GENERALIZED METHODS FOR USER-CENTERED BRAIN-COMPUTER
INTERFACING

GENERALIZED METHODS FOR USER-CENTERED BRAIN-COMPUTER
INTERFACING

By JASKIRET (KIRET) DHINDSA, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of
the Requirements for the Degree of Doctor of Philosophy (Ph.D.)

Descriptive Note

McMaster University DOCTOR OF PHILOSOPHY (2017)

Hamilton, Ontario, Canada (Computational Science and Engineering)

TITLE: Generalized Methods for User-Centered Brain-Computer Interfacing

AUTHOR: Jaskiret (Kiret) Dhindsa

SUPERVISOR: Professor Suzanna Becker

NUMBER OF PAGES: xxi, 224

For my little sister, Roopy

Abstract

Brain-computer interfaces (BCIs) create a new form of communication and control for humans by translating brain activity directly into actions performed by a computer. This new field of research, best known for its breakthroughs in enabling fully paralyzed or locked-in patients to communicate and control simple devices, has resulted in a variety of remarkable technological developments. However, the field is still in its infancy, and facilitating control of a computer application via thought in a broader context involves a number of a challenges that have not yet been met.

Advancing BCIs beyond the experimental phase continues to be a struggle. End-users have rarely been reached, except for in the case of a few highly specialized applications which require continual involvement of BCI experts. While these applications are profoundly beneficial for the patients they serve, the potential for BCIs is much broader in scope and powerful in effect. Unfortunately, the current approaches to brain-computer interfacing research have not been able to address the primary limitations in the field: the poor reliability of most BCIs and the highly variable performance across individuals. In addition to this, the modes of control available to users tend to be restrictive and unintuitive (*e.g.*, imagining complex motor activities to answer "Yes" or "No" questions). This thesis presents a novel approach that addresses both of these limitations simultaneously.

Brain-computer interfacing is currently viewed primarily as a machine learning problem, wherein the computer must learn the patterns of brain activity associated with a user's mental commands. In order to simplify this problem, researchers often restrict mental commands to those which are well characterized and easily distinguishable based on *a priori* knowledge about their corresponding neural correlates. However, this approach does not fully recognize two properties of a BCI which makes it unique to other human-computer interfaces. First, individuals can vary widely with respect to the patterns of activation associated with how their brains generate similar mental activity and with respect to which kinds of mental activity have been most trained due to life experience. Thus, it is not surprising that BCIs based on predefined neural correlates perform inconsistently for different users. Second, for a BCI to perform well, the human and the computer must become a cohesive unit such that the computer can adapt as the user's brain naturally changes over time and while the user learns to make their mental commands more consistent and distinguishable given feedback from the computer. This not only implies that BCI use is a skill that must be developed, honed, and maintained in relation to the computer's algorithms, but that the human is the fundamental component of the system in a way that makes human learning just as important as machine learning.

In this thesis it is proposed that, in the long term, a generalized BCI that can discover the appropriate neural correlates of individualized mental commands is

preferable to the traditional approach. Generalization across mental strategies allows each individual to make better use of their own experience and cognitive abilities in order to interact with BCIs in a more free and intuitive way. It is further argued that in addition to generalization, it is necessary to develop improved training protocols respecting the potential of the user to learn to effectively modulate their own brain activity for BCI use. It is shown through a series of studies exploring generalized BCI methods, the influence of prior non-BCI training on BCI performance, and novel methods for training individuals to control their own brain activity, that this new approach based on balancing the roles of the user and the computer according to their respective capabilities is a promising avenue for advancing brain-computer interfacing towards a broader array of applications usable by the general population.

Acknowledgments

I begin with overwhelming gratitude to my supervisor, Sue Becker, whose wise guidance, compassionate support, and diligent commitment to my success have shaped me into a scientist. She made me feel like I could work on whatever was important to me, and that she would guide me in pursuing my own goals. I owe her much for her continued confidence in me. It is difficult to imagine any way to repay her, but at the very least I aim to one day write a manuscript so flawless that not even Sue will be able to improve on its grammar and writing style. More than just helping me grow professionally, Sue has had an enormous impact on me as a person, and that has immeasurable and unforgettable value to me.

My committee members, Jim Reilly and John Connolly, who have shared with me a wealth of experience and insight into research, academia, and the path before me, also deserve great thanks. The time spent together has been both valuable and enjoyable. Much thanks to both of them for guiding me through the PhD process and for their support over the years. They helped make graduate school a pleasure.

I would like to show appreciation to my friends and fellow lab members, with whom I have had useful talks and fun conversations. Thanks Craig Hutton, Saurabh Shaw, Nadia Wong, Jeff Bruce, Nick Dery, Brandon Aubie, Rory Finnegan and Ranya Amirthamanoharan. I also can't forget the undergraduate students who have been extremely helpful in bringing my studies to completion. Thanks Vlad Drobinin, Dean Carcone, Kristen Marszalek, Braeden Terpou, Gabrielle Herman, and Kyle Gauder.

Without the support of my family, this process would have been much more difficult. Much thanks to mom and dad for their continued love and support over the years. Another big thanks to Neetu for making sure I grew up with a head on my shoulders and for being much more than a big sister. I love you very much.

Ryan Marr has been my best friend for just about my entire life. I wouldn't have been the same person if I hadn't grown up with him, and I doubt I would have studied what I did if it wasn't for his advice during our undergradute days. Guy Groisman and Yale Reinstein have also been extremely important through this process. All three of them made sure I didn't forget that there exists an entire world outside of academia and my work, and they've been with me since my childhood. Thank you.

Finally, I can't begin to thank my partner, Magdalena, enough. Her consistent encouragement, warm companionship, and endless energy carried me from the beginning to the end of these six years of my life. Now we begin a new chapter of our lives together. I'm looking forward to many more years together, and to spending more time with Gabe (thanks for helping me take my mind off work so I could gg this PhD), Raf, Mike, and Renata.

Contents

List of Figures

List of Tables

List of Abbreviations

AA	:	Alpha Asymmetry
ADJUST	:	Automatic EEG artifact Detector based on the Join Use of Spatial and Temporal features
ALS	:	Amyotrophic Lateral Sclerosis
ANFB	:	Adaptive Neurofeedback
BAI	:	Beck Anxiety Inventory
BBB	:	Byrne, Becker, Burgess
BCI	:	Brain-Computer Interface (also, Brain-Computer Interfacing)
BDI	:	Beck Depression Inventory
BDT	:	Boosted Decision Tree
BMI	:	Brain-Machine Interface
BSS	:	Blind Source Separation
CFC	:	Cross-Frequency Coupling
CRLS	:	Conventional Recursive Least Squares
CSD	:	Cross-Spectral Density
CSP	:	Common Spatial Patterns
CV	:	Cross-Validation
DC	:	Divide and Conquer
DRL	:	Driven Right Leg
ECoG	:	Electrocorticography
EEG	:	Electroencephalograph, Electroencephalography, or Electroencephalogram
EMG	:	Electromyograph, Electromyography, or Electromyogram
EOG	:	Electrooculograph, Electrooculography, or Electrooculogram
ERP	:	Event-Related Potential
FAA	:	Frontal Alpha Asymmetry
FASTER	:	Fully Automated Statistical Thresholding for EEG artifact Rejection
FBAR	:	Filter-Bank Artifact Rejection
FBCSP	:	Filter-Bank Common Spatial Patterns
FDR	:	False Discovery Rate
FFT	:	Fast Fourier Transform
fMRI	:	functional Magnetic Resonance Imaging
fNIRS	:	functional Near-Infrared Spectroscopy
FNR	:	False Negative Rate
FPR	:	False Positive Rate
FWE	:	Family-Wise Error
FWHM	:	Full-Width at Half-Maximum
GAA	:	Global Alpha Asymmetry
GLM	:	General Linear Model
HC	:	Hippocampal Cortex
HCI	:	Human-Computer Interface
HDR	:	Hemodynamic Response

Hz	:	Hertz
IAF	:	Individual peak Alpha Frequency
ICA	:	Independent Components Analysis
INVIS	:	Pointing while invisible condition
ITI	:	Inter-trial Interval
ITR	:	Information Transfer Rate
JRD	:	Judgement of Relative Direction
LOO	:	Leave-One-Out
LOSO	:	Leave-One-Subject-Out
LOVO	:	Leave-One-Video-Out
LR	:	Logistic Regression
LRE	:	Logistic Regression with Elastic Net Regularization
LSV	:	Laryngeal Sub-Vocalization, or Laryngeal Sub-Vocal
MEG	:	Magnetoencephalography
MI	:	Mental Imagery
ML	:	Machine Learning
MLP	:	Multilayer Perceptron
MRMR	:	Minimum Redundancy Maximum-Relevance (also appears as mRMR)
MTL	:	Medial Temporal Lobe
NFB	:	Neurofeedback (also appears as NF)
OHQ	:	Oxford Happiness Questionnaire
OVR	:	One-Versus-Rest
PC	:	Parahippocampal Cortex
PCA	:	Principle Components Analysis
PNFB	:	Progressive Neurofeedback
PSD	:	Power Spectral Density
PSS	:	Cohen's Perceived Stress Scale
PT	:	Progressive Thresholding
PTAA	:	Percent-Time with Asymmetric Alpha
QDA	:	Quadratic Discriminant Analysis
QP	:	Question Period
QPC	:	Quadratic Phase Coupling
RBF	:	Radial Basis Function
REF	:	Pointing from reference viewpoint condition
ROT	:	Viewpoint rotation condition
RSC	:	Retrosplenial Cortex
RSP	:	Response Period
SCP	:	Slow Cortical Potential
SMR	:	Sensorimotor Rhythm
SF	:	Spectral Features
SFBA	:	Single Feature of Brain Activity
ST	:	Standard Thresholding
SVM	:	Support Vector Machine
TAA	:	Temporal Alpha Asymmetry

TIM	:	Three-Item Mood Inventory
TMS	:	Transcranial Magnetic Stimulation
VIS	:	Pointing while visible condition
VR	:	Virtual Reality
WHI	:	Waterloo Handedness Inventory
WPLF	:	Weighted Phase-Locking Factor

Declaration of Academic Achievement

This thesis includes five published or submitted manuscripts that address the problem of generalizing and advancing methods for brain-computer interfacing. A list of full papers I have authored and my specific contributions to each are included below.

Chapter 1:
Author: Kiret Dhindsa

Chapter 2:
Author: Kiret Dhindsa

Chapter 3:
Manuscript:
Dhindsa K, Drobinin V, King J, Hall GB, Burgess N and Becker S (2014) Examining the role of the temporo-parietal network in memory, imagery, and viewpoint transformations. *Frontiers in Human Neurosciences* 8:709. doi: 10.3389/fnhum.2014.00709
Comments:
I collaboratively designed the experiment with S.B., J.K., and N.B (S.B., J.K., and N.B. defined the original research question). I independently wrote the experiment software. I led software testing, data collection, and data analysis with assistance from V.D. and N.B. I co-authored the manuscript with S.B., taking the lead on drafting the manuscript.

Chapter 4:
Manuscript:
Dhindsa, K., Carcone, D., & Becker, S. (In Press: 2017). Towards an Open-Ended Brain-Computer Interface: A User-Centred Co-Adaptive Design Approach. *Neural Computation.*
Comments:
I independently developed the research question, designed the experiment, and designed the brain-computer interface. I wrote most of the experimental software independently, and D.C. wrote software modules for displaying instructions and images to participants. I analyzed the data independently and wrote the manuscript with advice and edits from S.B.

Chapter 5:
Manuscript:
Dhindsa, K. & Becker, S. (In Press: 2017). Emotional Reaction Recognition from EEG. (2017). *International Workshop on Pattern Recognition in NeuroImaging (PRNI) 2017.* IEEExplore.

Comments:

I independently developed the research question and designed the experiment in collaboration with S.B. I analyzed the data independently and wrote the manuscript with edits by from S.B.

Chapter 6:
Manuscript:
Dhindsa, K., Gauder, K. D., Marszalek, K. A., Terpou, B., & Becker, S. (Revision Submitted). Progressive Thresholding: Incorporating Shaping and Specificity into Automated Neurofeedback Training.
Comments:
I developed the research question independently and designed the experiment collaboratively with S.B. I developed the idea and software for the core algorithm being tested independently and wrote the experimental software with K.D.G. I analyzed the data independently and wrote the manuscript with edits from S.B.

Chapter 7:
Author: Kiret Dhindsa
Comments:
Experimental design developed in collaboration with Suzanna Becker and John F. Connolly.

Chapter 8:
Author: Kiret Dhindsa

Appendix A:
Manuscript:
Dhindsa, K. (2017). Filter-Bank Artifact Rejection: High Performance Real-Time Single-Channel Artifact Detection for EEG. *Biomedical Signal Processing and Control.* 38: 224-235.
Comments:
This was independent work.

Papers Not Appearing in this Thesis
Manuscript:
Engchuan, W., Dhindsa, K., Lionel, A. C., Scherer, S. W., Chan, J. H., & Merico, D. (2015). Performance of case-control rare copy number variation annotation in classification of autism. *BMC medical genomics*, 8(1), S7.
Comments:
I collaborated with W.E. and D.M. to develop a machine learning strategy for classifying autism from genomic information. I independently performed machine learning analysis and feature level analysis, contributing the neural network results, support-vector machine results, and information theory based feature relevance results. In addition, I wrote the sections corresponding to my contributions.

I also assisted with editing of the manuscript and preparing the final submission, including supplementary data.

Manuscript:

Dhindsa, K., Carcone, D. & Becker, S. (2017). A Brain-Computer Interface Based on Abstract Visual and Auditory Imagery: Evidence for an Effect of Artistic Training. *International Conference on Augmented Cognition 2017*, 313-332. Springer, Cham

Comments:

I independently developed the research question, designed the experiment, and designed the brain-computer interface. I wrote most of the experimental software independently, and D.C. wrote software modules for displaying instructions and images to participants. I analyzed the data independently and wrote the manuscript with edits from S.B. This paper does not appear in the thesis because it is a conference version of the work presented in Chapter 4. The paper presented in Chapter 4 includes additional data and results, making the inclusion of this conference paper unnecessary.

Chapter 1

An Introduction to Brain-Computer Interfacing

1.1   From Brains to Brain-Computer Interfaces

The mammalian brain has evolved over millions of years as an organ capable of controlling complex biological systems, such as the human body. With this remarkable system, we are able to plan and execute multistage goal-oriented tasks, assess our performance, adjust our behaviour, and respond to unforeseen obstacles along the way. The brain, in collaboration with the sensory organs with which it interfaces, enables for us several channels of experiential life, including the senses, emotions, and most profoundly, consciousness. Together, these functions provide feedback to the brain about its environment, its actions, and itself.

Though the brain performs a plethora of survival, social, and cognitive tasks with impressive proficiency, one particular function of the brain stands out as not only having an exceptionally high degree of utility, but as being profoundly powerful. That is its ability to learn. The adaptability of the brain, and thus the organism, is unprecedented in known biology. It is precisely this quality of the brain and the degree to which it is realized in humans that makes possible the spectacle that is the ever-receding boundary of human potential in all areas of life. Yet the power of the brain's ability to learn is not limited to control over one's body and mind. The evolution of the human brain in particular has allowed our species to cross a critical threshold, whereby we have developed arts and sciences which have enabled us to transform our environments, our societies, and ultimately ourselves in ways that are at least partially conscious and directed.

The adaptability of our brains has allowed us to thrive in human-made environments that are in many ways unlike the environments in which the brain originally evolved. Rather than becoming separated from nature, humans have extended its boundaries to include what could previously only be envisioned. We do this primarily by creating new technologies, which in many ways become extensions of ourselves and of nature while magnifying our ability to transform both.

In just the last few decades, revolutionary technological advances have profoundly changed human life. One of the greatest examples of this is the digital computer. The computer empowers the brain by allowing us to compute and manipulate symbolic data beyond our biological capabilities. Combined with the internet, the computer also grants us unparalleled access to information and communications.

As we continue to use our brains to develop information technologies that have exponentially increasing capabilities, neuroscientists and biomedical engineers turn those technologies back towards ourselves in order to gain a deeper and more comprehensive understanding of the very brains which created those technologies. This evolving cycle has recently led us to somewhere unprecedented,

for at around the time the majority of the developed world became connected to the internet, the interdisciplinary field of brain-computer interfacing also began to emerge [1].

Brain-computer interfaces (BCIs) are systems which create direct communication and control pathways between brains and computers. Whether achieved through non-invasive wearable devices or through computer components implanted directly onto the brain, a BCI and its user form a new kind of biocybernetic system. BCIs have been envisioned as clinical tools which can be used to replace and restore lost or damaged functionality for their users [1], whether cognitive [2, 3], motor [4, 5], or communicative [6]. Beyond restoring and replacing damaged functions, BCIs have also been viewed as a pathway towards the enhancement of existing functions [7, 8]. Perhaps in the future, BCIs will enable entirely new functionality for humans. Research into this field has only just begun, and the impact brain-computer interfacing will have on human life can only barely be imagined.

While the kinds of biocybernetic modifications of humans found in science fiction might remain a far-off and inaccurately depicted future, the progress made en route and during tangential explorations of brain-computer interfacing and BCI-related technologies is poised to make a significant impact on human society [9]. Given that the field is still in its infancy, the current state of some BCI applications is interesting and useful enough that there are high expectations for the future. In the clinical setting (see Figure 1.1), BCIs have already been used to enable computer use [10, 11, 12, 13] and wheelchair driving [14, 15] for the severely paralyzed. Research in this field has also led to brain-controlled prostheses for amputees [16], paraplegics [17, 18], and tetraplegics [19], which have even been shown to induce some degree of restorative neuroplasticity [20, 21]. While brain-computer interfacing in non-clinical settings has only recently received significant attention, BCIs have been used to enhance memory encoding [22, 23, 24] and response time [25] in humans, as well as in gaming [26, 27, 28]. We can only speculate on what brain-computer interfacing will bring in the coming decades.

1.2   The Central Theme of this Thesis

Several major breakthroughs in brain-computer interfacing have been made over the last decade. Since BCIs make heavy use of machine learning, modern brain-computer interfaces have been greatly empowered by the rapid advancement of artificial intelligence (AI). In fact, it is not uncommon for brain-computer interfacing to be conceptualized as an intersection of neuroscience and AI. Despite this rapid development, there remain many challenges that must be met before BCIs

(a) Internet Browsing   (b) BCI Wheelchair   (c) Neuroprosthesis
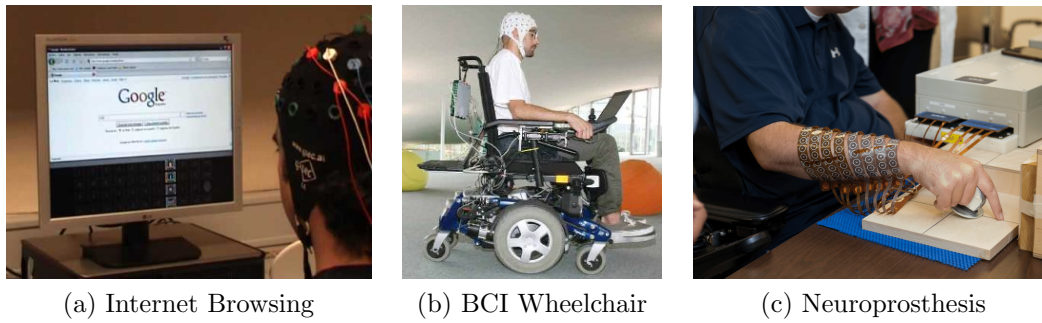
Figure 1.1: Three prominent examples of clinical brain-computer interfaces. **a)**: Using a P300-based BCI to browse the internet (see Section 2.2.3 for more detail on P300-based BCIs). Image reproduced from a publicly available video associated with [11] (link provided in [29]). **b)**: A BCI-controlled wheelchair [30]. **c)**: A tetraplegic patient using a BCI which stimulates muscles based on predicted user intentions in order to grasp an object. Image reproduced with permission from [31].

can be adopted as mainstream tools in broader society. The central issue preventing BCIs from moving beyond the laboratory is the high variability in reliability and usability across users and over time [9].

In this thesis it is argued that a paradigm shift is required to fully address the main limitations in current BCI technologies. Most progress in brain-computer interfacing thus far has been made through advancements in the technical components of a BCI. However, these advancements have had an insufficient impact on the fundamental usability of BCIs for the general population. Significant progress in making BCIs useful beyond highly specialized applications that are fine-tuned for specific users necessitates moving away from thinking of BCI users as system operators in the same way that they are operators of other human-computer interfaces (HCIs), such as a computer mouse or a keyboard. A BCI is a fundamentally different kind of HCI where the user is in fact the central component and not just an external force that inputs commands. Re-framing the relationship between a BCI and its user in this way has two major implications that are explored in this thesis.

The first implication of refocusing the BCI on its user is that the user must be the primary consideration when determining how a BCI is to be operated. The type of BCI that is of primary interest here is that which is controlled by specific willful mental activity, referred to as mental commands (*e.g.*, one might imagine pointing up with their hand, without physically doing so, as a mental command for scrolling up in an internet browser). For reasons which will be more deeply discussed in Chapter 2, the choice of mental command(s), which also determines the specific neurophysiological signals the BCI must recognize, is extremely important for successful control of a BCI. Considering the user when

choosing mental commands means that those mental commands must be based on the user's background and cognitive profile, as these factors influence which brain signals they can most easily learn to control. The challenge is that facilitating this kind of user-centered design requires that different users be able to employ different brain signals when controlling the same BCI. Meeting this challenge requires a generalized solution to the problem of translating mental commands into actionable commands in a computer. Without generalization across a variety of possible mental strategies, a BCI will not perform reliably for all of its users. However, this level of generalization has not yet been achieved.

The second implication is that improving methods for training the user to control a BCI by modulating their own brain activity may be just as important as developing more generalized translation algorithms. Situating the user at the center of a BCI comes with the recognition that using a BCI is a skill which must be learned, honed, and maintained. A BCI will not operate reliably for a user if that user cannot generate consistent and distinct mental commands, regardless of how advanced the BCI's AI. Moreover, in order to be compatible with the aforementioned need for personalized mental commands, any new approach to user training must be able to adapt to each user's unique style of operation rather than force a user to generate those brain signals that the BCI is pre-designed to expect, as is the current norm. The challenge, then, is to develop user-training methods which generalize to user-specific mental commands, even if the system is blind to what mental commands were chosen, while advancing the ability to provide specific and useful feedback to the user.

In combining these requirements, a strategy for putting this paradigm into practice can be put forth. The approach pursued throughout this thesis is to move towards a fully generalized BCI. This involves more than generalizing the technical components of a BCI so that it can be controlled using the wide variety of mental commands that might be employed by its various users. The user of a BCI must learn to operate the system by modulating their own brain activity directly, and the system must simultaneously learn to interpret the user's brain activity. While this defines two prongs to BCI control that can be addressed independently to a degree, the approach to generalization proposed here recognizes that these prongs are ultimately unified to form a single co-adaptive system. Therefore, the approach to developing a generalized BCI pursued through this thesis is founded on the perspective that solutions for generalized BCIs must be user-centered and must solve the generalized problem of co-adaptive human and machine learning.

The progress towards a generalized BCI described in this thesis should be taken as only the first steps towards solving the generalized problem of co-adaptive human and machine learning. There is much work to be done to achieve this goal, and many years of research may be required before BCIs, generalized or not, evolve to be useful tools for the general public. What this thesis sets out to demonstrate

is that generalized methods for user-centered brain-computer interfacing are both a feasible and a promising pathway towards improving the reliability and usability of BCIs in a way which may lead to BCI applications that can accommodate a wide variety of users. As such, the primary objective set out for this body of research was to develop a BCI which could be controlled by a wide variety of mental commands according to the preferences of each user, and then to propose methodology for enhancing a user's ability to learn self-regulation of the neural activity associated with those mental commands.

1.3   Thesis Overview

This thesis presents a series of studies which aim to put in place the basic components needed for a generalized BCI. Studies focused on generalizing the algorithms which translate brain activity into usable outputs for a computer are presented first. Following this, the focus is shifted towards the development of improved user training methods which can be applied to a generalized BCI. Together, these satisfy the original objectives pursued by this body of work. Following this, a pilot study is presented that integrates the newly developed generalized brain signal processing and user training methods into an updated generalized BCI. A more detailed overview is given below.

Chapter 2 provides the background needed to understand the studies presented in subsequent chapters. This includes more thoroughly defining what a BCI is and how a BCI works, as well as providing detailed descriptions of the primary methods used in this thesis. Technical and conceptual aspects of relevant algorithms and user-training methods are discussed here, with an emphasis on how these methods facilitate or impede progress towards a generalized BCI. While some of these methods are also described in subsequent chapters, the most complete descriptions are found in Chapter 2.

Chapter 3 presents a study which helps to contextualize the need for a generalized BCI. Many modern BCIs are based on the use of intentional mental imagery, the generation of perceptual experience within the mind, as mental commands. These are user-driven BCIs, also called spontaneous BCIs [32, 33, 34, 35], as opposed to BCIs driven by the user's reactions to external stimuli or changes in passive mental states [36, 37, 1, 38]. For a standard, *i.e.*, non-generalized, BCI, the neural correlates corresponding to a user's mental commands must be studied and built into the algorithms (BCIs are often defined by the type of mental commands they allow, as in a "motor imagery BCI"). The study presented in Chapter 3 explores the neural correlates of visuospatial mental imagery and a discussion on how those findings could be applied in a BCI follows. This study is used to

illustrate the traditional approach to designing a new BCI and to explain why this approach is inadequate if the field of brain-computer interfacing is to become more widely applicable and impactful in society. Thus this chapter motivates the pursuit of a generalized BCI starting from the viewpoint of traditional BCI research.

Chapters 4 and 5 present studies which establish and validate two methods for generalizing the BCI machine learning pipeline and for the recognition of mental commands and mental states more broadly. Chapter 4 in particular gives much attention to establishing the paradigm proposed in this thesis as well as the rationale upon which it is based. More importantly, this chapter presents an empirical validation of a first-generation generalized BCI which, for the first time, allowed users to choose their own mental commands within a number of different sensory modalities. In addition, this research presents evidence of a connection between the large individual differences in the ability to control a BCI using mental commands derived from different sensory modalities to training outside of the BCI context. Such evidence further demonstrates the need for generalized methods in brain-computer interfacing.

Chapter 5 presents a different way of generalizing the BCI machine learning pipeline. This approach was developed primarily for BCI implementations that use hardware for which the methods used in Chapter 4 may not be appropriate. However, the study presented in Chapter 5 helps to verify the value provided by generalized methods for BCI, because this alternative generalized method was successfully applied to the very different problem of detecting emotional states rather than mental commands. Together these studies show that generalized methods can be used to expand the possibilities of brain-computer interfacing.

With the generalized methods for identifying mental commands and mental states empirically validated, Chapter 6 returns to the question of how to train BCI users to effectively use mental commands to control a BCI. A new method for implementing neurofeedback, *i.e.*, training control over one's brain activity via direct sensory feedback, is defined and empirically validated in the context of controlling a single feature of brain activity. The rationale behind this approach and its implications for BCI user training are discussed.

Chapter 7 integrates the findings of all of the previous research chapters into an updated generalized BCI for answering "Yes" or "No" questions. In particular, an extension of the user training methodology presented in Chapter 6 to the BCI context is is presented. The system presented here makes use of all of the generalized methods developed throughout this thesis, and is presented as proposed next steps, since the main goals set out for this thesis had already been achieved. The results of a pilot study used to fine-tune the design of the system are presented and discussed.

Finally, Chapter 8 summarizes the key contributions and the overall implications of the research presented in this thesis to the field of brain-computer interfacing. Future directions for further development of a generalized BCI are outlined in some detail.

Bibliography

[1] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine.," *Mayo Clinic proceedings. Mayo Clinic*, vol. 87, pp. 268–79, mar 2012.

