are unable to directly emulate devices or be connected together. The MBOD also provides physical switch outputs for interfacing with AT device controllers that are normally switch-operated.

The MBOD consists of (1) hardware and firmware that receives USB input from a BCI Input Device and outputs the desired keyboard, mouse or switch state to a Device Controller and; (2) a “companion” program, running on the BCI Input Device that performs ASCII-to-USB translation and interfaces with the MBOD. An overview of the MBOD hardware is provided in **Figure 3**; briefly, a SiLabs CP2102 USB-to-UART bridge receives input from the BCI Input device, and an Atmel AT90USB1287 microcontroller receives the translated commands, generates the corresponding USB or switch output, and handles all associated protocols and timing concerns. Switch output is accomplished through the use of optorelays. The hardware includes three status LEDs that provide feedback of the device’s state to aid in troubleshooting. Descriptions of the design decisions that shaped the details of the implementation are provided in the following sections.

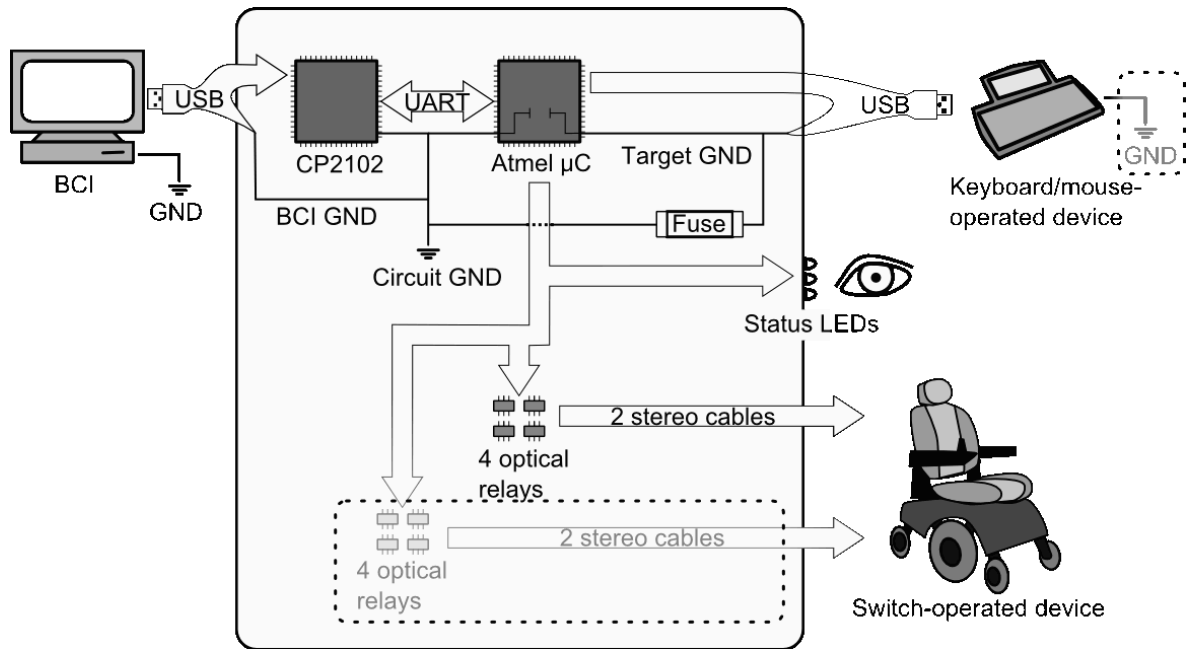


Figure 3: Overview of the MBOD hardware

Design Goal 1 - Device Controller compatibility

The first design goal for the MBOD was to ensure its compatibility with the maximum number of Device Controllers that prospective BCI users might employ for the purposes of augmentative and alternative communication (AAC), environmental control systems (ECS), computer access (CA) and movement (M). **Table 2** lists a number of common assistive technology devices used by individuals with disabilities and the type (e.g. switch, mouse, keyboard) of input required by their device controllers.

Table 2: Supported inputs of common assistive technology devices

Device	Assistive Technology Categories	Supported input		
		Switch	Mouse	Keyboard
Vantage ²	AAC	✓	✓	
DynaWrite ³	AAC	✓		✓
Imperium ⁴	ECS	✓		
Relax II ⁵	ECS	✓		
DASHER ⁶	CA		✓	
REACH Interface Author ⁷	AAC CA	✓	✓	✓
Cowriter ⁸	CA	✓		✓
PointSmart ⁹	CA		✓	
Assistive Mouse Adapter ¹⁰	CA		✓	
Smart Mustang Motorized Wheelchair ¹¹	M	✓		

² Prentke Romich Company, Wooster, OH, USA

³ DynaVox Mayer-Johnson, Pittsburgh, PA, USA

⁴ Tash, Inc, Richmond, VA, USA

⁵ Tash, Inc, Richmond, VA, USA

⁶ University of Cambridge, Cambridge, United Kingdom

⁷ Applied Human Factors, Inc., Helotes, TX, USA

⁸ Don Johnston Incorporated, Volo, IL, USA

⁹ Infogrip, Inc, Ventura, CA, USA

¹⁰ Montrose Secam Limited, Iver, Bucks, United Kingdom

¹¹ Smile Rehab Ltd, Newbury, Berkshire, United Kingdom

As illustrated in **Table 2**, seven of the ten listed devices have controllers that can be operated from switch input and seven of the ten can be operated with either keyboard or mouse input. To maximize Device Controller compatibility, the MBOD was designed to be able to emulate all three devices (USB keyboard, USB mouse, and physical switch).

The MBOD was programmed to identify itself as a keyboard and/or mouse using the USB Human Interface Device (HID) standards. Using these standards allows the MBOD to be recognized as a valid input device without modifications to the device controller (i.e. without the installation of drivers or software). USB ports are ubiquitous on personal computers regardless of operating system, and are frequently present on assistive technology devices. USB output is additionally attractive as USB specifications have a high degree of backward compatibility (e.g. both low- and high-speed USB 1.0 devices are supported by USB 2.0 and 3.0 hosts). By employing the USB HID standards, the MBOD is automatically compatible with all USB hosts that accept external keyboard or mice, without the need for installing drivers or software on the target system. This is particularly important as many assistive technology devices do not allow the installation of custom software, including drivers.

To emulate switch output, the MBOD also includes optorelays controlled by the microcontroller. To date, both Toshiba's TLP222A and Clare's PAA132 have been used successfully to generate switch output to assistive devices such as wheelchairs, including an Invacare TDX SP and Pride Quantum 600. As illustrated in **Figure 3**, the MBOD board layout can accommodate up to eight optorelays, though only units with four optorelays have been built and tested to date. The MBOD provides configurable timing for the duration of both keyboard and switch outputs, as some programs and devices cannot recognize short key-presses and precision timing of switch outputs is useful for interfacing with wheelchairs. The combination of USB and optorelay output enables the MBOD to interface with every assistive device identified in **Table 2**.

Design Goal 2 - BCI Input Device compatibility

While there have been a number of attempts to develop unifying frameworks and platforms for BCI research, a significant proportion of the field consists of BCI systems that are designed idiosyncratically [9]. With the introduction of BCI2000, an open-source

BCI research and development platform [10], many research laboratories presented investigations and developments based on this platform. While the MBOD was designed to accommodate input from BCI2000, care was taken to ensure compatibility with custom BCIs as well. This was done by creating a companion program to handle USB communication with the MBOD and thus abstract the hardware considerations away from the BCI. While the companion program could have been integrated into BCI2000, separating the functionality ensures that a custom BCI could easily work with the MBOD by implementing the simple BCI2000-style outputs described below.

BCI2000 and potentially other BCIs output user datagram protocol (UDP) packets that are human-readable strings in ASCII format containing (1) an identifier, (2) a whitespace, and (3) an output string. The companion program catches these UDP packets, translates them into a sequence of USB and switch commands, and forwards these commands to the MBOD hardware via a USB port on the BCI Input Device. Consider, for example, a BCI user who has just input the letter “A” using a BCI system running BCI2000 and one of its modules, the P300 speller. The output of this BCI Input Device is a UDP packet containing the string “P3Speller_Output A”. The companion program would parse this UDP packet sent by the BCI, translate the content into the keystroke ‘A’ in USB codes, and subsequently send this information onto the MBOD, which forwards it to a device controller, e.g. a communication system. The companion program is designed to accommodate key-presses (including special characters and strings of arbitrary length), mouse movements and switch commands in a similar manner. The system for handling special characters is described under design goal 3. For mouse and switch commands, the configurable identifier in the UDP packet is different, and the output string is numeric for mouse movements. If used with BCI2000's ConnectorModule, the companion program will listen for the Signal values for X and Y coordinate movement while ignoring all other state variable output, though this default behavior may be modified.

Design Goal 3 - Convenience

The MBOD was designed to enable BCIs to emulate a USB plug-and-play device, wherein the device facilitates its own discovery and installation into the USB host system, without the need for device configuration or user intervention. In other words, the

MBOD was designed to maximize convenience for the user and to offer immediate functionality once it was connected between the BCI Input Device and the Device Controller. To facilitate this ability, three features were considered: drivers, power, and flexible output capabilities. Ideally, the MBOD would be able to interface with Device Controllers without requiring driver installation; this is addressed in detail under design goal 1 via compliance with the USB HID standard. To maximize convenience, the MBOD should not necessitate the use of an external power source or batteries. Consequently, the MBOD is designed to operate from power taken from the USB port of the BCI Input Device. The MBOD's maximum current draw, with all three status LEDs and eight optorelays simultaneously active, is less than 100 mA; this is easily supplied by any standard-compliant USB port. This specification allows the MBOD to work from a battery-operated laptop; if the BCI amplifier is battery-powered or powered from the laptop as well, the entire BCI system can be mounted to a mobile platform, e.g., a powered wheelchair. In addition, the low current draw of the MBOD should not unduly affect the battery life of laptop computers. Finally, the MBOD is designed to accommodate a wide range of output configurations, rendering hardware configuration unnecessary for the end-user. The MBOD firmware is capable of handling simultaneous keyboard, mouse, and switch commands, as a simple packet structure supports sending different types of data to the microcontroller.

Design Goal 4 – Intuitive Command Structure

The MBOD was designed such that all three output modalities mapped intuitively from the BCI Input Device to the connected Device Controller. In the case of switch output, this simply involved mapping a single switch command from the BCI Input Device to a single switch selection through the Device Controller, a function that was easily implemented in the MBOD companion program as described above. To interface with the widest variety of devices, the MBOD is capable of translating the BCI Input Device's switch commands into variable-length switch presses. Similarly, intuitive mapping of mouse movements from the BCI Input Device, which might be used with continuous-output BCI modalities (e.g. sensorimotor rhythms), to the Device Controller involved mapping changes in the X and Y coordinate commands to corresponding

changes in mouse position on the output device. Once again, this functionality was implemented easily in the MBOD companion program.

By contrast, generating a full intuitive mapping of the keyboard required special consideration; while the UDP packet output of the BCI Input Device easily represented letters of the keyboard in ASCII symbols, keys without ASCII equivalents (e.g. F1 or arrow keys) were not directly represented. These special keys were therefore encoded in human-readable and user-configurable format using a command language in the companion program to handle modified key-presses and combinations (e.g. the Windows 'copy' command 'CTRL-C'). In this command language, configurable start and end characters (default '[' and ']') enclose tokens indicating the special characters. Tokens are defined for every key on a full keyboard (including a full numeric keypad). The default tokens are verbose to avoid conflicts, but can be redefined with a single line in a.ini file on the computer running the BCI. For example, the default token "KEYBOARD_F1" can be redefined to just "F1" by adding the line "KEYBOARD_F1=F1". To minimize interference with phrases, if a substring does not parse to a command, it is passed onto the Device Controller unchanged. There is also an "ECHO" keyword that instructs the command parser to ignore substrings in a single packet that would normally be considered commands, leaving a character-by-character ASCII-to-USB translation in place.

Creation of a custom ASCII extension to include special keys was rejected because the above scheme provides a more intuitive and efficient setup (e.g. typing "[CTRL-C]" instead of looking up an arbitrary hex value). It is also useful to the user, as target text is often displayed as the result on the BCI screen and these tokens are more interpretable than, e.g., the meaningless square that is currently produced on the BCI2000 screen when a non-standard ASCII value is output. Finally, such a custom extension would be likely to conflict with existing extensions used in, e.g., some non-English languages.

Results

The functionality of the MBOD was tested by using it to interface an EEG-based BCI (BCI2000 running on Windows XP with a g.USBamp amplifier) with various assistive technologies, operating systems and power wheelchairs. The interface was considered successful if users were able to use the BCI to control the functionality of the output device. The results of the test are presented in **Table 3**. The non-Windows operating systems were tested with simulated brain activity.

Table 3: Testing results for the MBOD with different output devices, using BCI control. 'Y' indicates successful operation, 'N' indicates unsuccessful operation, and 'N/A' indicates that the input mode is not supported by the device. Asterisks indicate notes in the following subsections. Non-Windows operating systems were tested with simulated brain activity.

MBOD Functionality	Assistive Technologies		Operating Systems			Power Wheelchairs	
	DynaWrite (AAC)	Scanning Director II ¹² ECS	Windows (XP, 7)	Mac OS	Ubuntu Linux	Invacare TDX	Pride Quantum
Keyboard Output	Y	N/A	Y	Y*	Y	N/A	N/A
Mouse Output	N/A	N/A	Y	Y	Y	N/A	N/A
Keyboard / Mouse Combo	N/A*	N/A	Y	Y*	Y	N/A	N/A
Switch Output	N/A	Y	N/A	N/A	N/A	Y	Y

¹² Lakeville Communications, Lakeville, OH, 44638

Assistive Technologies

The impact of plug-and-play use enabled by the MBOD on users' abilities to operate a BCI underwent rigorous testing with 24 individuals using it to interface BCI2000 with an assistive technology device (the DynaWrite AAC device) and a computer running Windows XP (unpublished data). Results from three of the individuals are available in a preliminary report [11].

One limitation of an early version of the MBOD came to light during this testing. The DynaWrite AAC device did not recognize the keyboard if the MBOD was placed in keyboard/mouse-combo mode, though it retained full functionality when the MBOD was placed in keyboard-only mode. The MBOD's USB operating mode (i.e. keyboard, mouse, or keyboard/mouse combo) can now be changed at runtime, which has proved to be a useful feature for troubleshooting compatibility issues.

Testing of the MBOD switch function was performed using a mu-rhythm BCI to generate switch closures, as well as with a P300 interface in the power wheelchair tests described below. This mode was successfully tested with the Scanning Director II environmental control system, although timing of outputs of the mu-rhythm BCI for operation of the scanning interfaces was challenging (unpublished data).

Operating Systems

The MBOD was tested on its ability to interface with three different Operating System families: Windows (XP and 7); Mac and Ubuntu Linux. Testing was considered successful if the MBOD functioned on the Operating System without the need for device configuration or user intervention. The MBOD functioned successfully on all three Operating Systems, though the Mac OS required a special key sequence upon the first installation.

Power Wheelchairs

The MBOD can produce switch outputs, and many wheelchair systems are designed for switch operation. However, feed-forward control is unstable and unsafe unless the wheelchair includes obstacle avoidance (see e.g. [12]). Using a 3-D

accelerometer to measure the current tilt position, we developed a system to allow control of the tilt-in-space system of powered wheelchairs [13]. Using this system, the MBOD allowed BCI control of two different powered wheelchair tilt-in-space systems - a Pride Quantum 600 and an Invacare TDX SP. The Invacare chair was used in a laboratory experiment to measure the effect of rotational movement on the P300 response [14]. Since the interaction was in a controlled environment, the MBOD switch was plugged directly into the chair through the standard device controller. The Pride chair was used in a subject's home, so ensuring that the switch only controlled tilt was more important than in the laboratory environment. The MBOD was thus used to interface with a dedicated tilt control board produced by the seating system manufacturer.

Discussion

This paper presents the design of the MBOD - an intervening layer of hardware which enables a BCI Input Device to emulate a keyboard, mouse, or a switch output and therefore interface with many Device Controllers. The MBOD is compatible with a wide variety of assistive technology, can receive commands from BCI Input Devices, and provides intuitive mapping of keyboard, mouse or switch functions. Additionally, the commands are configurable, allowing researchers to use terms intuitive to their lab and thus reduce errors when interfacing with the BCI.

Limitations and Future Work

The choice of USB communication did bring up an additional consideration: grounding. USB ground is typically connected directly to case ground on personal computers. Connecting two case grounds together forms a ground loop, which could potentially damage equipment. Although unlikely to be a problem in a user's home, ground loops can be a problem in large office buildings or hospitals. The user is protected from this unlikely event by the electrical isolation in BCI signal acquisition amplifiers, but such a ground loop could possibly damage equipment. In the MBOD, this issue has been addressed by separating the grounds with a fuse. The fuse has non-negligible resistance, alleviating the ground loop issue. Additionally, the fuse will protect the equipment on both sides of the MBOD from high current flow.

However, this solution is not ideal. In the event of failure, the fuse must be replaced before the MBOD will function again. Additionally, though the resistance is non-negligible, it is still small. Since the design and construction of existing MBOD units, a better solution has become commercially available: a USB opto-isolator chip. This chip, if placed between the Atmel microcontroller and the Device Controller, would completely prevent the ground loop, operate even if the grounds were at different potential, and not need to be replaced like the fuse. While it is not possible to add the chip to our existing MBOD units, future designs will incorporate this feature. Alternatively, wireless communication from the MBOD to a USB dongle could be implemented. The

USB dongle to convert the wireless signal to USB would be necessary since current AT devices do not directly accept wireless protocols such as Bluetooth.

While the MBOD allows interfacing a BCI to any AT that accepts a USB keyboard, USB mouse, or physical switch, it does introduce a second computer screen when used with a visual BCI and computer-based AT devices. Careful design of a commercial BCI with plug-and-play capabilities could reduce this concern, however. Minimizing BCI display size is an obvious step, since a small display requires less eye movement, provides more positioning options, and provides similar accuracy to a large display [15]. In the future, plug-and-play BCIs could use a transparent screen or retinal projection so the display appears at the same location as the target device [16]. Alternatively, a plug-and-play BCI could use auditory presentation of options [17],[18] or a visual BCI could operate an AT device with auditory feedback. For now, if the AT to be controlled is a piece of software that can be installed on the BCI computer, the separate display can be avoided by connecting the plug-and-play BCI back to the BCI host computer. While this configuration has been successfully tested with the MBOD, the special case of controlling AT software on the BCI computer can also be addressed in software rather than hardware.

A final limitation of the MBOD is that it only provides information flow from the BCI to the AT; only extremely limited information about USB configuration is returned to the BCI. In many cases, particularly when the BCI is used for communication through an AT device, this is unimportant; the AT device is designed to present feedback to the user. In situations where the BCI needs information from the device or device controller, additional sensors or communication channels may be added, such as the 3-D accelerometer used in the wheelchair tilt control application described earlier. In the case of switch-scanning AT devices, the lack of feedback may prove particularly problematic for system-paced BCIs.

Advantages for the BCI Community

As BCI research moves towards practical implementation in real-world settings with individuals with severe motor impairments, there is a growing need for theoretical frameworks, as well as hardware and software solutions to facilitate this transition. BCIs

are merely one aspect of a complex technological and social network that may enable individuals with disabilities to maintain autonomy and the ability to live in relationship with others. The integration of this technology into the existing networks of prospective BCI users is crucial for its acceptance and use [19]. The MBOD presents an intermediate solution to allow researchers to interface laboratory BCIs with the users' preferred assistive technologies and output devices, enabling the systematic study of the effect of BCIs in these settings. The time and resources necessary to integrate with a device such as the MBOD are considerably less than would be needed to replicate all the functionality of existing assistive technology and output devices for each new BCI user. The advantages offered by the MBOD interface would ideally be built into any commercial BCI system; however, in lieu of such technology, this device is presented to the BCI community as a tool for the development of BCIs that are fully integrated into the technological and social lives of individuals who rely on non-motor channels to act and communicate. Source code and hardware design files (e.g. GERBER files) are available to any member of the BCI community upon request to the corresponding author.

Conclusion

As BCIs transition from laboratory technologies to systems used by individuals with disabilities in real-world situations, it is important to begin systematically investigating the technical and social effects of integrating BCIs into the practical contexts of users' day-to-day lives. The MBOD is designed to facilitate this transition and has been successfully used as an intervening layer of hardware to enable BCIs to emulate keyboard, mouse or switch inputs to assistive technologies, computer operating systems and mobility devices. While its functionality should eventually be built into commercial BCI systems, this interface technology is currently available as a tool to the BCI research community to facilitate research toward the ultimate goal of enabling individuals with severe motor impairments to maintain their autonomy and their relationships with others in a real-world, day-to-day context.

Acknowledgements

The authors would like to thank Stefanie Blain for her assistance and advice with manuscript preparation. Also D. Camera, the creator of LUFA, for the availability of his library and his advice during the design phase of the project; finally, the creators and maintainers of BCI2000 for the availability of the platform and their assistance in getting up to speed quickly.

This work was supported in part by the National Institute of Child Health and Human Development (NICHD), the National Institutes of Health (NIH) under Grant No. R21HD054697, and the National Science Foundation (NSF) Graduate Research Fellowship under Grant No. DGE 0718128. The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD, NIH, or NSF.

References

- [1] D. E. Thompson and J. E. Huggins, "A multi-purpose brain-computer interface output device," *Clinical EEG and Neuroscience.*, vol. 42, no. 4, pp. 230–235, Oct. 2011.
- [2] A. Craig, Y. Tran, P. McIsaac, and P. Boord, "The efficacy and benefits of environmental control systems for the severely disabled," *Medical Science Monitor*, vol. 11, no. 1, pp. RA32–RA39, 2005.
- [3] S. G. Mason and G. E. Birch, "A general framework for brain-computer interface design," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 1, pp. 70–85, Mar. 2003.
- [4] S. G. Mason, A. Bashashati, M. Fatourechi, K. F. Navarro, and G. E. Birch, "A comprehensive survey of brain interface technology designs," *Annals of Biomedical Engineering*, vol. 35, no. 2, pp. 137–169, Nov. 2006.
- [5] F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A. Cherubini, M. G. Marciani, and F. Babiloni, "Non-invasive brain-computer interface system: Towards its application as assistive technology," *Brain Research Bulletin*, vol. 75, no. 6, pp. 796–803, Apr. 2008.
- [6] G. Xiaorong, X. Dingfeng, C. Ming, and G. Shangkai, "A BCI-based environmental controller for the motion-disabled," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 137–140, 2003.
- [7] E. Bergman and E. Johnson, "Towards accessible human-computer interaction," *Advances in Human-Computer Interaction*, J. Nielson, Ed. New Jersey: Ablex Publishing Corporation, 1995.
- [8] D. A. Ross, "Implementing assistive technology on wearable computers," *Intelligent Systems*, IEEE, vol. 16, no. 3, pp. 47–53, 2001.
- [9] J. R. Wolpaw, "Brain-computer interfaces (BCIs) for communication and control: a mini-review," *Supplements to Clinical Neurophysiology*, vol. 57, pp. 607–613, 2004.
- [10] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A General-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
- [11] D. E. Thompson, J. J. Baker, W. A. Sarnacki, and J. E. Huggins, "Plug-and-play brain-computer interface keyboard performance," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, New York, 2009, pp. 426–428.
- [12] B. Rebsamen, E. Burdet, C. Guan, H. Zhang, C. L. Teo, Q. Zeng, M. Ang, and C. Laugier, "A brain-controlled wheelchair based on P300 and path guidance," in *2006, 1st IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechanics*, 2006, pp. 1101–1106.
- [13] D. E. Thompson, M. T. McCann, C. C. Lee, and J. E. Huggins, "Impact of wheelchair tilt on P300 brain-computer interface accuracy," presented at the *Neuroscience 2010*, San Diego, CA.
- [14] D. E. Thompson, M. T. McCann, and J. E. Huggins, "Controlling wheelchair tilt with a brain-computer interface," presented at the *4th International BCI Meeting*, Asilomar, California, 2010.
- [15] M. Salvaris and F. Sepulveda, "Visual modifications on the P300 speller BCI paradigm," *Journal of Neural Engineering*, vol. 6, no. 4, p. 046011, Aug. 2009.
- [16] H. Takahashi and S. Hirooka, "Stereoscopic see-through retinal projection head-mounted display," in *Proceedings of SPIE*, San Jose, CA, USA, 2008, p. 68031N–68031N–8.

- [17] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kubler, “An auditory oddball (P300) spelling system for brain-computer interfaces,” *Psychophysiology*, vol. 46, pp. 617–625, 2009.
- [18] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, “A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients,” *Annals of the New York Academy of Sciences*, vol. 1157, no. 1, pp. 90–100, Mar. 2009.
- [19] M. J. Scherer, *Living in the State of Stuck: How technology impacts the lives of people with disabilities*. Cambridge, MA: Brookline Books, 1993.

Chapter 3

Evaluation of the Plug-and-Play BCI for Communication

In the previous chapter, I introduced the Multi-purpose BCI Output Device, or MBOD. The MBOD was designed with the goal of allowing a BCI to interface with a variety of assistive technology, whether hardware or software. This chapter presents an experiment that investigates whether doing so adversely affects performance. This chapter is in preparation for submission, with Jane E. Huggins as senior author, and Carmela Lee and Kirsten Gruis as co-authors.

Introduction

Brain-computer interfaces (BCIs) are intended to enable people with the most significant physical impairments to communicate and to operate technology without moving their body. However, BCI performance has yet to reach either user expectations [1] or, for most users, the level of performance available through commercial assistive technology (AT) [2]. To address this issue, researchers have begun incorporating AT techniques such as word prediction or symbolic communication into BCIs (see e.g. [3]).

This incorporation can take three forms. In the first approach, BCIs are implemented as stand-alone assistive technology (AT) devices with specific, limited functionality. In this approach, research time and effort is often spent in re-creating functionality already available through commercial products (see e.g. the programmable IR controller in [4]). Another approach which has gained recent popularity is integrating existing AT products into BCIs (see e.g. [5–8]). This second approach has the advantage of allowing access to the capabilities of proven products, though the effort and expense required for integration grows with the number of products integrated. Because of these costs, some products are likely to remain un-integrated, limiting user choice.

A third approach, which we recommended in [9], is to have the BCI act as a traditional input device through the principle of input emulation. In assistive technology

design, input device emulation has been emphasized as a design criterion, because input emulation can maximize interconnectivity with special input devices [10], and provide a way to accommodate the changing needs of users [11]. For individuals with progressive conditions such as amyotrophic lateral sclerosis (ALS), who often experience changes in their ability to interact with their environment, this ability could be quite important. For these users, the availability of multiple access pathways to control existing AT devices could lower adoption barriers to BCIs.

This chapter focuses on the practicality of a BCI keyboard that can replace physical keyboards. Such a keyboard must connect through the same universal serial bus (USB) port that the physical keyboard uses and also provide sufficient function for practical use. Performance categories include: 1) text generation accuracy and 2) text generation speed. Sufficient text generation accuracy, in the BCI field, is typically defined as 70% [12]. Sufficient text generation speed depends on competing AT options.

The P300 BCI design [13] is appropriate for keyboard replacement because it provides direct access to multiple options, frequently a visual letter grid. Each row and column of the grid flashes in a random order and the user mentally counts flashes of the desired option. Flashes of the desired option produce a P300 waveform in the user's electroencephalogram (EEG), while flashes of rows and columns without the desired option produce little response. The desired option can be determined from the EEG after several sequences of flashes (each row and column flashes exactly once per sequence). For people with vision impairments, P300 BCIs using auditory presentation of options are under development [14]. Paired with USB output emulation, a P300 BCI may be used as a keyboard replacement. In this chapter, our lab's Multi-purpose BCI Output Device (MBOD) [9] was used to provide the USB output capability.

While the technical capability of the MBOD to interface with a variety of devices was explored in the previous chapter and [9], it remains to be shown that the resulting system can be operated without impacting performance of the BCI itself. Using the system to interface with a standalone AT device presents several barriers to use that are absent when simply interacting with the BCI. For example, attention must be split between the AT device and the BCI display, which is operating under system timing regardless of the user's wishes. At a minimum, this operation mode could be considered

a secondary task, which is known to affect the amplitude and shape of the P300 response which the BCI depends upon [15]. Another barrier is that this operation requires gaze shift, though eventual adjustments to display modalities could reduce or eliminate this requirement (see the limitations section for further discussion).

This study evaluates the performance impact of using the MBOD to operate commercial devices, as compared to simply generating text with the BCI itself. While the MBOD itself is a stopgap measure designed for research such as this study rather than commercial applicability, the results should generalize to any standalone P300 BCI keyboard using a traditional monitor for display.

Methodology

Participants

Initial testing was performed with participants without amyotrophic lateral sclerosis (ALS). Participants experienced with P300 BCIs were recruited from the Wadsworth Center BCI Group participant list based on schedule availability. Novice participants without BCI experience were recruited first-come-first-serve through public postings at the University of Michigan. Inclusion criteria were: age 18 or older, able to read text on a computer screen, and able to understand participation instructions. Exclusion criteria were: inability to give informed consent, uncontrolled neck movements interfering with electroencephalogram (EEG) recording, photosensitive epilepsy, or open sores on the scalp.

Participants with ALS were then recruited from the patient list of the University of Michigan Motor Neuron Disease Clinic. Participants with ALS had additional recruitment criteria of ALS diagnoses and symptoms affecting a hand or arm. Functional impairments were measured with the ALS functional rating scale-revised (ALSFRRS-R), validated for telephone administration [16]. The ALSFRS-R rates function such that a score of 48 indicates no functional impairment while a score of zero indicates an individual who has loss of speech and purposeful limb movement; and requires a ventilator and feeding tube.

Additional participants without ALS were recruited for age-matched data. For these participants, an additional inclusion criterion was added: age within one year of a study participant with ALS.

Participants either signed consent forms approved by the appropriate institutional review boards (IRB) or gave consent and had a caregiver sign on their behalf. The subject compensation rate was \$12 per hour, rounded up to the next quarter hour and payable after each session.

Experimental Setup

Participants wore an EEG cap and sat approximately 0.8 meters from a 17-inch monitor showing the BCI display. The BCI was used to operate three devices: the BCI itself in standalone mode, a commercial communication aide¹³, and a separate laptop computer¹⁴. The communication aide device was below the BCI display; the laptop computer was beside the BCI display. Center-to-center distances between the BCI and communication aide displays and the BCI and computer displays were about 0.2 meters (14°) and 0.5 meters (32°) respectively. All BCI display variations contained the alphabet plus appropriate space, punctuation or command selections in a grid with 6 rows and 6 columns.

Experimental Design

The complete protocol involved 3 sessions of 1.5-2 hours, starting with about 30 minutes of time to set up the BCI. On the first session, a 19-character training set was collected without feedback; the EEG data from that training set were used to configure the BCI for each participant. After that, and during each following session, the participant used the BCI to reproduce a different sentence on each of the 3 devices; the device currently in use will be referred to as the "target" device. To require participants to interact with the target device and not merely use the BCI display, we provided feedback only through the target device; letters selected using the BCI did not appear on the BCI screen unless the BCI was being operated in standalone mode.

Regardless of target device, participants used a backspace selection to correct mistakes, so time and number of selections to reproduce sentences varied realistically. To limit user frustration, we restricted the time spent on any sentence to 15 minutes, ending the run manually even if the sentence was not complete.

In each target device, participants were asked to reproduce a 23-character sentence which was printed in 46-point Arial font and taped to the top of the BCI display. For the communication aide device, participants were also asked to make a special

¹³ DynaWrite from DynaVox Mayer-Johnson, Pittsburgh, PA, USA

¹⁴ Running Windows XP Professional, Service Pack 3, Microsoft Corporation, Redmond, WA, USA

selection which made the device speak the text that had been produced. A different sentence was used for each target device, and a different set of sentences was used in each session. The 9 sentences came from the sentence bank of the Compass interface evaluation software¹⁵. Participants reproduced the sentences in the same order, but device order was counterbalanced across sessions to minimize bias from fatigue, learning or sentence variation.

Equipment and Configuration

During BCI use, EEG from electrodes F3, Fz, F4, T7, C3, Cz, C4, T8, CP3, CP4, P3, Pz, P4, PO7, Oz, and PO8 was sampled at 256 Hz with a 16-channel electrode cap¹⁶ and a g.USBamp¹⁷ with reference and ground on the right and left mastoids respectively.

BCI2000¹⁸, was configured for each participant in the first session. The configuration was determined from EEG recordings while the subject focused on each character in “THE QUICK BROWN FOX” for 15 complete sets (sequences) of flashes. Each flash highlighted a row or column for 31.25 milliseconds with a pause of 125 milliseconds before the next row or column flashed. Time between characters was 3.5 seconds. The P300_GUI tool distributed with BCI2000 was used for configuration with settings of least squares weight selection, 800-ms EEG segments, and sample rate decimation of 20. For testing, the number of sequences for a participant was set to one greater than the number predicted to provide maximum BCI accuracy for that participant. This set the selections-per-minute for the participant since the BCI only selects a letter after the prescribed number of sequences.

The BCI was connected to the communication aide or computer USB port with the Multi-Purpose BCI Output Device (MBOD) [9] providing USB compatibility. An MBOD communication program on the BCI computer interpreted the non-standard BCI2000 output and transmitted it through the MBOD to the target device as standard

¹⁵ Koester Performance Research, Ann Arbor, MI, USA

¹⁶ Electro-Cap International, Inc, Eaton, OH, USA

¹⁷ Guger Technologies OEG, Graz, Austria

¹⁸ BCI2000 v2.0 build 2104, www.bci2000.org, Albany, NY, USA

USB keyboard codes. Thus, BCI2000 was recognized by these devices as a standard plug-and-play USB keyboard, making these devices BCI-operable without modification.

Data Analysis

BCI performance measurements were 1) text generation accuracy, 2) corrected text throughput, as measured by the BCI-Utility metric; see Chapter 5 for the reasoning behind the choice of this metric.

Determining BCI selection accuracy (**Figure 4**) required an accurate record of selections intended by the participant. Because participants were instructed to correct errors, the result of previous selections determined the next intended selection. A list of intended targets was created by examining the target sentence and the actual selections. If a participant informed us of an error of intent, such as losing place in a sentence or accidentally attending the wrong character, the intent was recorded according the participant's account. Once the intended targets were determined, accuracy was calculated as the percentage of attempted selections producing the intended result. The selection commanding the communication aide device to speak the completed sentence was included in accuracy calculations.

Sentence	I	T	_	I	S	_	Q	U	I	T	E	_	W	I	N	D	Y	_	T	O	D	A	Y						
Selected	I	Z	◀	T	_	I	S	_	W	◀	Q	U	I	Z	◀	T	E	_	W	I	N	D	Y	_	T	O	D	A	Y
Target	I	T	◀	T	_	I	S	_	Q	◀	Q	U	I	T	◀	T	E	_	W	I	N	D	Y	_	T	O	D	A	Y
Correct	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 4: Example sentence, actual letters selected, target text including error correction, and correctness of each selection (1 indicates correct, 0 indicates incorrect). Note that if no incorrect selections were made, Sentence and Target would be a perfect match and the backspace (◀) would not appear. In this example, 26 of 29 selections were correct, producing an accuracy of 89.7%.

Statistical analysis was performed for both participant groups (those with and without ALS) in one cohort. Ninety-five percent confidence intervals for the text generation accuracy and throughput were calculated using a linear mixed model [17]. The following factors were included as fixed effects: the number of sequences of stimuli presented, diagnosis of ALS, and subject age. Accuracy variations by participant and correlations among observations on the same participant were expected, so participant

was included as a random effect. Because each device was tested during each session, device was included as a repeated measures factor within session. The same model was used to calculate confidence intervals for throughput as measured by BCI-Utility, except that number of sequences was not included as a fixed effect because the term appears directly in the formula. Age was originally included in both models as a fixed effect, but was discarded due to correlations with diagnosis (see Limitations) and a dramatic difference in Akaike Information Criterion, a goodness-of-fit measure. Statistical analyses were performed with SAS¹⁹.

Similar analyses were performed with only the cohort diagnosed with ALS, but the results were not different enough to merit separate reporting.

¹⁹ Release 9.2 for Windows, SAS Institute, Inc., Cary, NC, USA.

Results

Recruitment and Data Characteristics

Twenty-nine participants without ALS, 3 experienced [18] and 26 novices, were recruited. Twenty-two participants without ALS, 13 men and 9 women, completed the 3 session protocol with 3-29 days between first and last sessions. Seven novice participants were excluded from analysis for not completing the protocol: reasons given included headache and nausea after participation and frustration with text generation accuracy and speed (one each), while some participants simply did not respond to further scheduling attempts. Mean age of participants without ALS completing the protocol was 43.1 years (range 18-79).

Thirteen novice participants with ALS were recruited. Eleven participants with ALS, (7 men and 4 women) completed the protocol. ALSFRS-R scores ranged from 18 to 41 with mean 28 ± 7 . Two participants with ALS were excluded from analysis for not completing the protocol. BCI accuracy was low (17%) after initial configuration for a 63 year-old subject with ALSFRS-R of 19. The subject did not want to repeat the configuration. A 56 year-old subject with ALSFRS-R of 11 had accuracy of 97.3% in the first session, but dropped out when accuracy in the second session was 50%. Mean age of participants with ALS completing the protocol was 61.7 years (range 45-78).

Of the 306 sentences attempted by the participants who completed the study, 214 were completed without errors remaining. Another 35 sentences were completed, but errors were left in the text, in contradiction with the instruction given to the participants. With corrections, completed sentences took a mean of 31.2 selections in 7.67 minutes. Sixteen participants (4 with ALS) completed at least one sentence without needing corrections, i.e. 100% accuracy. Three participants did not finish 23 characters of typing with one accidentally skipping a letter, one running out of session time, and one subject stopping early out of frustration with low accuracy.

The per-participant averages for each environment are shown in **Figure 5**.

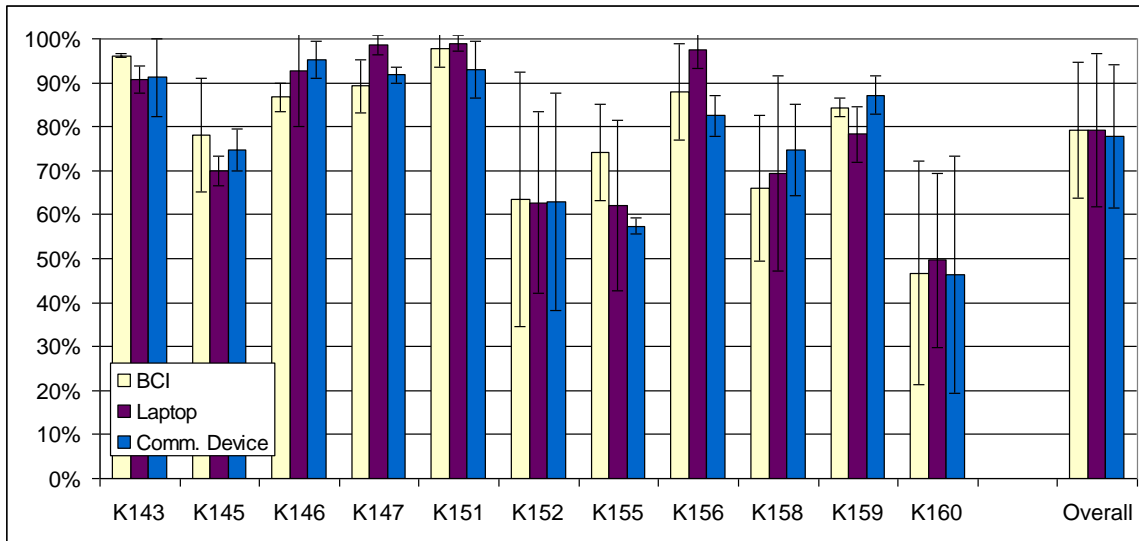
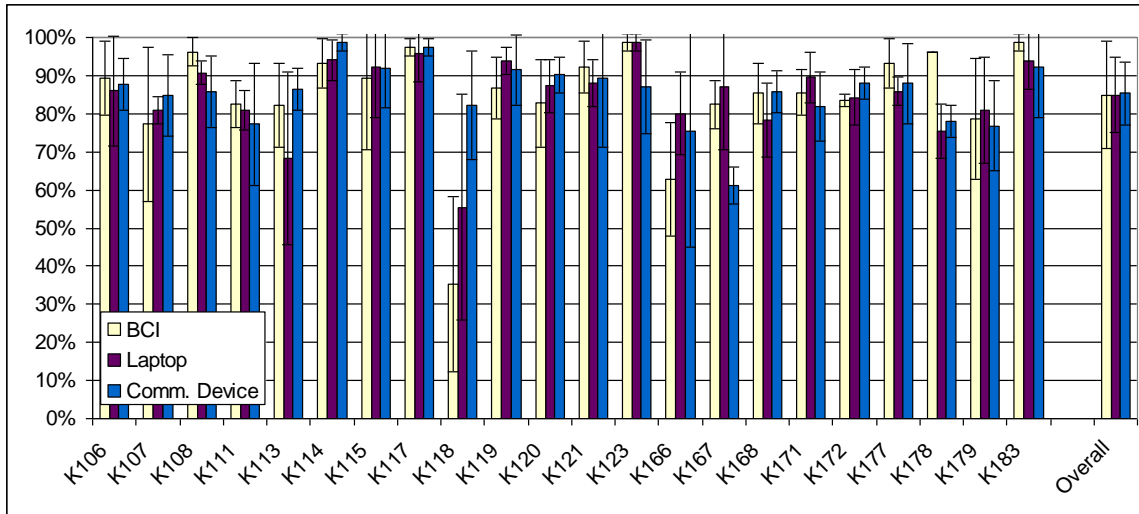


Figure 5: Average accuracy by device for each participant. Top) participants without ALS; Bottom) Participants with ALS. Error bars shows standard deviation across sessions.

Statistical Analysis

In the accuracy analysis, none of the fixed effects were found to be significant, though number of sequences trended toward significance ($F_{1,259}=3.43, p \sim 0.06$). This relationship is likely to be real but weak; the number of sequences was set based on training accuracy, which is an estimate of BCI performance, albeit a poor one (see Chapter 6 for details). Effects of session, diagnosis, and device were not significant ($p \sim 0.13, 0.15, \text{ and } 0.99$, respectively).

In the throughput analysis, the fixed effect of diagnosis was found to be significant ($F_{1,31}=4.27$, $p < 0.05$), with participants with ALS scoring on average 0.74 corrected selections/minute slower than able-bodied controls, or about 25% of the mean throughput for the control group. Fixed effects for session and device were far from significance ($p > 0.3$ for session, $p > 0.95$ for device). Relevant means and 95% confidence intervals are shown in **Table 4: Means and 95% confidence intervals for effects of environment and diagnosis..** A plot of BCI accuracy versus ALSFRS-R appears in **Figure 6.**

Table 4: Means and 95% confidence intervals for effects of environment and diagnosis.

	Accuracy (%)		
	Estimate	LB	UB
BCI	82.11	77.61	86.61
Laptop - BCI	-0.046	-3.401	3.31
Comm. - BCI	-0.145	-3.574	3.284

	Throughput (corrected selections/minute)		
	Estimate	LB	UB
BCI	2.751	1.974	3.144
Laptop - BCI	0.0035	-0.284	0.286
Comm. - BCI	-0.0306	-0.300	0.238
ALS - Control	-0.7428	-1.476	-0.009

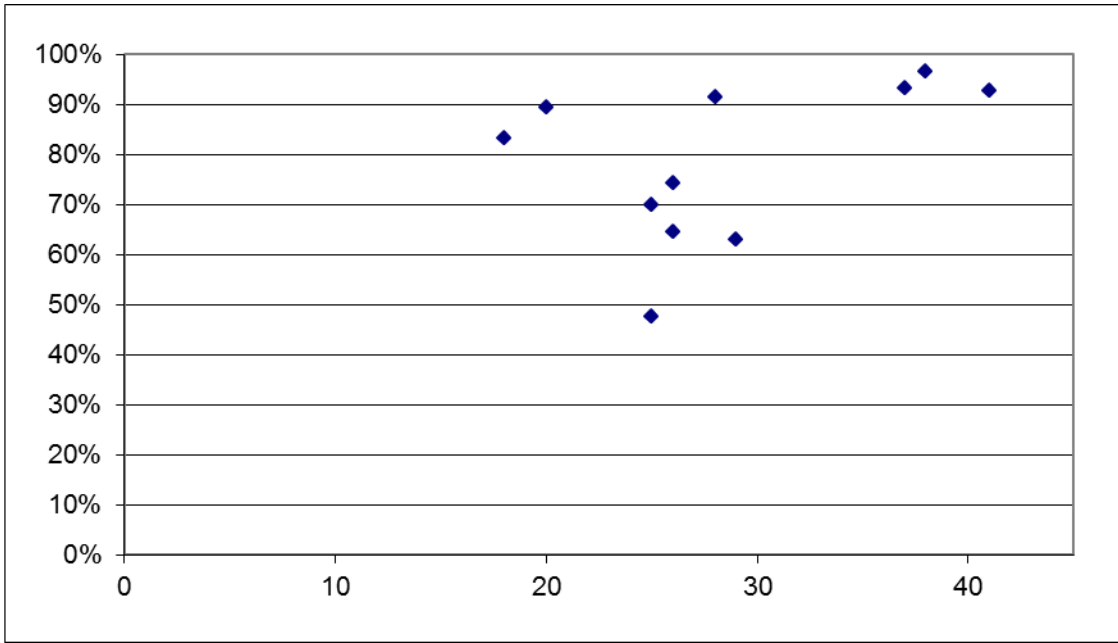


Figure 6: Average BCI accuracy vs. ALS-FRS.

Discussion

The results support the hypothesis that unmodified commercial assistive and mainstream technology can be functionally operated through plug-and-play replacement of a physical keyboard by a BCI keyboard. While some participants reported difficulty with screen positioning, particularly in the standalone communication device, the results show that for most users the performance was extremely similar across devices. The confidence bounds on differences in accuracy and throughput due to device are relatively tight compared to inter-participant differences. For example, from **Figure 5**, the standard deviation of mean BCI accuracy across subjects is 14%, as compared to the 3.5% bounds on device differences. Intra-participant differences due to variations in daily performance are also larger than device differences—the mean accuracy difference between session one and two was 5%. Even if the underlying device differences lie close to the lower 95% confidence bound, the boost from assistive technology is likely to be larger than the penalty associated with its use; in [3], predictive spelling increased corrected output speeds by 1.5 characters/minute, which would more than offset the worst-case 0.3 selections/minute penalty allowed by this study (see **Table 4**).