[2] M. D. Serruya and M. J. Kahana, "Techniques and devices to restore cognition," *Behavioural brain research*, vol. 192, no. 2, pp. 149–165, 2008.

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

[4] N. Birbaumer and L. G. Cohen, "Brain–computer interfaces: communication and restoration of movement in paralysis," *The Journal of physiology*, vol. 579, no. 3, pp. 621–636, 2007.

[5] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *The Lancet Neurology*, vol. 7, no. 11, pp. 1032–1043, 2008.

[6] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 113, pp. 767–91, June 2002.

[7] K. P. Thomas, A. Vinod, and C. Guan, "Enhancement of attention and cognitive skills using EEG based neurofeedback game," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, pp. 21–24, IEEE, 2013.

[8] B. Blankertz, L. Acqualagna, S. Dähne, S. Haufe, M. Schultze-Kraft, I. Sturm, M. Ušćumlic, M. A. Wenzel, G. Curio, and K.-R. Müller, "The Berlin brain-computer interface: Progress beyond communication and control," *Frontiers in Neuroscience*, vol. 10, 2016.

[9] C. Brunner, N. Birbaumer, B. Blankertz, C. Guger, A. Kübler, D. Mattia, J. d. R. Millán, F. Miralles, A. Nijholt, E. Opisso, *et al.*, "BNCI Horizon

2020: towards a roadmap for the BCI community," *Brain-computer interfaces*, vol. 2, no. 1, pp. 1–10, 2015.

[10] E. M. Mugler, C. A. Ruf, S. Halder, M. Bensch, and A. Kübler, "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, 2010.

[11] J. L. Sirvent, J. M. Azorín, E. Iáñez, A. Úbeda, and E. Fernández, "P300-based brain-computer interface for internet browsing," in *Trends in practical applications of agents and multiagent systems*, pp. 615–622, Springer, 2010.

[12] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih, and A. Kübler, "P300 brain computer interface: current challenges and emerging trends," *Frontiers in neuroengineering*, vol. 5, p. 14, 2012.

[13] V. Martinez-Cagigal, J. Gomez-Pilar, D. Alvarez, and R. Hornero, "An asynchronous P300-based brain-computer interface web browser for severely disabled people," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2016.

[14] T. Carlson, R. Leeb, G. Monnard, A. Al-Khodairy, and J. d. R. Millán, "Driving a BCI wheelchair: A patient case study," in *Proceedings of TOBI Workshop lll: Bringing BCIs to End-Users: Facing the Challenge*, pp. 59–60, 2012. EPFL-CONF-174373.

[15] T. Carlson and J. d. R. Millan, "Brain-controlled wheelchairs: a robotic architecture," *IEEE Robotics & Automation Magazine*, vol. 20, no. 1, pp. 65–73, 2013.

[16] G. S. Dhillon and K. W. Horch, "Direct neural sensory feedback and control of a prosthetic arm," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 13, no. 4, pp. 468–472, 2005.

[17] J. L. Contreras-Vidal and R. G. Grossman, "NeuroRex: A clinical neural interface roadmap for EEG-based brain machine interfaces to a lower body robotic exoskeleton," in *Engineering in medicine and biology society (EMBC), 2013 35th annual international conference of the IEEE*, pp. 1579–1582, IEEE, 2013.

[18] S. R. Chang, R. Kobetic, M. L. Audu, R. D. Quinn, and R. J. Triolo, "Powered lower-limb exoskeletons to restore gait for individuals with paraplegia–a review," *Case orthopaedic journal*, vol. 12, no. 1, p. 75, 2015.

[19] J. L. Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. McMorland, M. Velliste, M. L. Boninger, and A. B. Schwartz, "High-performance neuroprosthetic control by an individual with tetraplegia," *The Lancet*, vol. 381, no. 9866, pp. 557–564, 2013.

[20] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "Thought–control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neuroscience letters*, vol. 351, no. 1, pp. 33–36, 2003.

[21] M. Grosse-Wentrup, D. Mattia, and K. Oweiss, "Using brain–computer interfaces to induce neural plasticity and restore function," *Journal of neural engineering*, vol. 8, no. 2, p. 025004, 2011.

[22] T.-S. Lee, S. J. A. Goh, S. Y. Quek, R. Phillips, C. Guan, Y. B. Cheung, L. Feng, S. S. W. Teng, C. C. Wang, Z. Y. Chin, *et al.*, "A brain-computer interface based cognitive training system for healthy elderly: a randomized control pilot study for usability and preliminary efficacy," *PloS one*, vol. 8, no. 11, p. e79419, 2013.

[23] J. F. Burke, M. B. Merkow, J. Jacobs, M. J. Kahana, and K. A. Zaghloul, "Brain computer interface to enhance episodic memory in human participants," *Frontiers in human neuroscience*, vol. 8, 2014.

[24] J. Lumsden, E. A. Edwards, N. S. Lawrence, D. Coyle, and M. R. Munafò, "Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy," *JMIR Serious Games*, vol. 4, no. 2, 2016.

[25] P. Yuan, Y. Wang, W. Wu, H. Xu, X. Gao, and S. Gao, "Study on an online collaborative bci to accelerate response to visual targets," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 1736–1739, IEEE, 2012.

[26] E. M. Holz, J. Höhne, P. Staiger-Sälzer, M. Tangermann, and A. Kübler, "Brain–computer interface controlled gaming: Evaluation of usability by severely motor restricted end-users," *Artificial intelligence in medicine*, vol. 59, no. 2, pp. 111–120, 2013.

[27] M. Ahn, M. Lee, J. Choi, and S. C. Jun, "A review of brain-computer interface games and an opinion survey from researchers, developers and users," *Sensors*, vol. 14, no. 8, pp. 14601–14633, 2014.

[28] Y. Iidal, D. Tsutsumi, S. Saeki, Y. Ootsuka, T. Hashimoto, and R. Horie, "The effect of immersive head mounted display on a brain computer interface game," in *Advances in Affective and Pleasurable Design*, pp. 211–219, Springer, 2017.

[29] VR2-nBio Lab, "P300-based BCI for internet Browsing." `https://www.youtube.com/watch?v=GuuhQur7Tjk`. Accessed: 2017-05-04.

[30] Defitech chair in brain-machine interface, "Neuroprosthetics: the mind is the pilot." `http://cnbi.epfl.ch/page-34064-en.html`. Accessed: 2017-05-04.

[31] M. Bockbrader, N. V. Annetta, G. Sharma, D. A. Friedenberg, H. S. Bresler, W. J. Mysiw, and A. R. Rezai, "Standardized tests of upper limb motor function inform the promise and pitfalls of brain-computer-muscle neuro-prosthetics for tetraparetic patients," *Archives of Physical Medicine and Rehabilitation*, vol. 97, no. 10, p. e20, 2016.

[32] N. Achtman, A. Afshar, G. Santhanam, M. Y. Byron, S. I. Ryu, and K. V. Shenoy, "Free-paced high-performance brain–computer interfaces," *Journal of neural engineering*, vol. 4, no. 3, p. 336, 2007.

[33] M. Sugiyama, M. Krauledat, and K.-R. MÃžller, "Covariate shift adaptation by importance weighted cross validation," *Journal of Machine Learning Research*, vol. 8, no. May, pp. 985–1005, 2007.

[34] P. Von Bünau, F. C. Meinecke, F. C. Király, and K.-R. Müller, "Finding stationary subspaces in multivariate time series," *Physical review letters*, vol. 103, no. 21, p. 214101, 2009.

[35] P. Von Bünau, F. C. Meinecke, S. Scholler, and K.-R. Müller, "Finding stationary brain sources in EEG data," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 2810–2813, IEEE, 2010.

[36] 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, 2003.

[37] S. Halder, D. Agorastos, R. Veit, E. M. Hammer, S. Lee, B. Varkuti, M. Bogdan, W. Rosenstiel, N. Birbaumer, and A. Kübler, "Neural mechanisms of brain–computer interface control," *Neuroimage*, vol. 55, no. 4, pp. 1779–1790, 2011.

[38] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review.," *Sensors (Basel, Switzerland)*, vol. 12, pp. 1211–79, Jan. 2012.

Chapter 2

Background and Methods

2.1   A Brief History of Brain-Computer Interfacing

The very beginnings of the scientific study of brain-computer interfacing can be traced to a pair of publications in 1969 by Delgado [1] and Fetz [2]. Delgado created an implantable radio-controlled electrode capable of recording and wirelessly transmitting information about brain activity to a computer while stimulating brain cells using radio waves. In a famous demonstration, Delgado used this chip to stop a charging bull in its tracks by stimulating its caudate nucleus and basal ganglia, brain structures which are important for controlling goal-oriented motor functions [3]. As Delgado reports, repeatedly stopping the bull mid-charge with the press of a button had the added effect of making the bull less aggressive for longer and longer periods of time thereafter. Concurrently, Fetz demonstrated that monkeys could be trained to control the activity of a particular brain cell via neurofeedback and an implanted electrode, and thus control a machine that dispensed food rewards. These demonstrations inspired scientists and engineers to explore the possibility of human control over electronic devices via brain activity alone.

The modern conceptualization of BCIs intended for human use was first proposed by Vidal in 1973 [4], and thus he is often credited for initiating the academic enterprise that is now called brain-computer interfacing. Vidal exhibited an impressive degree of foresight in this seminal paper, in which he proposed a human BCI based on visual evoked potentials (electrical field potentials resulting from particular visual stimuli) as recorded from the scalp using electroencephalography (EEG). Vidal pointed out that at the time of writing his paper he did not have the computing technology necessary to create a useful version of the device (though he did produce a prototype a few years later in 1977 [5]), and that in the years it would take for such technology to be developed his proposals might already be obsolete. However, what he proposed included design schematics and an analytical approach which is in principle close to what is used today for BCIs based on visual evoked potentials [6]. One notable difference is that a statistical decision making algorithm was used in place of a more modern machine learning approach, as machine learning in its modern form did not exist in those days.

Following these very early days of research in brain-computer interfacing, very few publications on the topic appeared until the 1990s, when a handful of laboratories mainly situated in Germany and the United States began to experiment with rudimentary BCIs. A confluence of interest, exposure to this pioneering work, improved computer technology, better neuroscientific equipment and understanding, and the advancement of machine learning led to an explosion of research in brain-computer interfacing starting in the early 2000s (see Figure 2.1). Suddenly,
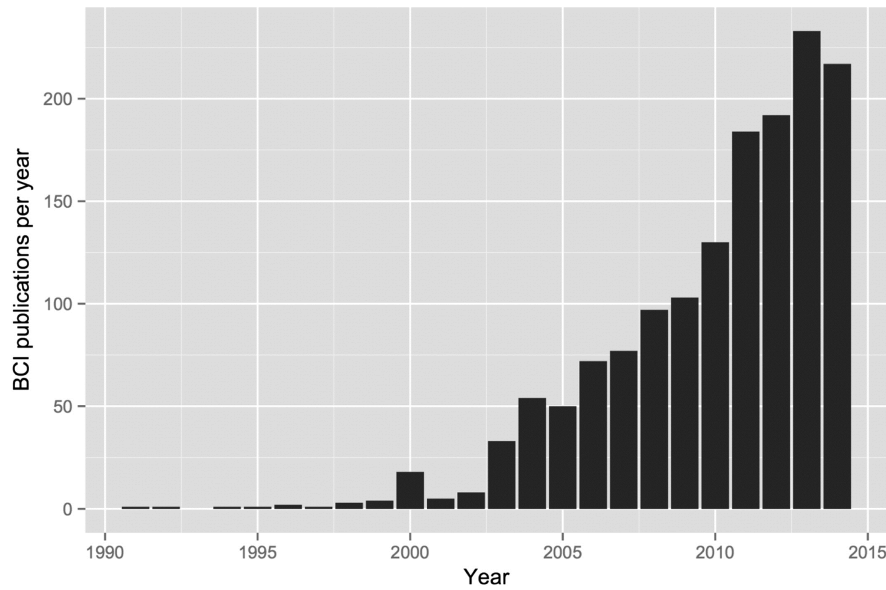
Figure 2.1: The number of BCI publications per year, obtained by searching for "brain-computer interface" on PubMed. The modest spike which occurs in the year 2000 is a result of the First International BCI Meeting held in 1999. Figure reproduced from [7] under the Creative Commons Attribution - NonCommercial-NoDerivatives License.

the idea of humans using BCIs for a variety of tasks became a realistic possibility for the not-so-distant future.

Despite the possibility of BCI technologies for humans, the justification for investing in what might have been a scientific and engineering curiosity with unclear long-term utility was not yet readily apparent. That began to change when collaborations between medical scientists, psychologists, and engineers, combined with increased social awareness, led to a broader recognition that individuals paralyzed or locked-in due to severe neuromuscular disorders, such as amyotrophic lateral sclerosis (ALS), often retain their cognitive ability [8]. Suddenly there was great interest in assistive technologies that could enable communication and control for these individuals. Because these individuals tend to suffer a complete loss of voluntary control over all muscles, whether gradually or suddenly, only a system with an interface that did not rely on any muscle activity could be used. Filling this need became the primary focus of brain-computer interfacing.

Brain-computer interfacing has now become an established field of science and engineering. There currently exist several classes of BCIs, each with a number of variations best suited for different applications. The utility of BCIs outside of the clinical setting is also being actively explored, and commercial recording devices intended for commercial BCI applications are beginning to enter the consumer market. BCIs have also become more precisely defined over the last several years.

15

A detailed definition and description of a BCI and its parts are given next.

2.2 The Defining Features of a BCI

A brain-computer interface, sometimes referred to as a brain-machine interface (BMI), is a special kind of communications and control system which interprets brain signals in order to allow its user to input commands into a computer without activity from the peripheral nervous system or muscles [8, 9, 10]. Due to the interdisciplinary nature of these systems, the field of brain-computer interfacing requires input from several different areas of research, including neuroscience, psychology, biomedical engineering, computer science, and statistics. The importance of each of these disciplines in brain-computer interfacing is expressed in the design, implementation, and evaluation of a BCI as a whole as well as for its distinct components.

2.2.1 The Essential Components of a BCI

Several key ingredients are required to create a BCI. Broadly speaking, there are two main parts to a BCI. The BCI transducer encompasses the signal acquisition, signal processing, and machine learning components of the system. Connected to this is the NFB module, which provides feedback to the user and executes the action of the BCI. These components and their relationships are illustrated by the BCI loop in Figure 2.2.

A BCI utilizes the mental activity of its user in order to communicate with a computer. Therefore, the central component of a BCI is the brain that uses it and the type of mental activity used to control the BCI. The type of mental activity employed by the user to issue commands through the system determines what kind of neurophysiological signals the BCI must learn to detect and interpret. This, in turn, determines how the technical components should be implemented, especially which specific signal processing and machine learning methods should be used.

In order for the BCI to access the user's mental activity, it must have some means of observation in real time through some brain signal recording hardware. Most commonly, electroencephalography (EEG) is used, though there are several other technologies that are used in the field. The type of information about brain activity obtained through brain recordings depends substantially on the recording technology used. In turn, the signal processing methods used depend heavily on
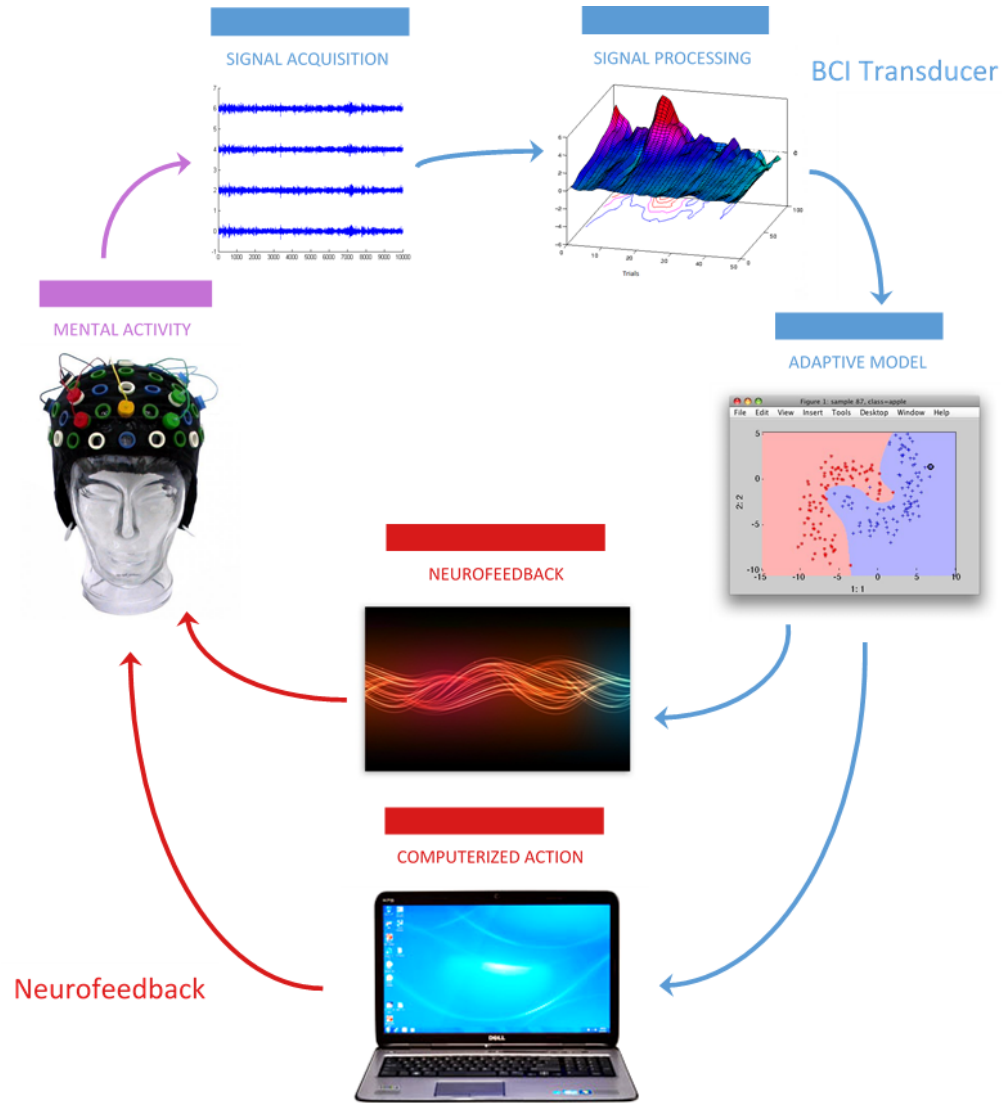
Figure 2.2: The BCI loop. The main components of a BCI, including the user, operate together and form a closed-loop system. One complete iteration through this loop describes the user issuing a single mental command, which is translated by the BCI into a computer-controlled action whose result and corresponding feedback is observed by the user. The BCI transducer is displayed with blue text and blue arrows, while the NFB module is displayed with red text. The user is displayed with purple text because they tie together the BCI transducer and the NFB module.

the recording technology. Likewise, the recording technology constrains the types of mental activity that can feasibly be detected and interpreted by a BCI. For example, it is difficult to precisely localize brain signals with EEG recordings, especially if the signal is generated deep in the brain. Because of this, an EEG-based BCI controlled with brain signals which are generated very close together in space, or which are generated deep in the brain, is likely to perform poorly compared to the same BCI controlled with brain signals whose sources are spatially far apart and near the surface of the brain. The recording technologies used in this thesis are described in Section 2.4.

Given a recording device which can be reasonably expected to provide the discriminative information needed to distinguish the user's mental commands, it remains a complicated task to extract the relevant information and separate it from any extraneous information. This is the job of the BCI's signal processing pipeline. Signal processing serves the dual purpose of preprocessing and denoising the brain signals (details in Section 2.5) as well as extracting informative values, or features, from the preprocessed signal (details in Section 2.6).

Once the brain signal has been preprocessed and the relevant information extracted, the resulting data can be translated into meaningful commands for a computer. This is accomplished with an adaptive model, typically in the machine learning sense. The role of the adaptive model is to compare the information provided by the signal processing pipeline to previous data and make a determination as to which mental command was issued by the user. The application of machine learning to BCI and the core methods used in this thesis are discussed in Section 2.7.

The mental command predicted by the adaptive model is passed to a software module which is designed to interpret each unique mental command as a unique instruction to the computer. Through this module, the interpretation of the user's intentions through the BCI transducer results in an action performed by the BCI. This computerized action is essential to a BCI both because it closes the BCI loop by providing feedback to the user regarding how the system had interpreted their mental activity, and because the set of actions which a BCI can produce defines its application.

While the previously mentioned BCI components technically form a complete BCI, the inherent difficulty of learning to control a BCI due to the difficulty in learning how to generate consistent brain signals as mental commands makes an explicit neurofeedback (NFB) component a practical necessity for most BCIs. NFB is especially important for BCIs which are driven by voluntary mental activity as opposed to those controlled by reactive brain activity (the distinctions between the main types of BCIs is further discussed in Section 2.2.3). NFB refers to a class of techniques used to train an individual to regulate or modulate their

own brain activity. This is accomplished by providing real-time feedback of a user's brain activity with respect to some target brain state [11, 12]. In BCI, NFB is used to help train the user to produce more consistent and distinct mental commands, thus making control more efficient and reliable [13, 14, 15]. The target brain state is defined both by the type of information extracted by the signal processing pipeline and by what patterns of brain activity the adaptive model expects for each mental command. The topic of neurofeedback and how it relates to BCI is discussed further in Section 2.3 as well as in Chapters 6 and 7.

2.2.2 BCI Model Training and Evaluation

An adaptive model or a machine learning model requires training before it can make accurate decisions. Due to this fact, there are typically three main phases of usage which a user must undergo with any BCI based on mental imagery. First, a training phase is necessary so that individualized models can be learned for the user. An evaluation phase is usually recommended after training in order to estimate BCI performance in a context closer to the use-case scenario. Finally, the BCI can be deployed for its intended application for the given user.

The training phase is the most important part of BCI development in the research setting, as BCI performance during training can be used to demonstrate improvements in BCI methodology. While the training phase can be implemented in a variety of ways, it typically simulates the BCI's intended application and is operated as described by the BCI loop illustrated in Figure 2.2. There is however one key difference. During training, the user is instructed when to perform each mental command. This allows the system to compile a dataset for which it is known which segments of recorded brain activity are associated with each mental command. This provides the required labeled data to train an individualized machine learning model.

At the start of the training phase, the models usually begin untrained (although some promising work has been done recently using inter-subject models to seed and enhance individualized models [16, 17, 18, 19]) and thus the BCI will respond in a random fashion to the user's mental commands. However, the models are updated periodically during training, and thus the accuracy of mental command recognition should increase over the course of training. Furthermore, the feedback provided by the system during training should help the user produce more consistent and distinct mental commands, which contributes significantly to reliable BCI control.

Sometimes a BCI is evaluated after training by having the user operate the BCI freely or again by following specific instructions. However, this step is not

Class	Ex. Brain Signal	Control Type	Control Dimensions	ITR
ERP	P300	Reactionary	Very High	20-30 bits/min
SFBA	SCP	Voluntary	Very Low (2 or 4)	5-12 bits/min
MI	SMR	Voluntary	Very Low (2 to 4)	5-35 bits/min

Table 2.1: A summary of BCI types and the most common brain signals used for control in each category.

always necessary in the research setting, especially if the goal is to evaluate new BCI methodology rather than a specific end-user application. In either case, performance metrics are required in order to evaluate and compare BCI implementations. Classification accuracy (the percentage or proportion of correctly recognized mental commands) is the most common performance metric used. However, when an evaluation phase is conducted for a specific BCI application, it is also common to report the Information Transfer Rate (ITR) [20], where

$$\text{ITR} = \frac{\log_2 K + P \log_2 P + (1 - P) \log_2 \left[(1 - P)(K - 1)\right]}{T} \tag{2.1}$$

for K classes (equivalent to the number of usable mental commands or BCI outputs) and probability of correct mental command recognition P over time period T. The ITR, measured in bits/min. is useful because it takes into account both classification accuracy and the speed of the BCI. It should be noted that there is still active discussion as to what standard metrics for evaluating BCI performance should be adopted in the field [21].

2.2.3   Types of BCIs

Some BCIs are not completely independent of muscle activity. For example, some early BCIs determined which on-screen object to select based on brain activity associated with looking at one object over other objects [22, 8], thus requiring eye movements to help drive the system. However, this thesis is restricted to the discussion of BCIs which rely solely on activity within the central nervous system. Such BCIs can be divided into three main categories: those based on event-related potentials (ERPs), those based on voluntary control over a single feature of brain activity (SFBA), such as the Slow Cortical Potential (SCP) [23], and those based on conscious mental imagery (MI). Each of these can be further subdivided based on the specific brain signal used. The types of BCIs differ in how they are controlled, how many control dimensions they can support, how frequently commands can be issued through the system, and which types of applications they support. This information is summarized in Table 2.1 for the three common categories of BCI.
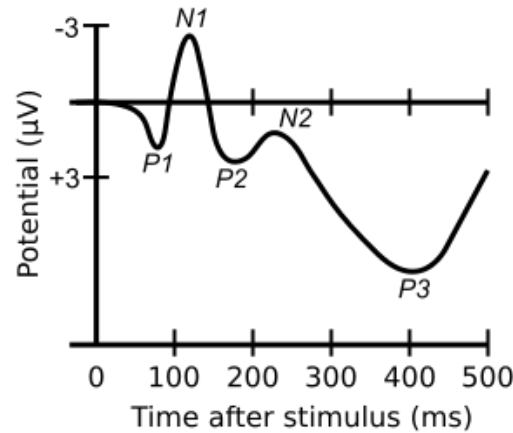
Figure 2.3: An illustration of a common ERP with its different components. The P300 (also called the P3) is commonly used in BCI. Note that this plot has a reversed y-axis, which is common in the ERP literature. Image obtained from [24].

This thesis is primarily concerned with BCIs which utilize mental imagery. However, P300-based BCIs are briefly summarized here because the field of brain-computer interfacing was established in large part through these systems, and they led to the later development of mental imagery BCIs.

BCIs Based on Event-Related Potentials

An ERP is a brain response measured with EEG (the MEG equivalent is called an event-related field, or ERF) and has an onset that occurs within a defined time interval following a specific sensory, motor, or cognitive event. These events have stereotypical morphologies which make them particularly amenable to detection (see for example the P300 in Figure 2.3). In addition, because ERPs can be produced in reaction to external stimuli in an automatic manner, a user does not need to be trained to control his/her brain activity in order to use an ERP-based BCI effectively, though the system itself does still need to be trained for each user. The main challenge in implementing an ERP-based BCI is that the characteristic forms of ERPs are usually only visible after averaging several trials that are temporally aligned to the stimulus event (*i.e.*, time-locked EEG trials).

The P300 evoked potential is the most common ERP used in BCIs. The P300, illustrated in Figure 2.3, occurs roughly 300ms after a subject is presented with an unexpected visual, auditory, or somatosensory stimulus [25]. One of the quintessential BCI applications is the P300 speller, which is displayed in Figure 2.4. The user observes a grid of characters while rows and columns flash at random. The user simply has to focus on the character they would like to have typed

next. A P300 potential is evoked when that character's corresponding row or column flashes. Detecting a P300 for both a row and a column uniquely identifies a character, and therefore provides a means for typing on a computer using attention-driven brain responses [26]. Moreover, the P300 BCI is particularly useful for immobilized patients because it does not depend on gaze direction, as long as the grid of characters is within view.

While BCI applications using P300 ERPs are usually simple to imagine (for example, a very similar setup to the P300 speller could be used to play chess [27] or control a digital painting program [28]) the implementation is challenging. The main limitation is accurately detecting the P300 ERPs while minimizing the number of repeated trials needed for averaging [29]. Thus, even though P300-based BCIs can reach nearly perfect accuracies, they can be slow to use. Very recently, near-perfect accuracies were achieved with single-trial classification of P300 ERPs using a novel machine learning method [30]. However, when perfect single-trial detection of the P300 ERP is achieved, the upper limit of P300-based BCIs will also have been reached. Therefore, other kinds neurophysiological phenomena are actively explored as avenues for new BCI applications.