One participant in the non-ALS group, K118, had dramatic device differences. The participant's accuracy on the communication device was within the range of other able-bodied controls, but three standard deviations below the mean of the others in the standalone BCI device. We considered excluding the data from analysis based on lack of effort when using the other devices, as the participant reported wanting to hear the communication aide speak the sentence. This highlights the growing interest in the impact of motivation on BCI accuracy [19]. Individuals with impairments who need a BCI to accomplish their top priority tasks should be more motivated than research participants. The apparent importance of motivation supports incorporating plug-and-play capabilities into commercial BCIs to maximize the number of BCI-operable tasks, since users may be less successful with a BCI that limits or reduces the tasks they can perform.

The reduced throughput for participants with ALS is a concern because these are people from the target user population. It is interesting to note that some participants with lower ALSFRS-R scores had BCI accuracy as high as any participants without ALS.

However, participants with ALS who had lower accuracies also had lower ALSFRS-R scores (**Figure 6**). This may indicate that BCI performance is affected by a particular subset of ALS symptoms, not overall ALS progression as measured by the ALSFRS-R. Unfortunately, the number of individuals with ALS in this study is too small to address the question of which subset is important, if any.

Additionally, it is worth mention that the 95% confidence bounds for accuracy in each environment were well above the 70% commonly cited in BCI research as a threshold for useful communication [12]. However, if the confidence bounds were calculated including only participants with ALS, the lower bounds were 69.97, 70.32, and 69.51% for the BCI, laptop computer, and communication device, respectively. While this threshold is commonly cited, it is not a hard threshold based on mathematics or theory, but a soft threshold based on intuition of usability. In the throughput analysis, the lower bound in each device was about 1.8 corrected selections/minute; this throughput measure is more meaningful than simple accuracy, as it can be weighed against available options. These lower bounds are on the overall average performance, and individuals may still fall outside these limits.

The speed and accuracy at which a BCI could become useful to an individual depends on the speed and accuracy of alternative interfaces, most of which are faster than current BCIs [2]. For users with few interface options, a plug-and-play BCI as a keyboard replacement could reduce accessibility barriers by allowing any keyboard-operable device to be BCI-operated. The MBOD is a stopgap measure to study this functionality with current BCIs. A plug-and-play BCI would enable participants with other options to alternate between interfaces controlled by physical movements and by a BCI depending on their situation, fatigue level, and preferences. This would preserve resources invested in a current functional system while maintaining its availability if their impairments increase. Likewise, assistive technology (AT) providers could prescribe a BCI as an alternative access method for familiar commercial AT. If BCIs cannot interface with AT, then the functionality that the AT provides to someone with a progressive impairment may be lost when the person's physical condition progresses to the point that they need a BCI. In effect, their disablements would be increased due to the unnecessary incompatibility of two technologies both designed to enable people with impairments.

Another way to handle alternation between interfaces is implementing a hybrid BCI, such as that presented in [23]. A primary goal of hybrid BCIs is to address the issue of fluctuating user conditions using a control fusion approach, with online systems determining which input device (BCI or physical interface) is providing the most useful control at any given moment. While the system in [23] is capable of operating most computer software, it cannot interface with standalone AT devices. If the ability to interface with such devices could be added to hybrid BCI systems, the result could be an elegant method for handling multiple inputs to existing AT devices.

Study Limitations

This study presents the first plug-and-play USB BCI, therefore the focus was on the effect of target device instead of participant characteristics. While age-matched participants were recruited and included in the study, the original group of young, able-bodied controls causes a correlation between age and diagnosis that could be obscuring the relative contribution of each factor. The ages distribution of the two samples is statistically different ($p < 0.05$ with a Wilcoxon rank-sum test). Also, participant motivation was not measured, and motivation has recently been shown to affect BCI performance [19].

Although BCIs are intended for operation without physical movement, several recent studies may indicate that the most common P300 BCI “keyboards” utilize eye movements [2], [20]. Plug-and-play BCIs that provide a separate technology interface and therefore a separate display are particularly susceptible; however careful design should reduce this reliance. Minimizing BCI display size is an obvious step, since a small display requires less eye movement, provides more positioning options, and provides similar accuracy to a large display [21]. Users who want to operate computer software can eliminate the separate display by connecting the plug-and-play BCI back to the BCI host computer; a configuration tested successfully using the MBOD. Eventually, BCIs could use a transparent screen or retinal projection so the display appears at the same location as the target device [22]. Alternatively, a plug-and-play BCI could use auditory presentation of options as in [14], or a visual BCI could operate an AT device with auditory feedback.

Conclusion

A plug-and-play BCI can be used as a functional keyboard replacement to operate AT and mainstream technology for people with and without ALS. Differences in average accuracy and throughput between devices appear to be quite small, and can be offset by the advantage of using AT technology.

Acknowledgements

This work was supported in part by the National Institute of Child Health and Human Development (NICHD), the National Institutes of Health (NIH) under Grant No. R21HD054697, and the National Science Foundation (NSF) Graduate Research Fellowship under Grant No. DGE 0718128. The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD, NIH, or NSF.

References

- [1] J. E. Huggins, P. A. Wren, and K. L. Gruis, "What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis*, vol. 12, no. 5, pp. 318–324, Sep. 2011.
- [2] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based brain-computer interface," *Behavioral and Brain Functions*, vol. 6, no. 1, p. 28, May 2010.
- [3] D. B. Ryan, G. E. Frye, G. Townsend, D. R. Berry, S. Mesa-G, N. A. Gates, and E. W. Sellers, "Predictive spelling with a P300-based brain-computer interface: Increasing the rate of communication," *International Journal of Human-Computer Interaction*, vol. 27, no. 1, pp. 69–84, 2011.
- [4] Xiaorong Gao, Dingfeng Xu, Ming Cheng, and Shangkai Gao, "A BCI-based environmental controller for the motion-disabled," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 137–140, 2003.
- [5] F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A. Cherubini, M. G. Marciani, and F. Babiloni, "Non-invasive brain-computer interface system: Towards its application as assistive technology," *Brain Research Bulletin*, vol. 75, no. 6, pp. 796–803, Apr. 2008.
- [6] C. Zickler, A. Riccio, F. Leotta, S. Hillian-Tress, S. Halder, E. Holz, P. Staiger-Sälzer, E.-J. Hoogerwerf, L. Desideri, D. Mattia, and A. Kübler, "A brain-computer interface as input channel for a standard assistive technology software," *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 236–244, Oct. 2011.
- [7] F. Karmali, M. Polak, and A. Kostov, "Environmental control by a brain-computer interface," in *Proceedings of the 22nd Annual International Conference of the EMBS*, vol. 4, pp. 2990–2992, 2000.
- [8] E. W. Sellers, T. M. Vaughan, and J. R. Wolpaw, "A brain-computer interface for long-term independent home use," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 5, pp. 449–455, Oct. 2010.
- [9] D. E. Thompson and J. E. Huggins, "A multi-purpose brain-computer interface output device," *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 230–235, Oct. 2011.
- [10] E. Bergman and E. Johnson, "Towards accessible human-computer interaction," in *Advances in Human-Computer Interaction*, J. Nielson, Ed. New Jersey: Ablex Publishing Corporation, 1995.
- [11] D. A. Ross, "Implementing assistive technology on wearable computers," *IEEE Intelligent Systems & Their Applications*, vol. 16, no. 3, pp. 47–53, 2001.
- [12] A. Kubler, V. K. Mushahwar, L. R. Hochberg, and J. P. Donoghue, "BCI Meeting 2005 - Workshop on clinical issues and applications," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 131–134, Jun. 2006.
- [13] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, Dec. 1988.
- [14] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, "An auditory oddball (P300) spelling system for brain-computer interfaces," *Psychophysiology*, vol. 46, no. 3, pp. 617–625, 2009.
- [15] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007.
- [16] P. Kaufmann, G. Levy, J. Montes, R. Buchsbaum, A. I. Barsdorf, V. Battista, R. Arbing, P. H. Gordon, H. Mitsumoto, B. Levin, and J. L. P. Thompson, "Excellent inter-rater, intra-rater, and telephone-administered

reliability of the ALSFRS-R in a multicenter clinical trial,” *Amyotrophic Lateral Sclerosis*, vol. 8, no. 1, pp. 42–46, Feb. 2007.

- [17] B. T. West, K. B. Welch, and A. T. Galecki, *Linear Mixed Models: A Practical Guide Using Statistical Software*, 1st ed. Chapman and Hall/CRC, 2006.
- [18] D. E. Thompson, J. J. Baker, W. A. Sarnacki, and J. E. Huggins, “Plug-and-play brain-computer interface keyboard performance,” in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, New York, 2009, pp. 426–428.
- [19] S. C. Kleih, F. Nijboer, S. Halder, and A. Kuebler, “Motivation modulates the P300 amplitude during brain-computer interface use,” *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1023–1031, Jul. 2010.
- [20] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk, “Does the ‘P300’ speller depend on eye gaze?,” *Journal of Neural Engineering*, vol. 7, no. 5, Oct. 2010.
- [21] M. Salvaris and F. Sepulveda, “Visual modifications on the P300 speller BCI paradigm,” *Journal of Neural Engineering*, vol. 6, no. 4, p. 046011, Aug. 2009.
- [22] H. Takahashi and S. Hirooka, “Stereoscopic see-through retinal projection head-mounted display,” in *Proceedings of SPIE*, San Jose, CA, USA, 2008, p. 68031N–68031N–8.
- [23] G. R. Müller-Putz, “Tools for brain-computer interaction: a general concept for a hybrid BCI,” *Frontiers in Neuroinformatics*, vol. 5, 2011.

Chapter 4

Evaluation of the Plug-and-Play BCI for Control of a Wheelchair Tilt-In-Space System

The previous chapters introduced the plug-and-play BCI, and evaluated its use in a communication task. This chapter evaluates the use of the BCI in an environmental control task, specifically, the control of a wheelchair tilt-in-space seating system.

Introduction

While communication is an important task, environmental control tasks can provide important independence benefits to users with amyotrophic lateral sclerosis (ALS). One such task that is highly interesting to potential brain-computer interface (BCI) users is control over the wheelchair, both driving and tilt-in-space [1]. However, P300-based BCI performance is insufficient to control the position of a standard wheelchair safely except in extremely controlled conditions (see [2], [3] for the control of a smart wheelchair in a home environment). By contrast, a wheelchair tilt-in-space system, used to relieve the pressure of sitting still for long periods of time, presents fewer safety obstacles. Single-switch controls for tilt-in-space systems are commonly installed on wheelchairs for those with limited mobility, but as ALS progresses, users will eventually find themselves unable to operate even these simple interfaces. These switch interfaces are available for nearly all powered wheelchairs, whereas more powerful interfaces are often proprietary and specific to each wheelchair manufacturer or even model. The Multi-purpose BCI Output Device (MBOD), presented in Chapter 2 and [4], can produce physical switch outputs under BCI control. Thus, a BCI could be used to provide tilt-in-space control to the target population, with their existing chairs.

However, providing the user with such control may present a barrier to the operation of the P300 BCI itself. A tilt-in-space system, by design, causes rotational movement of the user. Aside from movement artifact, this movement produces three

known effects that are likely to directly interfere with the detection of the P300 response. First, rotational movement has been shown to cause vestibular evoked potentials at various latencies under 30 ms [5]. Second, since the P300 visual display needs to be visible to the user at all times, the display must rotate with the user. This configuration requires the user to suppress their vestibulo-ocular reflex, eye movement that would normally compensate for rotation to keep the gaze fixed on a target. The resulting setup is very similar to the experiment performed in [6], which reported fairly large-amplitude evoked potentials with latencies of approximately one second. Third, rotating the user causes movement of the entire visual field. This could be considered similar to the "visual noise" in [7], which was shown to reduce P300 amplitude and increase its latency. Reduced amplitude affects signal-to-noise ratio, making classification more difficult. See Chapter 6 for a discussion of the effects of variable latency.

In addition to the direct effects on P300 detection listed above, rotating the user may form a cognitive distracter - a secondary task that the user has to attend to. Such distracters are known to reduce the amplitude of the P300 response and increase its latency, particular if the distracters are sensory or perceptual in nature [8], [9].

At least two groups have investigated the effects of linear, as opposed to rotational, movement on the P300 response to both audio [10] and visual [11] stimuli. Control of wheelchair position and the resulting linear movement did not significantly affect P300 accuracy in [3], though that study did not directly address acceleration as a distracter. However, to date no investigation of the effects of rotational movement on the P300 has been published.

Methods

Equipment

In order to perform the experiment in this chapter, I designed a system to interface a generic wheelchair tilt-in-space system with BCI2000. This system relied on our MBOD to translate BCI2000 selections into physical switch closures. These switch closures were used to interface with wheelchair controls for tilt-in-space designed to be operated by a physical switch, described briefly in the introduction of this chapter.

However, the MBOD can only provide feed-forward control of the tilt-in-space system. To provide feedback and stability, I mounted a 3-axis accelerometer to the chair to detect the current tilt angle. I also wrote software that uses the feedback from the accelerometer to enable both absolute and relative control of chair position through plain-text commands sent by BCI2000. Michael McCann, an undergraduate in our lab at the time, modified BCI2000 to send the movement commands at deterministic times that could be controlled by parameter settings. The resulting system was published as an abstract at the 4th International BCI Meeting [12].

Experiment

Each of twelve participants (8 women, 4 men, ages 19-63, mean 40.5) operated the P300 interface on three separate days (sessions), while sitting in a wheelchair. There were three 24-character "runs" per session, one in each case from the following list: no tilting, user-controlled tilting, and "rocking chair" tilting. The order of the tilt cases was counterbalanced across participants, similar to the experimental design in the previous chapter. In all cases, the text to be typed was a common city name, interleaved with numbers. Unlike the experiment in chapter 3, participants were not allowed to correct mistakes, so each run contained the same amount of data. In the user-controlled tilting case, the numbers in the corresponded to movement commands. In this tilt case, any selected movement command, whether correctly or accidentally selected, produced a movement, similar to how an end-user tilt control system would operate. In other cases, user commands had no effect on the wheelchair tilt.

In the rocking chair case, the chair was put through a series of 4° (~1 second) tilts in alternating directions, maximizing the number of starts and stops of movement. This movement method was chosen since, based on anatomy, any vestibular responses are presumably linked to acceleration rather than velocity. The stimulus presentation order was adjusted so that starts and stops would occur in an approximately uniform distribution between 0 and 6 stimuli from the next target stimuli. Data were recorded in an identical fashion to the previous chapter, although precisely 11 sequences of stimuli were presented to each participant in every tilt case, without optimizing the speed for the participant. All human subject work was conducted under the review and guidelines of the Institutional Review Board.

Data Irregularities

One participant (M141, a 26-year-old female) was excluded from analysis because her BCI accuracy in the non-tilting environment was below 50% on average. Since this accuracy does not allow for correcting of errors in online communication, the effects of movement on her BCI accuracy was deemed irrelevant.

An unfortunate programming error led to the MBOD ceasing to function after a large number of characters were selected through the BCI, and the method of failure was irregular enough that the problem was difficult to reproduce offline. While the problem has since been found and addressed, the online data from three participants was affected to varying degrees. M133, a 19-year-old male, experienced the failure just two characters from the end of one of his user-controlled movement sessions; the resulting data file is thus one character shorter than the others. M136, a 21-year-old male, experienced a similar failure. M142, a 59-year-old female, was affected the most severely. During her first session, the rocking chair environment failed approximately halfway through the run.

The above three participants are still in the study, as the majority of their data is still valid. The data from the affected runs was still used in analysis, accounting for the different number of characters when combining results from multiple runs.

Analysis

The statistical analysis in this experiment is meant to address the following hypothesis: That movement significantly affects BCI accuracy and/or corrected throughput.

However, of note is the fact that the same typing speed was used for all participants, regardless of the optimal speed for each person. Put another way, for most participants the BCI was artificially slowed by presenting extra stimuli. These extra stimuli not only slow the overall performance of the BCI in terms of throughput, but may serve to obscure any movement-related effects, since the additional data should improve the robustness of classification. Given that actual BCI use would be optimized to a speed suitable to each individual, the analysis was performed in an offline manner, artificially reducing the number of sequences to an ideal speed. The number of sequences for each participant was optimized based on a throughput measure known as BCI-Utility [13] calculated in the non-movement case; this approach mirrors an experienced BCI user being supplied with seating system control, who would already have a number of sequences set based on non-movement data. See Chapter 5 for the reasoning behind the selection of BCI-Utility from the available metrics.

Statistical analysis was performed with SAS²⁰, using a linear mixed model (LMM) [14] to produce confidence intervals for the difference in accuracy and throughput in the different tilt cases. Tilt case and session were included as categorical fixed effects. Since accuracy variations by participant and correlations among observations on the same participant were expected, participant was included as a random effect. The repeated measures design of the experiment was incorporated into the analysis.

A second analysis was performed breaking the data into just two conditions: those selections made without movement (from the no tilt and user-controlled tilt cases) and those that were interrupted by movements in any way (from the rocking chair and user-controlled tilt cases). The analysis was otherwise identical to the above.

²⁰ Release 9.2 for Windows, SAS Institute, Inc., Cary, NC, USA.

Results

Figure 7 shows the mean BCI accuracy across participants as a function of the number of sequences used for typing. The number of sequences used in the offline analysis, along with accuracy and throughput at that number of sequences, is shown in **Table 5**.

Table 6 shows the estimates and confidence intervals for the difference in accuracy and throughput in different tilt cases, calculated at the optimal number of sequences for each participant. As can be inferred from the confidence intervals, the only statistically significant difference observed was accuracy in the rocking chair environment.

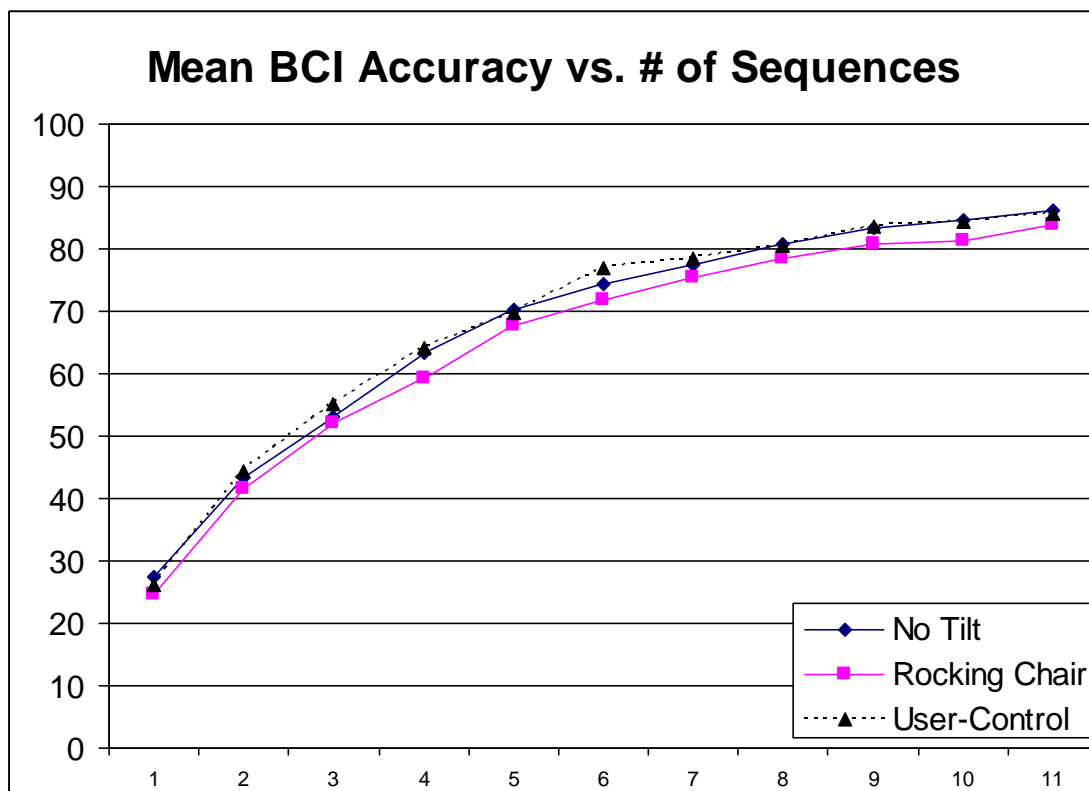


Figure 7: Mean BCI Accuracy vs. # of Sequences in the three tilt cases.

Table 5: Number of sequences used in the offline analysis, with accuracy and throughput at that number of sequences. All numbers were calculated from data in the no tilt case, and ideal sequences and throughput were calculated using BCI-Utility.

Participant Code	Ideal Sequences	Mean Accuracy (%)	Mean Throughput (correct char/min)
M132	4	93	4.70
M133	4	88	4.09
M135	5	79	2.72
M136	5	82	2.98
M138	9	82	1.88
M139	5	89	3.62
M140	10	79	1.57
M142	8	89	2.52
M144	4	76	2.88
M148	10	86	1.95
M149	8	82	2.07

Table 6: Estimates and confidence intervals for the differences between tilt environments. Lower and upper bounds are based on 95% confidence intervals.

Difference in % Accuracy	Estimate	Lower Bound	Upper Bound
User-Controlled Tilt – No Tilt	-3.986	-9.132	1.161
Rocking Chair Tilt – No Tilt	-8.333	-15.37	-1.298
Difference in Correct Characters/Minute	Estimate	Lower Bound	Upper Bound
User Control – No Tilt	-0.4212	-0.9714	0.1298
Rocking Chair Tilt – No Tilt	-0.5214	-1.1424	0.0994

The results from the second analysis appear in **Table 7**. As may be inferred from the confidence bounds, the difference in accuracy is significant at the 0.05 level, whereas the difference in throughput is not.

Table 7: Estimates and confidence intervals for the differences between characters interrupted by movements and those that were not. Lower and upper bounds are based on 95% confidence intervals.

Difference in % Accuracy	Estimate	Lower Bound	Upper Bound
Movement – Non-Movement	-5.41	-10.7	-0.1
Difference in Correct Characters/Minute	Estimate	Lower Bound	Upper Bound
Movement – Non-Movement	-0.373	-0.872	0.125

Discussion

The rocking chair environment, the worst-case distracter, produced the worst BCI performance, but the difference was only statistically significant in accuracy (not in throughput). The confidence bounds indicate that the difference in accuracy is likely to be less than 15 percentage points, while the impact on throughput is probably small but possibly important (<1.14 characters per minute for rocking chair). The user-controlled tilt case, which was a model for how a system like this might be implemented in practice, produced performance somewhere between the other two tilt cases; the bounds for user control are about half that for the rocking chair on accuracy but similar on throughput (9 percentage points, about 1 character per minute). The second set of values corresponds to about 10% of mean accuracy and 30% of mean throughput in the no-tilt case. If the underlying difference is near to these upper bounds, tilt operation may not be suitable for some users, particularly those for whom the BCI does not work well. Further testing would be required to narrow the confidence bounds and achieve a better estimate of the performance difference, but most participants in the study were able to use the BCI to control a wheelchair seating system even without modification to the underlying classification methods.

Only two participants dropped below the commonly-cited 70% accuracy threshold [15] in the user-controlled tilt case – M136 (mean accuracy 69% instead of 82%) and M144 (mean accuracy 62.5% instead of 76%). Both of these participants were among those for whom the BCI performed the best in terms of speed (the number of sequences was 5 and 4 respectively), and the difference in accuracy was substantially less at 11 sequences (5 and 1.1 percentage points, respectively). This highlights my earlier statement that the large number of sequences can substantially reduce accuracy differences, and suggests that simply increasing the amount of data collected could help minimize the accuracy differences caused by tilt operation. It should also be noted that the 70% threshold is based on corrected communication, where errors must be backspaced and then the original selection repeated. In this particular control task, one could usually simply attempt to select the desired chair angle a second time, so lower accuracies may be acceptable. Conversely, the annoyance and distraction of moving to

an undesired tilt may cause users to be less forgiving of low accuracies, depending on personality and preference.

The second analysis gave results nearly identical to the first: statistical significance only on accuracy (not throughput), and similar bounds. The bounds are more difficult to interpret, however, as they represent the average difference in BCI performance between no movement and some mix of minor and large movement interruptions.

This second analysis was primarily performed because the data from the first analysis suggest a "dose effect" of movement; the mean performance was worst in the rocking chair case, best in the no tilt case, and average in the user-controlled tilt case. As these results echoed the amount of movement present in each case, the idea was formed that perhaps the user-controlled tilt case was simply a mix of the other two cases. While the statistical significance indicates there is a difference in accuracy in this analysis, it should be noted that there is no statistical difference between the two halves of the data in the user-control case ($p > 0.5$ on all metrics). Given this fact, the statistically significant difference in accuracy is likely due only to the difference between no tilt and rocking chair (shown in the first analysis); this second analysis cannot be said to prove a dose effect.

Study Limitations

While the statistics indicate that rocking chair tilt does interfere with BCI operation, the confidence bounds on the effect of user-controlled tilt (the case of interest) are somewhat broad. This could be interpreted as indicating that insufficient data was collected in this study. Ultimately, technical difficulties related to data collection led to us canceling this study: the wheelchair seating system was on loan, the MBOD-related issues required some participants to repeat runs, etc.

This study only includes data from able-bodied individuals. Individuals with moderate to advanced ALS typically use a wheelchair for mobility, which poses a substantial challenge. Our able-bodied participants all used the loaner chair that we had instrumented for purposes of the study; participants who normally use a wheelchair would have to either transfer to the instrumented chair, or have instruments added to their chair. Transfer to the instrumented chair would be the easiest option for the experimenters, but could be a considerable burden for participants. While the system used in this chapter is capable of controlling tilt on most wheelchairs, instrumenting a chair requires a fair amount of effort and investment to ensure participant safety and experimental reliability. While that amount of effort might be reasonable when providing a user with an in-home system, it is difficult to justify for a short research study. Additionally, the tilt speed of two chairs could be quite different, forcing researchers to choose between consistent timings and consistent angles. Given these issues, the difficulties with data collection, and the relatively minor differences observed between tilt cases in the data we had, we decided not to recruit participants with ALS or other conditions; the outcome of this decision is that whether the results will generalize to potential BCI users is unclear.

The wheelchair seating system used in the study was a realistic example of a chair that might be used by the target population, but from a scientific perspective had several non-ideal characteristics. For example, the system only had one speed for tilt; the speed was different in each direction, and dependent on the weight of the person in the chair.

One original goal of this experiment was to determine the effect of rotational movement on the shape of the P300 waveform itself. This investigation does not appear

in this chapter because preliminary analysis indicated that insufficient data were collected. While the study was designed to ensure that an approximately equivalent number of target stimuli appeared at each time interval from a movement onset, the amount of data was limited by a desire to keep participant time requirements under two hours for the full session. The number of starts and stops that could be accomplished in a given amount of time was determined by the chair used, which required nearly two seconds to reverse directions. Due to these factors, each time lag from movement onset had, on average, 20 observations, which is already somewhat below the 36 artifact-free trials suggested for calculating P300 characteristics in clinical studies [16]. Furthermore, additional data should likely have been collected given that BCI operation is substantially faster than that suggested for clinical recordings; in this chapter stimuli were presented at approximately 6 Hz, as opposed to the clinical guideline of 0.5-1 Hz. In addition, substantial movement artifact appears to be present in the recordings. Finally, because the chair is designed for user comfort, the starts and stops are smooth rather than abrupt, making the onset of movement difficult to determine from the tilt angle. These factors combine to make the analysis untenable; future investigators are cautioned to collect substantially more data than they think they need, presumably by limiting the time offsets that are studied, and probably also by increasing the length of the experiment or dropping the user-controlled tilt case.

Conclusion

This chapter presented the design and testing of a system interfacing a BCI with a generic wheelchair seating system, creating the first BCI-controlled wheelchair seating system. The system could be operated by all users, though accuracy was significantly less in the worst-case scenario of near-constant movement. Statistical analysis indicates that the performance penalty of user-controlled tilt is likely to be less than 9 percentage points or 30% of corrected throughput, based on the 95% confidence intervals. Unlike in the previous chapter, these costs would not be offset by access to assistive technology; the system should likely be evaluated for individual users before being installed for in-home use. Presenting additional stimuli seems to offset the accuracy differences between tilt cases, so care providers may wish to increase the number of sequences presented to users who have difficulty controlling their wheelchair tilt angle through the BCI.

Acknowledgements

This work was supported in part by the National Institute of Child Health and Human Development (NICHD), the National Institutes of Health (NIH) under Grant R21HD054697 and R21HD054913. The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD, or NIH. This material is based upon work supported by the National Science Foundation Graduate Student Research Fellowship under Grant DGE0718128. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NSF.

References

- [1] J. E. Huggins, P. Wren, and K. L. Gruis, "What do brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis," *Journal of Amyotrophic Lateral Sclerosis*.
- [2] B. Rebsamen, E. Burdet, Cuntai Guan, Chee Leong Teo, Qiang Zeng, M. Ang, and C. Laugier, "Controlling a wheelchair using a BCI with low information transfer rate," in *2007, 10th International IEEE Conference on Rehabilitation Robotics*, 2007, pp. 1003–1008.
- [3] B. Rebsamen, C. Guan, H. Zhang, C. Wang, C. Teo, M. H. Ang Jr., and E. Burdet, "A brain controlled wheelchair to navigate in familiar environments," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 99, p. 1, 2010.
- [4] D. E. Thompson and J. E. Huggins, "A multi-purpose brain-computer interface output device," *Clinical EEG and Neuroscience.*, vol. 42, no. 4, pp. 230–235, Oct. 2011.
- [5] E. Leibner, J. Elidan, S. Freeman, M. Sela, M. Nitzan, and H. Sohmer, "Vestibular evoked potentials with short and middle latencies recorded in humans," *Electroencephalogram and Clinical Neurophysiology Supplements*, vol. 41, pp. 119–123, 1990.
- [6] J. Hood and A. Kayan, "Observations upon the evoked responses to natural vestibular stimulation," *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, vol. 62, no. 4, pp. 266–276, Jul. 1985.
- [7] G. McCarthy and E. Donchin, "A metric for thought: a comparison of P300 latency and reaction time," *Science*, vol. 211, no. 4477, pp. 77–80, Jan. 1981.
- [8] G. Hajcak, A. MacNamara, and D. M. Olvet, "Event-related potentials, emotion, and emotion regulation: an integrative review," *Developmental Neuropsychology*, vol. 35, no. 2, pp. 129–155, Feb. 2010.
- [9] T. W. Picton, "The P300 wave of the human event-related potential," *Journal of Clinical Neurophysiology*, vol. 9, no. 4, pp. 456–479, Oct. 1992.
- [10] H. Seidel, U. Schuster, G. Menzel, N. Nikolajewitsch Kurerov, J. Richter, E. J. Schajpak, R. Blüthner, A. Meister, and P. Ullsperger, "Changes in auditory evoked brain potentials during ultra-low frequency whole-body vibration of man or of his visual surround," *European Journal of Applied Physiology and Occupational Physiology*, vol. 61, no. 5–6, pp. 356–361, 1990.
- [11] H. Nolan, R. Whelan, R. B. Reilly, H. H. Bulthoff, and J. S. Butler, "Acquisition of human EEG data during linear self-motion on a Stewart platform," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, 2009, pp. 585–588.
- [12] D. E. Thompson, M. T. McCann, and J. E. Huggins, "Controlling wheelchair tilt with a brain-computer interface," presented at the *4th International BCI Meeting*, Asilomar, California, 2010.
- [13] B. Dal Seno, M. Matteucci, and L. T. Mainardi, "The Utility Metric: A novel method to assess the overall performance of discrete brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 1, pp. 20–28, Feb. 2010.
- [14] B. T. West, K. B. Welch, and A. T. Galecki, *Linear Mixed Models: A Practical Guide Using Statistical Software*, 1st ed. Chapman and Hall/CRC, 2006.
- [15] A. Kubler, V. K. Mushahwar, L. R. Hochberg, and J. P. Donoghue, "BCI Meeting 2005 - Workshop on clinical issues and applications," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 131–134, Jun. 2006.

- [16] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, R. Johnson, G. A. Miller, W. Ritter, D. S. Ruchkin, M. D. Rugg, and M. J. Taylor, "Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria," *Psychophysiology*, vol. 37, no. 2, pp. 127–152, Mar. 2000.

Chapter 5

Measuring Performance

As noted in the previous chapters, for most users the information throughput of BCIs is low relative to existing assistive technology. However, the issue of how to measure that throughput is far from resolved. Many metrics are used in the BCI field, several of which have been created within the field to try to account for the way text is generated in specific BCI paradigms.

This chapter is currently in submission to *IEEE Transactions in Neural Engineering and Rehabilitation*, with Stefanie Blain-Moraes as a co-first author and Jane E. Huggins as senior author. As such, the material below (figures and text) is copyright IEEE pending acceptance. The paper is written to suggest a standard choice of metrics and place of measurement for the BCI field. Although the criteria were developed for the communication task, one of the recommended metrics, unlike many contending metrics, is applicable to environmental control as well. Under the framework presented here, environmental control modules could be considered *selection enhancement* modules (defined below), and the results would follow.

Introduction

Brain-computer interface (BCI) technology can be used as a form of augmentative and alternative communication (AAC) by individuals who do not have any voluntary muscle control. The architecture of a BCI-based AAC system can be represented by many different frameworks [1]; herein, we model BCI-based AAC systems as two interconnected modules, each of which is comprised of a number of functional components, depicted in **Figure 8**. The *BCI Control Module* translates a BCI user's brain state into a logical control output. Its functional components may include a stimulus presentation paradigm which causes the BCI user to elicit particular brain states (e.g. the

presentation of options in P300-based BCIs); electrodes and amplifiers; feature extractors; and classification algorithms. A comprehensive review of the variations of each of these functional components is available in [2]. The BCI control module makes discrete selections from a system-dependent number of possible options. These selections are made independent of any semantic knowledge of the AAC interface, and the resulting logical control signal is sent to the Selection Enhancement module. A *Selection Enhancement* module translates this logical control to semantic control, using techniques ranging from direct association (e.g. one output option corresponds to one specific communicative symbol) to more complicated algorithms such as error correction and word prediction. These two modules work in tandem to provide a means of communication for individuals who have severe motor impairments that limit their ability to speak and to access traditional AAC devices.

Many variations of each of the functional components of both BCI Control and Selection Enhancement modules exist, and can be combined together in multiple ways to produce unique BCI-based AAC system configurations. To develop an optimal BCI-based AAC technology, researchers must be able to compare each of these configurations to assess the relative benefit of each component to the overall communication capacity of the system. In other words, the quest for the best BCI requires efficient evaluation criteria for the performance of each component of the communication system.

As is the case with the evaluation of any AAC system, the issue of *where* to measure performance is paramount. There are three locations, or *levels*, at which BCI-based AAC performance can be measured, as depicted in **Figure 8**. Level 1 performance is measured directly at the output of the BCI Control Module. Here, the effective generation of a logical control output is commonly assessed by measures of speed, accuracy, or a combination thereof, such as information transfer rate. To date, measurement of BCI performance has typically occurred at this level. However, as BCI systems begin to explore improved user interfaces (e.g. integration of word prediction in spelling applications, innovative spelling systems, adaptive user interfaces [3], [4]), Level 2 measures of communication capacity at the output of the Selection Enhancement Module have become more common [3–6]. Level 2 measures of BCI performance account for the fact that a single selection by the user may have different degrees of

“power” in terms of what it can accomplish when the logical control signal has been interpreted by the Selection Enhancement Module. While the rarity of in-home BCIs being used by the target population have delayed the need to identify a higher-level measure of BCI performance, the AAC literature indicates that it is also possible to measure performance of a communication system, and therefore of a BCI, at the level of the user. This would be considered a Level 3 measurement of BCI performance, and can be assessed by determining whether the presence of a BCI leads to fuller, richer communication with a partner [7], or an improved quality of life [8].

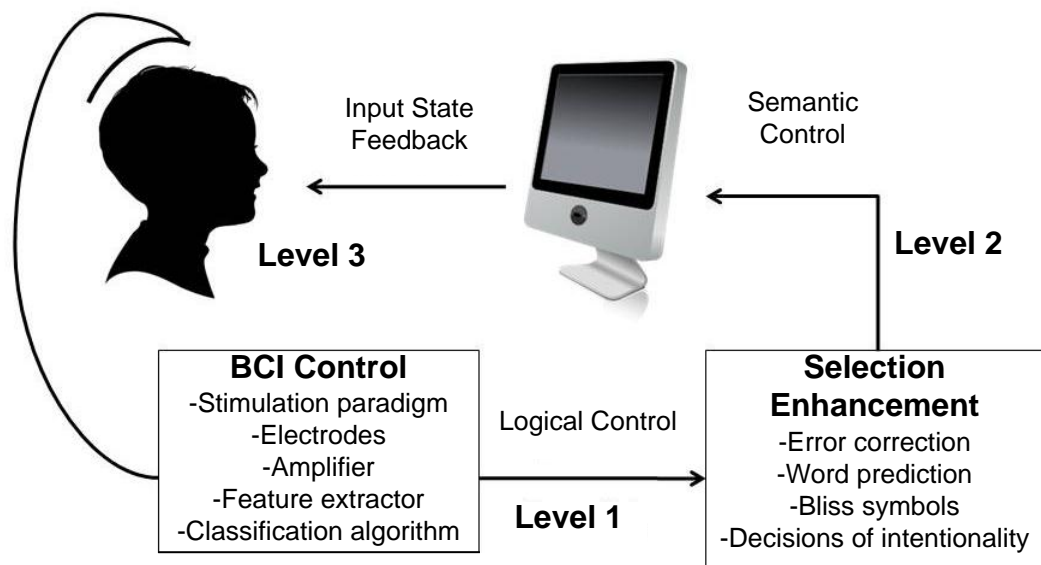


Figure 8: Architecture of a BCI-based AAC system that is comprised of two modules: (1) a BCI Control module that translates brain signals into logical control outputs and (2) a Selection Enhancement module that translates logical control to semantic control. Performance of BCI-based AAC systems can be measured at three levels (labeled Level 1, Level 2, Level 3) within this architecture; each level of measurement is currently assessed by a variety of often incommensurable performance metrics.

Within each level of assessment of BCI-based AAC performance, the issue of *how* to measure performance requires consideration. The way in which a user employs a BCI to accomplish a task significantly affects this process. BCIs can be used in two different ways – to control a process (“process control”) or to select a goal (“goal selection”) [9]. Whereas the path towards the goal is important in process control BCIs, it is of no consequence in goal selection BCIs.

A large number of performance metrics have been used in BCI research studies to quantify the communication capacity of a specific BCI system. As a standard metric does not currently exist, research groups are developing and publishing their own independent performance metrics to capture similar phenomena. For example, in order to accurately capture the performance of a BCI where users were given the option of correcting mistakes that they had made in typing a sentence, Townsend et. al (2010) [10] developed the “practical bit rate”. Jin et. al (2011) [11] used the practical bit rate, with the addition of the “written symbol rate”, which also claims to account for error correction.

To determine the variety of metrics in use, we conducted a literature review in Web of Science, combining the keywords “*brain-computer interface*” and “*communication*”. The search was limited to English communications in peer-reviewed journals dating between January 2005 and October 2011. Articles were included if they described the performance of synchronous BCIs used by human participants for communication. According to these criteria, 65 articles were retained and included in the appraisal.

Within these 65 articles, 10 different combinations of metrics were used to describe BCI-based AAC performance. These combinations included:

- (1) Accuracy [12–46]
- (2) Accuracy and information transfer rate (ITR): [10], [11], [47–59]
- (3) Information transfer rate (ITR): [60–66]
- (4) True and false positives: [67]
- (5) Accuracy and written symbol rate (WSR): [68]
- (6) Accuracy and speed: [69]
- (7) Accuracy and mutual information: [70]
- (8) Accuracy and number of errors: [71]
- (9) Accuracy and selections per minute: [35]
- (10) Accuracy, bit rate, selections per minute, output characters per minute: [3]

The distribution of metrics used to describe BCI performance in these articles is presented in **Figure 9**.

Not only are the metrics presented in **Figure 9** quite varied, many of them are also incommensurate. Additionally, many are based on digital communication theory and thus on assumptions that do not necessarily hold for human-based communication. Human-based communication is a dynamic process which has been measured in the AAC field using speed, efficiency and accuracy [72]. Metrics that do not include the same variables from this set cannot be compared. Furthermore, the omission of a variable in this set gives an incomplete account of the performance of the BCI as a communication system, which is the ultimate goal of BCI-based AAC technology. The large number of incommensurable metrics that are currently used in this field precludes the comparison of the performance of different BCI systems and hinders rapid growth and development.

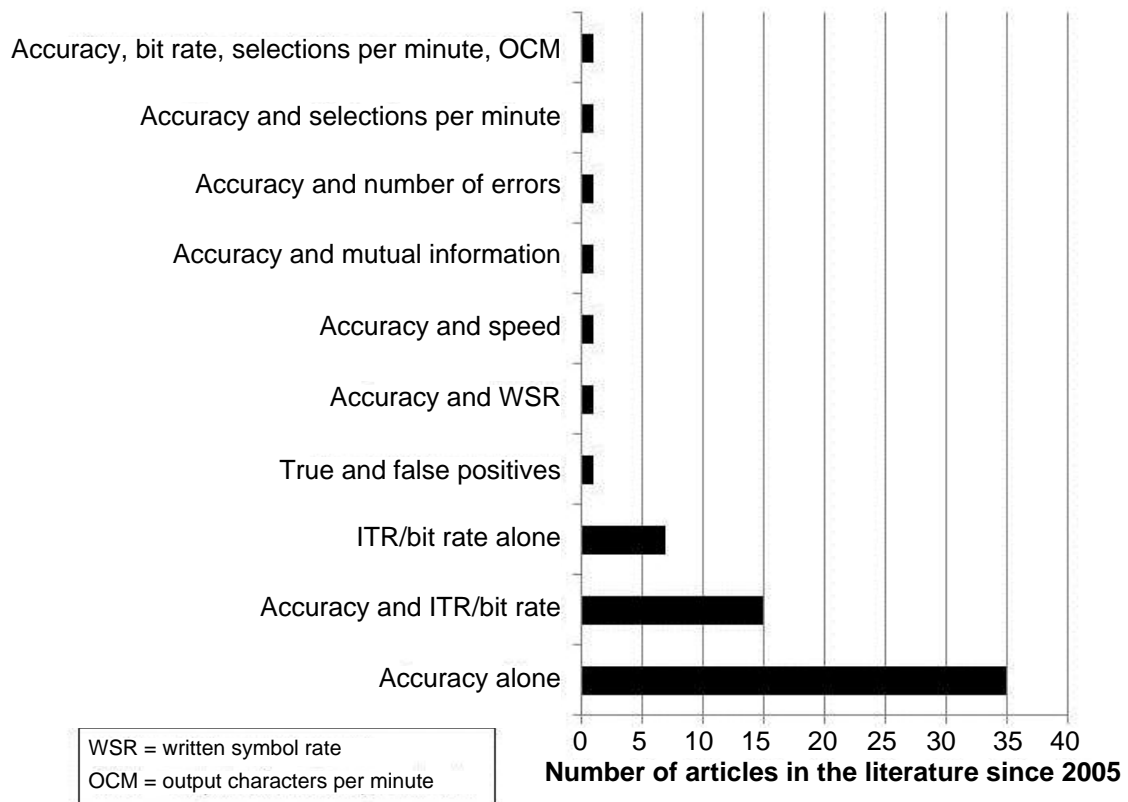


Figure 9: Distribution of the many metrics combinations that have been used in the literature to report performance of a BCI used by human participants for communication from January 2005 – October 2011.

As BCI-based AAC research continues to grow in popularity, there is a pressing need for the acceptance of standardized BCI evaluation metrics that can be used to report

performance in any study using a BCI for AAC. Such metrics would enable the efficient comparison of various BCI components, accelerating the development of a practical, efficient BCI that can be used by individuals with severe motor impairments for the purposes of communication. This manuscript will compare the performance metrics that have been used for Level 1 and Level 2 measurement of BCI-based AAC performance, and recommend a standard metric for each level. Level 3 metrics will not be addressed; the interested reader is referred to the literature regarding measuring assistive technology impact on the user [73–75].

Level 1 Performance Metrics

Types of BCI Control Modules

BCI Control Modules use pattern recognition techniques to translate the electrical signals generated from the brain states of BCI users into selections from variable numbers of discrete output options. The Control Module functions under the assumption that (1) specific mental operations or (2) responses to specific sensory stimuli result in reproducible frequency or event-related potential patterns. Thus, two types of BCI Control Modules can be distinguished: (1) Endogenous Control Modules, which respond to spontaneous control signals from the user (e.g. motor imagery to generate sensorimotor rhythms (SMRs)), and (2) Exogenous Control Modules, which respond to control signals evoked from the user by a stimulus (e.g. event-related potentials such as the P300 response, or visually-evoked potentials (VEPs)) [2]. Effective Level 1 performance metrics should enable comparison within and between both types of BCI Control Modules.

Evaluation Criteria for Level 1 Performance Metrics

We define six criteria (described in detail below) for the evaluation of common Level 1 performance metrics. An effective Level 1 metric would be able to capture the performance of a maximal number of BCI-based AAC systems. While future systems may be created that cannot be measured with existing metrics, we offer the following criteria aimed at measuring a metric's ability to capture performance of existing BCI systems with the intent to also accommodate future BCI systems. The metric should have the ability to capture (1) throughput (*throughput*), (2) the performance of a BCI with a variable number of categorical outputs (*categorical outputs*), (3) unbiased performance (*unbiased*) and (4) the performance of an exogenous BCI when combining data from variable numbers of stimuli (*# stimuli*). Furthermore, the metric should (5) enable prediction of the performance of BCIs built from different combinations of BCI Control

and Selection Enhancement modules (*prediction*), and (6) be accessible to researchers and clinicians from various disciplines working in the field of BCI (*accessibility*).

Throughput: BCI Control Modules must balance a tradeoff between system speed and system accuracy. While accuracy is commonly reported, the time per decision varies widely between different BCIs, and can even be manipulated within the same study through offline analysis. Effective Level 1 metrics must therefore capture system throughput (information per time). Metrics that report throughput can allow direct comparisons of varied BCI types; such metrics also allow comparisons between different configurations of the same BCI, such as those used to optimize parameter settings.