Mental Imagery BCIs

Mental imagery refers to any imagined sensory or cognitive experience, such as mentally replaying a song, or viewing visual scenes in the mind. BCIs based on mental imagery have become of greater interest as the field has developed. While they tend to be more complicated in design and implementation compared to ERP-based BCIs, MI-BCIs also have the potential to pave the way towards a greater variety and depth of applications. The reason is simple: the variety and richness of mental commands based on mental imagery are theoretically only limited by human imagination. However, interpreting what is being imagined based on recorded brain activity alone is extremely difficult from neuroscientific, signal processing, and machine learning perspectives, so there is still a long way to go before the potential of MI-BCIs are seen.

In order to simplify the problem of mental imagery discrimination, BCI researchers have focused most of their attention on motor imagery, which refers to kinaesthetically imagining performing actions using the body. For example, it is common for motor imagery BCIs to be controlled by imagining performing movements or actions using the hands or feet [32]. BCIs based on motor imagery have dominated the field to such an extent that much of the literature refers to motor imagery or SMR-BCIs specifically rather than to mental imagery BCIs [8, 10]. The neural correlates of motor imagery are well understood and well characterized in the neuroscience literature, in part because very similar activity is seen in

Figure 2.4: A typical P300 speller. On a two dimensional grid, a row and column together identify a specific grid position. Flashing a row or column which contains the letter attended to by the user elicits a P300 brain response. Therefore, a system which can perform reliable real-time detection of a P300 ERP can be used in conjunction with an interface which flashes rows and columns of characters on a grid in order to make a typing application based on brain responses. Image reproduced from [31] under the Creative Commons Attribution License.

the sensorimotor cortex when bodily actions are either physically performed or merely imagined.

Having sufficient *a priori* neurophysiological knowledge about motor imagery simplifies the problem of classification. With this kind of knowledge, researchers can develop specialized strategies for extracting and classifying relevant information from brain signals (*i.e.*, specialized implementations of the signal processing pipeline and adaptive modeling steps illustrated in Figure 2.2). Motor imagery involving distinct parts of the body is highly spatially localized along the sensorimotor cortices [33, 34], especially when considering contralateral control pathways between each hemisphere of the brain and the body [35]. Moreover, motor imagery produces μ rhythm (activity in the 7-13 Hz band seen in the sensorimotor cortex during motor planning and execution [36, 37, 38]) and associated β (specifically 18-26 Hz) band desynchronization [39, 40, 41]. Therefore, discriminative spatial patterns in these frequency bands are usually sought for motor imagery discrimination [8, 32]. These factors combine to make motor imagery an ideal starting point for studying and developing the underlying technology for BCIs based on mental imagery.

While the vast majority of MI-BCIs use motor imagery, recent studies have demonstrated a clear potential for mental commands based on auditory and visual imagery [42, 43]. There is growing interest in non-motor forms of mental imagery for BCI control because continued reliance on motor imagery is limiting. These limitations are discussed in detail in Chapter 4, but the underlying problem with an over-reliance on motor imagery is that the ability to use motor imagery effectively in the BCI context varies considerably across individuals [44, 45, 46]. This leaves a sizable proportion of the population unable to easily use a BCI based on motor imagery, not only because of insufficiently advanced motor imagery recognition or user-training methodology, but because their cognitive and neurobiological profiles are simply not well-suited to modulating those specific brain signals [47, 48, 49, 50, 46, 43].

Despite such significant limitations, BCIs based on motor imagery have become the standard and the benchmark for mental imagery BCIs in general. Hence the importance of motor imagery to MI-BCIs when considering the expansion to other forms of mental imagery cannot be understated. Furthermore, the user-training methods and experimental designs used for motor imagery BCIs are still well suited for BCIs based on other forms of mental imagery. For these reasons, motor imagery is included in Chapter 4 where alternative forms of mental imagery are evaluated for their use in BCI control.

2.3   Using a BCI is a Skill: Neurofeedback for BCI Training

It has been widely acknowledged for more than a decade that "BCI use is a skill" [8, 51]. This means that the user of a BCI, especially a mental imagery BCI, must be trained to produce consistent and reliable EEG patterns in order to successfully and easily control the system [52, 44]. If the user is not able to generate the brain activity corresponding to each mental command willfully and in such a way that they are distinct from one another, then the BCI will fail regardless of the sophistication of its signal processing and machine learning algorithms. Thus, BCI users typically develop specific mental strategies (*i.e.*, specific mental activity or imagery) and undergo training with neurofeedback in order to learn to generate useful brain activity.

Even though the importance of the user's skill has been acknowledged and explicitly stated in most of the widely cited reviews of brain-computer interfacing to date [8, 13, 9, 10], the BCI community has allocated most of its time and effort to the perfection of signal processing and machine learning techniques. However, some argue that perhaps the human user is as important as the machine and that the ability of humans to learn is under-exploited in BCIs [52, 51, 53, 54].

The idea that the user, or the brain, in the brain-computer interface is as important as the computer is, from the perspective of psychology and neuroscience, not a far-fetched one. The brain can be viewed as a biological computer capable of learning mental and physical behaviours extremely efficiently and often after only one or two attempts. In contrast, machine learning algorithms typically require many repetitions or examples of data before they are able to learn meaningful patterns (machine learning algorithms capable of "one-shot" learning, or learning from just a few examples, are quite novel in the machine learning literature [55, 56] and have not yet been applied to BCI).

The current standard in BCI user training is the Graz Protocol [57, 15], with most user training protocols used in BCI being variations of this approach [54]. The Graz Protocol uses a BCI training phase in which users are instructed when to produce a mental command and which command to produce using on-screen cues. After several trials (this can vary between 20 or more than 100 trials), machine learning algorithms are trained to classify each mental command using the recorded EEG data. Following this 'pre-training' period, the focus is shifted towards training the user. Here, the trials are repeated as previously described, but neurofeedback is provided to the user. This feedback can be concurrent in real time with the trial, or be provided after the mental imagery period of the trial is completed. In a motor imagery BCI intended to control a cursor on the left-right

axis, for example, neurofeedback might be displayed as an arrow extending in either direction, informing the user of which mental command the machine interprets the user as performing, and optionally, how confident the machine is in its interpretation. This protocol can be repeated across several sessions on different days in order to train the system and the user to be robust to nonstationarity within the EEG.

It has recently been shown that the Graz Protocol is a suboptimal method for training skills, even outside of the BCI context (*e.g.*, for training the motor skills of drawing triangles versus circles [53, 54]). Several theoretical limitations have been identified from the instructional design literature [51], but due to the inherent complexity of brain-computer interfacing, along with the unusual goal of training a user to control their own brain activity with respect to a linear or non-linear combination of features directly makes it difficult to incorporate what is known about training humans in other contexts. Because of the unique context of brain-computer interfacing, no specific recommendations have been made for improving user training for BCIs. As a result, there have yet to be new and better training protocols developed. However, one promising approach to augmenting BCI user training is discussed in Chapters 6 and 7, and is based on the concept of *shaping* from Learning Theory [12], which refers to training an individual towards a desired goal or behaviour in an incremental and progressive way.

2.4 Reading Brain Activity

Methods of recording brain activity can be categorized by what kind of biological phenomena they measure. BCIs are generally designed using recording technologies which either record electrophysiological signals or hemodynamic signals [10]. While functional near-infrared spectroscopy (fNIRS) and, to a lesser degree, magnetoencephalography (MEG) are recording technologies which are seen in the BCI literature, only functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) are discussed here, as only these two technologies were used in the work presented in this thesis.

2.4.1 Functional Magnetic Resonance Imaging (fMRI)

Functional magnetic resonance imaging (fMRI) non-invasively measures changes in oxygenated blood flow, most commonly using the blood-oxygen-level dependent (BOLD) contrast [58]. Associating the BOLD signal with neural activation relies on data suggesting that the flow of oxygenated blood to a region of the brain

Figure 2.5: A subject laying on the bench of the fMRI prior to being moved into the machine's tube. Stock Photo retrieved from [60].

follows neural activation within that region in order to support the metabolic needs of the activated neural tissue [59]. The magnetic resonance signal measured as a result of this process, called the hemodynamic response (HDR), does not appear until one second or longer after the neural activity of interest because of the time it takes for the vascular system to respond. Therefore the temporal resolution of fMRI recordings is quite low ($\approx 1 - 2s$). Combined with the fact that fMRI machines are extremely expensive and large, requiring a dedicated electromagnetically shielded room with the subject lying on a bench inside the scanner (see Figure 2.5), it is not a practical brain recording device for most brain-computer interfacing applications. However, the advantage of fMRI is its ability to acquire signals throughout the brain with fine spatial resolution ($\approx 0.5 - 2\text{mm}^3$, depending on the strength of the magnet) [10]. Therefore, fMRI can be very useful for spatial localization of the neural correlates of various kinds of mental activity. Chapter 3 presents one study of this nature. Once the neural correlates of specific mental activity are discovered via fMRI, the mental activity becomes potentially usable as a mental command in a BCI.

2.4.2   Electroencephalography (EEG)

Electroencephalography (EEG), first demonstrated in 1929 [61], is the most popular method of recording brain activity in the BCI literature [10]. EEG allows for non-invasive recordings of electrical activity that is produced by the brain and which reaches the scalp (what are often called brain waves) using an array of scalp electrodes as illustrated in Figure 2.6. As such, relatively to most recording methods, EEG is inexpensive and convenient and therefore most often chosen for laboratory studies involving healthy participants or patients who are not in ex-
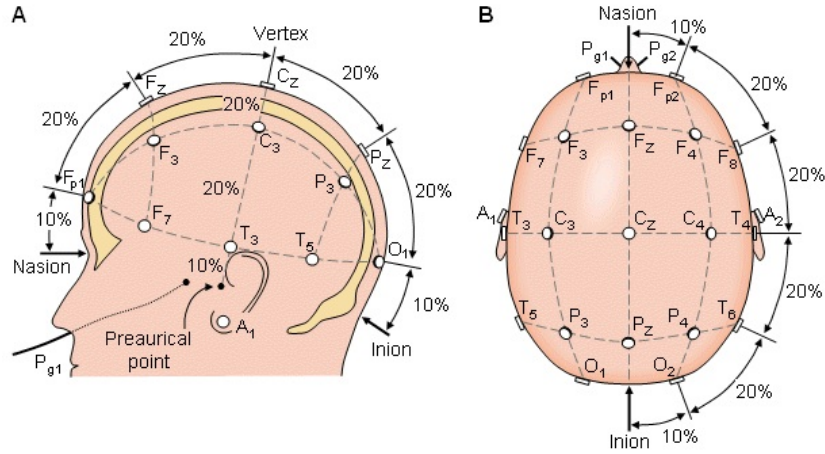
Figure 2.6: A grid of 21 EEG electrodes placed according to the International 10-20 system. Image reproduced from [63] with permission. (Disclaimer: re-use of this image will require separate permissions obtained from Oxford University Press.)

treme need. Despite the low spatial resolution afforded by EEG ($\approx 1\text{cm}^3$, with decreasing resolution for brain regions further away from the scalp) due to volume conduction through the cerebrospinal fluid, the skull, and the scalp, as well as any intermediary brain tissue (see Figure 2.7), EEG signals are useful for real-time applications because of their high temporal resolution (in the order of $10 - 20\text{ms}$). Note that while the temporal resolution of EEG is indeed high, conventional estimates of the temporal resolution of EEG may be somewhat overestimated because it can also be impacted by volume conduction [62].

EEG signals are reflective of the flow of electrical activity generated by at minimum many thousands of neurons firing nearly simultaneously, producing an electromagnetic local field potential large enough to measured at the scalp. Given the aforementioned limitation of low spatial resolution due to volume conduction, it is easiest to record activity from brain regions near the surface of the brain (though this limitation can be partially overcome with spatial localization techniques [65, 66]). Fortunately, much of the neocortex, which is responsible for most high level conscious mental activity [67, 68] and therefore is of primary interest for many BCI applications, makes up a large portion of the surface of the brain.

Given the importance of EEG in the field of brain-computer interfacing and for this thesis in particular, the proceeding sections on signal analysis and machine learning assume an EEG-based BCI.
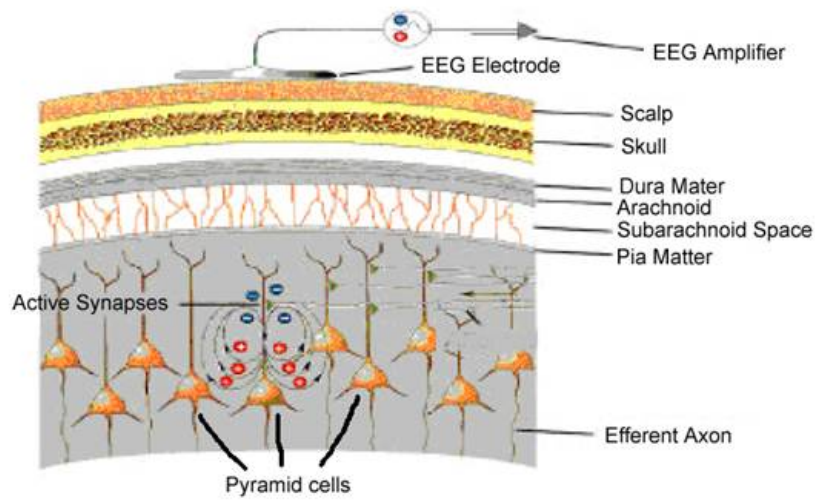
Figure 2.7: An illustration of EEG being produced by the cumulative field potentials of many active synapses of pyramidal cells. The summative local field potential must pass through several layers of tissue between the cerebral cortex and the scalp before reaching the EEG electrode [64].

2.5 EEG Signal Processing for BCI

Signal processing serves two essential functions for BCIs. The first is preprocessing, which involves denoising the raw brain signal and identifying where in the brain the signals of interest are generated. This is done to maximize the signal-to-noise ratio (SNR) and facilitate the second function, which is feature extraction. The second aim of feature extraction is to compute a set of features (also called variables or functions) which carry the information required for accurate and reliable recognition of the different patterns of brain activity used to control the BCI. This comprises the signal processing pipeline referred to in Figure 2.2.

EEG-based BCIs use a variety of signal processing tools depending on the specific EEG hardware and the intended application in order to perform the two functions outlined above. With respect to noise reduction, two key steps are involved. First is the reduction of environmental noise, various sources of low frequency noise, and extraneous information using digital filters. Second is the removale of artifacts which arise mainly from ocular and facial movements.

(a) Low-Pass Filter   (b) High-Pass Filter   (c) Notch Filter   (d) Band-Pass Filter
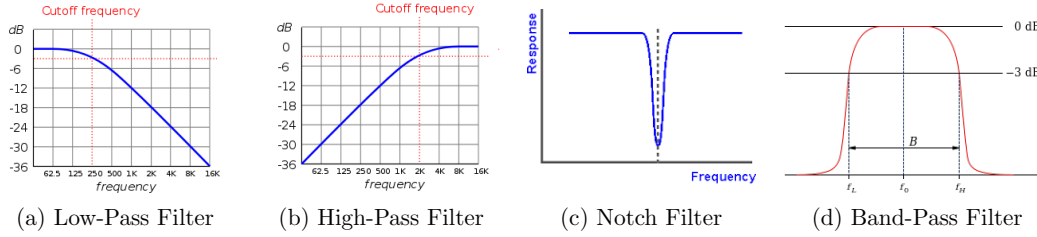
Figure 2.8: Exemplar shapes of common digital filters. Response is given on the Y-axis, while frequency is given on the X-axis. Where the response is high, the signal at those frequencies is retained. Where the response is low, power at those frequencies is diminished. Images obtained from the public domain [69].

Digital Filters

Digital filters are used almost ubiquitously in EEG signal processing. Digital filters are effective as a set of first-pass noise removal tools because they are able to reduce the magnitude of a signal within specified frequency ranges in order to enhance the signal in the frequency ranges of interest. This is extremely valuable for EEG studies, because it is often the case that the brain activity of interest occurs in a specific frequency band. Moreover, certain sources of noise occur in specific frequency bands, allowing digital filters to minimize their influence on subsequent analyses.

Digital filters can be categorized into four main types defined by how they alter the frequency spectrum of a signal. Low-pass filters retain frequency components of a signal below a specified frequency while reducing the magnitude of frequency components above the cutoff. High-pass filters are the opposite; they retain higher frequency components and reduce low frequency components. Notch filters reduce the influence of a signal in a narrow frequency band and are used to reduce the 50 Hz or 60 Hz power line noise that is picked up by EEG sensors. Finally, band-pass filters diminish the signal outside of a specified frequency range. These are commonly used in EEG data analysis in order to focus on the frequency bands of interest. Examples of the response curves of common digital filters used for noise reduction are shown in Figure 2.8.

There are a variety of methods for constructing digital filters. Butterworth filters were chosen for the work presented in this thesis because they have uniform response to the passband frequencies and a smooth response dropoff outside of the passband [70]. One tradeoff, however, is that they have a relatively slow roll-off, thus allowing more of the signal beyond the frequency cutoffs to be retained. For the purposes of the data analyses presented here, this drawback is sufficiently resolved by using, typically, a fourth order Butterworth filter, which performs

additional computations in order to achieve a much steeper roll-off.

Artifact Detection

Artifacts are extraneous signals that are mixed into EEG signals and distort or mask the neurophysiological signal of interest. Artifacts in the EEG can be generated from a number of sources, both physiological and non-physiological. The most common sources of artifacts are summarized in Table 2.2.

The presence of artifacts in EEG when training or using a BCI system is highly detrimental to the performance of the system. Because artifacts affect the brain signal of interest, they have a significant impact on the feature extraction and machine learning components of a BCI (discussed in more detail in Sections 2.6 and 2.7). Thus, their presence, if correlated with the issuance of mental commands, could be mistakenly used to control the BCI. This is especially problematic if, for example, a BCI is being experimentally validated with healthy participants before being deployed for patients suffering from paralysis. If the experimental group succeeds in controlling the BCI by using artifacts, then it remains unclear whether the system will be successful with a patient group who cannot generate such artifacts. On the other hand, if the artifacts are not particularly correlated with the mental commands being used, then they tend to mask those mental commands and prevent the system from learning useful patterns from the EEG.

Different types of artifacts can be distinguished by their spectral profile and sometimes their morphology in the time series. Non-physiological signals, such as power line noise, can typically be precisely characterized because their EMF outputs are constrained by design. This makes such signals significantly easier than other types of artifacts to detect and remove from the EEG, *e.g.*, by applying a 50 Hz or 60 Hz notch filter. On the other hand, physiological signals tend to be much more difficult to remove because they are much more variable and more often overlap in frequency with the frequency bands commonly used in EEG analyses. Furthermore, a greater variety of physiological artifacts compared to non-physiological artifacts are present in the EEG; some extremely large, and others very subtle. These artifacts cannot be reliably controlled at the source, because they are produced by physiological processes which are often involuntary (*e.g.*, blinking).

Many techniques have been developed to detect and remove artifacts from EEG signals, both for offline EEG experiments and in the real-time BCI setting [71]. While most non-physiological artifacts are successfully handled using standard signal processing tools, there is still active research in eliminating physiological artifacts, which are the most problematic in BCI. In particular, artifacts generated
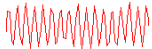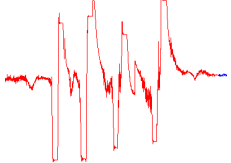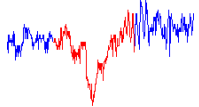
Source	Type	Frequency Range	Morphology
Power Line	Non-Physiological	50/60 Hz	
Electrode Displacement	Non-Physiological	Full Spectrum	
Eye Blinks	Physiological (EOG)	0-12 Hz	
Eye Movements	Physiological (EOG)	0-8 Hz	
Jaw Clenches / Facial Expressions	Physiological (EMG)	>20 Hz	

Table 2.2: Some common artifacts which contaminate EEG signals and their dominant characteristics. Examples of typical artifact morphologies are highlighted in red with surrounding EEG in blue. Note that frequency ranges are estimates based on standard EEG experimental setup and may vary according to the hardware and software used when recording the EEG.

from the electro-oculogram (EOG), such as eye blinks and eye movements, as well as artifacts generated from the electromyogram (EMG), such as jaw clenches and facial expressions, continue to be the most difficult and disruptive kinds of artifacts in EEG.

Two main approaches are used to handle artifacts: first, detection of an artifact and subsequent removal of the entire affected segment in the time series is used if the underlying neurophysiologically-derived EEG cannot be recovered with high confidence, and second, given suitable recording hardware, computational resources, and a sufficiently large dataset, regression and/or source separation methods can be used to filter artifacts and recover the underlying EEG. For the first approach, a common technique is statistical thresholding (deleting segments of the EEG which exceed a multiple of the variance in amplitude or power) of the EEG time series or the power spectrum (*e.g.*, the popular FASTER algorithm [72]). A specialized method for artifact rejection in the case of real-time single-channel EEG analysis is described in Appendix A.

Blind source separation (BSS) is commonly used in the second approach to artifact detection. In general, BSS methods, such as independent components analysis (ICA) [73, 74], seek a linear transformation between channel space and source space [75]. With a sufficient number of electrodes (the exact number depends on the method and the data, but usually a minimum of 14 EEG electrodes is recommended to achieve a useful level of precision [76]), EOG and EMG sources can be separated from brain sources. These sources can then be removed from the computed spatial filter and the remaining sources can be projected back into channel space with much of the artifact contamination removed [77].

2.6   EEG Feature Extraction

Given that EEG signals are typically highly noisy, nonstationary, and high-dimensional multichannel time series, even the cleaned signals cannot usually be effectively used for classification as they are. Instead, a feature extraction stage is necessary in order to reduce the dimensionality of the input space in a way which accentuates any discriminative information carried by the signals. In fact, in typical BCI applications, most of the information contained within the signals is irrelevant to the classification problem at hand, so feature extraction is used in order to isolate only those signal characteristics which are important for the application [9].

The approach to feature extraction used for a generalized BCI is necessarily different than what is commonly used in the BCI literature. In fact, the feature extraction stage is a key area which distinguishes a generalized BCI from standard BCIs. It is usually desirable to compute only those features which are likely to be valuable for the machine learning task at hand. Minimizing the amount of computation required is particularly important for real-time BCI applications. However, knowing which features to compute requires a significant degree of *a priori* information about the problem and the nature of the data being analyzed. A generalized BCI does not rely upon this kind of information, because it would require restricting users to predefined mental commands. Instead, the strategy used in this body of work is to compute a wide variety of features that have been found to be useful for different kinds of machine learning problems involving EEG and BCI, and then to select *a posteriori* which of those candidate features are most useful for a given subject at a given time.

Broadly speaking, there are two main classes of methods for feature extraction in the absence of sufficient *a priori* knowledge for feature engineering. The first is the most direct implementation of the approach to generalized feature extraction described above. That is, a long list of features is extracted from the EEG based on general domain knowledge (*i.e.*, knowledge about how information can

be extracted from EEG in general) rather than specific domain knowledge (*i.e.*, knowledge about how information related to the specific mental activity being studied can be extracted from EEG). In this thesis, this approach is explored using spectral analysis, which employs methods developed by the statistical signal processing community. Since there is a wide variety of spectral methods which yield different information about a signal, this approach requires minimal assumptions about what kind of information will be useful for classification or regression.

The second class of methods is based on feature learning. Feature learning methods use machine learning algorithms to find discriminative features. Because features are learned from data rather than computed prior to considering the data, feature learning methods are often more efficient in the sense that fewer features need to be computed. However, choosing an appropriate feature learning method also requires some general domain knowledge. For example, the basic feature learning method used throughout this thesis and that is most widely used in the BCI literature is Common Spatial Patterns (CSP) [78]. CSP, when applied to EEG, assumes that there exists some spatial structure to the signal topology in at least one frequency band which can be exploited in classification, and thus the use of CSP in BCI assumes that separable spatial patterns can be found for each mental command.

2.6.1   Spectral Features: Power Spectral Density

The power spectral density (PSD) is defined for a time-varying signal as the power in each frequency component, where power is the amount of work performed or energy consumed per unit time, as it is used in physics [79]. The PSD of a time-varying signal $x(t)$ is a function of its Fourier Transform. Here the finite-time Fourier Transform of $x(t)$, $\hat{x}_T$, is used because real-world EEG signals are finite in time. Thus, at frequency f, the finite time Fourier Transform over the interval $[0, T]$ can be written as

$$\hat{x}_T(f) = \frac{1}{\sqrt{T}} \int_0^T x(t)e^{-ift}dt, \tag{2.2}$$

where $i = \sqrt{-1}$. The PSD at a frequency component f is then defined as

$$P_{xx}(f) = \lim_{T \to \infty} \mathbf{E}\left[|\hat{x}_T(f)|^2\right], \tag{2.3}$$

where $\mathbf{E}$ is the expectation operator. Note that this is the time average of the squared magnitude of the Fourier Transform of $x(t)$. Equivalently, $P_{xx}(f)$ is the limit of the second moment of the magnitude of $\hat{x}_T(f)$ (in fact, the central second moment of a signal, or its variance, is simply the integral over P_{xx}). This fact is

important for understanding the bispectrum and bicoherence of a signal, which is discussed in Section 2.6.4 below.

In practice, it is of course impossible to exactly compute the PSD because this would require integration to infinity at an infinite sampling rate. Instead, the Fast Fourier Transform (FFT) is commonly used to efficiently compute the discrete Fourier Transform, which is a close approximation to the true Fourier Transform using discrete sampling and windowing [80]. In all computational analyses performed in this thesis, the PSD refers to the magnitude-squared of the FFT.

For an EEG signal, the PSD at a given electrode site or source can be interpreted as a measure of how much synchronous neural activity is taking place there at different frequency components. Changes in the PSD are known to be associated with certain neurophysiological and behavioural phenomena and are used extensively with EEG [81, 82, 83, 84]. For example, entering a relaxed mental state is known to increase power in the alpha band (generally 8-12 Hz, but the precise band varies across individuals), whereas beginning to fall asleep can be detected by a rise in theta power (usually 4-7 Hz). PSD features are used in a variety of types of BCIs [85, 10], most notably for detecting mental commands based on motor imagery [86, 87].

The PSD can be used in a variety of ways. In this thesis, PSD features are sometimes extracted from wide bands in order to reduce the number of features produced (*e.g.*, in Chapter 5), and at other times narrow bands are used to extract more fine-grained detail about a signal (*e.g.*, Chapter 4). Yet another approach, used in Appendix A, takes statistics of the PSD in various frequency bands. Hence, wide applicability of the PSD and functions of the PSD make its use popular in BCI and EEG analysis, and have led to it being used throughout the work presented in this thesis.

2.6.2   Spectral Features: Cross-Spectrum and Coherence

The cross-spectrum is simply a generalization of the power spectrum for a pair of signals. Given signals $x(t)$ and $y(t)$, the cross-spectral density (CSD) is defined as

$$P_{xy}(f) = \lim_{T \to \infty} \mathbf{E}\left[\hat{x}_T(f)\hat{y}_T(f)\right]. \tag{2.4}$$

Note that this is the mixed moment, or the cross-moment, of $\hat{x}_T$ and $\hat{y}_T$, which is also the Fourier Transform of the cross-correlation function of $x(t)$ and $y(t)$.

The coherence, also called magnitude-squared coherence, between a pair of signals is the magnitude-squared CSD normalized by the product of each signal's

PSD:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}.$$ (2.5)

Thus, coherence can be interpreted as a measure of linear similarity between two signals which approaches one as $x(t) \to y(t)$.