Categorical outputs: In BCI-based AAC systems, the discrete output options selected by the BCI user are either (1) categorical (e.g. letters from the alphabet in a P300 speller) or (2) ordinal (e.g. targets in a one-dimensional SMR-based BCI whose labels indicate their distance from each other). Metrics that are compatible with categorical outputs can be used with ordinal outputs, but the converse is not true; thus, Level 1 metrics must support categorical outputs to allow comparison between varied BCI types.

Unbiased: The reported performance of BCI Control Modules can be biased by two factors. The first factor is a variable number of discrete outcomes. P300-based BCI spellers enable the user to select from many options within a single trial (e.g. 4 options [76], 36 options [77], and 72 options [10]), whereas some mu-rhythm based BCI selection tasks only permit selection between two discrete outputs within a single trial [78]. Chance performance of the BCI Control Module is inversely related to the number of options. The second factor is the marginal distribution of the intended BCI outputs determined by the experiment; in other words, the potential bias that is introduced if a BCI user is instructed to select one output option more frequently than others or if a Control Module preferentially selects one class over others. To enable efficient comparison between different BCI Control Modules, Level 1 metrics must be unbiased by either factor.

stimuli: The previous three evaluation criteria apply to both endogenous and exogenous BCI Control Modules. A fourth evaluation criterion applies only to exogenous BCI Control Modules. Event-related potentials (ERPs) such as the P300 may be used to determine which stimulus, within a large set of possibilities, a BCI user is

attending. Classification accuracy can be improved by averaging results across multiple trials; however, this improved accuracy comes at the cost of an increase in the overall amount of time required to make a selection [79]. To facilitate comparison between BCI Control Modules that use different numbers of presentations of the ERP-evoking stimuli, an effective Level 1 metric should also have the ability to assess performance across a variable number of stimuli presentations.

Prediction: A powerful consequence of using standard, commensurable metrics to report BCI performance is the potential to predict the performance of a BCI system without needing to build and test it. A metric that enables researchers to predict the performance of combinations of BCI Control Modules and Selection Enhancement Modules can save significant time and financial resources; instead of needing to physically build the BCI system of interest and recruit participants to test its performance, such a metric would enable an estimate of performance with a simple calculation.

Accessibility: Finally, an effective metric must be practicably communicable between various research groups, accessible to individuals from various disciplinary backgrounds working in the BCI field. The metric must present BCI performance in a form that is practical for journal articles, and sufficiently simple to be understood by those without engineering expertise.

Common Level 1 Performance Metrics

In light of the six criteria defined above, we present and discuss five common Level 1 performance metrics in this section: (1) error rate or classification accuracy; (2) Cohen's Kappa coefficient; (3) confusion matrix; (4) mutual information; and (5) information transfer rate or bit rate. The discussion is summarized in **Table 8**. It is possible to address the limitations of some metrics through relatively minor adjustments (e.g. in addition to accuracy, one can report time per sequence and time between characters in a P300-speller BCI to enable the derivation of ITR). However, as these metrics are often reported without the information necessary for such conversions, they will be evaluated according to the six criteria under the assumption that no further information about performance is provided. Several other Level 1 BCI performance metrics exist that are typically used to measure continuous BCI output, such as the

correlation coefficient and mean square error [80]. These metrics are often used in SMR-based BCIs, but as they cannot be used with the categorical output generated by some BCI-based AAC systems, they will not be discussed further in this paper.

(1) Error rate or classification accuracy [81]

This metric determines how often the BCI makes a correct selection; in other words, the percentage of total selections that are correct. While it is the most intuitive metric of BCI performance, it does not account for time, often suggesting that BCI performance increases monotonically with time per decision. Furthermore, this metric is biased by the chance performance of AAC configurations with different numbers of discrete outcomes, and assumes the existence of a single accuracy which is uniform across all possible outputs.

(2) Cohen's Kappa coefficient [82], [83]

Cohen's Kappa is a measure of the agreement of two observers; for a BCI-based AAC system, it is used as a measure of agreement between the correct output and the BCI Control Module output. Like classification accuracy, this metric does not account for the time required to make a selection, and does not give a measure of throughput. Unlike accuracy, Cohen's kappa factors in chance agreement, however, it can still be biased by the distribution of samples used to test the Control Module [84].

(3) Confusion matrix [85]

For BCI-based AAC systems, a confusion matrix is a matrix with correct output as rows, BCI Control Module outputs as columns, and the number of occurrences in the intersections. The diagonal therefore represents the number of correct outputs. The confusion matrix does not account for time and thus does not measure throughput. One advantage of this metric is that the relative sums of the rows reveals the frequency of each intended output in the experiment, thus, confusion matrices can show error bias (e.g.

row-column errors in the P300 Speller). The tradeoff for this additional information, however, is the need for enough data to sufficiently populate the matrix. Every entry in the matrix is proportional to a probability density estimate for a particular combination of correct and actual outputs; the number of density estimates that are required thus grows as the square of the number of states. Particularly in P300 experiments where 36 or more possible outputs are typical, this amount of data is rarely available.

Confusion matrices contain more information than other Level 1 metrics. In addition to showing the frequency of each intended output of the experiment, these matrices can show non-uniformities in accuracy between the possible outputs and also error biases, if present. These aspects could make Confusion matrices a powerful tool for prediction. However, Confusion matrices are 2-D when representing the performance of one specific BCI configuration; representing the performance across a varying number of stimulus presentations would require reporting a 3-D matrix, which would be impractical in journal articles. In addition, while the 2-D matrices may be easily reported for SMR-based systems or other BCIs with a small number of total possible outputs, it will be much more difficult for ERP-based spellers or other BCIs with a large number of total possible outputs. (As examples: Farwell and Donchin's 36-class P3 speller would require reporting a matrix with 1296 entries [77]; Townsend's 72-class speller would require reporting 5184 entries [10]). Furthermore, most of the entries are small in value, and therefore difficult to measure accurately. The combination of these factors makes the Confusion matrix inaccessible.

(4) Mutual information [86], [87]

Mutual information is a measure of the overlap between the correct output and the output of the BCI Control Module; it is a measure, in bits, of the throughput of information from the BCI. Since its formula includes marginal and error probabilities, it is robust with respect to experimental and system bias. However, to account for these sources of bias, the calculation of mutual information requires the estimation of the joint statistical distribution of the input and output; the amount of information needed for this estimation scales as the square of the number of total possible outputs of the system,

making it impractical for use in a realistic setting with a BCI with a high number of possible outputs, such as a P300 Speller.

(5) Information transfer rate (ITR) or bit rate [87].

Information transfer rate (ITR), also called bit rate, is another measure of the amount of information passing through a device per unit time. It is derived from mutual information, so it works with categorical outputs and accounts for variable numbers of stimulus presentations. In the derivation, Wolpaw et al. [87] assumed that the probability of error is uniform across all possible outputs, and that errors (when made) are uniformly distributed among the available choices. While the violation of these assumptions can produce unexpected results, ITR has far lower data requirements than Mutual Information; the two are equivalent when the assumptions hold.

Level 1 Performance Metric Recommendations

As illustrated in **Table 8**, none of the metrics that are currently used to report performance of a BCI Control Module satisfy all six criteria of an effective Level 1 metric. The Confusion Matrix is the only metric that enables prediction of the performance of a BCI system; however, its lack of accessibility makes it impractical as a standard Level 1 metric. Mutual information and information transfer rate are the only two metrics that measure throughput. Mutual information satisfies five of the six evaluation criteria, its robustness with respect to bias making it a better metric than information transfer rate (ITR). However, the amount of data required to estimate the joint statistical distributions of the input and output required to calculate mutual information can be very difficult to attain in BCI experiments due to practical considerations such as subject fatigue and motivation. Fortunately, ITR and mutual information are commensurable metrics; in fact, ITR was derived from mutual information. We therefore recommend the use of **mutual information** when bias is expected or deliberately introduced, and the **ITR** approximation in other situations, as the standard metrics to report Level 1 performance of any BCI Control Modules.

Table 8: Comparison of common Level 1 BCI-based AAC performance metrics. Check marks indicate that the metrics fulfills the evaluation criterion.

Metric	Description / Equation	Evaluation Criteria					
		Throughput	Categorical Outputs	# stimuli (exogenous)	Unbiased	Prediction	Accessibility
Accuracy/Error Rate	$Acc = P; Err = (1 - P)$		✓	✓			✓
Cohen's Kappa	$\kappa = \frac{P - 1/N}{1 - 1/N}$		✓	✓			✓
Confusion Matrix	A matrix with intended (true) outputs as rows, actual outputs as columns, and the number of occurrences in the intersections.		✓		✓	✓	
Mutual Information	$I(X;Y) = \sum_Y \sum_X p(x, y) * \log\left(\frac{p(x, y)}{p(x) * p(y)}\right)$ (Note this can be used to measure throughput rate by simply dividing by the time per trial)	✓	✓	✓	✓		✓
Information Transfer Rate (ITR)	$ITR = \frac{1}{c} \left(\log_2(N) + P \log_2(P) + (1 - P) \log_2\left(\frac{1 - P}{N - 1}\right) \right)$	✓	✓	✓			✓

The following abbreviations were used in the above table: **P**: probability of correct selection; **N**: number of choices; **p(x)**: marginal distribution of X; **p(x,y)**: joint distribution of X and Y; **c**: time per selection.

Level 2 Performance Metrics

Types of Selection Enhancement Modules

Three types of Selection Enhancement modules can be defined based on their respective mechanisms for enhancing the logical output they receive from the BCI controller: (1) error correction; (2) rate enhancement; and (3) control state detection.

1. Error correction mechanisms are ways by which either the user or the system can recover from errors. These mechanisms range from providing the user with an option to undo a previous selection, to detection of an “error potential” brainwave that would automatically undo an erroneous selection [88].
2. Rate enhancement mechanisms map the discrete logical selection received from the BCI controller into selections with a larger unit of semantic meaning. Such mechanisms range from populating selection options with communicative signs and symbols, e.g. Blissymbols [89], to enhancing a P300 speller with a word prediction program, which enables users to complete full words in a single selection [3].
3. Control state detection mechanisms monitor the attention of the user, and abstain from making a selection when the user is not paying attention to the BCI, thus preventing selections which are likely to be erroneous [90].

While all three mechanisms operate on different principles, they are each designed with the common purpose of enhancing the effectiveness of BCI-based communication beyond what is possible with a BCI Control Module alone.

Evaluation Criteria for Level 2 Performance Metrics

To develop an evidence-base for the informed choice of Selection Enhancement modules for a BCI, it is imperative to have a metric that enables comparison within and across all three Selection Enhancement mechanisms described in section 3.1. In addition, the metric should be usable with different subject instructions regarding the handling of errors specified in an experimental protocol (e.g. user required to correct errors, users required to ignore errors). Finally, the criteria of prediction and accessibility as defined for Level 1 metrics also apply for Level 2 performance metrics. The accessibility criterion is even more important for Level 2 metrics than for Level 1 metrics, as it is likely that a broader range of disciplines will be interested in Level 2 measures of performance. Thus, a Level 2 metric must be compatible with 1) error correction, 2) rate

enhancement, 3) control state detection, 4) experimental protocol with and without error correction, and should allow 5) prediction, and 6) accessibility.

Common Level 2 Performance Metrics

In the current literature, six Level 2 metrics are used: (1) the Written Symbol Rate, (2) practical bit rate (3) the Extended Confusion Matrix, (4) the system efficiency, denoted " Eff_{SYS} ", (5) Output Characters per Minute and (6) the BCI-Utility Metric. We describe and discuss these metrics with respect to the six criteria presented in section 3.2. The results of the comparison are presented in **Table 9**.

1) Written Symbol Rate [4]

Written Symbol Rate (WSR) is primarily applicable to error correction mechanisms in BCI Selection Enhancement. The formula accounts for the cost of selecting an erroneous character – selecting a backspace, then selecting the correct character – and for the fact that each selection involved in correcting the error is subject to error itself. However, the WSR strictly underestimates system performance, especially for low accuracies, as the formula uses ITR to derive the symbol rate. ITR already includes theoretical error correction; thus WSR accounts for each error twice, making it an invalid measure.

2) Practical bit rate [10]

Like the WSR, practical bit rate is primarily applicable to error correction mechanisms in BCI Selection Enhancement. To determine an ecologically valid metric of performance, the formula adds a penalty of two additional selections for every error incurred, accounting for the same likelihood and making an error during the correcting process as in the original attempt. The formula used to calculate this metric is the same as that used for the BCI-Utility metric (presented below), but is less flexible.

3) Extended Confusion Matrix [6]

The extended confusion matrix (ECM) is an extension of the confusion matrix (described in section 2.3) that accounts for abstentions, or situations where the BCI system deliberately decides not to output a selection. ECM enables the prediction of system performance with classifiers and control interfaces from different studies. However, this requires the collection of sufficient data to provide estimates of each probability of misclassification. Thus, like Mutual Information or Confusion Matrices among the Level 1 metrics, ECM requires more information than is available in many BCI experiments (e.g. ECMs for spellers would include thousands of entries and be impractical to both report and interpret). ECM also does not currently have a mechanism for capturing selection enhancements such as word prediction or symbolic communication.

The problem of inaccessibility could be reduced by reporting aggregate data from all subjects; however, this approach introduces subtle biases into the data. As examples: the backspace option is more likely to be selected by individuals with poor performance; in time limited trials, only participants with good performance will complete the sentence, thus characters appearing earlier in the sentence are likely to show a bias towards poor performance. These subtle factors mean that even aggregate data must be reported and interpreted with caution.

4) Eff_{SYS} [6]

Eff_{SYS} is a measure of the efficiency of a BCI system. Eff_{SYS} is based on ECM, but differs in that it (a) includes calculations for the cost of errors; (b) is a scalar metric, and therefore accessible for publication; and (c) integrates information from the whole system, and thus cannot be used to predict performance of new systems. Eff_{SYS} is designed to account for the fact that different outputs may have different probabilities of correct classification; however, its derivation assumes that the probability of selecting a ‘backspace’ option is equal to the probability of selecting all other outputs. This inconsistency leads to erratic behavior in this metric; if the accuracy of every potential

output of the BCI is not greater than 50%, Eff_{SYS} is zero. This behavior can be corrected by a modification to the formula, which we present as Eff_{SYS}' .

5) Eff_{SYS}'

Eff_{SYS}' is a modification of Eff_{SYS} that accounts for the fact that different outputs may have different probabilities of correct classification. The formula presented in **Table 9** is derived for the condition of a BCI user selecting outputs (e.g. letters) with the option of undoing erroneous selections, and re-selecting the output of choice. This formula allows all outputs but the 'undo' option to have any non-zero accuracy; only the 'undo' option is required to have an accuracy greater than 50%

6) Output Characters per Minute [3]

Output Characters per Minute (OCM) is calculated by dividing the final length of the error-corrected output by the time required to accomplish the task. The metric is intuitive, and has the ability to capture the performance of all three types of selection enhancement modules. However, as currently presented, the metric is not applicable to experimental protocols where errors remain in the final text, and it unduly penalizes system performance in situations where users did not notice an error immediately and continued typing before returning to correct the mistake. Furthermore, OCM is restricted to character-based communication. Although the majority of BCI research has focused on spelling applications, there are a wide variety of AAC systems that take advantage of symbolic or pictorial communication. To improve communication efficiency, BCI-based AAC systems will likely adapt these well-established conventions from the AAC field; the performance of such BCI systems is impossible to capture with OCM.

7) BCI Utility metric [5]

The BCI Utility metric was designed to capture the performance achieved by BCI systems. BCI Utility is able to effectively measure the performance of all three types of selection enhancement modules, and is applicable to experimental protocols with and

without error correction. Dal Seno et al. [5] present several forms of the metric; in **Table 9**, we present the form that is directly comparable to ITR. If none of the presented forms are appropriate, researchers may derive a new form from its basic principles to account for the specific implementation of their BCI-based AAC system. A simple example would be adding a modification for non-uniform accuracy across all possible outputs. Consequently, the BCI Utility metric can be extended to performance enhancements that the authors did not anticipate. Indeed, the Utility metric may also be appropriate for BCIs designed for purposes other than communication.

To further illustrate the differences between these six Level 2 metrics, a comparison is provided in **Figure 10**. Data was collected in a 3-session experiment performed using the methodology of [91]. Briefly, participants ($n = 22$, including 9 with amyotrophic lateral sclerosis) were asked to copy a total of 9 sentences, each 23 characters in length. Participants corrected errors using a backspace option in the BCI. Sentences were excluded if the participant did not complete the full sentence, correcting all errors, within 15 minutes. The 75 sentences with the lowest OCM are reported in the figure. The metrics were calculated from online results using a simple least-squares classifier.

In this dataset, OCM and BCI-Utility differed only on two datapoints (circled). In both cases, the user noticed an error only after typing several correct letters, and had to erase those letters to correct the mistake. Thus, the BCI-Utility represents an estimate of the OCM the system would have achieved without user error. It can also be noted that WSR severely underestimates performance, while ITR overestimates performance (particularly for low accuracies). Eff_{SYS} estimates the performance of 40 of the sentences to be zero. Eff_{SYS}' takes similar values to the non-zero Eff_{SYS} estimates, but is still less accurate than BCI-Utility.

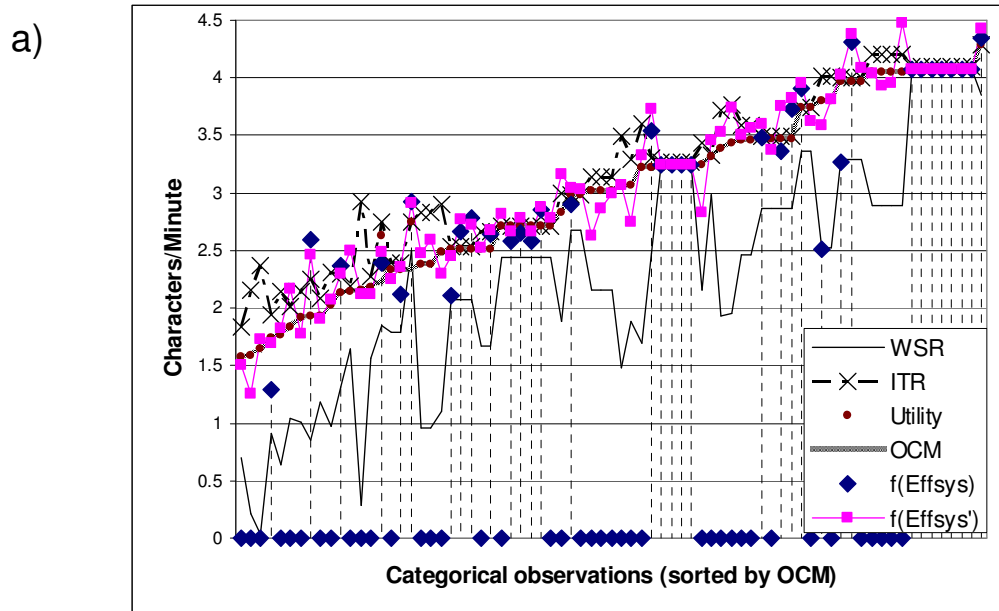
Table 9: Comparison of common Level 2 BCI-based AAC performance metrics. Check marks indicate that the metric fulfils the evaluation criterion.

Metric	Description / Equation	Evaluation Criteria					
		Error Correction	Rate Enhancement	Control State Detection	Experimental Protocol	Prediction	Accessibility
Written Symbol Rate (WSR)	$WSR = \begin{cases} (2SR - 1), & SR > 0.5, \text{ where } SR = ITR * c / \log_2(N) \\ c & , \text{ else} \\ 0 & \end{cases}$	✓			✓		✓
Practical Bit Rate (PBR)	$PBR = \begin{cases} (2P - 1) * \log_2(N), & P > 0.5 \\ c & , \text{ else} \\ 0 & \end{cases}$ The original reference did not report a formula, this was back-calculated from the results	✓			✓		✓
Extended confusion matrix	Confusion matrix, as in Table 8, extended to allow for an extra detected class, "abstention."	✓		✓	✓	✓	
Eff _{sys}	$Eff_{sys} = \frac{1}{L_{CW} * ESC}$ Where L _{CW} is the mean codeword length and ESC is the expected selection cost, as per [6].			✓			✓
Eff _{sys} '	As Eff _{sys} , but with $ESC = \sum_{i=1}^{N_{LA}} \hat{p} * \left(\frac{1}{p(i)} + \left(\frac{1}{p(i)} - 1 \right) \left(\frac{1}{2 * p_b - 1} \right) \right)$ Where L _{CW} is the mean codeword length, N _{LA} is the number of symbols in the logical alphabet, and \hat{p} is the probability of a logical output appearing, as per [6]. The symbols p(i) and p _b are the probabilities of correctly identifying output i and the backspace option, respectively.	✓		✓	✓		✓
Output characters per minute (OCM)	$OCM = \frac{CorrectCharacters}{TimeTaken}$	✓	✓	✓			✓
BCI-Utility Metric	$U = \begin{cases} (2P - 1), & P > 0.5 \\ c & , \text{ else} \\ 0 & \end{cases}$ Note several forms exist in the original reference (no claims are made that this particular form generalizes to all enhancements; this form is characters/minute)	✓	✓	✓	✓		✓

The following abbreviations were used in the above table: **c**: time per selection; **N**: number of choices; **ITR**: information transfer rate; **P**: probability of correct selection.

Level 2 Performance Metric Recommendations

As illustrated in **Table 9**, the BCI-Utility Metric meets the most selection criteria. Unlike the Extended Confusion matrix, it cannot be used to predict performance of different combinations of BCI Control Selection Enhancement modules, which limits the power of this metric. However, it is compatible with all types of Selection Enhancement modules and experimental protocols; it is also accessible. Thus, we recommend the use of the BCI-Utility Metric as the standard to report Level 2 performance of any BCI, enabling the efficient comparison of BCI Selection Enhancement Modules. Level 1 metrics should also be reported in any research involving Selection Enhancement, so that the effect of the Selection Enhancement Module can be clearly seen, and underlying experimental differences due to different BCI Control Modules can be identified.



b)

TO	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	.	<BS>	!	Space	Abstention			
A																																								
B		2																																						
C																																								
D				2																																				
E					4							1																												
F																																								
G								1																																
H									2																															
I										1																														
J																																								
K																																								
L																																								
M																																								
N																1																								
O																	2																							
P																																								
Q																																								
R																																								
S																																								
T																																								
U																																								
V																																								
W																																								
X																																								
Y																																								
Z																																								
1																																								
2																																								
3																																								
4																																								
5																																								
6																																								
.																																								
<BS>									1																													10		
!																																								
Space																																						4		

Figure 10: a) Comparison of each scalar Level 2 metric on data from a P300 copy-spelling task with correction, 75 sentences from 22 users, sorted by OCM. All metrics were converted into characters per minute. Eff_{SYS} and Eff_{SYS}' were calculated on the assumption that the accuracy for untested outputs was equal to the mean of the accuracies of all tested outputs. b) The ECM for the first datapoint presented in a). For space reasons, the 73 ECMs corresponding to the other datapoints are not presented. This illustrates the lack of accessibility of this metric.

Discussion

Recommendations

The continued popularity of research in and development of BCIs has created a pressing need for the adoption of standardized BCI Evaluation Metrics that can be used in any BCI study to report performance. Without such metrics, BCI studies that demonstrate the performance of various BCI Control module or Selection Enhancement module components remain incommensurable, preventing comparisons of BCI function between labs. This severely limits progress toward developing a practical, efficient BCI that can be used for communication by individuals with severe motor impairments. Based on criteria chosen to maximize comparability between all variations of BCI-based AAC systems, we make the following recommendations:

1. Using Mutual Information/ **Information Transfer Rate (ITR)** as the standard metric for reporting Level 1 BCI performance, and the **BCI-Utility Metric** as the standard metric for reporting Level 2 BCI performance.
2. Supplementing these standard metrics with specific metrics typically used for a particular BCI paradigm. For example, in the P300-Speller BCI, the accuracy of the system versus the number of stimulus presentations is typically reported; in this situation, we recommend reporting accuracy versus time, with ITR overlaid on top, as presented in **Figure 11**. Note that if the speller involved any Level 2 enhancements, BCI-Utility should also be reported. Such a graph is not applicable for endogenous BCIs such as those controlled by SMRs, where the BCI user is presented with constant feedback; for these systems, reporting the accuracy and ITR of the system using online settings is sufficient.
3. Reporting both Level 1 and Level 2 metrics in Selection Enhancement module studies. The performance of BCI systems with Selection Enhancement modules is dependent upon the performance of the BCI Control Module as well as the performance of the Selection Enhancement module. Reporting both metrics enables the performance of each module to be assessed independently. Similarly, when BCI systems are eventually assessed at the level of the user, it will be important to report Level 1, Level 2 and Level 3 metrics simultaneously, so that effective comparisons can be drawn between different BCI systems.

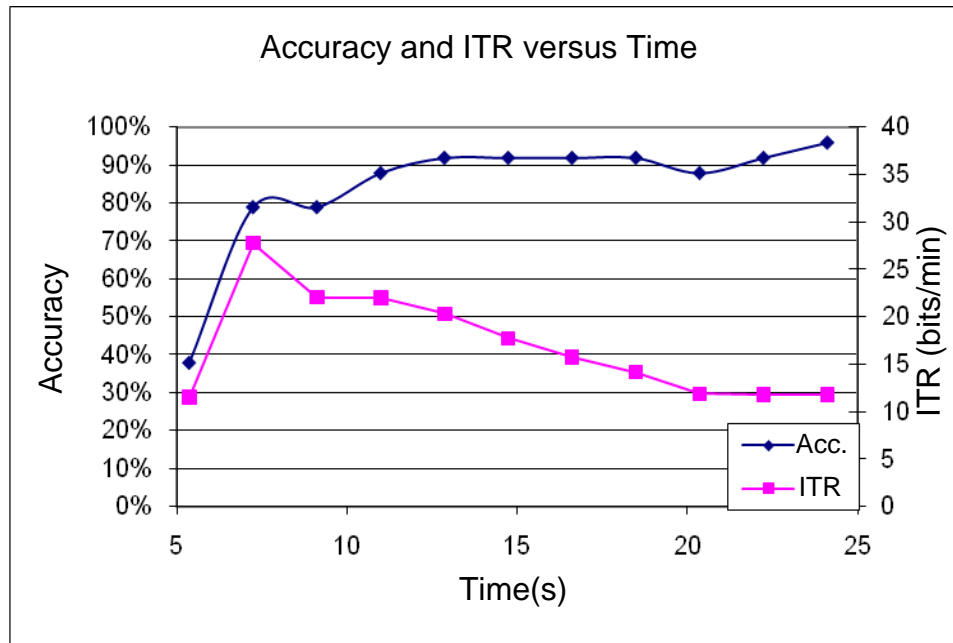


Figure 11: Example of augmented Level 1 performance metric for a P300-speller BCI. Both the ITR and the accuracy are reported with respect to time, enabling comparison with other BCI Control Modules. Note that ITR was calculated including the time between selections.

Limitations

While the recommended metrics enable efficient comparison of most existing BCI systems, they may be limited in their ability to measure the performance of BCI systems that are developed under control paradigms other than those mentioned in this paper. For example, theoretical self-paced BCIs are continuously available to the user, and aware of when the user is engaging with the BCI interface or paying attention to something else (i.e. they support no-control states) [92], [93]. Neither of the recommended metrics would be adequate to measure performance of such a system. Further, throughput is not necessarily a key metric for self-paced BCIs, which incorporate potentially long periods of subject inaction. Other metrics, such as receiver-operator characteristics (ROC) [94] may be more suitable [90]. However, if self-paced BCIs are used to communicate frequently, this would likely be accomplished through a switch scanning system. In this case, the BCI-Utility Metric could be extended through careful measurement of the average time per selection and accuracy of selection achieved.

The recommendations in section 4.1 are specific to BCIs used for AAC. BCIs used for the purposes of mobility or environmental control (e.g. to drive a power wheelchair or to operate a call-button) may use different evaluation criteria to select efficient performance metrics, as system accuracy is often more important than throughput. In such situations, the benefit of each selection (a critical concept in BCI-Utility) may be difficult to define. Therefore, while we recommend reporting the above metrics in any BCI research that includes communication, we do not expect the metrics to capture all aspects of system performance outside the realm of BCI-based AAC systems.

The selection of standard metrics to report Level 1 and Level 2 BCI-based AAC performances is a critical first step in enabling effective comparison of various BCI systems used for communication. The adoption of these metrics as the standard in the field is necessary, but not sufficient, to achieve this goal. A set of guidelines must also be established within the BCI field that detail the appropriate ways of presenting and using each of the metrics recommended in this review. Examples of issues to be resolved in future guidelines are: the Level 1 ITR metric has sometimes been reported with the time between trials artificially removed, which makes comparison of actual user performance

very difficult; the Level 2 BCI-Utility metric will only be effective when comparing symbol-based versus letter-based selections if the relative benefit of symbolic-based communication is provided. This review provides a foundation for the development of such guidelines; future work in this direction is encouraged in order to develop widely-accepted standards that are used to report BCI-based AAC performance using these recommended metrics. Finally, as BCIs transition from laboratory-based technologies to home-based technologies, the development of standard Level 3 metrics will be necessary to facilitate the comparison and development of effective BCI-based AAC systems that can be used by individuals with severe motor impairments in a naturalistic communication setting.

Finally, it is important to recognize that in spite of our best efforts, there are experimental factors that potentially bias comparisons that cannot be corrected for by any single metric. Information about performance is always obtained under a restricted set of parameters that may favor one device over another. Standardizing the metrics used by the BCI field is advantageous to all involved, however, researchers must be vigilant against the biases inherent in each metric to ensure fair comparison of the performance of different BCI systems.

Conclusion

Based on the criteria proposed in this paper, we recommend that when results of BCI-based AAC studies are disseminated: (1) the ITR should be used to report Level 1 BCI performance and the BCI-Utility Metric should be to report Level 2 BCI performance; (2) these metrics should be supplemented by information specific to each unique BCI configuration (see **Figure 11** as an example); and (3) studies involving Selection Enhancement Modules should report performance at both Level 1 and Level 2 in the BCI system. Following these recommendations will enable efficient comparison between both BCI Control and Selection Enhancement modules, accelerating the development of a practical, efficient BCI that can be used by individuals with severe motor impairments for the purposes of communication.

Acknowledgements

This work was supported in part by the National Institute of Child Health and Human Development (NICHD), the National Institutes of Health (NIH) under Grant R21HD054697 and by the National Institute on Disability and Rehabilitation Research (NIDRR) in the Department of Education under Grant H133G090005 and Award Number H133P090008. The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD, NIH, NIDRR or the Department of Education. This material is based upon work supported by the National Science Foundation Graduate Student Research Fellowship under Grant DGE0718128. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NSF.

References

- [1] S. G. Mason and G. E. Birch, "A general framework for brain-computer interface design," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 1, pp. 70–85, Mar. 2003.
- [2] S. G. Mason, A. Bashashati, M. Fatourehchi, K. F. Navarro, and G. E. Birch, "A comprehensive survey of brain interface technology designs," *Annals of Biomedical Engineering*, vol. 35, no. 2, pp. 137–169, Nov. 2006.
- [3] D. B. Ryan, G. E. Frye, G. Townsend, D. R. Berry, S. Mesa-G, N. A. Gates, and E. W. Sellers, "Predictive spelling with a P300-based brain-computer interface: increasing the rate of communication," *International Journal of Human-Computer Interaction*, vol. 27, no. 1, pp. 69–84, 2011.
- [4] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kubler, "An auditory oddball (P300) spelling system for brain-computer interfaces," *Psychophysiology*, vol. 46, pp. 617–625, 2009.
- [5] B. Dal Seno, M. Matteucci, and L. T. Mainardi, "The Utility Metric: A novel method to assess the overall performance of discrete brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 1, pp. 20–29, 2010.
- [6] L. Bianchi, L. R. Quitadamo, G. Garreffa, G. C. Cardarilli, and M. G. Marciani, "Performances evaluation and optimization of brain computer interface systems in a copy spelling task," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 207–216, 2007.
- [7] A. Waller, F. Dennis, J. Brodie, and A. Y. Cairns, "Evaluating the use of TalksBac, a predictive communication device for nonfluent adults with aphasia," *International Journal of Language & Communication Disorders*, vol. 33, no. 1, pp. 45–70, 1998.
- [8] M. J. Scherer, D. B. Gray, L. A. Quatrano, and M. L. Lieberman, "The impact of assistive technology on the lives of people with disabilities," in *Designing and Using Assistive Technology: The Human Perspective*, Baltimore, MD: Paul H. Brookes Publishing Co., 1996, pp. 99–115.
- [9] J. R. Wolpaw, "Brain-computer interfaces as new brain output pathways," *The Journal of Physiology*, vol. 579, no. 3, pp. 613–619, Mar. 2007.
- [10] G. Townsend, B. LaPallo, C. Boulay, D. Krusienski, G. Frye, C. Hauser, N. Schwartz, T. Vaughan, J. Wolpaw, and E. Sellers, "A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109–1120, 2010.
- [11] J. Jin, B. Z. Allison, E. W. Sellers, C. Brunner, P. Horki, X. Y. Wang, and C. Neuper, "An adaptive P300-based control system," *Journal of Neural Engineering*, vol. 8, no. 3, 2011.
- [12] E. V. C. Friedrich, R. Scherer, K. Sonnleitner, and C. Neuper, "Impact of auditory distraction on user performance in a brain-computer interface driven by different mental tasks," *Clinical Neurophysiology*, vol. 122, no. 10, pp. 2003–2009, 2011.
- [13] W. Y. Hsu, C. Y. Lin, W. F. Kuo, M. Liou, Y. N. Sun, A. C. H. Tsai, H. J. Hsu, P. H. Chen, and I. R. Chen, "Unsupervised fuzzy c-means clustering for motor imagery EEG recognition," *International Journal of Innovative Computing Information and Control*, vol. 7, no. 8, pp. 4965–4976, 2011.
- [14] W. Y. Hsu, "EEG-based motor imagery classification using enhanced active segment selection and adaptive classifier," *Computers in Biology and Medicine*, vol. 41, no. 8, pp. 633–639, 2011.
- [15] P. Bobrov, A. Frolov, C. Cantor, I. Fedulova, M. Bakhnyan, and A. Zhavoronkov, "Brain-computer interface based on generation of visual images," *PLoS ONE*, vol. 6, no. 6, p. e20674, Jun. 2011.

- [16] M. Arvaneh, C. T. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1865–1873, 2011.
- [17] S. Ikegami, K. Takano, N. Saeki, and K. Kansaku, "Operation of a P300-based brain-computer interface by individuals with cervical spinal cord injury," *Clinical Neurophysiology*, vol. 122, no. 5, pp. 991–996, 2011.
- [18] D. J. Krusienski and J. J. Shih, "Control of a visual keyboard using an electrocorticographic brain-computer interface," *Neurorehabilitation and neural repair*, vol. 25, no. 4, pp. 323–331, 2011.
- [19] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby, and J. Mattout, "A robust sensor-selection method for P300 brain-computer interfaces," *Journal of Neural Engineering*, vol. 8, no. 1, 2011.
- [20] Y. Q. Li, C. S. Nam, B. B. Shadden, and S. L. Johnson, "A P300-based brain-computer interface: Effects of interface type and screen size," *International Journal of Human-Computer Interaction*, vol. 27, no. 1, pp. 52–68, 2011.
- [21] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk, "Does the 'P300' speller depend on eye gaze?," *Journal of Neural Engineering*, vol. 7, no. 5, p. 056013, Oct. 2010.
- [22] E. W. Sellers, T. M. Vaughan, and J. R. Wolpaw, "A brain-computer interface for long-term independent home use," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 5, pp. 449–455, Oct. 2010.
- [23] Y. Hashimoto, J. Ushiba, A. Kimura, M. G. Liu, and Y. Tomita, "Change in brain activity through virtual reality-based brain-machine communication in a chronic tetraplegic subject with muscular dystrophy," *BMC Neuroscience*, vol. 11, p. 9, 2010.
- [24] D. Zhang, A. Maye, X. R. Gao, B. Hong, A. K. Engel, and S. K. Gao, "An independent brain-computer interface using covert non-spatial visual selective attention," *Journal of Neural Engineering*, vol. 7, no. 1, 2010.
- [25] A. F. Cabrera, D. Farina, and K. Dremstrup, "Comparison of feature selection and classification methods for a brain-computer interface driven by non-motor imagery," *Medical & Biological Engineering & Computing*, vol. 48, no. 2, pp. 123–132, 2010.
- [26] C. Guger, S. Daban, E. W. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger, "How many people are able to control a P300-based brain-computer interface (BCI)?," *Neuroscience Letters*, vol. 462, no. 1, pp. 94–98, 2009.
- [27] K. Takano, T. Komatsu, N. Hata, Y. Nakajima, and K. Kansaku, "Visual stimuli for the P300 brain-computer interface: A comparison of white/gray and green/blue flicker matrices," *Clinical Neurophysiology*, vol. 120, no. 8, pp. 1562–1566, 2009.
- [28] R. Fazel-Rezai and K. Abhari, "A region-based P300 speller for brain-computer interface," *Canadian Journal of Electrical and Computer Engineering-Revue Canadienne De Genie Electrique Et Informatique*, vol. 34, no. 3, pp. 81–85, 2009.
- [29] T. A. Kayagil, O. Bai, C. S. Henriquez, P. Lin, S. J. Furlani, S. Vorbach, and M. Hallett, "A binary method for simple and accurate two-dimensional cursor control from EEG with minimal subject training," *Journal of Neuroengineering and Rehabilitation*, vol. 6, 2009.
- [30] E. V. C. Friedrich, D. J. McFarland, C. Neuper, T. M. Vaughan, P. Brunner, and J. R. Wolpaw, "A scanning protocol for a sensorimotor rhythm-based brain-computer interface," *Biological Psychology*, vol. 80, no. 2, pp. 169–175, 2009.
- [31] A. Sano and H. Bakardjian, "Movement-related cortical evoked potentials using four-limb imagery," *International Journal of Neuroscience*, vol. 119, no. 5, pp. 639–663, 2009.

- [32] F. Guo, B. Hong, X. Gao, and S. Gao, "A brain-computer interface using motion-onset visual evoked potential," *Journal of Neural Engineering*, vol. 5, no. 4, pp. 477–485, 2008.
- [33] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Müller, "The Berlin brain-computer interface: Accurate performance from first-session in BCI-naïve subjects," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 10, pp. 2452–2462, Oct. 2008.
- [34] O. Bai, P. Lin, S. Vorbach, M. K. Floeter, N. Hattori, and M. Hallett, "A high performance sensorimotor beta rhythm-based brain-computer interface associated with human natural motor behavior," *Journal of Neural Engineering*, vol. 5, no. 1, pp. 24–35, 2008.
- [35] F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N. Birbaumer, and A. Kuebler, "An auditory brain-computer interface (BCI)," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 43–50, 2008.
- [36] A. Chatterjee, V. Aggarwal, A. Ramos, S. Acharya, and N. V. Thakor, "A brain-computer interface with vibrotactile biofeedback for haptic information," *Journal of Neuroengineering and Rehabilitation*, vol. 4, 2007.
- [37] X. Liao, D. Yao, and C. Li, "Transductive SVM for reducing the training effort in BCI," *Journal of Neural Engineering*, vol. 4, no. 3, pp. 246–254, 2007.
- [38] Q. Wei, M. Fei, Y. Wang, X. Gao, and S. Gao, "Feature combination for classifying single-trial ECoG during motor imagery of different sessions," *Progress in Natural Science*, vol. 17, no. 7, pp. 851–858, 2007.
- [39] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274–2281, 2006.
- [40] B. Mahmoudi and A. Erfanian, "Electro-encephalogram based brain-computer interface: improved performance by mental practice and concentration skills," *Medical & Biological Engineering & Computing*, vol. 44, no. 11, pp. 959–969, 2006.
- [41] M. Phothisonothai and M. Nakagawa, "EEG-based classification of new imagery tasks using three-layer feedforward neural network classifier for brain-computer interface," *Journal of the Physical Society of Japan*, vol. 75, no. 10, p. 104801–10401–6, 2006.
- [42] N. F. Ince, S. Arica, and A. Tewfik, "Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time-frequency tilings," *Journal of Neural Engineering*, vol. 3, no. 3, pp. 235–244, 2006.
- [43] C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller, "Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG," *Cognitive Brain Research*, vol. 25, no. 3, pp. 668–677, 2005.
- [44] D. R. Burke, S. R. Kelly, P. de Chazal, R. B. Reilly, and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 12–17, 2005.
- [45] B. Kamousi, Z. M. Liu, and B. He, "Classification of motor imagery tasks for brain-computer interface applications by means of two equivalent dipoles analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 2, pp. 166–171, 2005.
- [46] S. P. Kelly, E. C. Lalor, R. B. Reilly, and J. J. Foxe, "Visual spatial attention tracking using high-density SSVEP data for independent brain-computer communication," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 2, pp. 172–178, 2005.
- [47] D. W. Kim, H. J. Hwang, J. H. Lim, Y. H. Lee, K. Y. Jung, and C. H. Im, "Classification of selective attention to auditory stimuli: toward vision-free brain-computer interfacing," *Journal of Neuroscience Methods*, vol. 197, no. 1, pp. 180–185, 2011.

- [48] G. Pires, U. Nunes, and M. Castelo-Branco, "Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis," *Journal of Neuroscience Methods*, vol. 195, pp. 270–281, 2011.
- [49] E. M. Mugler, C. A. Ruf, S. Halder, M. Bensch, and A. Kubler, "Design and Implementation of a P300-Based Brain-Computer Interface for Controlling an Internet Browser," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 6, pp. 599–609, 2011.
- [50] P. L. Lee, J. J. Sie, Y. J. Liu, C. H. Wu, M. H. Lee, C. H. Shu, P. H. Li, C. W. Sun, and K. K. Shyu, "An SSVEP-actuated brain computer interface using phase-tagged flickering sequences: A cursor system," *Annals of Biomedical Engineering*, vol. 38, no. 7, pp. 2383–2397, 2010.
- [51] C. S. Nam, Y. Q. Li, and S. Johnson, "Evaluation of P300-based brain-computer interface in real-world contexts," *International Journal of Human-Computer Interaction*, vol. 26, no. 6, pp. 621–637, 2010.
- [52] C. W. Chen, M. S. Ju, Y. N. Sun, and C. C. K. Lin, "Model analyses of visual biofeedback training for EEG-based brain-computer interface," *Journal of Computational Neuroscience*, vol. 27, no. 3, pp. 357–368, 2009.
- [53] G. Bin, X. Gao, Z. Yan, B. Hong, and S. Gao, "An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method," *Journal of Neural Engineering*, vol. 6, no. 4, 2009.
- [54] D. S. Klobassa, T. M. Vaughan, P. Brunner, N. E. Schwartz, J. R. Wolpaw, C. Neuper, and E. W. Sellers, "Toward a high-throughput auditory P300-based brain-computer interface," *Clinical Neurophysiology*, vol. 120, no. 7, pp. 1252–1261, 2009.
- [55] P.-L. Lee, J.-C. Hsieh, C.-H. Wu, K.-K. Shyu, and Y.-T. Wu, "Brain computer interface using flash onset and offset visual evoked potentials," *Clinical Neurophysiology*, vol. 119, no. 3, pp. 605–616, 2008.
- [56] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 115–125, 2008.
- [57] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance," *Biological Psychology*, vol. 73, no. 3, pp. 242–252, 2006.
- [58] P.-L. Lee, J.-C. Hsieh, C.-H. Wu, K.-K. Shyu, S.-S. Chen, T.-C. Yeh, and Y.-T. Wu, "The brain computer interface using flash visual evoked potential and independent component analysis," *Annals of Biomedical Engineering*, vol. 34, no. 10, pp. 1641–1654, 2006.
- [59] P. L. Lee, C. H. Wu, J. C. Hsieh, and Y. T. Wu, "Visual evoked potential actuated brain computer interface: a brain-actuated cursor system," *Electronics Letters*, vol. 41, no. 15, pp. 832–834, 2005.
- [60] I. Volosyak, D. Valbuena, T. Luth, T. Malechka, and A. Graser, "BCI demographics II: How many (and what kinds of) people can use a high-frequency SSVEP BCI?," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 3, pp. 232–239, 2011.
- [61] H. Takahashi, T. Yoshikawa, and T. Furuhashi, "Error control for performance improvement of brain-computer interface: Reliability-based automatic repeat request," *IEICE Transactions on Information and Systems*, vol. E94D, no. 6, pp. 1243–1252, 2011.
- [62] I. Volosyak, "SSVEP-based Bremen-BCI interface-boosting information transfer rates," *Journal of Neural Engineering*, vol. 8, no. 3, 2011.
- [63] Y. T. Wang, Y. J. Wang, and T. P. Jung, "A cell-phone-based brain-computer interface for communication in daily life," *Journal of Neural Engineering*, vol. 8, no. 2, 2011.
- [64] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Mueller, and G. Curio, "The non-invasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects," *Neuroimage*, vol. 37, no. 2, pp. 539–550, 2007.