The CSD and coherence are included as part of spectral feature extraction in this thesis because they are simple ways of representing linear pairwise interactions between two EEG channels or brain sources [88, 89]. These measures, in practice, can be used in similar ways as the PSD, though they relay a different kind of information. Coherence has been used in BCIs based on motor imagery (*e.g.*, [90]), but is not commonly seen in the BCI context because other kinds of features are typically considered sufficient for the types of mental commands commonly used. However, certain kinds of mental activity and states do result in changes in the CSD and coherence (*e.g.*, [91, 92]). Therefore it is reasonable to include these features in a generalized BCI, where such mental activity might be used as control signals. In fact, in the work presented in Chapter 5, the CSD and coherence were found to be useful for emotional reaction recognition with EEG.

2.6.3   Spectral Features: Cross-Frequency Coupling

Cross-frequency coupling (CFC) describes linear phase-amplitude relationships between brain oscillations in different frequency ranges and reflects neural integration and information transfer across spatial and temporal scales [93, 94, 83]. A particular measure of CFC known as the weighted phase-locking factor (WPLF) [95] is used in this thesis and is shown to contribute to detection of emotional states from EEG in Chapter 5. WPLF was chosen over alternative methods because it measures both the instantaneous coupling strength and the related phase angle for a pair of signals in a way which is robust to nonstationarity in comparison to coherence. In addition, it is a measure of phase-to-phase coupling, rather than phase-to-amplitude coupling like coherence [96]. Thus WPLF contributes unique information which can be used to discriminate between mental commands or mental states compared to the other spectral features included in the SF method.

WPLF is computed as follows: if T is time and θ is the phase difference between two signals or two frequency components of the same signal, then

$$\mathrm{WPLF} = \frac{1}{T} \int_0^T e^{i\theta(t)} dt.$$ (2.6)

This measure has been used to study neurological conditions such as epilepsy [95]

and schizophrenia [97], but to date, CFC measures have not appeared in the BCI literature.

2.6.4 Spectral Features: Bispectrum, Bicoherence, and Quadratic Phase Coupling

The bispectrum of a signal is derived from a mathematical generalization of the PSD. As noted earlier, the power spectrum is simply the limit of the second moment of the magnitude of the Fourier Transform of a signal. Theoretically, related characteristics of a signal could be obtained using its moment-generating function, $M_x(t) = \mathbf{E}[e^{tx}]$ or by the cumulant-generating function of a pair of signals, $Cu_{x,y}(t) = M_{x+y}(t)$. The bispectrum is one particular case of generalizing time-frequency analysis using such generating functions from theoretical statistics (higher order spectra can be referred to generally as polyspectra) [98].

The bispectrum B of a signal is obtained by taking a two-dimensional Fourier Transform of the third-order cumulant generating function [99], where the pair of signals in $Cu_{x,y}(t)$ is represented by the signal $x(t)$ at two frequency components f_1 and f_2. Thus, for the Fourier Transform $\mathcal{F}$ its complex conjugate $\mathcal{F}^*$,

$$B(f_1, f_2) = \mathcal{F}^*(f_1 + f_2)\mathcal{F}(f_1)\mathcal{F}(f_2). \tag{2.7}$$

Bicoherence [100], like coherence, can be defined with respect to n time bins as

$$B_c(f_1, f_2) = \frac{|\sum_n \mathcal{F}_n(f_1)\mathcal{F}_n(f_2)\mathcal{F}_n^*(f_1 + f_2)|}{\sum_n |\mathcal{F}_n(f_1)\mathcal{F}_n(f_2)\mathcal{F}_n^*(f_1 + f_2)|}. \tag{2.8}$$

Since the numerator is the summed magnitude of the bispectrum across the binned segments of time, it increases monotonically with increasing phase coupling between the two frequency components and approaches zero for uncorrelated phases. The denominator is a normalization term equal to the maximum value of the numerator (*i.e.*, the numerator and denominator are equal if the pair of frequency components exhibit perfect phase coupling).

Quadratic Phase Coupling (QPC) features are obtained by performing an autoregressive analysis of the bicoherence of a signal [101]. The QPC features are simply the autoregression coefficients. These provide information about how bicoherence changes with respect to frequency.

The bispectrum, bicoherence, and QPC capture non-linear interactions between pairs of frequency components in a signal. The bispectrum gives information about non-linear interactions with respect to the signal's magnitude, while

bicoherence measures the extent of quadratic phase-amplitude coupling (as opposed to linear phase coupling as measured by WPLF and related measures). QPC, on the other hand, provides non-linear phase-phase coupling information which is not represented in the bicoherence explicitly.

Like the WPLF features described above, features based on the bispectrum have not previously been used in brain-computer interfacing. However, these measures have previously been used to analyze sleep EEG [101], EEG signals recorded from epileptic patients [102], and elsewhere in the clinical EEG literature [103]. These features are interesting and distinct from the other spectral features and from CSP due to their construction using higher order statistics and their ability to capture information about non-linear interactions in brain activity. Hence these were also included as experimental features in parts of this thesis. These features played an interesting role in the work presented in Chapter 5.

2.6.5   Common Spatial Patterns

Common Spatial Patterns (CSP) is one of the most important methods used in BCI. While it was originally designed to extract spatially localized abnormalities in clinical EEG data [104], it is now better known as the most prominent feature extraction method for BCIs based on mental imagery [105, 78]. CSP has traditionally been described as a feature extraction method for motor imagery classification. However, this association appears to be a result of the historical use of CSP rather than being based on theory. In this thesis, CSP is used for a variety of mental imagery classification tasks, showing that it can be applied more broadly than it has been in the past.

In principle, CSP is appropriate for any BCI controlled by mental commands that have distinct spatial distributions of activity over the neocortex. This is because CSP is simply a supervised statistical machine learning method for feature learning that attempts to find discriminative spatial patterns in EEG data. While it fits well with motor imagery classification because motor imagery is spatially separable in EEG, there is nothing inherent about CSP that should limit it to motor imagery. This fact, while still underutilized, has received greater appreciation over the last few years. Most notably, CSP has been successfully applied to emotional imagery recognition [106].

CSP uses supervised learning to construct a spatial filter that maximizes the variance of a signal for one class while minimizing its variance for another class. Formally, the spatial filter W is obtained by solving the problem posed by the

following objective function:

$$W = \max_{W} \frac{\|WX_1\|^2}{\|WX_2\|^2}, \tag{2.9}$$

where X_i, $i \in \{1, 2\}$, is the $N \times T_i$ matrix of EEG data comprised of N channels and the T_i samples belonging to class i in the training dataset. Therefore, W is solved by maxmizing the ratio of variance in the spatially-transformed EEG between class 1 and class 2 in the training data.

The solution for W can be found via a simple linear program. In fact, CSP can be seen as an extension of the more widely known Principle Components Analysis (PCA) [107, 108] technique to a classification problem. Where PCA uses the eigendecomposition of a covariance matrix to find components, *i.e.*, linear combinations of channels, which carry the bulk of the variance of a signal, CSP uses the eigendecomposition of a ratio of covariance matrices to find components which instead maximize the ratio of variances between two classes. Therefore CSP uses a normalized spatial covariance matrix for each class of signals, given by

$$R_1 = \frac{X_1 X_1^T}{\text{trace}(X_1 X_1^T)}, \qquad R_2 = \frac{X_2 X_2^T}{\text{trace}(X_2 X_2^T)}, \tag{2.10}$$

where T denotes the transpose of a matrix. In order to make effective use of these ratios, an extra step is required.

CSP finds components of a signal that maximize the ratio of variances between two classes in order to differentiate those classes based on the variance in those components. However, there are two desired properties of covariance matrices if these ratios are to be meaningfully compared across components and in order to ensure that these components are not derived merely on the basis of spatial correlations that exist between the two classes, which can be particularly high for EEG signals. The whitening transform of the composite covariance R_c provides both of these properties, because it transforms a signal to have a covariance matrix equal to the identity matrix.

Given the composite spatial covariance matrix R_c, where

$$R_c = R_1 + R_2, \tag{2.11}$$

and its diagonal matrix of eigenvalues λ and matrix of eigenvectors V given by the eigendecomposition

$$R_c = V\lambda V^T, \tag{2.12}$$

the whitening transform Q is obtained by

$$Q = V\sqrt{\lambda^{-1}}. \tag{2.13}$$

The whitened composite spatial covariance matrix can then be computed by taking QR_cQ^T, which equals the identity matrix I.

The reason that whitening transform enables comparisons of variance ratios across CSP components is because it normalizes the variances across channels. However, the most important property it provides is derived from the fact that Q is computed using the sum of the class covariance matrices in Equation 2.11. This ensures that

$$R_1^* = QR_1Q^T \qquad \text{and} \qquad R_2^* = QR_2Q^T \tag{2.14}$$

have common eigenvectors in V^* such that

$$R_1^* = V^*\lambda_1 V^{*T} \qquad \text{and} \qquad R_2^* = V^*\lambda_2 V^{*T}. \tag{2.15}$$

Most importantly, it means that if the rows of V^* are sorted in descending order of the values in λ, then

$$\lambda_1 + \lambda_2 = I. \tag{2.16}$$

Achieving this relationship between the eigenvalues of class-specific covariance matrices is extremely important for two reasons: first, it means the variances across CSP components can be compared, which allows a machine learning classifier to use differences in variance to discriminate between classes, and second, it means that the largest eigenvalues of the whitened covariance matrix of one class correspond to the smallest eigenvalues of the whitened covariance matrix of the other class. Therefore, these eigenvalues maximize the ratio of variance between the classes, meaning the classes can be discriminated on the basis of the variance of the CSP components.

The CSP filter that satisfies Equation 2.9 is

$$W = (V^{*T}Q)^T, \tag{2.17}$$

and the CSP components of an EEG signal X are obtained by computing

$$C = WX. \tag{2.18}$$

Therefore, multiplying X by W whitens the EEG signal X and computes linear combinations of channels whose coefficients are the rows of V^*. These two steps are illustrated in Figure 2.9.

Since V^* has its rows sorted in descending order of its corresponding eigenvalues λ, its top M and bottom M eigenvectors will have the highest ratio of variances between the two classes to be classified, where $M \in \{1, 2, \ldots [N/2]\}$ is a parameter chosen by the experimenter. Therefore, discriminative CSP features
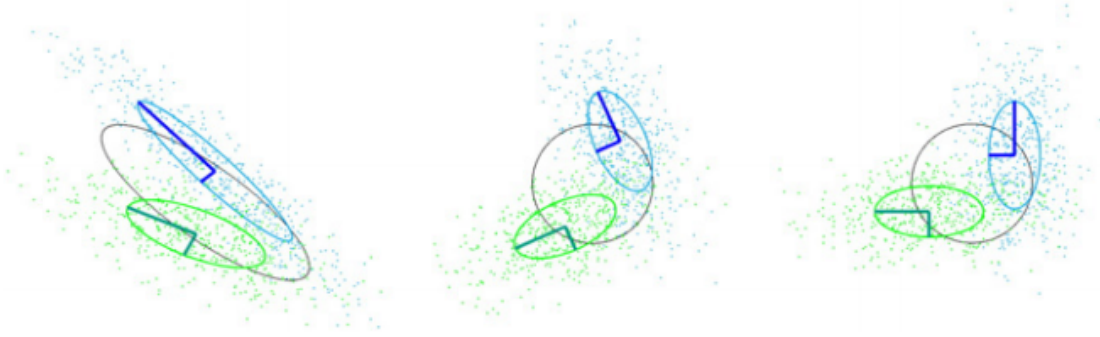
Figure 2.9: The two key steps of CSP encapsulated by the CSP filter W. **Left**: Toy data representing samples from two EEG channels are plotted as points with their corresponding covariance matrices plotted as ellipsoids. Green and blue data are used to differentiate between classes, while the grey ellipsoid represents the composite covariance matrix R_c. **Center**: Orthogonalization. The covariance matrices are orthogonalized by the whitening transform given in Equation 2.13. Since whitening the composite covariance matrix transforms its corresponding covariance ellipsoid into the unit circle, not only are the class-specific covariance matrices orthogonalized, they are also made congruent to one another. **Right**: Rotation. Transforming the whitened data with the eigenvectors in V^* projects the data into a new coordinate space where the principle axes of variance for each class are aligned to standard coordinate axes. Since the class-specific covariance matrices are both congruent and orthogonal, this rotation step ensures that on one axis, variance is maximum for one class and minimum for the other class, and vice versa for the other axis. In higher than two dimensions, these axes correspond to the directions given by the first and last eigenvectors. Image reproduced with permission from [109].

$f_j, j = 1, \ldots, 2M$, are computed by taking the log of the normalized variance for each of the $2M$ components in $Z = \{1, \ldots, M, N - M + 1, \ldots, N\}$:

$$f_J = \log \left[\frac{\text{var}(C_m)}{\sum_{i \in Z} \text{var}(C_i)} \right], \tag{2.19}$$

where $m \in Z$. It should be noted that in a machine learning experiment, W must be computed using training data and then applied later to separated test data in order to avoid overfitting.

Notice, however, that CSP can facilitate the extraction of more than just the standard CSP features defined by Equation 2.19. In fact, CSP can be used as a preprocessing step for other feature extraction methods as well. This is because the matrix $C = WX$ gives the EEG components derived from unique and discriminative spatial distributions over the scalp, and thus the components in C can be used as input for further feature extraction methods just like one might use a set of source signals or a set of channels (*e.g.*, one can make use
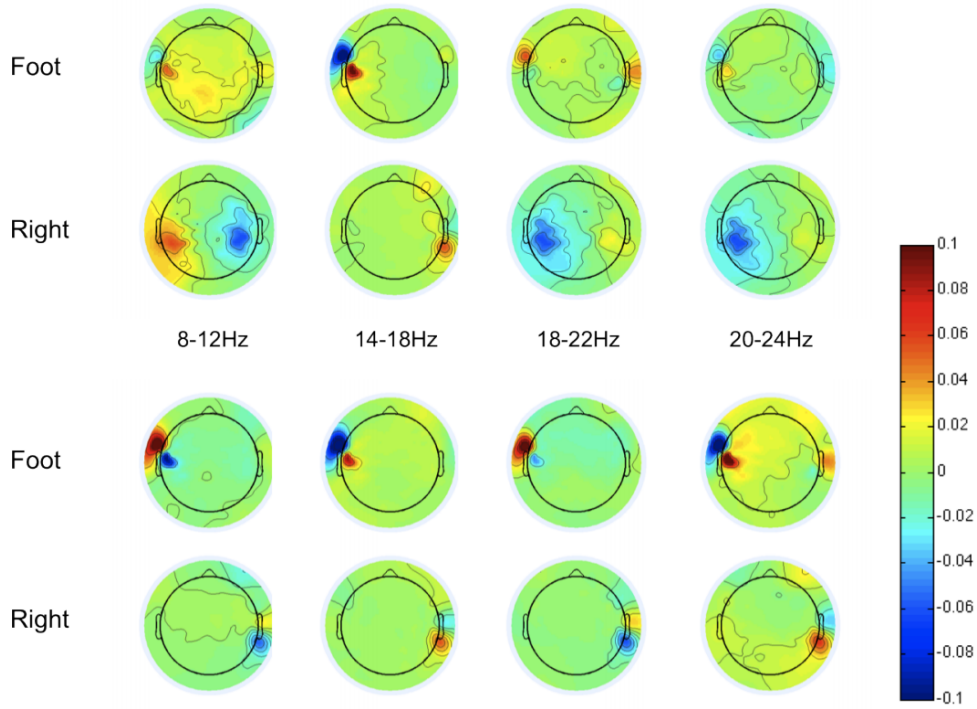
Figure 2.10: Common spatial patterns for foot and right hand motor imagery computed in different frequency bands. Common spatial patterns show the distribution of discriminative activity in the brain, but these patterns are sensitive to the frequency band chosen when constructing the filter, and typically require a fairly narrow frequency band. Figure reproduced with permission from [113] (2010 IEEE).

of PSD features from the CSP components or the original channel data, though these features would not carry the exact same information). Therefore, any feature extraction method which would normally be applied to the EEG signal in channel space, X, can also be applied to the EEG signal as represented in CSP space, C. As will be seen in Chapter 4, this fact is exploited since the BCI presented relied on PSD features extracted over the CSP components in C in order to better generalize across various kinds of mental commands.

As a linear spatial filter, CSP also provides a convenient means with which to visualize the discriminative EEG spatial patterns for each class. In particular, the columns of W^{-1} are called the common spatial patterns and are interpreted as time-invariant EEG source distributions [78, 110, 111]. Examples of these patterns are shown in Figure 2.10. Note that when $M < N$, it is more accurate to use as the spatial patterns the columns of $\overline{W}^{-1} = \text{cov}(x) \times W \times \text{cov}(C)^{-1}$ [112].

Robustness of the Common Spatial Patterns Algorithm

The CSP algorithm, while powerful for classifying spatially separable patterns in EEG, electrocorticography (ECoG), or MEG, has a tendency to overfit, or produce a model which is overly specific with respect to the training data and therefore generalizes unsatisfactorily to new test data [114]. There are several reasons for this problem [115]. For instance, CSP is sensitive to artifacts because artifacts typically cause spikes in variance and covariance in certain regions of EEG topology. If there is a correlation between the mental commands and certain artifacts, this can result in the CSP filter becoming dependent on those artifacts for discrimination. When those artifacts are not correlated with the mental commands, then they can introduce significant noise in those spatial filters. CSP components are also not stationary, and this is particularly problematic if only a small amount of training data is available (*e.g.*, less than 50 trials).

Since the introduction of CSP, several extensions have been developed which improve upon the base algorithm in one or more ways. For example, the algorithm has been modified to extend to multi-class analysis [116], to incorporate regularization [117], and to add robustness to nonstationary EEG signals [118]. In this thesis focus is placed on Filter-Bank Common Spatial Patterns (FBCSP) [119], which is an especially applicable method for the experiments discussed later in this thesis because it allows for the discovery of discriminant CSP features without *a priori* knowledge of the frequency band in which those features should be found. Therefore FBCSP is the most appropriate for a generalized approach to BCI.

Filter-Bank Common Spatial Patterns

The ability to classify EEG in a BCI using CSP is highly dependent on constructing W using EEG signals which have been filtered to a narrow band of frequencies just large enough to contain the brain activity relevant to the mental imagery performed by the BCI user. FBCSP allows one to capture useful CSP components in narrow frequency bands when the optimal band cannot be determined *a priori* [119, 120]. This property makes it especially useful for a generalized BCI. FBCSP accomplishes this simply by computing a CSP filter W_f for each of a set of bandpass filters $f \in \mathcal{F}$, also known as a filter-bank. CSP features are computed using each W_f, and a feature selection method (see Section 2.7.2) is applied to choose features useful for classification. Figure 2.11 shows a schematic of the FBCSP method.
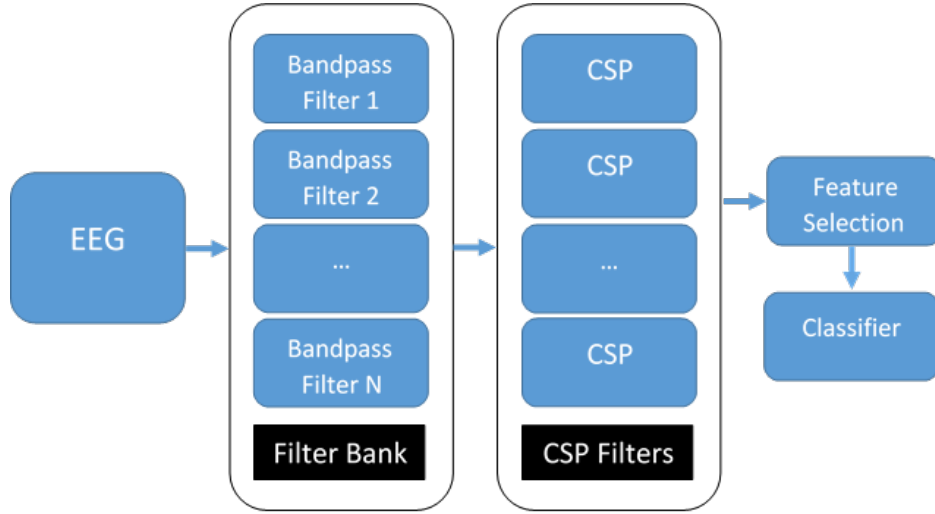
Figure 2.11: The FBCSP scheme.

Multiclass Extensions of Common Spatial Patterns

A machine learning method that can only be applied to binary classification would be extremely limited in its real world application. Since it is of great interest to develop BCIs with more than just two controls, several extensions of CSP algorithms to multiclass classification have been developed [116, 121, 122, 120]. Here, two such methods are described, each of which is best applied to one of two common BCI scenarios. Note that these two methods are not necessarily exclusive to CSP, but are simply the primary multiclass extensions used with CSP.

The One-Versus-Rest (OVR) approach to multiclass CSP is most applicable when a BCI uses three or more controls that cannot be organized hierarchically [122, 120]. For example, if a BCI allows the user to move a cursor in the four cardinal directions on a screen using four distinct mental commands, the OVR extension for CSP is a typical choice. In the OVR algorithm, K CSP filters are computed for a K-class classification problem. Each CSP filter (or set of filters in the case of FBCSP) is constructed using any version of CSP by treating one of the K classes as class one and the aggregate of all of the remaining $K-1$ classes as class two. Therefore, each CSP model is designed for binary classification such that it predicts whether the given data belong in the specific class associated with that model or one of the other classes. When performing classification, the choice of which $k \in K$ is the most likely true label for the given EEG signal x is determined by

$$k = \max_{k \in K} P(k|x), \tag{2.20}$$

where $P(k|x)$ is the probability estimate or confidence level of x belonging to class k. How this probability estimate is determined depends on the specific

classifier used. Alternatively, the CSP features from each of the K filters can be aggregated into a single feature vector and K-way classification can be performed using a single classifier.

In some cases it is more appropriate to use a hierarchical, or tree-based, classification approach. For example, Chapter 7 discusses a BCI designed to enable a user to answer "Yes" or "No" questions in real time without requiring time-locked questions. Therefore, the BCI must also determine when the user is attempting to answer a question and when the user is simply not engaging with the BCI at all. This is posed as a three-way classification problem, including the classes "Yes", "No", and "Rest". Since there is a natural hierarchy to this classification problem, where the system can first determine whether the user is trying to answer a question and then determine which answer they are trying to give, the Divide and Conquer (DC) approach to multiclass CSP is a good fit [123, 120]. Since classification is hierarchical, instead of K CSP filters as in the OVR approach, only $P < K$ CSP filters are required, where P is the number of levels in the hierarchy (in the example given, (in the example given, $K = 3$ and $P = 2$).

It is often the case that either the OVR or DC approaches to multiclass CSP can be used for the same problem. Returning to the first example given here, controlling a cursor along all four cardinal directions can be posed hierarchically by first determining along which axis the user is attempting to exert control, and then determining the direction along that particular axis. The choice is up to the experimenter. Since the DC approach is less computationally intensive, it might be preferred here. On the other hand, the DC approach requires two-step classification before the cursor can move, which might slow down control. It should also be considered that the user will often wish to move the cursor diagonally on the screen. This would make the initial choice of axis particularly difficult, as the desire to reach a button in the top right corner of the screen, for example, could illicit brain activity for upwards movements and right-wards movements simultaneously. In the OVR approach, the BCI might, in this scenario, be more likely to misclassify between up and right movements. However, both of these still help the user reach their goal. Both methods have their merits and drawbacks for this application. The choice will depend on what kind of user interface is desired.

2.7   Machine Learning for BCI

Machine learning algorithms are broadly categorized as either being supervised or unsupervised [124, 125]). Supervised learning means learning with respect to a known ground truth. In other words, the data are labeled and there are defined correct answers that the machine must learn to produce or approximate. Super-

vised learning problems are typically either classification problems or regression problems. Classification and regression are highly related in the sense that both require fitting a function which uses input features to predict an output. In the case of classification, the output is a discrete categorical variable, whereas in regression, the output is a continuous variable.

In contrast, unsupervised learning problems have no such labels from which to learn and it can be more difficult to objectively measure the algorithm's performance depending on the problem and the particular algorithm. The quintessential example of unsupervised learning is clustering, which refers to the problem of identifying naturally occurring subgroups of data points that are more similar to one another than to members in other groups. This thesis is focused almost exclusively on supervised learning problems, and so unsupervised machine learning methods will not be addressed in great detail. The following sections provide an overview of the supervised learning methods used throughout this thesis for feature selection, classification, and regression.

2.7.1   Feature Selection

Feature selection is a form of dimensionality reduction particular to classification problems. In other contexts where dimensionality reduction is used, the goal might be to find the set of variables, or combinations of variables, which contain the bulk of the information of the original dataset (*e.g.*, as in PCA). For classification, the goal is to find only the features that contribute to improved classification accuracy and to omit the other features.

Feature selection is often very important in classification problems due to an appropriately named problem called the "Curse of Dimensionality". Each feature in a dataset is a dimension in a feature space (*i.e.*, it forms an axis). If the number of samples is held constant and the number of dimensions is increased, the data become exponentially sparser in feature space. Machine learning algorithms attempt to model the relationships between data in feature space in order to partition this space into regions belonging to different classes. When the number of features is large compared to the number of number of training samples, the machine learning algorithm may not have the diversity and density of data needed to effectively and reliably learn how these data are distributed. Therefore, given a fixed number of training samples, the predictive power of a machine learning algorithm decreases as the number of dimensions increases. This is known as Hughes phenomenon [126].

Dealing with unfavourable ratios of features to training samples is unavoidable in practice. In most machine learning experiments, the data collection process
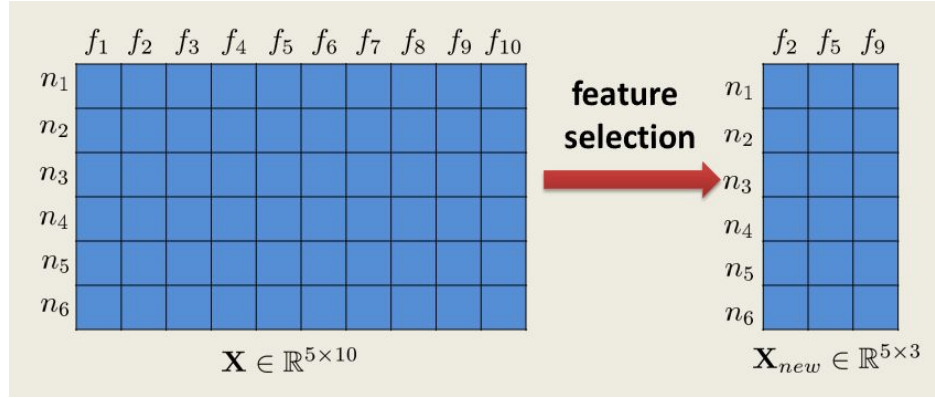
Figure 2.12: An illustration of feature selection.

usually does not consider the needs of the machine learning algorithm. Even when the data are collected for the purpose of machine learning and the properties of the algorithm are used to guide data collection, three common factors lead to the Curse of Dimensionality: 1) if the predictive features needed for classification are not known *a priori*, a common approach is to extract many features and find which of those are useful, 2) it is often simply not practical or possible to collect a large number of samples, and 3) the data may not be classifiable using only a small number of features. EEG data and data used in BCI experiments often suffer from all three of these deficiencies. Feature selection is used to address the first problem by finding the most predictive features and using only those for classification. Further attention is given to Minimum-Redundancy, Maximum-Relevance (MRMR) [127], which is used throughout this thesis.

Minimum-Redundancy Maximum-Relevance

MRMR [127] uses a supervised learning approach to feature selection, and thus requires a training set $F_{M \times N}$ comprised of M candidate features and N training samples, as well as a vector of class labels $Y_{1 \times N}$. MRMR goes beyond selecting just those features that best predict Y and also considers which predictive features contribute distinct discriminative information. As such, MRMR is particularly useful when the set of candidate features are highly correlated amongst one another, as is usually the case with EEG.