- [65] A. S. Royer, M. L. Rose, and B. He, "Goal selection versus process control while learning to use a brain-computer interface," *Journal of Neural Engineering*, vol. 8, no. 3, 2011.
- [66] R. J. Vlek, R. S. Schaefer, C. Gielen, J. D. R. Farquhar, and P. Desain, "Sequenced subjective accents for brain-computer interfaces," *Journal of Neural Engineering*, vol. 8, no. 3, 2011.
- [67] T. Solis-Escalante, G. Müller-Putz, C. Brunner, V. Kaiser, and G. Pfurtscheller, "Analysis of sensorimotor rhythms for the implementation of a brain switch for healthy subjects," *Biomedical Signal Processing and Control*, vol. 5, no. 1, pp. 15–20, Jan. 2010.
- [68] Y. Liu, Z. T. Zhou, and D. W. Hu, "Gaze independent brain-computer speller with covert visual search tasks," *Clinical Neurophysiology*, vol. 122, no. 6, pp. 1127–1136, 2011.
- [69] D. J. McFarland, W. A. Sarnacki, G. Townsend, T. Vaughan, and J. R. Wolpaw, "The P300-based brain-computer interface (BCI): Effects of stimulus rate," *Clinical Neurophysiology*, vol. 122, no. 4, pp. 731–737, 2011.
- [70] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "A fully on-line adaptive BCI," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1214–1219, 2006.
- [71] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based brain-computer interface," *Behavioral and Brain Functions*, vol. 6, no. 28, 2010.
- [72] A. Ratcliff, B. A. Sutton, and M. Lehman, "Metrics for comparing three word-based software programs used for augmentative and alternative communication," *Augmentative and Alternative Communication*, vol. 25, no. 3, pp. 176–186, Jan. 2009.
- [73] J. W. Jutai, M. J. Fuhrer, L. Demers, M. J. Scherer, and F. DeRuyter, "Toward a taxonomy of assistive technology device outcomes," *American Journal of Physical Medicine and Rehabilitation*, vol. 84, pp. 294–302, 2005.
- [74] M. J. Scherer, *Living in the State of Stuck: How technology impacts the lives of people with disabilities*. Cambridge, MA: Brookline Books, 1993.
- [75] M. J. Scherer, "Outcomes of assistive technology use on quality of life," *Disability & Rehabilitation*, vol. 18, no. 9, pp. 439–448, 1996.
- [76] E. W. Sellers and E. Donchin, "A P300-based brain-computer interface: Initial tests by ALS patients," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 538–548, 2006.
- [77] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.
- [78] J. R. Wolpaw, "An EEG-based brain-computer interface for cursor control," *Electroencephalography and Clinical Neurophysiology*, vol. 78, no. 3, pp. 252–259, 1991.
- [79] D. McFarland and J. Wolpaw, "EEG-based communication and control: speed-accuracy relationships," *Applied Psychophysiology and Biofeedback*, vol. 28, no. 3, pp. 217–232, 2003.
- [80] Y. Gao and M. H. Black, "A quantitative comparison of linear and non-linear models of motor cortical activity for the encoding and decoding of arm motions," 2003.
- [81] B. Blankertz, K. Muller, G. Curio, T. M. Vaughan, G. Schalk, J. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroder, and N. Birbaumer, "The BCI Competition 2003: Progress and perspectives in detection and discrimination of EEG single trials," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1044–1052, 2004.

- [82] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp. 37–46, 1960.
- [83] A. Schlogl and F. Y. Lee, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *Journal of Neural Engineering*, vol. 2, no. 4, pp. 14–22, 2005.
- [84] K. Gwet, "Inter-rater reliability: Dependency on trait prevalence and marginal homogeneity," *Statistical Methods for Inter-Rater Reliability Assessment Series*, vol. 2, pp. 1–9, 2002.
- [85] B. Obermaier, C. Neuper, C. Guger, and G. Pfurtscheller, "Information transfer rate in a five-classes brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 9, no. 3, pp. 283–288, 2001.
- [86] A. Schlogl, C. Neuper, and G. Pfurtscheller, "Estimating the mutual information of an EEG-based brain-computer interface," *Biomedizinische Technik*, vol. 47, no. 1–2, pp. 3–8, 2002.
- [87] J. Wolpaw and H. Ramoser, "EEG-Based Communication: Improved Accuracy by Response Verification," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 326–333, 1998.
- [88] A. Buttfield, P. W. Ferrez, and J. R. Millan, "Towards a robust BCI: Error potentials and online learning," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 14, no. 2, pp. 164–168, 2006.
- [89] B. Baker, "Minspeak: A semantic compaction system that makes self-expression easier for communicatively disabled individuals," *Byte*, vol. 7, pp. 186–202, 1982.
- [90] D. Zhang, Y. Wang, X. Gao, B. Hong, and S. Gao, "An algorithm for idle-state detection in motor-imagery-based brain-computer interface," *Computational Intelligence and Neuroscience*, 2007.
- [91] D. E. Thompson, J. J. Baker, W. A. Sarnacki, and J. E. Huggins, "Plug-and-play brain-computer interface keyboard performance," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, 2009, pp. 433–435.
- [92] S. G. Mason, J. Kronegg, J. Huggins, M. Fatourechi, and A. Schlogl, "Evaluating the performance of self-paced brain-computer interface technology." 2006.
- [93] R. C. Panicker, S. Puthusserypady, and Ying Sun, "An asynchronous P300 BCI With SSVEP-based control state detection," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1781–1788, Jun. 2011.
- [94] A. Schlogl and P. Anderer, "Artefact detection in sleep EEG by the use of Kalman filtering," 1999.

Chapter 6

Estimating and Predicting BCI Accuracy

Cursory study of the previous chapter will reveal that nearly every metric requires a value for BCI accuracy. For some metrics, this value is required per class (e.g., for a P300 Speller using a 6x6 matrix, 36 accuracy values would be required to calculate the mutual information metric).

Unfortunately, one cannot simply measure accuracy. From a statistical perspective, accuracy is an unknown parameter that can only be observed through a Bernoulli process (the outcome of which is a binomial random variable). While estimations can be made from even small datasets, the confidence bounds are large unless a large number of trials are obtained. Since BCIs are quite slow, this requires longer experiment lengths, which can be inconvenient and uncomfortable for participants.

An example will illustrate the severity of the problem. Based on our experience, two hours is a cut-off point at which most participants lost interest, and about half an hour is required for setup. Allowing time for rest periods, a participant whose BCI speed is 5 selections/minute could complete perhaps 300-400 selections in an experiment. A participant whose BCI speed is 2 selections/minute might only be able to complete 120-160 selections. Most BCI experimental designs involve comparing performance under multiple conditions, and an accuracy estimate will be required per condition; this effectively divides the number of selections available by the number of conditions.

The most favorable scenario, a fast BCI user in an experiment with only two conditions, thus results in a maximum of 200 selections per condition. The width of the confidence bounds depend on the accuracy achieved, but range from 1-14% (at an observed accuracy of 90%, the confidence bounds are [85%, 94%], nearly 10% wide). For a modest worst-case scenario, a slow BCI user with even 4 experimental conditions, the bounds range from 10%-37%. An observed accuracy of 90% produces a 95% confidence interval of [73.5%,93.9%], covering over 20% of the possible accuracy space.

If estimates are required per class, the situation is untenable. Best-case performance in even a single-condition experiment allows only ten observations per class (measurement granularity of 10%, confidence bounds cover 30-60% of the accuracy space). Worst-case performance in a single-condition experiment allows only four observations per class (measurement granularity of 25%, confidence bounds cover 60-85% of the accuracy space). Considering that a single-condition experiment doesn't have an experimental variable, this situation is unacceptable.

The following chapter is in submission to *IEEE Transactions in Neural Engineering and Rehabilitation*, with Jane E. Huggins as senior author and Seth Warschausky as co-author. Due to the venue of publication, more time is spent on method development than motivating the need for performance prediction. The introduction thus focuses on a particular aspect of EEG signals that will be exploited in the method proposed. However, the primary result is a method to estimate accuracy on small datasets. Confidence intervals will not be calculated, but it will be shown that the method produces estimates more correlated to daily accuracy than the observed accuracy on small datasets.

Introduction

Brain-computer interfaces (BCIs) are commonly defined as a technology that allows communication and control without requiring muscle movements by the user [1]. As such, BCIs offer hope of communication for people with the most severe muscle impairments, such as locked-in syndrome.

One of the most common BCIs researched today is the P300 Speller, first proposed by Farwell and Donchin [2]. The P300 or P3 Speller is so named because it utilizes the P300 response, a positive deflection approximately 300 ms post-stimulus that is evident in electro-encephalogram (EEG) recordings. In most modern methods, some form of automated data mining or machine learning is used. As a consequence, several event-related potentials (ERPs), including the P300, can be used in the classification process [3], [4]; accordingly, the system will be referred to herein as the ERP-Speller.

In this paradigm, the brain activity of users is recorded while the users are presented with a series of stimuli, often flashing rows and columns in a grid of letters. The system must then identify which stimulus the user is attending, by determining which brain responses contain the indicative ERPs. In the signal processing community, this process is known as classification. Particularly in EEG, the signal-to-noise ratio is so low that ERPs are buried within noise and difficult to detect or study from single flashes; single-response measurements are of interest in neuroscience and psychophysiology as well as BCI (see, for example, [5–8]). In most BCI systems, however, each stimuli is presented multiple times before making a decision. Signal averaging is then performed, either explicitly prior to classification, or implicitly through averaging the classifier scores (mathematically equivalent for linear classifiers).

Signal averaging is also commonly used in the neuroscience and psychophysiology fields, where ERPs such as the P300 have been and continue to be studied in detail – a Web of Science search for "P300" and "ERP" returns over 1900 results, 29 of them in the first four months of 2012. One aspect of ERPs which has been acknowledged and studied in these bodies of literature, but is nearly untouched in the BCI research which relies on ERPs, is that of latency jitter, also known as latency variation. Latency jitter is when the latency, the lag between stimulus and ERP, is not constant between trials.

Despite its name, the P300 response does not always appear precisely 300 ms post-stimulus. Latency has been shown to correlate with age, cognitive ability, and other factors (for reviews, see [9], [10]). However, since P300-based BCIs are typically trained with data from each user, variations between users are less important for BCIs than within-user variations. Even under tightly controlled conditions, such within-user variation is known to exist [11], [12]. Large deviations are observed when attention is divided between two tasks, with the magnitude of the latency change varying with the perceptual difficulty of the second task [10]. BCI use, particularly on-line, in-home use such as composition of text, could be accompanied by several secondary tasks requiring varying amounts of attentional and perceptual resources. Additionally, BCI systems often disregard the criteria recommended for clinical ERP measurement. For example, [13] recommends visual stimuli be presented one to two seconds apart, whereas in BCI

systems several stimuli are presented per second. Many of the differences between ERP laboratory studies and BCI system use can be expected to increase latency jitter or otherwise have undesirable effects on signal quality. Latency jitter is far more troubling than amplitude variations for systems relying on signal averaging, as noted in [2]. For a simulated example of the difference, see **Figure 12**.

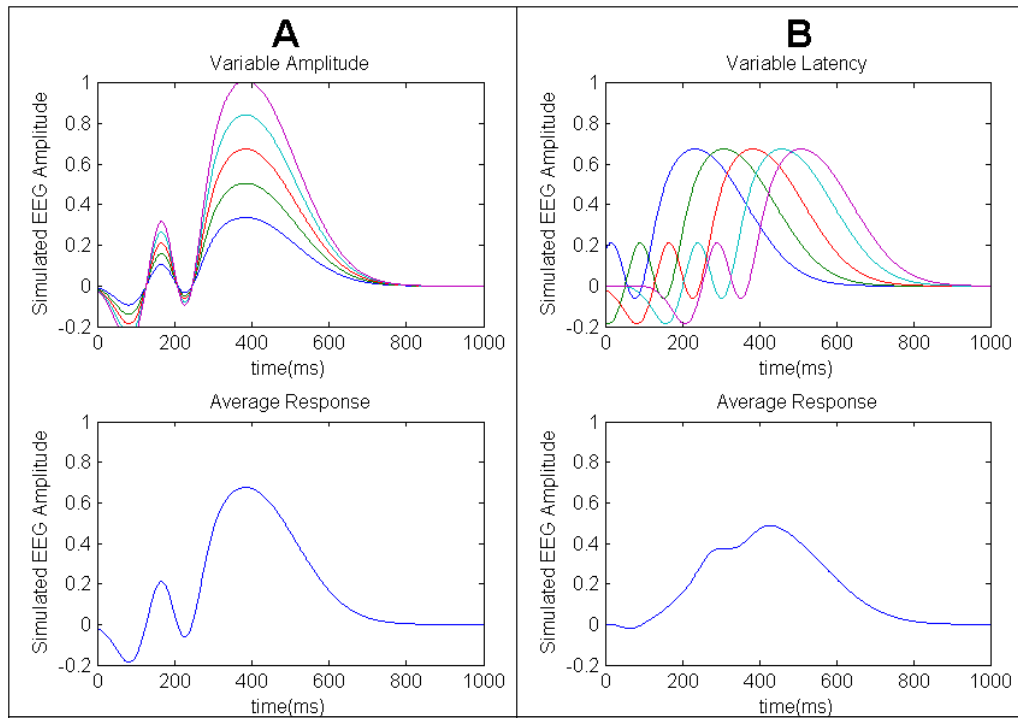


Figure 12: Simulated P300 responses demonstrating the effects of amplitude variation and latency jitter on the average response. The amplitude and latency of the response were varied by identical proportions. The distortion in the average is clearly visible when latency jitter is present.

Yet latency variation is nearly unstudied in BCIs; the only extant work appears to be Markazi, et al. [14], in which latency was adjusted in the time domain before wavelet filtering. We therefore investigated the role that latency jitter plays in BCI operation. We present a new method of measuring latency jitter, called "classifier-based latency estimation" or CBLE, that exploits the time sensitivity of the classifiers used in BCI research. The following is an initial report of our findings, including the results from the new method, which will be shown to be a useful way of estimating and predicting BCI accuracy.

Methods

Experimental Design

The data were collected according to the methodology presented in Thompson, et al. [15]. Briefly, each participant completed three sessions, on three separate days, using our row-column ERP-Speller. On the first visit, participants were asked to copy a 19-character phrase ("THE QUICK BROWN FOX"), which was used to train the online classifier. The training was performed using 15 intensifications of each row and column; the number of intensifications during the testing set was determined per participant based on training accuracy, varying from 3 to 15 with a mean of 6.7.

The five-character word "JUMPS" was used as an online test. An accuracy of 80% was required to continue the experiment; if an accuracy of 60% or less was observed, training was repeated until 80% was achieved. If 80% accuracy could not be achieved within an acceptable time, the participant was compensated for their time but not invited to continue the study. This test was performed because our protocol included the correction of errors, which requires greater than 50% BCI accuracy.

Once the desired accuracy was achieved, participants were asked to copy three 23-character sentences, correcting all errors using a backspace option in the BCI. Each sentence was typed into a different device using the BCI and our Multi-purpose BCI Output Device [16]. When using one device, participants were also asked to select a "speak" option from the matrix after correctly completing the sentence.

The second and third sessions were identical to the first, with the exception that the initial 19-character training was not performed, and the classifier from the first session was used. Device order was counterbalanced both within and between participants.

Participants

These analyses were conducted on data from 31 of the first 32 participants to successfully complete an ongoing study of BCI applications. Twenty participants (ages 18-79, mean 45.3) were recruited from the community as able-bodied controls; the data

from three of these users were previously reported in [15]. Ten participants (ages 45-78, mean 61.3) with Amyotrophic lateral sclerosis (ALS) were recruited from an ALS outpatient clinic. One participant (age 19) with muscular dystrophy (MD) was recruited from an MD outpatient clinic. Only the three users from [15] had previous BCI experience, the rest of the participants had not used a BCI prior to this study.

The excluded participant, an individual with ALS, lost track of his place in the sentence several times during the experiment. As such, the data is unusable, as his intent cannot be unambiguously determined.

All activities were performed under the review of our Institutional Review Board (IRB) and conformed to all institutional guidelines on human subject research. All participants gave informed consent and were offered compensation for their time. Several of the participants with ALS refused to accept compensation.

Classifier-Based Latency Estimation

Our new method works by exploiting the classifier's sensitivity to latency variability. The classifier is applied at different time shifts, and the shift that corresponds to the maximum score from the classifier is defined as the latency shift for the current observation. Then the statistical variance of this shift feature (among attended flashes) is defined as the classifier-based latency estimate (CBLE).

This new method can be considered a generalization of the Woody & Nahvi technique to align asynchronous responses [17]. Where Woody & Nahvi [17] used the statistical cross-correlation to find the optimal latency shift between a response and a template, our method uses a classifier score. Any linear classifier, which can be expected to show significant vulnerability to latency variability, should be suitable for this technique. Non-linear classifiers may vary in their vulnerability to latency variability; the benefit of this technique may be reduced for some non-linear classifiers. Note that if the mean response to attended stimuli is used as the classifier weights, the first iteration of the original Woody & Nahvi method is recovered.

To confirm that this method is relatively classifier-independent, this study will incorporate the use of two classifiers – one based on least-squares (LS), and another on stepwise linear discriminant analysis (SWLDA). Throughout this chapter, the subscripts

LS and SW will refer to values as calculated from least-squares and SWLDA-based data, respectively.

Data Characteristics

EEG data was collected using a g.USBamp from Guger Technologies, which sampled at 256 Hz after hardware filtering. The data was then broken into epochs consisting of the first 800 ms after each stimulus. These epochs were decimated by a factor of 13 using a moving-average-and-downsample operation, reducing the number of features to prevent overfitting of the classifiers. The classifier weights were upsampled by the same factor rather than applying the same operation to all data to be classified.

CBLE estimates were calculated by applying the classifier over the range [-100, 900] ms, thereby allowing the estimates to range from approximately -100 ms to +100 ms, in steps of approximately 4 ms. Due to rounding and the decimation operation, the range was not perfectly symmetrical around zero offset.

Hypotheses and Data Analytic Approach

We hypothesized that CBLE would correlate significantly with BCI accuracy. Initially, we examined the association between CBLE variability and BCI accuracy by computing the Pearson correlation coefficient. CBLE and BCI accuracy were both calculated on a per-sentence basis, including data from all days and users.

Using longitudinal data elements, we then investigated CBLE as a predictor of BCI performance. We hypothesized that the variance of CBLE on the five-character word "JUMPS" would be a good predictor of the first day's performance. This hypothesis was tested by comparing the prediction based on CBLE to that based upon accuracy on "JUMPS". We obtained the prediction from CBLE for each user in two steps:

1. Applying a curve fit to $\text{var}(\text{CBLE})$ and accuracy from all other users' full-length sentences (the current user's data were not used in the calculation)
2. Calculating the value of that curve fit at the $\text{var}(\text{CBLE})$ observed in the "JUMPS" data for the current user. This value was used as the predictor for the first day's BCI accuracy.

Statistical significance of the resulting prediction was tested based on the Pearson correlation coefficient. We also calculated confidence bounds on the difference between the two correlation coefficients, using methods presented in [18].

We then hypothesized that the CBLE of training data (the 19-character "THE QUICK BROWN FOX") would be a good predictor of the first day's performance. This hypothesis was tested by comparing CBLE predictions to those based on training accuracy, and also the gold-standard for training data: cross-validation accuracy on the training dataset. The prediction was performed in the same manner as the above test, though the CBLE was calculated on the training data. Confidence bounds were calculated on all pairwise difference between the three methods.

Results

Figure 13 is a scatter plot of the relationship between latency variation and BCI accuracy. As can be seen from the figure, CBLE is highly correlated with online BCI accuracy for both classifiers ($r_{LS} = -0.7451, p_{LS} < 10^{-49}$; $r_{SW} = -0.7044, p_{SW} < 10^{-42}$).

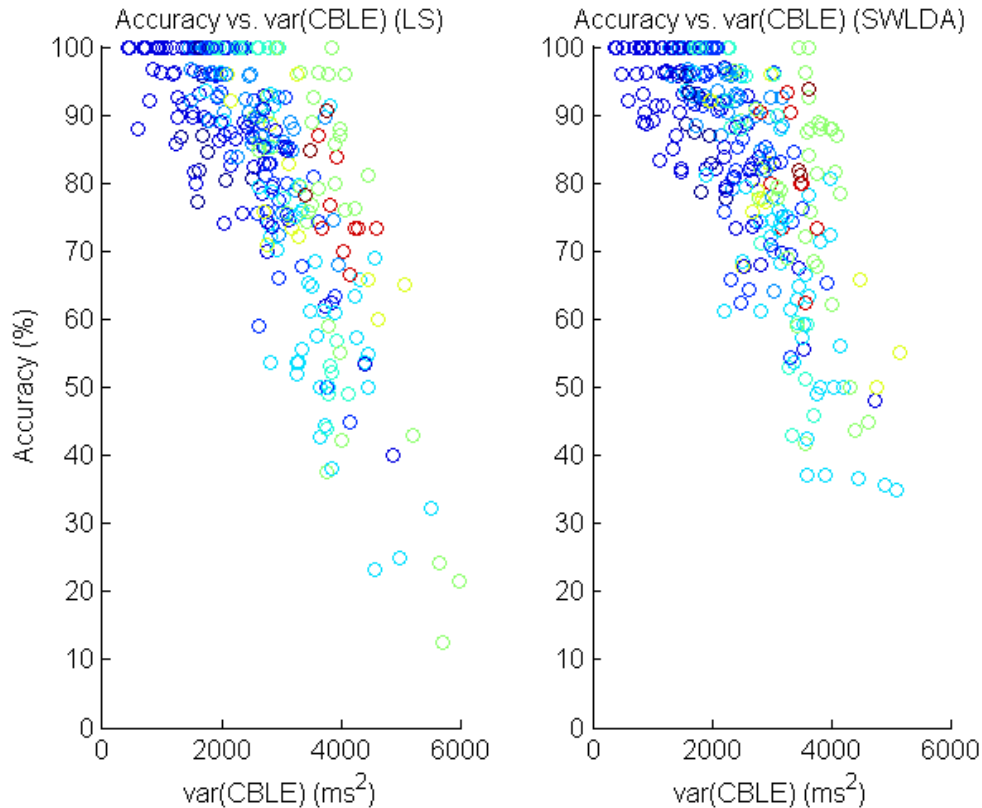


Figure 13: Accuracy plotted against classifier-based latency jitter estimates (CBLE) using two classifiers. Left - results from a classifier using least-squares (LS) regression; right - the popular step-wise linear discriminant analysis (SWLDA) classifier.

Figure 14 shows the outcomes from the first hypothesis test. Predictions based on online accuracy on the five-character word "JUMPS" are not satisfactory ($r_{LS} = 0.15953, p_{LS} = 0.39$; $r_{SW} = 0.45217, p_{SW} = 0.01$). Accuracy estimates based on CBLE variance produce stronger and more significant correlations from the same data ($r_{LS} = 0.74402, p_{LS} < 10^{-5}$; $r_{SW} = 0.68318, p_{SW} < 10^{-4}$). The 95% confidence interval for $\rho_{ACC,LS} - \rho_{CBLE,LS}$ is $[-0.94214, -0.23853]$, indicating that CBLE estimates are significantly more correlated than estimates based on accuracy for LS. The analogous confidence bound for

SWLDA is $[-0.54903, 0.051919]$, which includes zero, so the result, while suggestive, does not quite reach the usual threshold for statistical significance.