The objective is to find a subject of features $F' \in F$ that has maximum joint mutual information with the true class labels Y, thus satisfying the maximum-relevance criterion, while simultaneously keeping the mutual information between the features in F' low, satisfying the minimum-redundancy criterion. Furthermore, this min-max problem is solved given the experimenter-determined parameter $M' < M$, which is the number of features to include in F'. These two criteria

can be combined into one objective function by simply taking the difference

$$\max_{F'|M'}(D - R), \tag{2.21}$$

where D is the dependency of Y on F' and R is the redundancy within F'.

The optimal solution for F' given M' is obtained iteratively by selecting features one at a time. The first feature, f_1 is found by identifying the feature $x_i \in F$ which has the highest mutual information with the class labels in the training data, *i.e.*,

$$f_1 = \max_i I(F = \{x_i, i = 1, \dots, M\}; Y), \tag{2.22}$$

where I is the mutual information function defined for two random variables Z_1 and Z_2 as

$$I(Z_1; Z_2) = \int_{Z_1} \int_{Z_2} p(z_1, z_2) \log\left(\frac{p(z_1, z_2)}{p(z_1)p(z_2)}\right). \tag{2.23}$$

Subsequent features $f_2, \dots, f_{M'}$ are selected by iteratively choosing features from F which maximize $D - R$. Here D is defined for the mth matrix F'_m of selected features as

$$D = I(F'_m = \{f_i, i = 1, \dots, m < M'\}; Y), \tag{2.24}$$

and is estimated by

$$\bar{D} = \frac{1}{m} \sum_{f_i \in F'_m} I(f_i; Y). \tag{2.25}$$

Similarly, R is defined as

$$R = I(F'_m; F'_m) \tag{2.26}$$

and is estimated by

$$\bar{R} = \frac{1}{m^2} \sum_{f_i, f_j \in F'_m} I(f_i; f_j). \tag{2.27}$$

These estimates must be used in practice because they avoid computing the computationally expensive or even intractable multivariate joint probability densities with which D and R are defined. In fact, calculating joint probability densities with many variables given a finite amount of data is another instance where the Curse of Dimensionality poses a challenge and may lead to innaccurate estimates [127].

2.7.2   Classification and Regression

Much of this thesis revolves around classification of EEG signals, with some use of regression in Chapter 5. In this section the basic notions of classification and regression are made clear, and the methods used throughout this thesis are explained in detail. The classifiers described here are the Support Vector Machine (SVM) [128] and Logistic Regression with Elastic Net Regularlization (LRE) [129]. The specific regression method described here is the Boosted Decision Tree (BDT) [130].

Classification Analysis

The purpose of a classifier is to generalize from past or observed labeled data in order to assign a discrete categorical label to new data. Most often this is done by finding a 'classification boundary' in feature space which optimally separates previously observed data. For simplicity, begin with the linearly separable two-class scenario illustrated by Figure 2.13. In the two-class case, once a classification boundary is determined, classifying new data becomes a simple task. Any datum that lies on one side of the border is assumed to belong to the class that is most represented on that side of the border, and vice versa. This is true also for non-linear classification boundaries (see Figure 2.13), as the curve merely serves to partition the feature space into two non-overlapping boundaries. Because optimizing this partitioning is the essential basis for classification, the concept can easily be extended to more than two classes. In multi-way classification, as illustrated in Figure 2.13, multiple boundaries can be used to partition the feature space into one region per class. The question of how to compute these boundaries given labeled data will be discussed below.

Regression Analysis

Regression analysis is often thought of as a means to find a 'line of best fit', for some definition of 'best' (*e.g.*, least squares), to continuously distributed rational data. However, regression more broadly refers to the prediction along a continuous variable of an output value with respect to a realized set of input values [131]. One key property of regression that is made more apparent when using its broader definition is that regression analysis does not need to assume a continuous polynomial function over its domain. This flexibility is particularly exploited by machine learning regression methods as opposed to classical statistical regression, which typically seeks to find a continuous function over the domain of a set of dependent variables rather than seeking to attribute a continuous-valued number

(a) Linear Classification   (b) Non-Linear Classification   (c) Multi-Way Classification
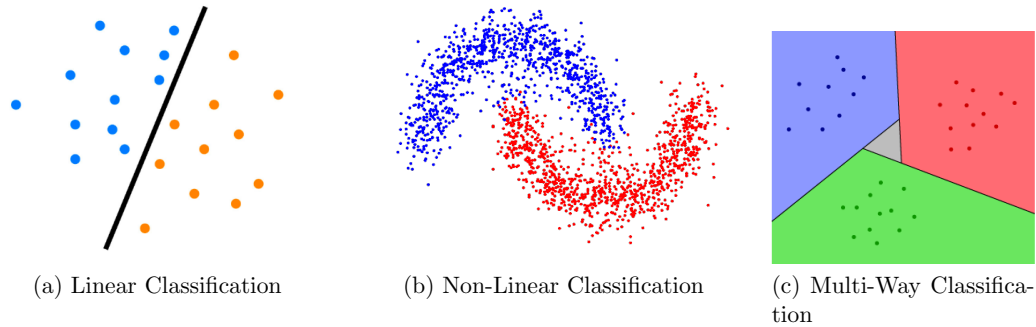
Figure 2.13: Illustrated examples of different classification problems. a) An illustration of a simple classification problem: data from two classes (shown in blue versus orange dots) are separated by a linear classification boundary. b) An illustration of the "horseshoe" problem for which non-linear classification is required. c) An example of multi-way classification, which, rather than finding one classification boundary, uses multiple boundaries in order to partition the feature spaces into class-defined regions.

to each particular observed data point. This difference is illustrated in Figure 2.14.

Cross-Validation

Cross-validation refers to a set of model validation methods used to evaluate the results of statistical or machine learning models on independent test data. Cross-validation can also be used for model selection if model parameters are adjusted across cross-validation runs [132]. Note that if cross-validation is used for model selection, an inner cross-validation loop is required to evaluate each unique set of model parameters.

For predictive models, such as classification and regression models, results obtained without cross-validation are likely to be biased due to overfitting [132]. This is unsurprising, because if the same data used to train a model is then used to test the same model, some degree of memorization is likely to be reflected in any evaluation metric. Cross-validation avoids this source of overfitting by using mutually exclusive partitioned training and test data. The model is trained using the training data and the evaluation criterion is estimated by comparing the predicted values for the withheld test data to its true values. In order to obtain a representative estimate of model performance, this process is repeated several times.

Figure 2.15 illustrates a common cross-validation method called K-fold cross-
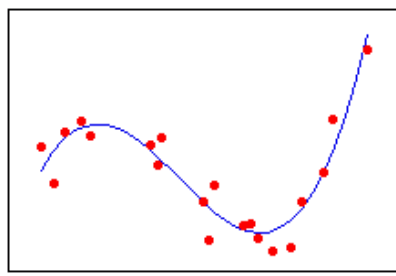
(a) Polynomial Regression   (b) ML Regression
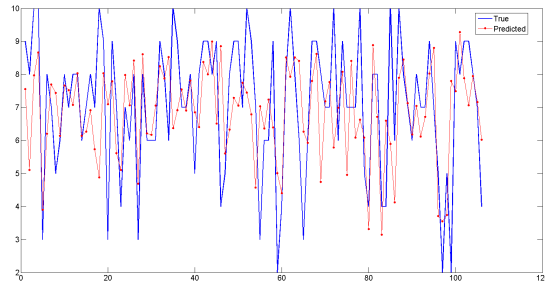
Figure 2.14: **a)** Illustration of classical regression analysis. b) Illustration of some machine learning regression models which do not assume continuity in the domain (*e.g.*, Boosted Decision Trees [130]). In classical regression, continuity is assumed along the domain. Therefore, it is possible to interpolate between two observed values on the x-axis (*e.g.*, if a new data point x_{new} is halfway between two adjacent previously observed values x_1 and x_2, then its predicted value y_{new} can be estimated to be the point on the regression curve halfway between x_1 and x_2). In contrast, certain machine learning regression models do not interpolate between adjacent values on the x-axis. That is, since continuity is not assumed in the domain, the predicted value between any two values on the x-axis is unrelated to the predicted values for those two points. While forgoing this assumption makes modeling more challenging, it allows for highly non-linear prediction.
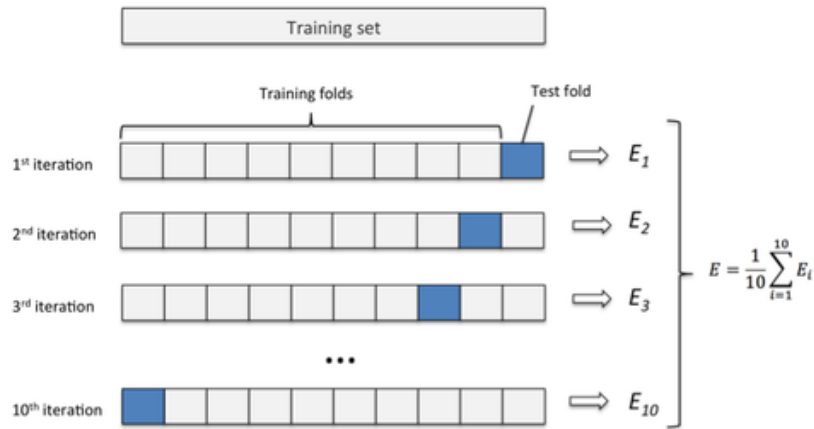
Figure 2.15: The K-fold cross-validation scheme [133]. Each of the K partitions is used as a test set once and the remaining K-1 partitions are used for training.

validation. In this approach, the full dataset is divided into K, which is often set to 10, partitions of approximately equivalent size. Each partition is withheld as a test set once and the remaining K-1 partitions are used for training. When the number of samples N is small (*e.g.*, N$\leq$20), K is often set to N so that each individual sample is used as a test set once. This is referred to as Leave-One-Out cross-validation. In addition, when the N is sufficiently large and particular proportions are desired when partitioning the data into training and test sets or a large number of cross-validation runs is desired, an alternative is to randomly partition data at the start of each cross-validation run. This latter approach is used throughout this thesis, typically with test sets making up 25% of the available data.

Classification with Support Vector Machines

A classifier that is used throughout this thesis is the Support Vector Machine (SVM) [128]. The SVM is a popular classifier because, for problems where a set of discriminative features are computed from prior knowledge, it is often one of the most effective and efficient methods. The SVM is effective because it goes beyond simply finding a hyperplane that separates two classes in two ways. First, it also maximizes the margin between the classification boundary and the nearest data samples on either side (these samples are called the support vectors, because they define the classification boundary). In addition, the SVM can compute non-linear classification boundaries by using a kernel, or a transformation function, applied to the input space in order to make the classes linearly separable in a new feature space. The linear classification boundary can then be projected back into the original input space in which it is non-linear using the inverse of the kernel.
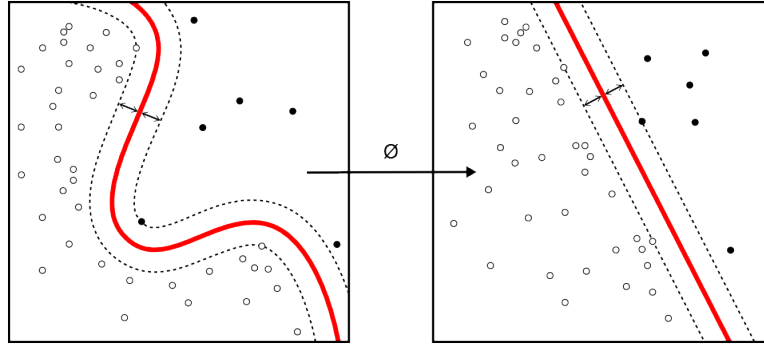
Figure 2.16: An illustration of the kernel method in the SVM, which transforms the data to become linearly classifiable, and then projects the classification boundary back into the original feature space. Figure obtained from [134].

Both of these principles are illustrated in Figure 2.16.

The current standard SVM algorithm uses a soft margin approach, because it generalizes to cases where the data are not linearly separable. Given a dataset of N samples and P features $x_i \in X_{N \times P}$, $i = 1, \ldots, N$ with labels $y_i \in Y_{N \times 1}$ (for the SVM algorithm, $Y \in \{-1, 1\}$), the goal is to find the maximum-margin hyperplane $w \cdot x - b = 0$ that separates the set $\{x_i \in X \mid y_i = 1\}$ from the set $\{x_i \in X \mid y_i = -1\}$. Here w is the normal vector to the hyperplane.

The solution for the maximum-margin hyperplane is given by an optimization problem with respect to the hinge loss function

$$h_i = \max\left(0, 1 - y_i(w \cdot x_i - b)\right). \tag{2.28}$$

The value of h_i is zero if x_i is on the correct side of the margin (*i.e.*, if the prediction $\hat{y}_i$ is correct given the margin, or if $y_i(w \cdot x_i - b) \geq 1$). Otherwise, h_i is proportional to the distance between x_i and the margin. Therefore, the maximum-margin hyperplane can be solved using the following optimization problem:

$$\min_{w,b} \text{mean}(h_i) + \lambda \|w\|^2 = \min_{w,b} \left[\frac{1}{N} \sum_{i=1}^{N} \max\left(0, 1 - y_i(w \cdot x_i - b)\right) \right] + \lambda \|w\|^2, \tag{2.29}$$

where λ is a parameter which controls the tradeoff between having a wide margin, which increases the number of samples that fall within the margin because the data are inherently not linearly separable, and maximizing the number of samples which fall on the correct side of the margin.

To extend the SVM to solve for non-linear classification boundaries, the so-called 'kernel trick' is used [135, 136, 128]. The kernel trick is implemented by transforming the data according to some nonlinear function ϕ, commonly a Gaus-

sian radial basis function $\phi(x) = exp(-\lambda\|x_i\|^2)$. Thus each x_i is projected in a nonlinear way into a new feature space using $\phi(x_i)$ and $w = \sum_{i=1}^{N} c_i y_i \phi(x_i)$ with c_i obtained by solving the maximization problem

$$\max \sum_{i=1}^{N} c_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i c_i (\phi(x_i) \cdot \phi(x_j)) y_i c_j \tag{2.30}$$

subject to $\sum c_i y_i = 1$ and $\frac{1}{2N\lambda} \geq c_i \geq 0 \ \forall i$. The kernel trick is represented by the inner product $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ and serves to linearize a nonlinear classification problem by projecting the data into a higher dimensional, and potentially non-Euclidean, space in which the data are separable by a hyperplane.

Having solved for the maximum-margin hyperplane in the transformed feature space, new data can be classified by measuring on which side of the classification boundary they fall:

$$x_{new} \mapsto \text{sign}(w \cdot \phi(x_{new}) + b) \tag{2.31}$$

Bibliography

[1] J. M. R. Delgado, *Physical control of the mind: Toward a psychocivilized society.* Harper & Row New York, 1969.

[2] E. E. Fetz and O. A. Smith, "Operant conditioning of precentral cortical cell activity in awake monkeys," in *Federation Proceedings*, vol. 28, p. 521, Federation Amer Soc Exp Biol 9650 Rockville Pike, Bethesda, MD 20814-3998, 1969.

[3] J. A. Grahn, J. A. Parkinson, and A. M. Owen, "The role of the basal ganglia in learning and memory: neuropsychological studies," *Behavioural brain research*, vol. 199, no. 1, pp. 53–60, 2009.

[4] J. J. Vidal, "Toward direct brain-computer communication," *Annual review of Biophysics and Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973.

[5] J. J. Vidal, "Real-time detection of brain events in EEG," *Proceedings of the IEEE*, vol. 65, no. 5, pp. 633–641, 1977.

[6] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, "Brain-computer interfaces based on visual evoked potentials," *IEEE Engineering in medicine and biology magazine*, vol. 27, no. 5, pp. 64–71, 2008.

[7] C. Brunner, N. Birbaumer, B. Blankertz, C. Guger, A. Kübler, D. Mattia, J. d. R. Millán, F. Miralles, A. Nijholt, E. Opisso, *et al.*, "BNCI Horizon 2020: towards a roadmap for the BCI community," *Brain-computer interfaces*, vol. 2, no. 1, pp. 1–10, 2015.

[8] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 113, pp. 767–91, June 2002.

[9] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine.," *Mayo Clinic proceedings. Mayo Clinic*, vol. 87, pp. 268–79, mar 2012.

[10] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review.," *Sensors (Basel, Switzerland)*, vol. 12, pp. 1211–79, Jan. 2012.

[11] H. Heinrich, H. Gevensleben, and U. Strehl, "Annotation: Neurofeedback - Train your brain to train behaviour," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 48, no. 1, pp. 3–16, 2007.

[12] L. H. Sherlin, M. Arns, J. Lubar, H. Heinrich, C. Kerson, U. Strehl, and M. B. Sterman, "Neurofeedback and Basic Learning Theory: Implications for Research and Practice," *Journal of Neurotherapy*, vol. 15, pp. 292–304, Oct. 2011.

[13] E. Curran, "Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems," *Brain and Cognition*, vol. 51, pp. 326–336, Apr. 2003.

[14] J. Pineda, D. S. Silverman, A. Vankov, J. Hestenes, *et al.*, "Learning to control brain rhythms: making a brain-computer interface possible," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, no. 2, pp. 181–184, 2003.

[15] C. Neuper and G. Pfurtscheller, "Neurofeedback training for BCI control," in *Brain-Computer Interfaces*, pp. 65–78, Springer, 2010.

[16] P.-J. Kindermans, M. Tangermann, K.-R. Müller, and B. Schrauwen, "Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training erp speller," *Journal of neural engineering*, vol. 11, no. 3, p. 035005, 2014.

[17] P. Yuan, X. Chen, Y. Wang, X. Gao, and S. Gao, "Enhancing performances of SSVEP-based brain–computer interfaces via exploiting inter-subject information," *Journal of neural engineering*, vol. 12, no. 4, p. 046006, 2015.

[18] R. J. Kobler and R. Scherer, "Restricted boltzmann machines in sensory motor rhythm brain-computer interfacing: a study on inter-subject transfer and co-adaptation," in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, pp. 000469–000474, IEEE, 2016.

[19] S. Saha, K. I. Ahmed, R. Mostafa, A. H. Khandoker, and L. Hadjileontiadis, "Enhanced inter-subject brain computer interface with associative sensorimotor oscillations," *Healthcare Technology Letters*, vol. 4, no. 1, pp. 39–43, 2017.

[20] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: improved accuracy by response verification," *IEEE transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 326–333, 1998.

[21] P. Yuan, X. Gao, B. Allison, Y. Wang, G. Bin, and S. Gao, "A study of the existing problems of estimating the information transfer rate in online brain–computer interfaces," *Journal of neural engineering*, vol. 10, no. 2, p. 026014, 2013.

[22] 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, 1992.

[23] N. Birbaumer, T. Elbert, A. G. Canavan, and B. Rockstroh, "Slow potentials of the cerebral cortex and behavior.," *Physiological reviews*, vol. 70, no. 1, pp. 1–41, 1990.

[24] "Event-related potentials." `https://en.wikipedia.org/wiki/Event-related_potential`. Accessed: 2017-05-01.

[25] J. Polich and A. Kok, "Cognitive and biological determinants of P300: an integrative review," *Biological psychology*, vol. 41, no. 2, pp. 103–146, 1995.

[26] J. Mak, Y. Arbel, J. Minett, L. 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, 2011.

[27] D. Marshall, D. Coyle, S. Wilson, and M. Callaghan, "Games, gameplay, and BCI: the state of the art," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 5, no. 2, pp. 82–99, 2013.

[28] J. I. Münßinger, S. Halder, S. C. Kleih, A. Furdea, V. Raco, A. Hösle, and A. Kübler, "Brain painting: First evaluation of a new brain-computer interface application with ALS patients and healthy volunteers.," *Frontiers in neuroscience*, vol. 4, no. November, p. 182, 2010.

[29] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 Speller," *Journal of neural engineering*, vol. 3, no. 4, p. 299, 2006.

[30] L. Korczowski, M. Congedo, and C. Jutten, "Single-trial classification of multi-user P300-based brain-computer interface using Riemannian geometry," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 1769–1772, IEEE, 2015.

[31] N. V. Manyakov, N. Chumerin, A. Combaz, and M. M. Van Hulle, "Comparison of classification methods for P300 brain-computer interface on disabled subjects," *Computational intelligence and neuroscience*, vol. 2011, p. 2, 2011.

[32] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, *et al.*, "Review of the BCI competition IV," *Frontiers in neuroscience*, vol. 6, p. 55, 2012.

[33] M. Jeannerod, "Mental imagery in the motor context," *Neuropsychologia*, vol. 33, no. 11, pp. 1419–1432, 1995.

[34] J. Decety, "The neurophysiological basis of motor imagery," *Behavioural brain research*, vol. 77, no. 1, pp. 45–52, 1996.

[35] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 316–325, 1998.

[36] G. Pfurtscheller, "Functional topography during sensorimotor activation studied with event-related desynchronization mapping.," *Journal of Clinical Neurophysiology*, vol. 6, no. 1, pp. 75–84, 1989.

[37] Z. A. Keirn and J. I. Aunon, "A new mode of communication between man and his surroundings," *IEEE transactions on biomedical engineering*, vol. 37, no. 12, pp. 1209–1214, 1990.

[38] W. Lang, D. Cheyne, P. Höllinger, W. Gerschlager, and G. Lindinger, "Electric and magnetic fields of the brain accompanying internal simulation of movement," *Cognitive brain research*, vol. 3, no. 2, pp. 125–129, 1996.

[39] G. Pfurtscheller and A. Berghold, "Patterns of cortical activation during planning of voluntary movement," *Electroencephalography and clinical neurophysiology*, vol. 72, no. 3, pp. 250–258, 1989.

[40] G. Pfurtscheller and F. L. Da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.

[41] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain topography*, vol. 12, no. 3, pp. 177–186, 2000.

[42] E. V. Friedrich, R. Scherer, and C. Neuper, "The effect of distinct mental strategies on classification performance for brain–computer interfaces," *International Journal of Psychophysiology*, vol. 84, no. 1, pp. 86–94, 2012.

[43] C. Jeunet, B. NKaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns," *PloS one*, vol. 10, no. 12, p. e0143962, 2015.

[44] B. Z. Allison and C. Neuper, "Could anyone use a BCI?," in *Brain-computer interfaces*, pp. 35–54, Springer, 2010.

[45] A. B. Randolph, "Not all created equal: Individual-technology fit of brain-computer interfaces," *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 572–578, 2011.

[46] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain–computer interface: A brief review," *Journal of neuroscience methods*, vol. 243, pp. 103–110, 2015.

[47] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance," *Neuroimage*, vol. 51, no. 4, pp. 1303–1309, 2010.

[48] A. B. Randolph, M. M. Jackson, and S. Karmakar, "Individual characteristics and their effect on predicting mu rhythm modulation," *Intl. Journal of Human–Computer Interaction*, vol. 27, no. 1, pp. 24–37, 2010.

[49] E. M. Hammer, S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, K. R. Müller, and A. Kübler, "Psychological predictors of SMR-BCI performance," *Biological Psychology*, vol. 89, no. 1, pp. 80–86, 2012.

[50] A. Vuckovic and B. A. Osuagwu, "Using a motor imagery questionnaire to estimate the performance of a brain–computer interface based on object oriented motor imagery," *Clinical Neurophysiology*, vol. 124, no. 8, pp. 1586–1595, 2013.

[51] F. Lotte, F. Larrue, and C. Mühl, "Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design.," *Frontiers in human neuroscience*, vol. 7, no. September, p. 568, 2013.

[52] B. Allison, "The I of BCIs: Next generation interfaces for brain–computer interface systems that adapt to individual users," *Human-Computer Interaction. Novel Interaction Methods and Techniques*, pp. 558–568, 2009.

[53] C. Jeunet, A. Cellard, S. Subramanian, M. Hachet, B. N'Kaoua, and F. Lotte, "How well can we learn with standard BCI training approaches? A pilot study.," in *6th International Brain-Computer Interface Conference*, 2014.

[54] C. Jeunet, E. Jahanpour, and F. Lotte, "Why standard brain-computer interface (BCI) training protocols should be changed: An experimental study," *Journal of neural engineering*, vol. 13, no. 3, p. 036024, 2016.

[55] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 594–611, 2006.

[56] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015.

[57] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[58] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional magnetic resonance imaging*, vol. 1. Sinauer Associates Sunderland, 2004.

[59] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, vol. 412, no. 6843, pp. 150–157, 2001.

[60] "An fMRI machine (Stock Photo)." `http://today.uconn.edu/2014/01/fmri-machine-will-expand-research-capabilities/`. Accessed: 2017-05-10.

[61] H. Berger, "Über das elektrenkephalogramm des menschen," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 87, no. 1, pp. 527–570, 1929.

[62] B. Burle, L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal, "Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view," *International Journal of Psychophysiology*, vol. 97, no. 3, pp. 210–220, 2015.

[63] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields.* Oxford University Press, USA, 1995.

[64] Neurofeedback Australia. `http://www.neurofeedbackaustralia.com.au/`. Accessed: 2017-01-16.

[65] R. D. Pascual-Marqui, "Review of methods for solving the EEG inverse problem," *International journal of bioelectromagnetism*, vol. 1, no. 1, pp. 75–86, 1999.

[66] R. Grech, T. Cassar, J. Muscat, K. P. Camilleri, S. G. Fabri, M. Zervakis, P. Xanthopoulos, V. Sakkalis, and B. Vanrumste, "Review on solving the inverse problem in EEG source analysis," *Journal of neuroengineering and rehabilitation*, vol. 5, no. 1, p. 25, 2008.

[67] J. H. Lui, D. V. Hansen, and A. R. Kriegstein, "Development and evolution of the human neocortex," *Cell*, vol. 146, no. 1, pp. 18–36, 2011.

[68] S. Lodato and P. Arlotta, "Generating neuronal diversity in the mammalian cerebral cortex," *Annual Review of Cell and Developmental Biology*, vol. 31, pp. 699–720, 2015.

[69] "Digital filters." `https://en.wikipedia.org/wiki/Digital_filter`. Accessed: 2017-05-03.

[70] S. Butterworth, "On the theory of filter amplifiers," *Wireless Engineer*, vol. 7, no. 6, pp. 536–541, 1930.

[71] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey.," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 118, no. 3, pp. 480–494, 2007.

[72] H. Nolan, R. Whelan, and R. Reilly, "FASTER: fully automated statistical thresholding for EEG artifact rejection," *Journal of neuroscience methods*, vol. 192, no. 1, pp. 152–162, 2010.

[73] R. N. Vigário, "Extraction of ocular artefacts from EEG using independent component analysis," *Electroencephalography and clinical neurophysiology*, vol. 103, no. 3, pp. 395–404, 1997.

[74] R. N. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings.," *IEEE transactions on bio-medical engineering*, vol. 47, no. 5, pp. 589–593, 2000.

[75] S. Choi, A. Cichocki, H.-M. Park, and S.-Y. Lee, "Blind source separation and independent component analysis: A review," *Neural Information Processing-Letters and Reviews*, vol. 6, no. 1, pp. 1–57, 2005.

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

[77] T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts: comparison between ica and pca," in *Neural Networks for Signal Processing VIII, 1998. Proceedings of the 1998 IEEE Signal Processing Society Workshop*, pp. 63–72, IEEE, 1998.

[78] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement.," *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 8, no. 4, pp. 441–446, 2000.

[79] J. Y. Stein, *Digital signal processing: a computer science perspective.* John Wiley & Sons, Inc., 2000.

[80] C. Van Loan, *Computational frameworks for the fast Fourier transform.* SIAM, 1992.

[81] G. Buzsaki, *Rhythms of the Brain.* Oxford University Press, 2006.

[82] L. M. Ward, "Synchronous neural oscillations and cognitive processes," *Trends in cognitive sciences*, vol. 7, no. 12, pp. 553–559, 2003.

[83] R. T. Canolty and R. T. Knight, "The functional role of cross-frequency coupling," *Trends in cognitive sciences*, vol. 14, no. 11, pp. 506–515, 2010.

[84] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain research reviews*, vol. 29, no. 2, pp. 169–195, 1999.

[85] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *Journal of Neural engineering*, vol. 4, no. 2, p. R32, 2007.

[86] H. Yuan and B. He, "Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1425–1435, 2014.

[87] L. Acqualagna, L. Botrel, C. Vidaurre, A. Kübler, and B. Blankertz, "Large-scale assessment of a fully automatic co-adaptive motor imagery-based brain computer interface," *PloS one*, vol. 11, no. 2, p. e0148886, 2016.

[88] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clinical neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.

[89] M. Murias, S. J. Webb, J. Greenson, and G. Dawson, "Resting state cortical connectivity reflected in EEG coherence in individuals with autism," *Biological psychiatry*, vol. 62, no. 3, pp. 270–273, 2007.

[90] E. Gysels and P. Celka, "Phase synchronization for the recognition of mental tasks in a brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 4, pp. 406–415, 2004.

[91] S. Makeig, T.-P. Jung, and T. J. Sejnowski, "Using feedforward neural networks to monitor alertness from changes in EEG correlation and coherence," *Advances in neural information processing systems*, pp. 931–937, 1996.

[92] T. Shibata, I. Shimoyama, T. Ito, D. Abla, H. Iwasa, K. Koseki, N. Yamanouchi, T. Sato, and Y. Nakajima, "The time course of interhemispheric EEG coherence during a GO/NO-GO task in humans," *Neuroscience letters*, vol. 233, no. 2, pp. 117–120, 1997.

[93] O. Jensen and L. L. Colgin, "Cross-frequency coupling between neuronal oscillations," *Trends in cognitive sciences*, vol. 11, no. 7, pp. 267–269, 2007.

[94] M. X. Cohen, "Assessing transient cross-frequency coupling in EEG data," *Journal of neuroscience methods*, vol. 168, no. 2, pp. 494–499, 2008.

[95] J.-P. Lachaux, E. Rodriguez, J. Martinerie, F. J. Varela, *et al.*, "Measuring phase synchrony in brain signals," *Human brain mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[96] A. Hyafil, "Misidentifications of specific forms of cross-frequency coupling: three warnings," *Frontiers in Neuroscience*, vol. 9, p. 370, 2015.

[97] B. J. Roach and D. H. Mathalon, "Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia," *Schizophrenia bulletin*, vol. 34, no. 5, pp. 907–926, 2008.

[98] D. R. Brillinger, "An introduction to polyspectra," *The Annals of mathematical statistics*, pp. 1351–1374, 1965.

[99] U. Greb and M. Rusbridge, "The interpretation of the bispectrum and bicoherence for non-linear interactions of continuous spectra," *Plasma physics and controlled fusion*, vol. 30, no. 5, p. 537, 1988.

[100] S. Hagihira, M. Takashina, T. Mori, T. Mashimo, and I. Yoshiya, "Practical issues in bispectral analysis of electroencephalographic signals," *Anesthesia & Analgesia*, vol. 93, no. 4, pp. 966–970, 2001.

[101] P. Venkatakrishnan, R. Sukanesh, and S. Sangeetha, "Detection of quadratic phase coupling from human EEG signals using higher order statistics and spectra," *Signal, Image and Video Processing*, vol. 5, no. 2, pp. 217–229, 2011.

[102] V. Chandran, R. Acharya, C. Lim, *et al.*, "Higher order spectral (HOS) analysis of epileptic EEG signals," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 6495–6498, IEEE, 2007.

[103] J. W. Johansen and P. S. Sebel, "Development and clinical application of electroencephalographic bispectrum monitoring," *The Journal of the American Society of Anesthesiologists*, vol. 93, no. 5, pp. 1336–1344, 2000.

[104] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalography and clinical Neurophysiology*, vol. 79, no. 6, pp. 440–447, 1991.

[105] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.

[106] C. A. Kothe, S. Makeig, and J. A. Onton, "Emotion recognition from EEG during self-paced emotional imagery," *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, pp. 855–858, 2013.

[107] K. Pearson, "LIII. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[108] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[109] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage*, vol. 56, no. 2, pp. 387–399, 2011.

[110] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[111] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components–a tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, 2011.

[112] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *Neuroimage*, vol. 87, pp. 96–110, 2014.

[113] G. S. G. Sun, J. H. J. Hu, and G. W. G. Wu, "A novel frequency band selection method for Common Spatial Pattern in Motor Imagery based Brain Computer Interface," *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pp. 18–23, 2010.

[114] N. J. Hill, T. N. Lal, M. Schröder, T. Hinterberger, G. Widman, C. E. Elger, B. Schölkopf, and N. Birbaumer, "Classifying event-related desynchronization in EEG, ECoG and MEG signals," in *Joint Pattern Recognition Symposium*, pp. 404–413, 2006.

[115] B. Reuderink and M. Poel, "Robustness of the Common Spatial Patterns algorithm in the BCI-pipeline," pp. 3–7, 2008.

[116] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction.," *IEEE transactions on bio-medical engineering*, vol. 55, pp. 1991–2000, aug 2008.

[117] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized Common Spatial Pattern with aggregation for EEG classification in small-sample setting," vol. 57, pp. 2936–2946, Dec. 2010.

[118] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing.," *Journal of neural engineering*, vol. 9, p. 26013, Apr. 2012.

[119] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2391–2398, 2008.

[120] Z. Y. Chin, K. K. Ang, and C. Guan, "Multiclass voluntary facial expression classification based on Filter Bank Common Spatial Pattern," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 1005–1008, IEEE, 2008.

[121] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Increase information transfer rates in BCI by CSP extension to multi-class.," in *NIPS*, pp. 733–740, 2003.

[122] W. Wu, X. Gao, and S. Gao, "One-versus-the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pp. 2387–2390, IEEE, 2006.

[123] 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*, vol. 2007, pp. 5–5, 2007.

[124] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," 2007.

[125] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach.* Springer Science & Business Media, 2013.

[126] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE transactions on information theory*, vol. 14, no. 1, pp. 55–63, 1968.

[127] H. C. Peng, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.

[128] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[129] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[130] T. Hastie, R. Tibshirani, and J. Friedman, "Boosting and additive trees," in *The Elements of Statistical Learning*, pp. 337–387, Springer, 2009.

[131] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[132] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, pp. 1137–1145, Stanford, CA, 1995.

[133] "K-fold cross-validation." `http://karlrosaen.com/ml/learning-log/2016-06-20/`. Accessed: 2017-02-09.

[134] "Support vector machine." `https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Kernel_Machine.svg`. Accessed: 2017-05-01.

[135] M. Aizerman, E. M. Braverman, and L. Rozonoer, "Theoretical foundations of potential function method in pattern recognition," *Automation and Remote Control*, vol. 25, pp. 917–936, 1964.

[136] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152, ACM, 1992.

Chapter 3

On the Neural Correlates of Visuospatial Mental Imagery and its Relation to Brain-Computer Interfacing

3.1   Introduction

There is a growing interest in the potential of BCIs based on mental imagery due to the wide variety of potential mental commands and related applications. Thus, there is a drive towards making available new kinds of mental imagery for use in BCIs. In the traditional approach to BCI design, a form of mental imagery must be selected *a priori*. If insufficient knowledge about its neural correlates can be found in the neuroscientific literature, it must be studied further from a neuroscientific perspective before it can be used in a BCI. This chapter discusses a study on the neural correlates of visuospatial imagery, which could be used to inform the fundamentals of a design of a BCI which uses spatial imagery as a means of control.

The interest in spatial imagery as a way to control a BCI arises from the enormous potential that might be realized if it can be exploited. For example, if it were possible to accurately estimate from brain activity just the direction in space an individual was thinking of, the design of BCIs for controlling wheelchairs, mouse cursors, exoskeletons, and other devices could be revolutionized. Currently such devices are typically controlled using motor imagery (*e.g.*, [1]), which is a more unnatural way to control a device like a wheelchair compared to focusing on a direction of movement, and at best provides only up to four discrete directions of control [2]. Therefore, even if a spatial direction could only be resolved with a wide margin of error of 45 degrees on either side, making it roughly as operable as a BCI based on motor imagery, it would likely be preferable over a wheelchair controlled using motor imagery.

Currently, spatial imagery and spatial navigation imagery are used in much the same way that motor imagery is used for BCIs; as an arbitrary but convenient distinct thought process that can be differentiated from a completely unrelated thought process by the brain activity it generates. For example, BCIs have been developed that can resolve spatial navigation imagery versus musical imagery in order to move a cursor left or right [3, 4]. Extracting useful information about more specific features of the user's spatial imagery, on the other hand, is a much more difficult problem. The work that follows is an example of the neuroscientific research that is required in order to determine what features of brain activity would need to be extracted from data reflecting brain activity in order to use spatial imagery involving spatial memory and viewpoint transformations.

Rather than translating this work into a BCI based on visuospatial imagery, however, generalized methods for BCI were pursued instead. This is in part because the ability to use visuospatial imagery seemed to be quite variable across individuals, suggesting that only some individuals would be able to use a BCI based on these brain networks effectively. Indeed, this prompted a review of the

literature which confirmed that a similar problem was being reported for a wide variety of brain signals used for BCI control, suggesting that if different users could employ different mental strategies of control, more people might be able to successfully control a BCI. In addition, the high-end EEG hardware needed to extract relevant features of visuospatial imagery from EEG was simply not available at the time, making it necessary to explore alternative forms of mental imagery. These factors combined to give rise to the idea that a generalized approach to BCI would be an important next step for a field. Thus the study presented here helps to tell the story of how generalized methods for brain-computer interfacing became a goal in the first place.

3.2 Examining the role of the temporo-parietal network in memory, imagery, and viewpoint transformations

Dhindsa, K., Drobinin, V., King, J., Hall, G. B., Burgess, N., & Becker, S. (2014). Examining the role of the temporo-parietal network in memory, imagery, and viewpoint transformations. Frontiers in human neuroscience, 8:709. doi: 10.3389/fnhum.2014.00709
Article reprinted under the Creative Commons License 4.0.

Examining the role of the temporo-parietal network in memory, imagery, and viewpoint transformations

Kiret Dhindsa[1,2], Vladislav Drobinin[2], John King[3], Geoffrey B. Hall[2], Neil Burgess[4] and Suzanna Becker[2]*

[1] School of Computational Science and Engineering, McMaster University, Hamilton, ON, Canada
[2] Neurotechnology and Neuroplasticity Lab, Department of Psychology Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada
[3] Psychology and Language Sciences, University College London, London, UK
[4] Institute of Cognitive Neuroscience, University College London, London, UK

The traditional view of the medial temporal lobe (MTL) focuses on its role in episodic memory. However, some of the underlying functions of the MTL can be ascertained from its wider role in supporting spatial cognition in concert with parietal and prefrontal regions. The MTL is strongly implicated in the formation of enduring allocentric representations (e.g., O'Keefe, 1976; King et al., 2002; Ekstrom et al., 2003). According to our BBB model (Byrne et al., 2007), these representations must interact with head-centered and body-centered representations in posterior parietal cortex via a transformation circuit involving retrosplenial areas. Egocentric sensory representations in parietal areas can then cue the recall of allocentric spatial representations in long-term memory and, conversely, the products of retrieval in MTL can generate mental imagery within a parietal "window." Such imagery is necessarily egocentric and forms part of visuospatial working memory, in which it can be manipulated for the purpose of planning/imagining the future. Recent fMRI evidence (Lambrey et al., 2012; Zhang et al., 2012) supports the BBB model. To further test the model, we had participants learn the locations of objects in a virtual scene and tested their spatial memory under conditions that impose varying demands on the transformation circuit. We analyzed how brain activity correlated with accuracy in judging the direction of an object (1) from visuospatial working memory (we assume transient working memory due to the order of tasks and the absence of change in viewpoint, but long-term memory retrieval is also possible), (2) after a rotation of viewpoint, or (3) after a rotation and translation of viewpoint (judgment of relative direction). We found performance-related activity in both tasks requiring viewpoint rotation (ROT and JRD, i.e., conditions 2 and 3) in the core medial temporal to medial parietal circuit identified by the BBB model. These results are consistent with the predictions of the BBB model, and shed further light on the neural mechanisms underlying spatial memory, mental imagery and viewpoint transformations.

Keywords: spatial cognition, navigation, learning, fMRI, hippocampus

1. INTRODUCTION

The precise role of the hippocampus in memory has been the subject of much debate. A large body of evidence points toward a crucial role for this structure in the formation of allocentric spatial representations, based on rodent and non-human primate hippocampal place cell recordings, as well as studies of humans with hippocampal lesions and implanted electrode hippocampal recordings (e.g., O'Keefe, 1976; King et al., 2002; Ekstrom et al., 2003). However, evidence also points toward allocentric representations outside of the hippocampus. For example, neuroimaging of healthy individuals and studies of individuals with lesions implicate the retrosplenial and parahippocampal cortices in memory for scenes and landmarks, navigation to goals and memory across changes of viewpoint (Bohbot et al., 1998; Epstein and Kanwisher, 1998; Aguirre and D'Esposito, 1999; Maguire,

2001; Lambrey et al., 2012; Zhang et al., 2012; Sherrill et al., 2013; Sulpizio et al., 2013).

Considering that information arrives at the sensory receptors in an egocentric frame of reference, e.g., retinocentric in the case of visual input, a transformation must be carried out to translate from egocentric to allocentric co-ordinates. Such a transformation of co-ordinates is a non-trivial calculation for a neural circuit. It is therefore likely that a hierarchy of multiple brain regions is involved in carrying out this transformation, with a gradual emergence of progressively more global, allocentric representations; this is a key assumption underlying the Byrne, Becker, and Burgess (BBB) model of spatial memory (Byrne et al., 2007). Moreover, the BBB model suggests a role for hippocampal neurons in learning conjunctions of allocentric boundary and landmark features, as well as

other non-spatial features, explaining the emergence of context-modulated place cells (Anderson and Jeffery, 2003). The BBB model thus sheds light on two major unresolved issues in the literature concerning the role of the hippocampus in memory: (1) What is the role of the hippocampus vs. extra-hippocampal structures in allocentric coding? (2) What is the role of the hippocampus in conjunctive/episodic (including non-spatial) encoding?

Central to the BBB model is an egocentric parietal window that maintains representations of objects, landmarks and boundaries. The parietal window is postulated to be located in the precuneus/medial parietal cortex. The contents of the parietal window can be maintained in working memory through reciprocal fronto-parietal connections. Additionally, object/landmark locations within the parietal window can be continuously updated during real or imagined self-movement through reciprocal connections with the medial temporal lobe. These head-centered and body-centered representations formed in posterior parietal cortex are mapped, via a transformation circuit, into allocentric spatial representations in the parahippocampal region and hippocampus. An egocentric parietal window thus allows one to integrate sensory inputs into an egocentric map, cueing the recall of spatial representations in long-term memory. Conversely, reciprocal connections from the hippocampus to posterior parietal regions allow the products of memory retrieval to generate mental imagery within the parietal window which can be manipulated for the purpose of planning ahead and imagining the future. Other non-spatial contextual features are also integrated at the level of the hippocampus, giving rise to configural memories for places and events.

The BBB model makes several empirical predictions. The first step in mapping from egocentric to allocentric representations involves combining head-centered object representations maintained in the parietal window with allocentric head-direction signals; retrosplenial cortex is anatomically well situated to carry out this computation, as it is reciprocally connected with parietal and medial temporal regions, and receives inputs from areas known to carry head-direction information (Wyss and Groen, 1992; Maguire, 2001; Kobayashi and Amaral, 2003). Thus, we predict that retrosplenial cortex would be engaged whenever egocentric-to-allocentric mappings (or the reverse) are required. Egocentric object/boundary representations modulated by allocentric head direction are in turn transformed into a map of allocentric representations of individual boundaries, objects and landmarks in the parahippocampal cortex. Finally, these object and boundary features are combined at the level of the hippocampus into allocentric representations of particular places (place cells). Thus, according to the BBB model, allocentric coding emerges in at least three levels of representation: in the retrosplenial cortex, in the parahippocampal cortex and in the hippocampus. Whether a given allocentric task requires the hippocampus should depend on whether a conjunction of object/boundary locations is required to solve the task. Thus, orienting to a single landmark might engage the retrosplenial and parahippocampal cortices but may not require the hippocampus. On the other hand, locating an object relative to a configuration of landmarks and other contextual features, thereby uniquely placing it in space and context, should be hippocampal-dependent.

Recent evidence from fMRI studies supports some of the predictions of the BBB model. When participants performed a change detection task while viewing object configurations, trials involving imagined changes in viewpoint (involving both a translation and rotation) were associated with activation of the precuneus, parieto-occipital sulcus/retrosplenial cortex and hippocampus (Lambrey et al., 2012). Similarly, performance of a judgment of relative direction (JRD) task activated the parahippocampal and retrosplenial cortices to a greater degree after learning routes through a virtual town relative to a map-learning condition (Zhang et al., 2012). To further test predictions of this model empirically, we investigated spatial memory retrieval under conditions that impose varying demands on the transformation circuit.

We employed a virtual reality implementation of the JRD task with several conditions, each providing less context and placing a progressively greater burden on memory, mental imagery, and viewpoint transformation: no viewpoint change (REF), a pure rotation of viewpoint (ROT), and (as in Lambrey et al., 2012; Zhang et al., 2012) combined translation and rotation (JRD). A "baseline" condition involving no viewpoint change but including the background scenery and visual feedback was also included for comparison. After learning a configuration of object locations in a virtual environment with easily distinguished distal landmarks, participants underwent fMRI scanning while performing spatial memory and imagery test trials. On each test trial the participant was asked point to an object from either a familiar or novel viewpoint. During REF trials participants were asked to imagine their position and viewpoint were identical to the familiar reference viewpoint they had learned previously before pointing to the cued object. During ROT trials, participants were asked to imagine their position being identical to the position in REF, but that they were instead facing one of the objects and asked to point to a second object. During JRD trials, commonly referred to as a judgment of relative direction (Shelton and McNamara, 1997), participants were asked to imagine they were standing in the position of object X, facing object Y, and then to point to a third object Z.

Thus, in ROT and JRD, participants were asked to imagine a configuration of objects from a changed perspective, which should require the egocentric to allocentric transformation circuit. We analyzed the brain areas which correlated with the execution of these different tasks, and how brain activity in key regions of interest correlated with accuracy in judging the direction of an object after a perspective shift. We hypothesized that all three conditions would require visuo-spatial imagery and therefore activate the parietal window/precuneus, but only ROT and JRD would activate the transformation circuit (retrosplenial cortex) and allocentric representation of objects (parahippocampal cortex) and object configurations (hippocampus). We further predicted that JRD would most strongly activate this circuit, since it requires the most complex transformation (involving both the transformation required in ROT as well as a translation).

2. MATERIALS AND METHODS

2.1. ETHICS STATEMENT

The study was approved by the ethics review boards at McMaster University and St. Joseph's Healthcare Hamilton. All participants gave written consent to participate in the behavioral selection experiment, written notice of interest to be considered for scanning, and additional written consent to take part in the fMRI scan on the day of scanning.

2.2. PARTICIPANTS

Fifteen participants were included in the final analysis after six were rejected due to excess motion in the scanner and technical issues with scanning, and one elected to withdraw from the study during scanning. All participants were male right-handed McMaster University students (14 undergraduate, 1 graduate) with normal or corrected vision and were classified as gamers (minimum of 10 h a week playing video games). Gamers were chosen due to their experience with operating and navigating in virtual environments, and males were preferred to avoid sex differences in spatial cognition and navigation (Voyer et al., 1995; Parsons, 2004; Levin et al., 2005).

2.3. SELECTION EXPERIMENT vs. SCANNING EXPERIMENT

Participants performed the experiment twice, first outside of the scanner (the "selection experiment") and second, within the scanner 3–5 weeks later. The initial selection experiment, used to select participants for the scanning session, was run in a quiet testing room on a Lenovo Thinkpad E430 laptop with a 14″ 1366 × 768 resolution display.

During the selection experiment, each participant performed three rounds, each involving a block of each task. A single round included five consecutive blocks of Collect and Replace for learning, followed by one block each of VIS, INVIS, REF, ROT, and JRD, in order (these tasks are defined in detail below). Final JRD accuracy was used to determine whether a participant was invited for a follow-up scanning session so as to select only those who could learn the arena map and infer object-to-object relationsihps well. Participants were required to have either an average JRD error less than one standard deviation below the mean on the final round of the experiment, or average JRD errors less that

one standard deviation below the mean on each of the first two rounds of the experiment.

The scanning experiment was conducted at St. Joseph's Hospital in Hamilton, Ontario, Canada. Prior to scanning, each participant performed two complete rounds outside the scanner as done in the selection experiment to refresh their memory of the arena map and instructions for each task. Only the last four pointing tasks (INVIS, REF, ROT, and JRD) were performed in the scanner.

2.4. EXPERIMENTAL STIMULI AND TASKS

A virtual environment was built in the open-source simulation platform OpenSimulator (Overte Foundation, 2007). All of the tasks took place in a circular arena on a flat grassy ground with visible distal landmarks distributed in the background (see **Figure 1**). These landmarks included two uniquely shaped hill formations, the sun in a fixed location, and a tree. The environment involved no variation in weather, brightness, or atmosphere.

The circular arena contained four distinct objects set in a consistent spatial configuration used for all tasks and participants. Object locations were chosen so they did not directly align with distal landmarks to encourage participants to encode each object relative to the configuration of landmarks. They were also set so they did not form the vertices of a simple polygon to discourage participants from learning object-to-object spatial relations. Finally, they were set so that all objects were within the field of view from the reference viewpoint used in learning and REF.

2.4.1. Experiment overview

The participant was required to perform several different tasks in this environment. Verbal instructions were given by the experimenter between tasks during the selection experiment. The selection experiment with verbal instructions was repeated immediately before the scanning session to refresh the participant's memory of the object locations and VR controls. No feedback was given for REF, ROT, or JRD, so object-to-object relations could not be learned directly.

All input from the participants involved either navigation using directional keys within the arena (in the selection experiment only) or pointing to objects using an arrow. When navigating (only during the task Collect and Replace, described below),
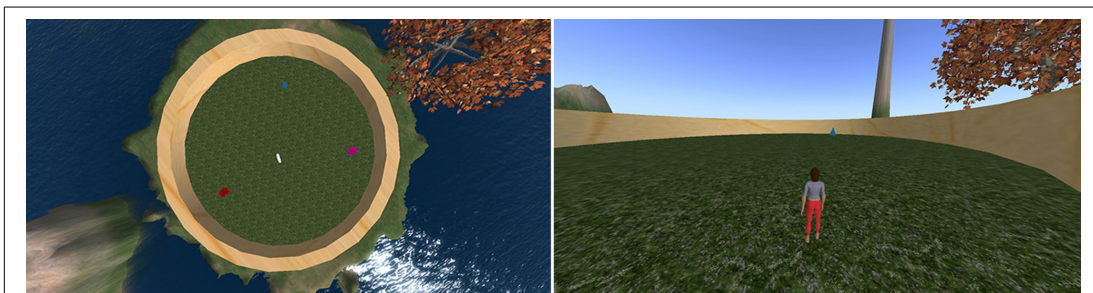
FIGURE 1 | **Left:** Bird's-eye-view of the layout of the environment used for all experiments (avatar not shown). Contains a red cube, a blue pyramid, a white taurus, and a pink sphere. **Right:** View of arena during navigation (only one object visible at a time).

participants controlled an avatar (a virtual character) from a third-person view using the keyboard's directional arrows. When pointing, the participants rotated an arrow, presented in the fronto-parallel plane, to identify the direction of the cued object. In the scanner, participants made pointing responses using buttons on a gamepad-like device that was safe for operation within the scanner. The pointing arrow was rotated clockwise or counterclockwise in the fronto-parallel plane using two buttons, and a response was made with the third button. On all pointing trials, the initial direction of the arrow was randomized to avoid the interference of proprioceptive memory over visuospatial memory.

During the scanning portion of the experiment, participants were cued using pictorial instructions in a heads-up display to inform them of the task that was starting and the goal of each trial. Prior to each trial onset, a blank screen was displayed for 3 s. Afterwards, the cue was overlaid on the blank screen for another 3 s (6 and 9 s respectively for the ROT and JRD tasks, whose cues are described more fully below), followed by another blank screen for 3 s. Following this second blank gray screen, the pointing response arrow appeared on the blank screen to indicate that a response could be made. Other than the differences in the cues visually and temporally, all trials across all tasks were presented identically. Cues and the arrow used for pointing response are presented in **Figure 2** for each condition.

Participants performed the tasks at their own pace, though trials which lasted longer than 25 s were rejected, as it was deemed likely that the participant had not been sufficiently engaged during the trial. Since each trial required the participant to imagine and reason about the spatial relationships of the objects in the arena before making a response, no strict response time could be imposed. Therefore, response time varied from trial to trial, as evident in **Table 1**. Given this necessity for self-paced trials, and the strict time limit on individual scanning sessions imposed by the institution housing the fMRI scanner (allowing us 21 min and 48 s per participant), we allowed each participant to complete as many trials as possible within their allotted time. Participants were not given any indication of a time limit or any suggestion that they should perform the trials quickly as to avoid rushing them. The minimum number of rounds completed was 2, and the maximum was 4, with a mean of 2.73 fully completely rounds.

Each round contained a block each of REF, ROT, and JRD (the first two rounds included a block of the baseline task INVIS to orient the participant within the scanner and to check the validity of their responses) with four randomly generated trials in each block. In REF, each object was pointed to once in each block in a random order, but in ROT and JRD, trials were completely randomly generated, i.e., the viewpoint direction and/or position was resampled on each trial. Each round ended with a screen indicating the end of the round, and participants were able to start the next round at their own pace by pressing a button on their input device. On average, 13.6 REF trials, 12.0 ROT trials, and 10.8 JRD trials were completed within the 25 s rejection threshold by each participant.

For each round of the experiment, REF, ROT and JRD were performed in the same order. It was required that REF come first in order to re-establish the reference viewpoint for the ROT condition, which must follow REF in order make use of this freshly re-established viewpoint since ROT uses the same position, albeit a different heading direction. Since the JRD breaks away from this reference viewpoint, it may interfere with performance of the ROT trials if it were to be interposed between REF and ROT. Therefore, there was no counterbalancing of the task order in this experiment. The order during the selection experiment was Collect and Replace, VIS, INVIS, REF, ROT, and JRD, while the order for the scanning eperiment was INVIS, REF, ROT, and JRD.