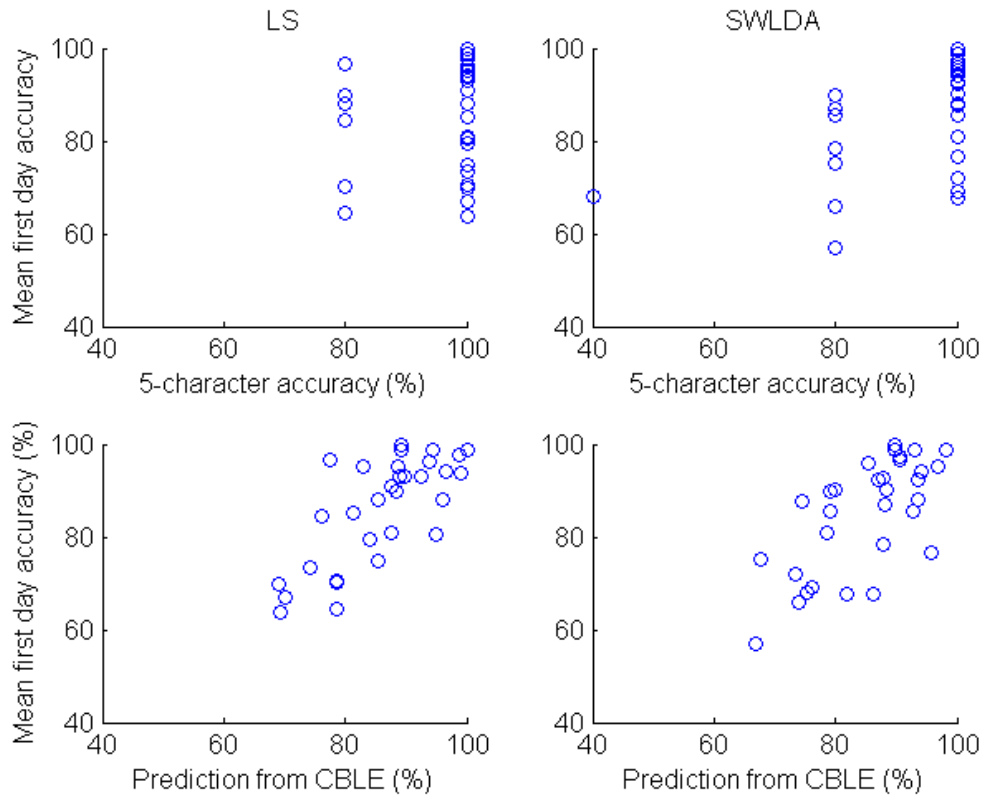


Figure 14: Predicting first-day accuracy based on a 5-character dataset, by classifier type and estimation method. Left - least-squares classification (LS); right - classification using step-wise linear discriminant analysis (SWLDA). Top - prediction based on 5-character accuracy; bottom – prediction based on classifier-based latency jitter estimates (CBLE).

Figure 15 shows the predictions from the second hypothesis test. Training accuracy was correlated with online accuracy ($r_{LS} = 0.20049$, $r_{SW} = 0.26146$), but the relationship was not significant ($p_{LS} = 0.28$, $p_{SW} = 0.16$). Cross-validation performs better ($r_{LS} = 0.53117$, $r_{SW} = 0.54158$), and is a significant predictor ($p_{LS}, p_{SW} < 0.01$). Predictions based upon CBLE were equivalent or better than both ($r_{LS} = 0.64836$, $r_{SW} = 0.63282$), and are significant ($p_{LS}, p_{SW} < 10^{-4}$). The 95% confidence intervals for each pairwise difference are presented in **Table 10**. The confidence intervals indicate that CBLE performs significantly better than training accuracy for both methods; other differences are non-significant, though the difference between predictions from training accuracy and cross-validation is very close to the significance threshold.

Table 10: Confidence Intervals for Correlation Coefficients.

	LS		SWLDA	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
$\rho_{ACC}-\rho_{CBLE}$	-0.803	-0.238	-0.697	-0.067
$\rho_{ACC}-\rho_{XVAL}$	-0.664	0.003	-0.393	0.065
$\rho_{XVAL}-\rho_{CBLE}$	-0.379	0.115	-0.543	0.093

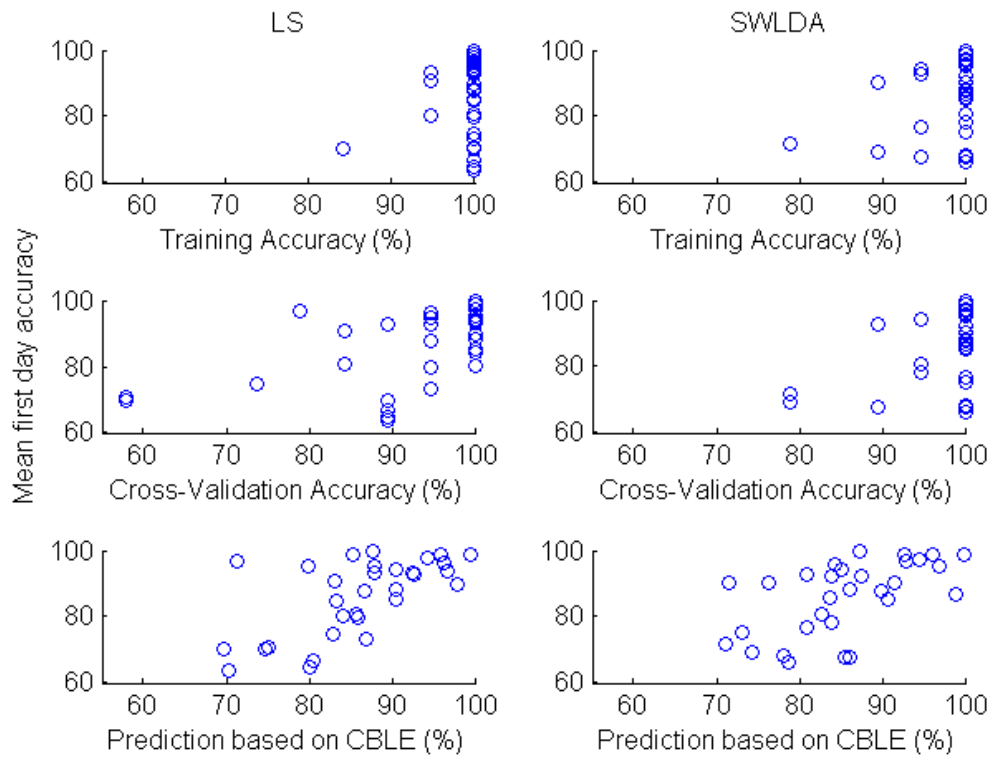


Figure 15: Predicting first-day accuracy based on training data, by classifier type and estimation method. Left - least-squares classification (LS); right - classification using step-wise linear discriminant analysis (SWLDA). Top - prediction based on training accuracy; middle – prediction based on leave-one-out cross-validation accuracy on training data; bottom – prediction based on classifier-based latency jitter estimates (CBLE).

Discussion

The method we have proposed, CBLE, has been shown to be a powerful way to estimate and predict BCI accuracy. The ability to predict performance from training data is of interest from a clinical viewpoint: this technique could be used to reduce the time needed to determine if a BCI could be useful for a particular patient. Alternatively, this method may be useful in indicating if additional training data is required – before performing online feedback. This could have important implications in patient morale and trust of a system.

Of particular interest to the BCI community is the ability to predict performance from small datasets, such as 5 characters (see [19], where both training and test data sets were 5 characters long). Statistically, measuring accuracy is a binomial parameter estimation problem. In addition to granularity issues (a test of 5 characters can produce only 6 possible accuracy estimates), the confidence bounds are extremely large; the smallest 95% confidence bounds are for 0 and 5 correct trials, and cover 52% of the accuracy space. The confidence bounds for the other values cover 71% and 80% of the accuracy space. While [19] is an extreme example of data scarcity, some metrics suggested for use with BCIs (e.g. [20]) require accuracy estimates for every class, increasing the data requirements by at least an order of magnitude. Few BCI studies have the data required for such estimates to be useful. CBLE could be used in larger studies to calculate the more powerful metrics, or in small studies to improve the reliability of accuracy estimates.

Making accurate estimates from these small datasets could improve the value gained from typical BCI experiments. Also, smaller data requirements could be particularly important for the study of non-stationarities in accuracy, such as fatigue effects or distractions. As it stands, state-of-the-art accuracy estimates simply are not precise enough to study these effects on short timescales. This study has shown that CBLE estimates are a better predictor of accuracy from small data sets than existing methods, indicating that the time resolution of such studies could be increased by using CBLE estimates rather than e.g. a windowed 5-character accuracy.

As CBLE variability is so highly correlated with accuracy, it is possible that CBLE might be used to improve BCI performance. Investigations into possible uses are already underway.

Study Limitations

While this method was designed to measure latency jitter of the ERP complex used for classification, the relationship has not yet been verified. If the relationship holds, BCI performance would be correlated to latency jitter, which is a strong argument for investigating latency correction in BCI operation.

The estimates in this work range from approximately 20-80 ms, which is comparable but somewhat larger than the circa 15-50 ms reported in [12] for P3b. However, our task is comparatively complex, so some deviation is to be expected; additionally, CBLE attempts to model a single latency shift for the entire ERP complex, which is a substantial simplification. Our next step is to implement a method for estimating latency of individual components and confirm if the measures of latency jitter correlate.

In this work, CBLE has only been demonstrated to work with LS and SWLDA classifiers. Many classifiers and pre-processing algorithms are currently used in the BCI community, including support vector machines, wavelet transformations, etc. No attempt has been made here to prove that CBLE works with every system ever used with BCI; however there is no obvious reason why it should not work with any system, provided said system is vulnerable to latency jitter. One of the strengths of CBLE is that it can be easily implemented by different groups, typically with a simple "for" loop around existing code. While this is the most computationally expensive approach, in terms of programmer time and effort it would be a reasonable investment for most laboratories. If an offline evaluation indicates that CBLE works with the researchers' preferred methods, more time could be invested to reducing the runtime. This may not be necessary, however, as accuracy estimation is most often done in retrospective offline analysis rather than online operation.

Even without proof that CBLE works with all classifiers, the popularity of LS and in particular SWLDA classifiers in the BCI literature suggests that the technique could be used in several laboratories.

Conclusion

We have developed a new method that we have termed CBLE. This method produces a feature that is significantly and strongly correlated with BCI accuracy in the ERP-Speller. Additionally, CBLE is strongly predictive of daily performance, even from small datasets or datasets that have already been used to train the classifier. The technique should be relatively classifier-independent, and the results were confirmed on two linear classifiers. CBLE may have uses in improving classification, as well as predicting performance from smaller datasets. Additional investigation into methods for correcting for latency jitter are also indicated.

Acknowledgements

This work was supported in part by the National Institute of Child Health and Human Development (NICHD), the National Institutes of Health (NIH) under Grant R21HD054697 and by the National Institute on Disability and Rehabilitation Research (NIDRR) in the Department of Education under Grant H133G090005 and Award Number H133P090008. The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD, NIH, NIDRR or the Department of Education. This material is based upon work supported by the National Science Foundation Graduate Student Research Fellowship under Grant DGE0718128. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NSF.

References

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, Jun. 2002.
- [2] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, Dec. 1988.
- [3] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based brain-computer interface," *Behavioral and Brain Functions*, vol. 6, no. 1, p. 28, May 2010.
- [4] L. Bianchi, S. Sami, A. Hillebrand, I. Fawcett, L. Quitadamo, and S. Seri, "Which physiological components are more suitable for visual ERP based brain-computer interface? A preliminary MEG/EEG study," *Brain Topography*, vol. 23, no. 2, pp. 180–185, 2010.
- [5] C. D'Avanzo, S. Schiff, P. Amodio, and G. Sparacino, "A Bayesian method to estimate single-trial event-related potentials with application to the study of the P300 variability," *Journal of Neuroscience Methods*, vol. 198, no. 1, pp. 114–124, May 2011.
- [6] D. Jarchi, S. Sanei, J. C. Principe, and B. Makkiabadi, "A new spatiotemporal filtering method for single-trial estimation of correlated ERP subcomponents," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 132–143, Jan. 2011.
- [7] D. Jarchi, S. Sanei, H. R. Mohseni, and M. M. Lorist, "Coupled particle filtering: A new approach for P300-based analysis of mental fatigue," *Biomedical Signal Processing and Control*, vol. 6, no. 2, pp. 175–185, Apr. 2011.
- [8] R. Li, A. Keil, and J. C. Principe, "Single-trial P300 estimation with a spatiotemporal filtering method," *Journal of Neuroscience Methods*, vol. 177, no. 2, pp. 488–496, Mar. 2009.
- [9] T. W. Picton, "The P300 wave of the human event-related potential," *Journal of Clinical Neurophysiology*, vol. 9, no. 4, pp. 456–479, Oct. 1992.
- [10] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007.
- [11] J. Polich, "P300 development from auditory stimuli," *Psychophysiology*, vol. 23, no. 5, pp. 590–597, Sep. 1986.
- [12] A. M. Fjell, H. Rosquist, and K. B. Walhovd, "Instability in the latency of P3a/P3b brain potentials and cognitive function in aging," *Neurobiology of Aging*, vol. 30, no. 12, pp. 2065–2079, Dec. 2009.
- [13] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, R. Johnson, G. A. Miller, W. Ritter, D. S. Ruchkin, M. D. Rugg, and M. J. Taylor, "Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria," *Psychophysiology*, vol. 37, no. 2, pp. 127–152, Mar. 2000.
- [14] S. A. Markazi, L. K. Stergioulas, A. Ramchurn, and D. Bunce, "Latency corrected wavelet filtering of the P300 event-related potential in young and old adults," in *2007 3rd International IEEE/EMBS Conference on Neural Engineering*, New York, 2007, pp. 582–586.
- [15] D. E. Thompson, J. J. Baker, W. A. Sarnacki, and J. E. Huggins, "Plug-and-play brain-computer interface keyboard performance," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, New York, 2009, pp. 426–428.

- [16] D. E. Thompson and J. E. Huggins, "A multi-purpose brain-computer interface output device," *Clinical EEG and Neuroscience.*, vol. 42, no. 4, pp. 230–235, Oct. 2011.
- [17] C. Woody and M. Nahvi, "Application of optimum linear filter theory to detection of cortical signals preceding facial movement in cat," *Experimental Brain Research*, vol. 16, no. 5, pp. 455–465, 1973.
- [18] G. Y. Zou, "Toward using confidence intervals to compare correlations," *Psychological Methods*, vol. 12, no. 4, pp. 399–413, Dec. 2007.
- [19] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger, "How many people are able to control a P300-based brain-computer interface (BCI)?," *Neuroscience Letters*, vol. 462, no. 1, pp. 94–98, Sep. 2009.
- [20] J. N. Mak, Y. Arbel, J. W. Minett, L. M. McCane, B. Yuksel, D. Ryan, D. Thompson, L. Bianchi, and D. Erdogmus, "Optimizing the P300-based brain-computer interface: current status, limitations and future directions," *Journal of Neural Engineering*, vol. 8, no. 2, Apr. 2011.

Chapter 7

Conclusion

Brain-Computer Interfaces (BCIs) have unique promise to allow people with total locked-in syndrome to communicate. Since several of the diseases that can cause this syndrome are progressive in nature, many potential BCI users have spent some amount of time with severe movement impairments, but not yet being locked-in. BCIs compete with existing assistive technology (AT) for users in this category, and the typical user will already own an AT device or program. BCIs can either be used as a replacement for these AT systems, or an interface to access them. This dissertation shows that BCIs can in fact be used as an interface, presenting the world's first (and only) plug-and-play BCI, along with experimental evidence that the performance cost associated with this use is small. In order to investigate these performance costs, the dissertation also included work on performance measurement, an unsolved problem of particular interest in the BCI field.

Contributions

A plug-and-play BCI

The Multi-purpose BCI Output Device, or MBOD, is part of the world's first plug-and-play BCI. This device acts as a virtual switch, USB keyboard, USB mouse, or any combination of the three, transforming non-standard BCI outputs into plug-and-play, standards-compatible outputs. Using the principle of input emulation, this device can allow any person using a BCI to access the capabilities of many existing AT solutions, including those that the person might already own. This may help reduce adoption barriers to BCI technology, as the BCI in fact acts as an "interface" instead of a full communication system.

Plug-and-play BCI communication testing

The feasibility of a plug-and-play BCI to interface with existing technology was tested with a large cohort including individuals with amyotrophic lateral sclerosis (ALS). Confidence bounds were established for the performance costs associated with this strategy; the data indicate that the performance costs, if any, are smaller than the reported gains from interfacing with AT in other studies. Taken together, these facts indicate that the MBOD can be used to increase the available communication options for BCI users, with minimal performance cost and likely performance benefits.

BCI control of a wheelchair tilt-in-space system

A novel system to control unmodified wheelchair seating systems was developed in order to determine the feasibility of BCI control of wheelchair tilt angle. A statistically significant difference in accuracy (8 percentage points, $p < 0.05$) was observed in the worst-case distraction of near-constant movement, but the differences in throughput were not statistically significant. The observed difference in performance for user control of tilt was smaller (4 percentage points) but not significant, though the bounds on performance cost are somewhat larger than ideal (about 9 percentage points of accuracy or 30% of corrected throughput). These performance costs, unlike those in the communication task, would not be offset by interaction with AT, indicating that the usefulness of the system may need to be evaluated for individual users.

Comparison of BCI performance metrics

Many incommensurate metrics are currently used in the BCI field, including several that are provably poor measures of achieved performance. To guide the field in the selection of standardized metrics, a criteria-based comparison and critique of many existing performance metrics was performed. In addition to being the basis for the metrics used in earlier sections, this critique addresses a critical issue in current BCI research – incommensurate results from different studies. Currently, researchers in BCI tend to work on one part of the system (e.g. the signal processing algorithms), and results

from studies in other area are often ignored due to the inability to determine which changes have actually increased performance relative to others.

Classifier-based latency estimation

A novel method called classifier-based latency estimation (CBLE) was developed and presented. CBLE was shown to be a powerful tool for estimating and predicting BCI accuracy, which is a necessary step in calculating almost any of the BCI performance metrics currently used. This estimation technique may be particularly useful for studies with small datasets, but could also be used in calculating some of the more powerful BCI metrics. Additionally, preliminary analysis indicates that CBLE has promise in improving BCI classification performance.

Study Limitations

Study limitations are provided in each chapter. However, the work as a whole has one limitation that could be important: while potential users such as individuals with ALS were included in the studies, none of the participants had progressed to the point of having difficulty controlling their eye movements. The population with impaired eye control is of particular interest in BCI research because for almost any individual with intact control of eye movements, AT systems such as eye trackers are faster than existing BCIs [1]. The inclusion criteria would have allowed such individuals to participate in the study, but no volunteers with this condition were found.

Some studies have indicated that eye gaze is involved in the P300 BCI [1], [2], and the use of a separate computer monitor to display the BCI would certainly increase such requirements. However, the alternative display methods mentioned in Chapter 3, such as retinal projection, auditory BCIs, etc., could help eliminate this problem.

Future Work

Additional subject groups for the keyboard replacement experiment

The experiment in Chapter 3 is still ongoing, recruiting participants with muscular dystrophy and cerebral palsy. While the primary question of usability has been answered in ALS, these participants will be asked to follow the same protocol so that group comparisons can be made. Previous results have indicated that BCI performance is lower for individuals with disabilities (see e.g. [3] for ALS), so initial investigations into these populations are warranted. Note furthermore that cerebral palsy, which commonly involves spastic or involuntary movements, may pose unique recording challenges for BCI research.

Movement Artifact Removal for Tilt Study

The wheelchair control study only included data from able-bodied participants, partially because during the experiment it became evident that the data quality was poor. One of the original purposes of the experiment, to study the effect of movement on P300 shape, was not accomplished due to insufficient data, particularly in the presence of substantial movement artifact and other noise. A suitable movement artifact removal algorithm, coupled with other artifact removal methods such as electro-oculogram (EOG) removal, could be used to condition the data to a point where the neuroscience questions of interest might be addressed. Artifact removal algorithms have shown success in removing large amounts of noise from data, see e.g. [4], though whether these algorithms would perform adequately with our lower number of channels is unknown (16 as opposed to 264). Even with such algorithms, a redesign of the experiment may be necessary.

A future study with individuals with ALS or other conditions would also be of interest, but the fact that these individuals mostly use a wheelchair for mobility poses a substantial challenge. Either participants would have to transfer to a chair instrumented for the study, or each participant's chair would have to be instrumented separately; neither is ideal. See Chapter 4's limitations section for further discussion of this issue.

Metric Guidelines

The metrics suggested in Chapter 5 are excellent metrics, but simple adoption of these metrics may not be sufficient to enable comparisons between all studies in the BCI field. For example, even Information Transfer Rate (ITR), which should be easy to calculate from its formula, has been reported both including and excluding the time between trials. The BCI-Utility metric in particular has a few "tuning" parameters and conventions that ought to be agreed upon as a field, particularly in the extension to environmental control. Perhaps the best way forward is to produce a set of guidelines for the field, not individually but at a workshop in the next BCI meeting. I currently am tentatively planning to run such a workshop, and plan on writing up and publishing the results.

Classifier Based Latency Estimation (CBLE)

The CBLE research is my most recent work, and has, in my opinion, the most promise for future research. Several characteristics of the CBLE estimates make me think that they could be useful in improving classification:

- Variance is much higher if calculated from responses to unattended stimuli; in fact choosing the stimuli with minimal variance produces ~50% accurate classification, even ignoring the actual classifier scores.
- The shape of the classifier scores as a function of time shift appears to be different between unattended and attended stimuli. See **Figure 16**.
- The difference between the unadjusted classifier score and the maximum score after latency adjustment tends to be larger for responses not containing a P300.

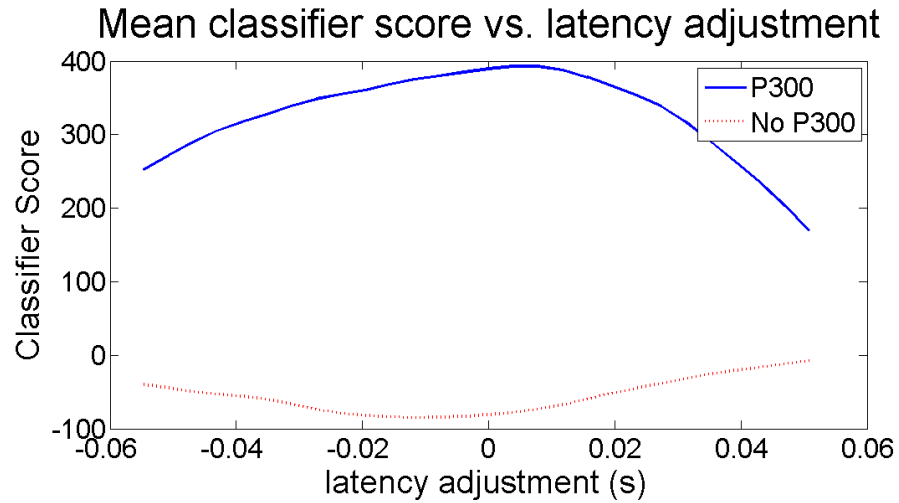


Figure 16: Mean classifier score as a function of latency shift.

In short, I think that further study in this direction will yield improved classification rates. Initial investigations into the use of a second-level classifier are already underway. If a method can be found to use CBLE to improve performance, the relatively classifier-independent nature of the technique gives the research a broader potential impact than many signal processing-based research projects.

References

- [1] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based brain-computer interface," *Behavioral and Brain Functions*, vol. 6, no. 1, p. 28, May 2010.
- [2] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk, "Does the 'P300' speller depend on eye gaze?," *Journal of Neural Engineering*, vol. 7, no. 5, Oct. 2010.
- [3] F. Piccione, F. Giorgi, P. Tonin, K. Priftis, S. Giove, S. Silvoni, G. Palmas, and F. Beverina, "P300-based brain computer interface: Reliability and performance in healthy and paralysed participants," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 531–537, Mar. 2006.
- [4] J. T. Gwin and D. P. Ferris, "An EEG-based study of discrete isometric and isotonic human lower limb muscle contractions," *Journal of Neuroengineering and Rehabilitation*, vol. 9, no. 1, p. 35, Jun. 2012.