2.4.2. Collect and replace
Participants performed five rounds of collecting the objects and returning them to their original locations in order to learn the

FIGURE 2 | Top Left: Cue for VIS, INVIS, and REF. **Top Right:** Cue for ROT. **Bottom Left:** Cue for JRD. **Bottom Right:** Pointing response arrow.
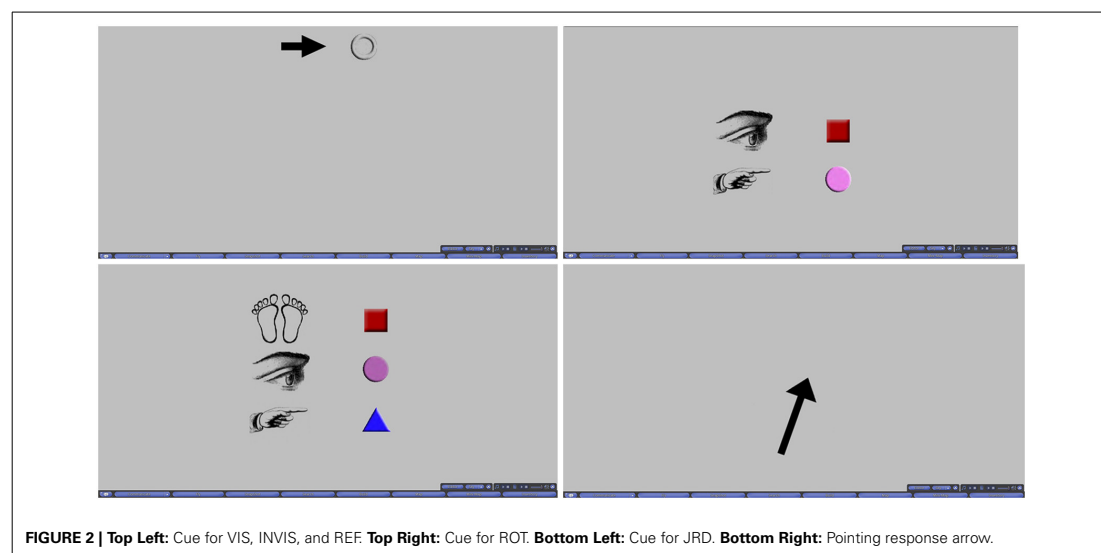
layout of the arena and distal landmarks. Participants were cued to collect each object in the arena one by one in a random order and each object was collected simply by walking to the object. Participants were instructed that they would need to replace each object during the next task and that no visual information aside from the distal landmarks would be avialable during that time, implying that these landmarks should be carefully observed during the collect phase. Only the cued object was visible during each collect trial so that participants were less able to encode the spatial relationships between the objects themselves and were needed to rely on the distal landmarks. All other potential cues were minimized by ensuring the shape and textures of the ground, arena, and sky were consistent throughout.

After each object had been collected, the avatar was teleported to a random location within the arena with a viewpoint in a random direction. With none of the objects visible, the participant was next required to walk close to the cued object's original location in order to replace it. The object appeared for 1 s when the avatar was close enough to the location (within two virtual meters to the object center, which is roughly the height of the avatar) to provide feedback for learning, and disappeared again before the next object was cued. After all of the objects were replaced, the avatar was randomly teleported and the collect phase repeated. After the fifth collect and replace round was completed, the next task was initiated.

2.4.3. Pointing While Visible (VIS) and pointing while invisible (INVIS)

Pointing While Visible (VIS) was a calibration task that simply asked the participant to point to each successively cued object using the black pointing arrow (here on a circular white background superimposed on the avatar to avoid the possibility of avatar acting as an additional directional cue). This served the purpose of allowing the participant to establish a familiar viewpoint and to become accustomed to the pointing controls. It also provided an indication of the baseline pointing error, in degrees, for that participant. These errors were checked to ensure that the participant correctly performed the task and that performance was higher on this task than on any of the other more challenging pointing tasks.

Pointing while invisible (INVIS) was identical to pointing while visible except that the objects were not visible (all other aspects of the scene remained visible). The participant was required to point to the location of each object from memory and was provided feedback after each response by the brief reappearance of the target object. Note that this reference viewpoint, used for VIS, INVIS, and REF, had all objects in the field of view, so visual feedback was always possible.

During scanning, VIS was only used as a means to orient the participant to viewing the screen and operating the controls within the scanner, and to check that they were still able to perform the trials with similar accuracy as outside of the scanner. INVIS was used in a similar way, but also for comparison to REF, since the task is similar but without as much reliance on mental imagery. For the remainder of the paper, INVIS will generally be referred to as the baseline task.

2.4.4. Pointing from Imagery (REF)

Pointing from imagery required the participant to point to the objects from the same viewpoint as in the previous pointing tasks (the *reference viewpoint*). However, the entire virtual environment was now occluded and only the pointing arrow and cues were visible. Pointing necessarily took place purely from memory without the assistance of distal landmarks and was done with a black arrow on a gray background occluding the entire arena (the same black arrow over a gray background was used for the ROT and JRD tasks as well).

2.4.5. Pointing with Rotation (ROT)

Pointing with Rotation differed from the previous pointing tasks in that the participant was instructed to imagine that their point of view was rotated from the reference view established during the previous pointing tasks to a view centered on one of the objects. The environment remained occluded, as in REF, and task structure was identical except for the extra 3 s given to interpret the cue. The cue was changed to reflect the need to illustrate two instructions (**Figure 2**): which object the participant should center their viewpoint on, and which object they should point to from that viewpoint. As in REF, pointing responses were made with a black arrow on a gray background.

On the first ROT trial of each round, the cue was displayed for 9 s to give the participant extra time to process the cue if they were not sufficiently prepared. On subsequent trials, the cue was presented for only 6 s (3 s for each instruction). As seen in **Figure 2**, all instructions of the cue were displayed together (similarly for JRD trials). This additional time was especially important (as well as the inclusion of this period in analysis of the fMRI data), because pilot studies and participant surveys both found that many participants engaged in adjusting their imagined viewpoint in steps as they read each instruction of the cue.

2.4.6. Judgment of relative direction (JRD)

The final task was a judgment of relative direction (JRD). This was almost identical to ROT except that a translation was added. Participants were required to imagine standing at the location of one object, facing a second object, and then to point to a third object. A third line of images was added to the cue to illustrate all three components of the instructions (**Figure 2**). This cue was presented for 9 s. On the first JRD trial of each round, the cue was displayed for 12 s for the same reasons given above in ROT.

2.5. fMRI DATA ANALYSIS

Scans were performed with a 3 Tesla General Electric fMRI scanner. A T1-weighted anatomical scan in the axial orientation was obtained prior to functional imaging. The scanning parameters for the anatomical image series were: 3D SPGR pulse; fast IRP sequence; prep time = 450; flip angle = 12; FOV = 240 mm; $TE = 2.2$ ms; $TR = 7.7$ ms; 80 slices; slice thickness = 2 mm, no skip.

Functional images were were collected in interleaved axial slices with a GRE-EPI pulse sequence. The field of view was 21 cm with a slice thickness of 2.9 mm and a slice gap of 0.1 mm There

75

were 40 slices per volume with a *TR* of 2600 ms, totalling 500 volumes and a functional scan time of 21 min and 48 s. The *TE* was 25 ms and the flip angle was 90°.

Image processing and statistical analysis were performed using BrainVoyager QX 2.6 (Brain Innovation, Maastricht, The Netherlands) (Formisano et al., 2006; Goebel et al., 2006). Anatomical data were remapped to an iso-voxel size of 1.0 × 1.0 × 1.0 mm with a cubic spline interpolation and a framing cube dimension of 256 points. Each data set underwent manual anterior commissure to posterior commissure alignment. The anatomical 3D data sets were then normalized to Talairach space using linear affine transformation.

The functional data sets were slice-time corrected, 3D motion corrected and realigned to the fifth volume in the series, high-pass filtered at 2 sines/cosines, and normalized to Talairach space (Talairach and Tournoux, 1988). Funtional data series with motion greater than the fMRI voxel size were discarded from analysis as recommended by Formisano et al. (2006). The functional data were then co-registered with the 3D anatomical data allowing for the creation of a 3D aligned time course. The 3D aligned time course data was smoothed with a 6 mm full-width at half-maximum (FWHM) Gaussian filter. Finally, the functional data was masked to filter out noise in the data that fell outside of brain tissue.

A general linear model (GLM) was used to model each participant's data individually. Due to the self-paced nature of each trial and the various possible strategies participants may have used for producing their responses (we found in a survey that most performed the necessary mental imagery during both the cue phase and response phase of the trial), we used the time window from cue onset to response input to measure brain activation. For the ROT and JRD conditions, a parametric model was built by using the standardized pointing errors as an additional regressor. This weighted the brain activity by trial accuracy (i.e., coefficients are found for performance/accuracy regressor). The unweighted brain activity was subtracted from the performance-weighted activity (the z-scores of all regressors needed to be used here so that the regressors were similarly scaled prior to subtraction) to find performance-related activations and to reduce the loss of information when averaging across strategy differences, individual skill differences, and trial-by-trial changes in attention, effort, strategy, and performance. This included the added benefit of filtering brain activations that may have been a result of superfluous

processing not contributing to task performance, including processing on-screen visuals. Data from all participants were combined in a random-effects GLM using a participant-averaged mask and an averaged anatomical.

2.5.1. Correction for multiple comparisons

Where possible, we used the false discovery rate correction (FDR) of $p_{FDR} < 0.05$, which is usually thought of as a stricter correction than the alternatives when activation is sparse, and less conservative if activated areas are large (Genovese et al., 2002). When the FDR correction was either too strict or too lenient for the specific analysis being run, we used a threshold corrected for family-wise error (FWE) of $p_{FWE} < 0.05$, which is another standard. This was done by using Monte Carlo simulations to find the minimum cluster size needed to achieve significance threshold based on an uncorrected per voxel threshold of $p < 0.005$ (Forman et al., 1995). The type of correction and the required minimum cluster size required for each contrast is given in the tables of results. We found the same, and often additional, areas of activation using both methods, but the statistically significant areas were either extremely large or extremely small when using either only FDR or FWE for all contrasts, making the results difficult to interpret without employing each where they are best-suited.

3. RESULTS

3.1. BEHAVIORAL RESULTS

To investigate the accuracy of the representations used by participants to perform the JRD and Rotation tasks, we examined participants' average absolute pointing errors from the true object location (denoted by the center of the object) and their response times from cue offset. Participants were highly accurate in both the REF and INVIS conditions, and less accurate in the Rotation task than the JRD task (see **Table 1** and **Figure 3**). While not

Table 1 | Means (standard error of means) of pointing errors and response times.

	JRD	ROT	REF	INVIS
Absolute pointing error (°)	25.56 (4.97)	38.91 (2.64)	6.00 (0.75)	5.27 (0.66)
Response time (s)	15.84 (1.63)	16.47 (1.81)	11.13 (1.08)	12.02 (0.93)
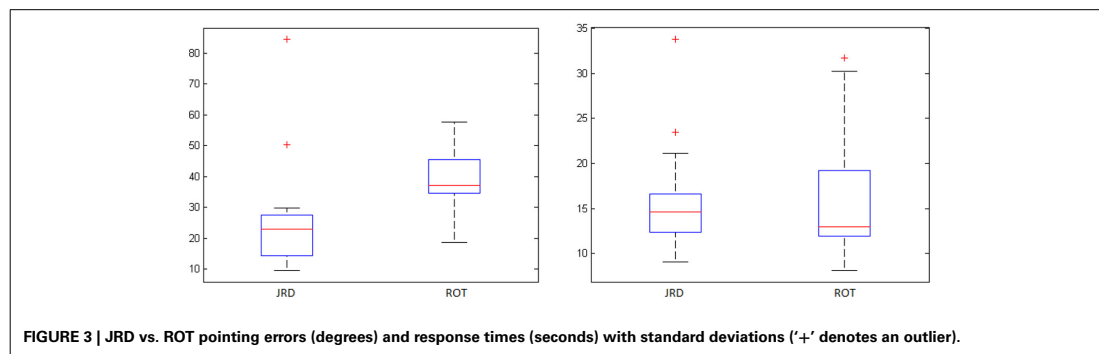
FIGURE 3 | JRD vs. ROT pointing errors (degrees) and response times (seconds) with standard deviations ('+' denotes an outlier).

76

statistically significant due to the very high variances, there was a clear trend toward both decreased accuracy and longer response times for ROT compared to JRD trials, suggesting that participants found the ROT task to be more difficult.

3.2. BRAIN ACTIVITY

Using the analytic methods described above and the parametric GLMs with pointing error as an additional regressor, we identified brain regions that were activated in proportion to performance on the different tasks performed in the scanner (except for the REF task, where the small errors and variance made the extra performance-related regressor unnecessary). Brain areas having significant task-related activations (a simple contrast with no parametric modeling) for REF vs. baseline (where participants pointed from the same viewpoint as REF but were still able to see the virtual environment, excluding the objects themselves, so that pointing was not completely from memory), and performance-related activity for the JRD task and the ROT task are given in **Tables 2–4** respectively.

For both the JRD and the ROT tasks when parameterized by performance, we found significant performance-related activations in the hippocampus (**Figure 4**), the parahippocampal cortex, the precuneus (however, for the JRD, precuneus activity was only seen in a non-parametric model where the pointing errors were not used as a regressor, as would be implied by the BBB model), the parietal cortex, and the retrosplenial cortex, all of which are predicted by the BBB model. In both conditions, significant task-related activations were also seen in many areas in

the occipital, temporal and parietal lobes associated with visual object processing, working memory and imagery, as well as areas in the frontal lobe and cingulate cortex associated with cognitive control. Somewhat surprisingly, task-related activation was seen in the caudate nucleus in the JRD task, and in the primary somatosensory cortex in the ROT task.

We also assessed the difference in performance-related activation between the JRD task and the ROT task (**Figure 5**). When the JRD task was contrasted with the ROT task, we saw greater performance-related activation during the JRD task in the left parahippocampal gyrus, the right and left precuneus, and the right retrosplenial cortex. Additionally, the inferior temporal gyrus, middle frontal gyrus, superior frontal gyrus, middle temporal gyrus, precentral gyrus, posterior cingulate, lingual gyrus, thalamus, medial frontal gyrus, superior and inferior parietal lobules, and the middle occipital gyrus were also significantly more active during the JRD task than the ROT task ($p_{FDR} < 0.05$).

There was greater activity in the ROT task contrasted with the JRD task in the superior temporal gyrus, the postcentral gyrus, the left middle temporal gyrus, the middle frontal gyrus, the medial frontal gyrus, the cuneus, and the anterior cingulate gyrus ($p_{FDR} < 0.05$). Therefore, the regions identified by the BBB model were more activated by the JRD task than the ROT task.

When comparing to REF, activity in both the JRD and the ROT tasks showed significant performance-related activation (i.e., we first subtracted the unweighted brain activity from

Table 2 | Activity during REF task relative to baseline ($p_{FWE} < 0.05$).

REGION	Coord. (mm)	Voxels	T-score
OCCIPITAL			
Inferior occipital	LH −40 −74 −6	1050	4.27
	LH −52 −68 −6	48	3.60
TEMPORAL			
Superior temporal	RH 44 −38 3	103	4.29
	LH −49 −23 6	2028	4.29
Inferior temporal	RH 40 −68 3	785	4.14
	RH 38 −41 −18	280	4.79
PARIETAL			
Inferior parietal	RH 56 −29 18	1471	4.91
	RH 53 −41 45	212	4.05
	LH −37 −29 42	333	4.28
	LH −43 −32 30	1363	5.38
	LH −67 −29 21	50	3.84
*Precuneus	RH 23 −56 48	134	3.73
	RH 8 −44 45	1769	5.70
	LH −28 −47 51	49	3.89
Superior parietal	LH −28 −56 45	336	5.41
FRONTAL			
Precentral gyrus	RH 47 −8 45	230	4.15
	LH −13 −20 60	1724	6.14
Inferior frontal	RH 47 4 14	342	5.19

RH, right hemisphere; LH, left hemisphere; * *Main areas of interest.*

Table 3 | Performance-related activity during JRD task ($p_{FDR} < 0.05$).

REGION	Coord. (mm)	Voxels	T-score
OCCIPITAL			
Cuneus	RH 8 −83 3	40	4.42
Lingual gyrus	RH 17 −74 −3	26	4.12
	RH −1 −92 −9	26	4.69
	RH 26 −62 3	201	5.34
	LH −25 −77 0	29	4.52
MEDIAL TEMPORAL			
*Hippocampus	RH 29 −38 3	21	4.32
*Parahippocampus	RH 26 −29 −3	86	4.35
	LH −49 −29 −12	34	4.53
BASAL GANGLIA			
Caudate	RH 23 −44 15	253	4.76
TEMPORAL			
Middle temporal	RH 69 −23 −9	17	6.20
	LH −61 −32 −9	17	4.09
PARIETAL			
Inferior parietal	RH 29 −56 21	721	5.86
CINGULATE CORTEX			
Posterior cingulate	RH 11 −26 24	103	4.64
	LH −22 47 3	894	5.63
*Retrosplenial cortex	RH 2 −53 24	45	4.24
Anterior cingulate	LH −4 31 12	34	4.15
	LH −1 40 15	5	3.89
	LH −7 46 3	5	3.92

RH, right hemisphere; LH, left hemisphere; * *Main areas of interest.*

Table 4 | Performance-related activity during the ROT task ($p_{FDR} < 0.05$).

Region	Coord. (mm)	Voxels	T-score
OCCIPITAL			
Cuneus	RH 11 −98 9	474	3.80
	LH −4 −86 15	1447	3.21
Lingual gyrus	LH −1 −92 −3	35	2.62
	LH −25 −77 −6	51	2.55
Fusiform gyrus	RH 56 −14 −24	229	3.62
Middle occipital	RH 43 −86 12	90	3.27
MEDIAL TEMPORAL			
*Hippocampus	LH −31 −14 −9	112	2.95
*Parahippocampus	RH 25 1 −10	107	2.89
	LH −37 −41 −3	434	3.18
	LH −25 −38 3	32	2.45
	LH −31 −53 9	308	3.21
	LH −43 −8 −15	2732	5.10
Uncus	RH 26 −27 −30	34	3.00
	LH −22 1 −30	50	2.69
	LH −22 −11 −30	47	2.74
TEMPORAL			
Superior temporal	RH 63 −56 18	1180	3.04
	LH −28 10 −33	877	3.60
Middle temporal	RH 53 7 −21	258	4.44
Inferior temporal	LH −37 −8 −42	59	3.47
	LH −67 −59 −9	98	3.27
PARIETAL			
Inferior parietal	RH 50 −65 45	797	3.40
*Precuneus	LH −7 −67 64	21	2.52
Postcentral gyrus	RH 29 −23 39	269	3.58
FRONTAL			
Insular cortex	RH 29 10 −12	414	3.74
Inferior frontal	RH 53 25 3	115	3.37
	LH −37 31 −3	103	2.83
	LH −43 28 −15	645	3.66
Middle frontal	RH 50 37 −6	248	3.33
	LH −16 43 −18	235	4.47
	LH −28 10 54	132	3.09
Medial frontal	RH 11 46 15	4589	4.48
	LH −43 16 24	54	2.57
Superior frontal	LH −16 43 36	3223	4.24
	LH −13 71 −3	40	3.21
	LH −16 52 48	98	2.91
	LH −19 64 9	127	3.29
CINGULATE CORTEX			
*Cingulate/Retrosplenial	RH 16 −47 18	28995	5.07
Posterior cingulate	LH −10 −26 36	137	3.38

RH, right hemisphere; LH, left hemisphere; * Main areas of interest.

the performance-weighted, as in our parametric model, and then subtracted the REF coefficients) in the right hippocampus, parahippocampal gyrus (left for JRD and right and left for ROT), right retrosplenial cortex, and right precuneus as predicted by the BBB model, as well as activity in the middle frontal gyrus (left and right), the anterior cingulate (left), and the superior and middle

temporal gyri (left). Visual comparisons between ROT and REF and JRD and REF are shown in **Figures 6**, **7** respectively.

4. DISCUSSION

4.1. THE TRANSFORMATION CIRCUIT

The brain regions predicted by the BBB model to be involved in transforming between egocentric and allocentric representations (intraparietal sulcus/retrosplenial cortex (RSC), parahippocampal cortex and hippocampus) were active during our mental transformation tasks. Importantly, these activations correlated with task performance, consistent with our hypothesis that this neural circuit subserves the transformation between egocentric and allocentric representations. Moreover, as predicted, this circuit was more strongly activated in imagined transformations involving both rotation and translation of viewpoint (JRD) relative to transformations involving only a rotation of viewpoint (ROT). Our results in the JRD condition are consistent with those of two other recent studies that asked participants to perform spatial memory tasks after imagined viewpoint changes that included both a rotation of viewpoint and a translation of the observers location (Lambrey et al., 2012; Zhang et al., 2012). Additionally, the involvement of the hippocampus in the tasks that involve retrieval from novel viewpoints, as shown here and elsewhere (Zhang and Ekstrom, 2013), is consistent with Eichenbaum's view of the hippocampal role in relational memory (Eichenbaum and Cohen, 2001), which could be implemented by a neural circuit that supports viewpoint transformations as in the BBB model.

As expected, there were additional areas of activation in all tasks in regions associated with visuo-spatial processing and cognitive control. Unexpectedly, task-related activation of the caudate nucleus was also observed in the JRD condition. The caudate is generally associated with motor planning and goal-directed behaviors (for a review see Grahn et al., 2008). It has been implicated in prospective coding of motor responses driven by egocentric sensory and/or working memory representations (Postle and D'Esposito, 2003). This could explain the involvement of the caudate in the JRD task, if participants were planning and imagining mental navigation to the goal location in egocentric co-ordinates. Although the medial temporal lobe and basal ganglia are often portrayed as having competitive, mutually exclusive roles in spatial memory and navigation, using allocentric vs. egocentric strategies respecitvely, it is possible that a combination of the two would be employed in a complex task such as the JRD. Once an allocentrically stored spatial representation of the goal location has been retrieved, one could map this into egocentric co-ordinates to perform planning and mental navigation.

Interestingly, even though activation in the transformation circuit was more closely related to performance in the JRD condition than in the ROT condition, the ROT condition seemed to be more challenging for participants: response times were somewhat longer, and errors were somewhat higher, in the ROT than in the JRD condition (**Table 1**). Imagining a rotated viewpoint while holding one's location constant may pose a particular challenge to participants due to the cue conflict between one's real and imagined heading directions, especially when the avatar position was identical to the reference position

FIGURE 4 | Performance-related MTL activations for JRD (top) and ROT (bottom) conditions independently.
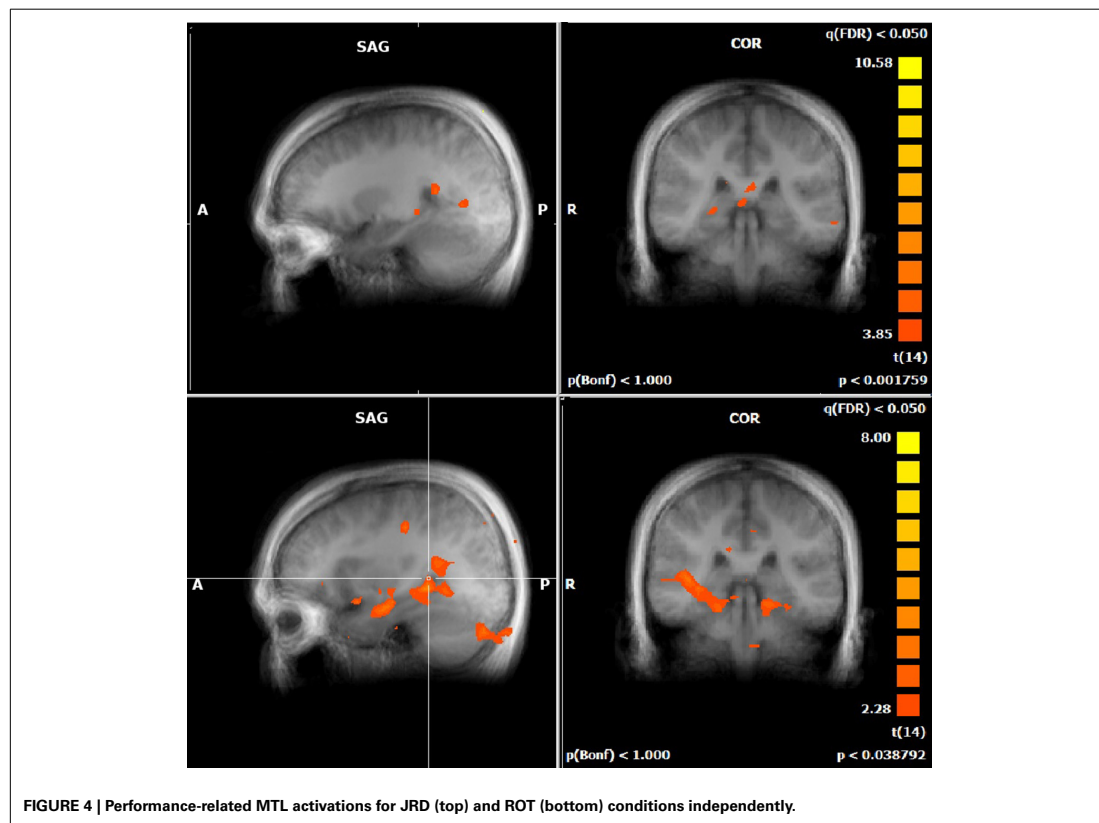
in the baseline condition, which is the only position from which the environment is explicitly learned visually with feedback. This could introduce interference between the remembered locations of objects from one's current location and the target locations of objects after the imagined view rotation (May, 2004). Consistent with this sensorimotor interference interpretation, participants' response times to point to the remembered locations of objects have been found to increase in proportion to the change in imagined viewpoint, irrespective of whether there was a change in location (May, 2004). Furthermore, the ROT condition was the only one in which task-related activation was observed in the primary somatosensory cortex. Among other things, somatosensory cortex represents the position of the eye relative to the head (Wang et al., 2007) and is activated in humans when processing changes in head position (Fasold et al., 2007). Thus, activation in this region may reflect the cue conflict noted above, as participants attempt to resolve the interference between their target and actual or imagined head direction.

4.2. A HIERARCHY OF ALLOCENTRIC REPRESENTATIONS

Much debate has been devoted to the precise role of the hippocampus in spatial memory. Is it uniquely responsible for creating allocentric spatial representations or are other areas

involved? While spatial memory deficits across viewpoint changes are observed in patients with hippocampal lesions (e.g., King et al., 2002), lesions to the parahippocampal cortex and retrosplenial cortex are also associated with topographic disorientation (Habib and Sirigu, 1987; Takahashi et al., 1997). The BBB model specifies the distinct contributions of these different regions to allocentric coding: (1) posterior parietal cortex forms an egocentric representation of landmarks and boundaries, (2) posterior cingulate/retrosplenial cortex (RSC) forms a map of landmark locations modulated by egocentric or allocentric head direction respectively, depending on whether the circuit is sensory- or memory-driven, (3) parahippocampal cortex (PC) forms an allocentric map of landmark locations, and (4) hippocampal cortex cells (HC) respond to places by encoding conjunctions of landmarks, boundaries and other contextual information. Thus, damage to either the RSC, PC, or HC could cause deficits in allocentric memory and orienting.

Another controversy surrounding the role of the hippocampus in spatial coding is whether its role is time-limited, or is it always required? A challenge for the BBB model, and more generally for cognitive map theory, is to explain the finding that KC, a dense amnesic with bilateral damage to the PC and HC, was able

79

FIGURE 5 | Significant performance-related activations for JRD vs. ROT (top) and ROT vs. JRD (bottom).
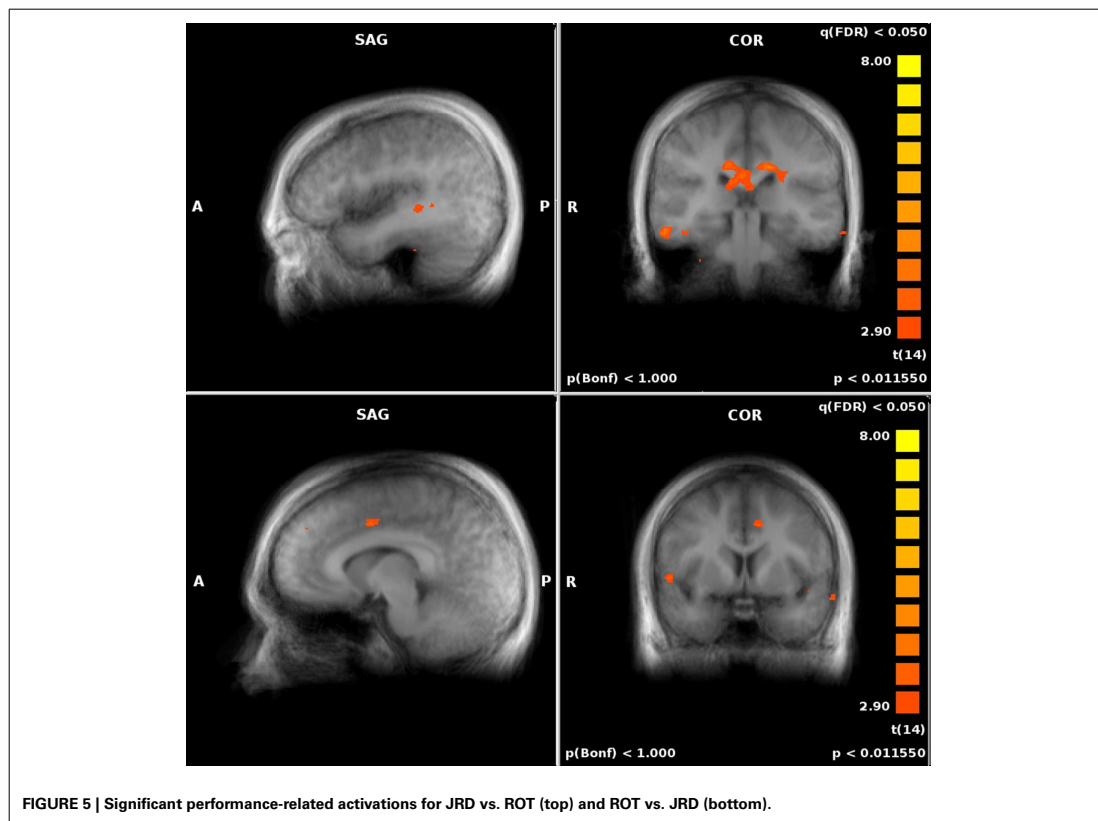
to perform several allocentric judgments from his remote spatial memories of a familiar environment (Rosenbaum et al., 2000), including specifying alternative routes between major landmarks when the direct route was blocked. However, his remote memory for less salient landmarks was highly impaired. These findings accord with the pattern of spared and impaired spatial abilities observed in TT, a London taxi driver with bilateral hippocampal damage, who was able to navigate in a virtual model of London via major routes but was highly impaired at navigating via alternative routes (Maguire et al., 2006). The ability to orient toward salient landmarks in KC and TT could be supported by spared regions of parahippocampal and retrosplenial cortices. This would mean that the parahippocampal region, and not only the hippocampus, encodes associations amongst landmarks.

4.3. LIMITATIONS AND FUTURE WORK

A limitation of the present study is that participants were drawn from a narrow demographic: male, right-handed university students with computer game experience. It was hoped that selecting for video game experience would minimize adverse reactions to immersive VR (e.g., nausea), and would keep learning time to a minimum, as gamers tend to show greater facility at traversing and learning VR layouts. The choice of right-handed males was made to minimize variability in functional activation due to sex differences and lateralization of functions. These benefits come at a cost to the generalizability of our results. One of the most well-documented sex differences is in spatial cognition (e.g., Voyer et al., 1995), a difference that often holds in VR experiments (e.g., Levin et al., 2005). However, sex differences do not manifest on all spatial tasks. For example, in a virtual 8-arm maze task, male and female participants were equally likely to report the use of spatial vs. response strategies, and both sexes performed equally well when there were multiple landmarks; only when the environment was devoid of landmarks was a male advantage evident (Andersen et al., 2012). Equally, females out-perform males in detecting changes in object locations, but this advantage is lost when the participant must move viewpoints between encoding and test (Burgess et al., 2004).

The environment used in the present study had distal landmarks clearly visible from all points in the environment that could serve as global orienting cues. However, to encourage participants to use global configural cues and form allocentric representations, there were no local landmarks intermixed with the objects within the virtual arena. The lack of local landmarks might confer a male

FIGURE 6 | Performance-related MTL activations for ROT vs. REF contrast.
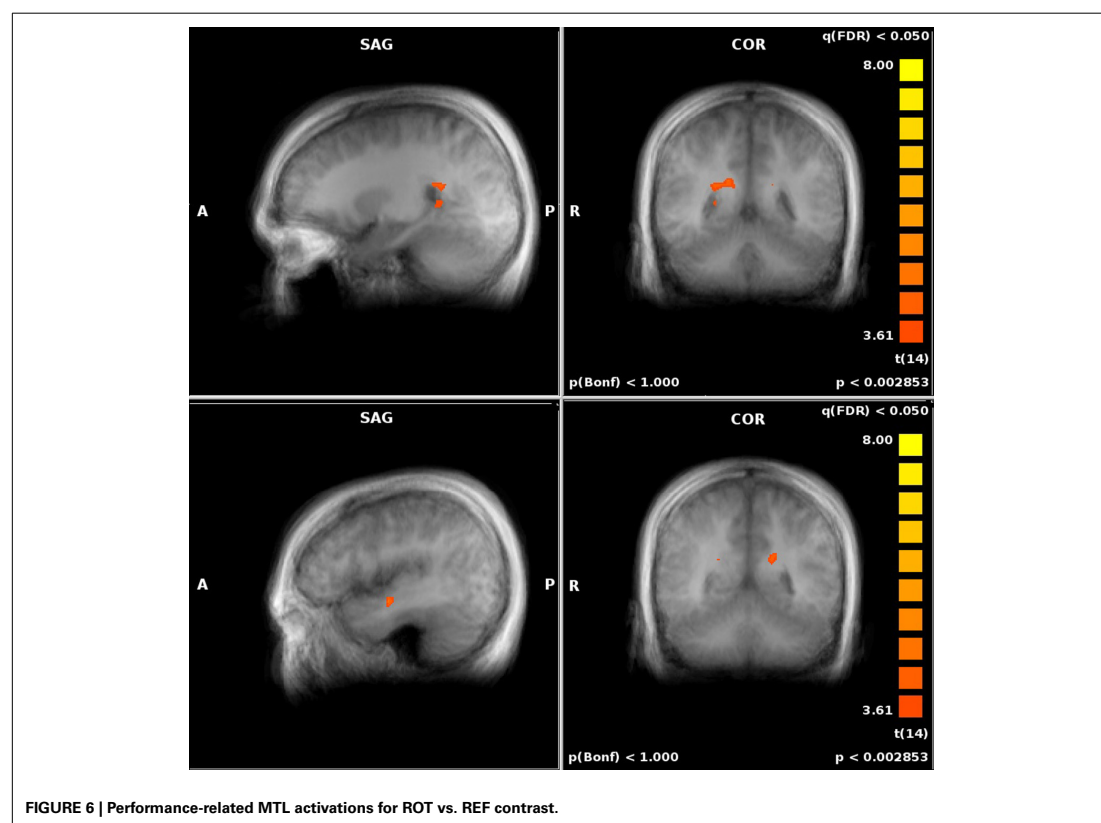
advantage on our task. Future work is required to explore whether there are sex differences and/or individual strategy differences in this specific task, and how they may correlate with use of the transformation circuit.

Another methodological limitation involves the use of a response pointer in the fronto-parallel plane. This method of response was chosen mainly because pilot experiments showed that pointers, cursors, or sliders in plane with the arena ground were difficult and/or time-consuming for the participant, especially with the smaller screen of the scanner, which would have interfered with both task accuracy and the number of trials available for analysis. This method of response, however, could have led to some undesired brain activity, as some participants may have mentally transformed their planned pointing response into the co-ordinates of the fronto-parallel plane. We attempted to minimize this problem by asking participants to imagine pointing as if they were immersed in the environment, which may not always eliminate undesired re-transformation, if it occurs, and may introduce an imagined transformation of the pointer on the horizontal plane. Further studies involving JRD-type responses in a visual or VR paradigm must take into consideration the pros and cons of different implementations of response input, or devise a new strategy.

Screen-based testing in a visual VR paradigm also involves one other important limitation. Since participants needed to watch the screen for cues, there may be some interference in brain activations between the actual visual information being seen by the participant in the scanner and the mental imagery we are attempting to access. We attempted to minimize this by leaving the screen blank as much as possible during time-windows of interest. However, cues and the response pointer were visible during some of this time, and there is likely to be at least some brain activation due to these, since cue complexity varied between REF, ROT, and JRD (potentially contributing to some increased activation in JRD compared to ROT). While activations due to the pointer itself are likely to be mostly filtered out by averaging and parametric modeling, it is possible that how the cues were processed mentally was correlated to some degree with the accuracy of the responses, introducing some interference. However, since surveys showed that participants generally began imagining their cued position, direction, and objects while the cue was present, this interference is unavoidable with the current visual paradigm and setup.

Due to the limitation in the scanning time available for each participant to perform a sufficient number of trials, we were not able to accommodate an additional task, pointing from a fixed

FIGURE 7 | Performance-related MTL activations for JRD vs. REF contrast.
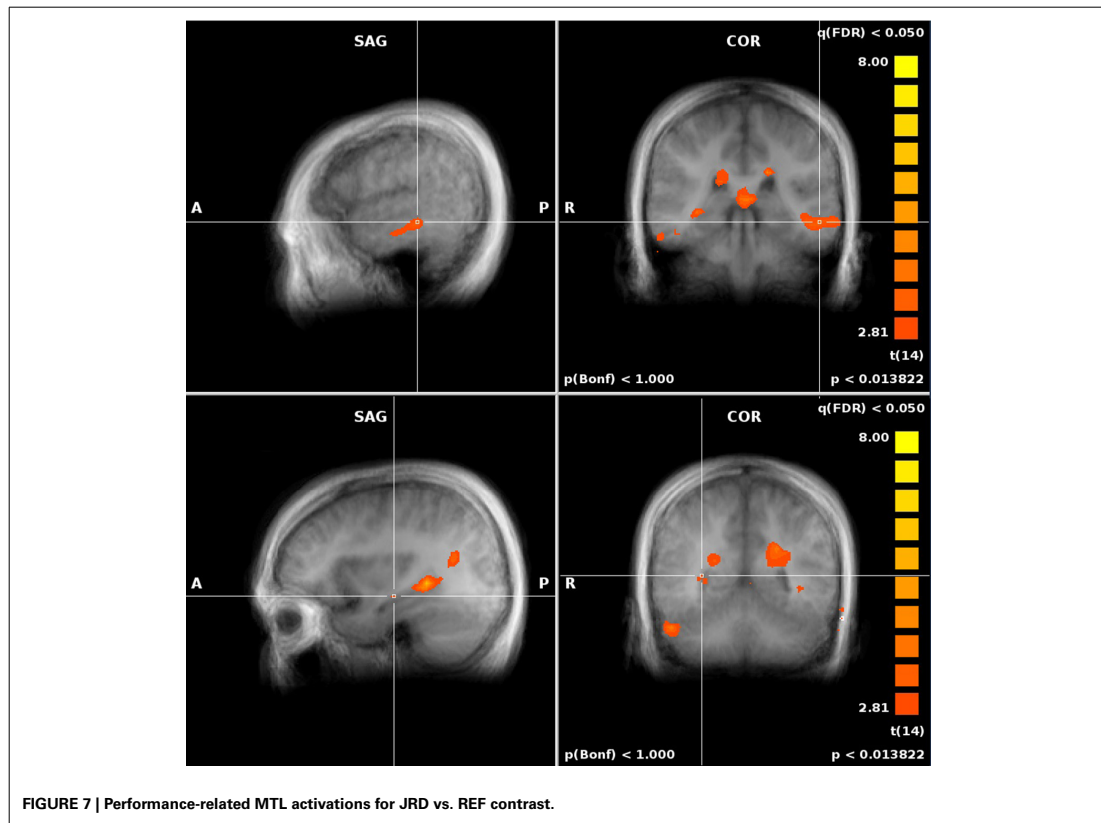
viewing direction with a translation only (i.e., the opposite half of the JRD that ROT encompasses). Though this task does not directly address the questions we have posed for this study, it would shed light on interesting and related questions on the differences between rotation-only and translation-only processing, and their relative contributions to the JRD task. Future work is required to tease apart the independent contributions of rotation vs. translation vs. rotation plus translation to behavioral performance in the tasks reported here, as well as the effects of rotation and translation magnitude on performance and brain activity.

We have focused in this paper on the effects of task demands such as rotation and translation on the neural circuits involved in imagined pointing responses. Other potential sources of variability in participants' responses in the ROT and JRD tasks are the degree of imagined heading disparity and object-target disparity. For example, a well established finding is that tasks requiring a rotation of imagined perspective incur a reaction time cost in proportion to the degree of heading direction disparity (e.g., Rieser, 1989). This has been taken as evidence of a mental self-rotation process that continuously updates in proportion to the degree of imagined rotation. On the other hand, others have called into

question this interpretation and suggested instead that it is the interference between actual and imagined heading directions that causes the cost in reaction time (May, 2004; Wang, 2005). Our current results cannot differentiate whether the mental transformation processes studied here involve a "jump in viewpoint" that is independent of degree of rotation or a gradual viewpoint rotation that takes more time for larger angular changes. All we can conclude is that the same transformation circuit is active in both the ROT and JRD conditions in proportion to error, suggesting that the process of transforming one's viewpoint engages the same circuit irrespective of the degree of cue conflict. Future work is required to systematically manipulate heading and object direction disparity and determine whether these variables would be additional modulators of activation in the spatial transformation circuit investigated here.

REFERENCES

Aguirre, G. K., and D'Esposito, M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain* 122, 1613–1628. doi: 10.1093/brain/122.9.1613

Andersen, N. E., Dahmani, L., Konishi, K., and Bohbot, V. D. (2012). Eye tracking, strategies and sex differences in virtual navigation. *Neurobiol. Learn. Mem.* 97, 81–89. doi: 10.1016/j.nlm.2011.09.007

Anderson, M. I., and Jeffery, K. J. (2003). Heterogeneous modulation of place cell firing by changes in context. *J. Neurosci.* 23, 8827–8835.

Bohbot, V. D., Kalina, K. S., and Spackova, N. (1998). Spatial memory deficits in patients with lesions to the right hippocampus and to the right parahippocampal cortex. *Neuropsychologia* 36, 1217–1238. doi: 10.1016/S0028-3932(97)00161-9

Burgess, N., Spiers, H. J., and Paleologou, E. (2004). Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences in spatial memory. *Cognition* 94, 149–166. doi: 10.1016/j.cognition.2004.01.001

Byrne, P., Becker, S., and Burgess, N. (2007). Remebering the past and imagining the future: a neural model of spatial memory and imagery. *Psychol. Rev.* 114, 340–375. doi: 10.1037/0033-295X.114.2.340

Eichenbaum, H., and Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain.* Oxford: Oxford University Press.

Ekstrom, A., Kahana, M., Caplan, J., Fields, T., Isham, E., Newman, E., et al. (2003). Cellular networks underlying human spatial navigation. *Nature* 425, 184–187. doi: 10.1038/nature01964

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598–601. doi: 10.1038/33402

Fasold, O., Heinau, J., Trenner, M. U., Villringer, A., and Wenzel, R. (2007). Proprioceptive head posture-related processing in human polysensory cortical areas. *Neuroimage* 40, 1232–1242. doi: 10.1016/j.neuroimage.2007.12.060

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., and Mintun, M. A. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI). *Magn. Reson. Med.* 33, 636–647. doi: 10.1002/mrm.1910330508

Formisano, E., Di Salle, F., and Goebel, R. (2006). "Fundamentals of data analysis methods in fMRI," in *Advanced Image Processing in Magnetic Resonance Imaging*, eds L. Landini, V. Positano, and M. F. Santarelli (New York, NY: Marcel Dekker), 481–504.

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. doi: 10.1006/nimg.2001.1037

Goebel, R., Esposito, F., and Formisano, E. (2006). Analysis of FIAC data with BrainVoyager QX: from single-subject to cortically aligned group GLM analysis and self-organizing group ICA. *Hum. Brain Mapp.* 27, 392–401. doi: 10.1002/hbm.20249

Grahn, J. A., Parkinson, J. A., and Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Prog. Neurobiol.* 86, 141–155. doi: 10.1016/j.pneurobio.2008.09.004

Habib, M., and Sirigu, A. (1987). Pure topographic disorientation: a definition and anatomical basis. *Cortex* 23, 73–85. doi: 10.1016/S0010-9452(87)80020-5

King, J. A., Burgess, N., Hartley, T., Vargha-Khadem, F., and O'Keefe, J. (2002). Human hippocampus and viewpoint dependence in spatial memory. *Hippocampus* 12, 811–820. doi: 10.1002/hipo.10070

Kobayashi, Y., and Amaral, D. (2003). Macaque monkey retrosplenial cortex: II. Cortical afferents. *J. Comp. Neurol.* 466, 48–79. doi: 10.1002/cne.10883

Lambrey, S., Doeller, C., Berthoz, A., and Burgess, N. (2012). Imaging being somewhere else: neural asis of changing perspective in space. *Cereb. Cortex* 22, 166–174. doi: 10.1093/cercor/bhr101

Levin, S., Mohamed, F., and Platek, S. (2005). Common ground for spatial cognition? A behavioral and fMRI study of sex differences in mental rotation and spatial working memory. *Evol. Psychol.* 3, 227–254.

Maguire, E. A. (2001). The retrospelnial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand. J. Psychol* 42, 225–238. doi: 10.1111/1467-9450.00233

Maguire, E. A., Nannery, R., and Spiers, H. J. (2006). Navigation around London by a taxi driver with bilateral hippocampal lesions. *Brain* 129, 2894–2907. doi: 10.1093/brain/awl286

May, M. (2004). Imaginal perspectives in remembered environments: transformations versus interference accounts. *Cogn. Psychol.* 48, 163–206. doi: 10.1016/S0010-0285(03)00127-0

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Exp. Neurol.* 51, 78–109. doi: 10.1016/0014-4886(76)90055-8

Overte Foundation. (2007). *Opensimulator.* Available online at: http://www.opensimulator.org

Parsons, T. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia* 42, 555–562. doi: 10.1016/j.neuropsychologia.2003.08.014

Postle, B. R., and D'Esposito, M. (2003). Spatial working memory activity of the caudate nucleus is sensitive to frame of reference. *Cogn. Affect. Behav. Neurosci.* 3, 133–144. doi: 10.3758/CABN.3.2.133

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 1157–1165. doi: 10.1037/0278-7393.15.6.1157

Rosenbaum, R. S., Priselac, S., Kohler, S., Black, S. E., Gau, F., Nadel, L., et al. (2000). Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. *Nat. Neurosci.* 3, 1044–1048. doi: 10.1038/79867

Shelton, A. L., and McNamara, T. P. (1997). Multiple views of spatial memory. *Psychon. Bull. Rev.* 4, 102–106. doi: 10.3758/BF03210780

Sherrill, K. R., Erdem, U. M., Ross, R. S., Brown, T. I., Hasselmo, M. E., and Stern, C. E. (2013). Hippocampus and retrosplenial cortex combine path integration signals for successful navigation. *J. Neurosci.* 33, 19304–19313. doi: 10.1523/JNEUROSCI.1825-13.2013

Sulpizio, V., Committeri, G., Lambrey, S., Berthoz, A., and Galati, G. (2013). Selective role of lingual / parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. *Behav. Brain Res.* 242, 62–75. doi: 10.1016/j.bbr.2012.12.031

Takahashi, N., Kawamura, M., Shiota, J., Kasahata, N., Hirayama, K., and Hirayama, K. (1997). Pure topographic disorientation due to right retrosplenial lesion. *Neurology* 49, 464–469. doi: 10.1212/WNL.49.2.464

Talairach, J., and Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain.* New York, NY: Thieme.

Voyer, D., Voyer, S., and Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol. Bull.* 117, 250–270. doi: 10.1037/0033-2909.117.2.250

Wang, R. F. (2005). Beyond Imagination: perspective change problems revisited. *Psicologica* 26, 25–38.

Wang, X., Zhang, M., Cohen, I. S., and Goldberg, M. E. (2007). The proprioceptie representation of eye position in monkey primary somatosensory cortex. *Nat. Neurosci* 10, 640–646. doi: 10.1038/nn1878

Wyss, J., and Groen, T. V. (1992). Connections between the retrosplenial cortex and the hippocampal formation in the rat: a review. *Hippocampus* 2, 1–11. doi: 10.1002/hipo.450020102

Zhang, H., Copara, M., and Ekstrom, A. D. (2012). Differential recruitment of brain networks following route and cartographic map learning of spatial environments. *PLOS ONE* 7:e44886. doi: 10.1371/journal.pone.0044886

Zhang, H., and Ekstrom, A. (2013). Human neural systems underlying rigid and flexible forms of allocentric spatial representation. *Hum. Brain Mapp.* 34, 1070–1087. doi: 10.1002/hbm.21494

3.3   Discussion

The understanding gained through this study of the combined roles of the hippocampus, parahippocampal gyrus, precuneus, parietal cortex, and retrosplenial cortex in computing viewpoint transformations via spatial imagery could, in theory, be used to drive a BCI. That activity in the temporo-parietal network was correlated with pointing accuracy after viewpoint transformations suggest information specifying pointing direction may be carried by these brain signals. If so, it would be possible to construct a BCI which is capable of resolving a direction of focus in 360 degrees around the user with better than chance accuracy.

As discussed at the beginning of this chapter, a BCI based on resolving direction from visuospatial imagery could provide more natural and more fine control for virtual or real-world navigation or cursor control than is currently available. However, a study would have to be carried out in order to determine to what extent activity in this network can be used to resolve imagined spatial directions, and furthermore, whether that information could be extracted from EEG as well as from fMRI. Since it is notoriously difficult to precisely resolve, from EEG, activity generated from multiple deep brain structures, like those involved in the temporo-parietal network identified in this study, extracting relevant features using EEG in order to facilitate a BCI could be a very challenging problem on its own. However, translating the results of this study into the BCI context might be feasible.

While conducting this study, proposals were made for developing a BCI based on visuospatial imagery in a virtual reality setting. The idea for the experiment was to have objects appear in a random location around the participant on a plane parallel to the ground one at a time. The participant would then have to focus on the direction of the object (an alternative strategy would be to imagine navigating to that object). In future iterations, the distance between the participant and the object could be varied, and objects could be made to appear within a sphere around the participant. This setup would allow for EEG data to be collected associated with focusing on pointing or moving in a direction in space. The goal would be to construct a BCI that could resolve a user's intended pointing direction. In test phases, a beam would point out from the partipant in the virtual reality environment, and the participant would be tasked with pointing to the object with the beam or selecting from a set of objects using the beam via their brain activity. Thus, the beam would act as real-time feedback for training the participant.

An outline of a strategy for analyzing the EEG data was also developed. First, a high density electrode cap would be needed in order to construct beamformers targetting the nodes in the medial temporal network revealed through the study presented in this chapter. Alternatively, a more robust approach might be to

use a technique like sLORETA (standardized low-resolution brain electromagnetic tomography) [5], which maps spatial EEG information into voxels. Such a method can therefore be used to discover additional locations relevant for classification in the EEG signal and to address the problem of individual variation in the precise locations of the hippocampus, parahippocampus, and other relevant brain regions. The raw EEG signal can then be transformed into a source-space multi-channel time series, from which features can be extracted using a CSP approach or the SF approach defined in Chapter 2 and a standard classifier could be used. If successful, such a system could be used for avatar control, object selection, and object manipulation in a virtual reality setting, and later adapted for similar real-world applications.

While a whole body of research could be undertaken in order to extend these findings to brain-computer interfacing, it was not pursued. Around the time this work was completed, the standard approach of utilizing such findings to design a BCI based on specific neurophysiological signals was becoming less attractive, because it was becoming more clear that the large variability in performance and reliability of BCIs across individuals was due in part to the fact that different individuals are better able to control different neurophysiological signals [6, 7, 8]. The large body of work which would have gone into designing a BCI specifically for a narrow aspect of visuospatial memory would at best be useful only for individuals who happen to have individual backgrounds and characteristics favourable for learning control over parts of their temporo-parietal network. It was becoming clear that a better way to make progress in brain-computer interfacing would be to move towards a generalized BCI rather than another BCI based on predefined mental commands.

Bibliography

[1] D. Huang, K. Qian, D.-Y. Fei, W. Jia, X. Chen, and O. Bai, "Electroencephalography (EEG)-based brain–computer interface (BCI): A 2-D virtual wheelchair control based on event-related desynchronization/synchronization and state control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 3, pp. 379–388, 2012.

[2] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review.," *Sensors (Basel, Switzerland)*, vol. 12, pp. 1211–79, Jan. 2012.

[3] A. F. Cabrera and K. Dremstrup, "Auditory and spatial navigation imagery in brain–computer interface using optimized wavelets," *Journal of neuroscience methods*, vol. 174, no. 1, pp. 135–146, 2008.

[4] E. A. Curran and M. J. Stokes, "Learning to control brain activity: a review of the production and control of EEG components for driving brain–computer interface (BCI) systems," *Brain and cognition*, vol. 51, no. 3, pp. 326–336, 2003.

[5] R. D. Pascual-Marqui, "Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details," *Methods & Findings in Experimental & Clinical Pharmacology*, vol. 24, no. Suppl D, pp. 5–12, 2002.

[6] B. Z. Allison and C. Neuper, "Could anyone use a BCI?," in *Brain-computer interfaces*, pp. 35–54, Springer, 2010.

[7] C. Neuper and G. Pfurtscheller, "Neurofeedback training for BCI control," in *Brain-Computer Interfaces*, pp. 65–78, Springer, 2010.

[8] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain–computer interface: A brief review," *Journal of neuroscience methods*, vol. 243, pp. 103–110, 2015.

Chapter 4

Open-Ended Brain-Computer Interfacing: Moving Beyond Prescribed Mental Commands

4.1   Introduction

In Chapter 3, some applications of spatial imagery to BCI were hypothesized and the appropriate neural targets were elucidated using fMRI. Studying the neurophysiological basis of a cognitive task and then using that knowledge to design a BCI is the standard approach used in the BCI literature. Using this approach, the BCI community has significantly expanded the number and variety of BCI applications (*e.g.*, [1, 2, 3, 4]), and, to a lesser extent, allowed for new types of mental imagery to be used in the control of BCIs [5, 6]. In order to facilitate the rise in the variety of BCIs, corresponding advances in signal processing and machine learning methods for BCIs have also taken place [7, 8, 9, 10].

Despite the rapid advancement in BCI technology over the last decade, BCIs have yet to make significant in-roads to the oft mentioned but rarely reached end user. BCIs have yet to achieve the speed, reliability, usability, and convenience needed for practical applications [11]. The most visible evidence for this is the BCI illiteracy rate [12, 13, 14], which refers to an estimated 15% to 25% of users who cannot control a given BCI reliably at all. Just as problematic is the fact that the majority of BCI users only achieve moderate levels of accuracy in BCI control (between 60% and 85%). While improvements in signal processing and machine learning methods have increased average BCI performance rates over time, they have not had as substantial an impact on reducing the BCI illiteracy rate or in making BCIs reliable enough to be useful whenever alternative methods of control are available [15].

Users who might have difficulty controlling a BCI with one set of mental commands can often control a BCI reliably using a different set of mental commands, even if the commands leading to successful control fall under the same category as commands which led to poor control (*e.g.*, using foot motor imagery instead of hand motor imagery) [16, 17]. There is evidence to suggest that there are individual differences in BCI performance with respect to the category of mental task as well. For example, some individuals perform well with motor imagery but not with visual imagery, and others perform well with visual imagery and not motor imagery [5, 6].

Both of the above sources of variability are related to differing abilities to voluntarily modulate certain neurophysiological signals, even with training [18, 14, 19]. The implications of these findings lead to the following conjecture: in order to design a BCI that is suitable for the majority of people, different individuals need to be able to use different kinds of mental commands. This is only achievable in any practical way with a generalized BCI.

In this chapter, a step toward a generalized BCI is taken by designing a BCI

transducer that allows each individual to use a personalized set of mental commands. This is named the open-ended BCI. An open-ended BCI achieves its flexibility by shifting the burden of finding a good set of personalized mental commands to the user and requiring the machine learning algorithms to find discriminative patterns of brain activity without access to *a priori* information about the user's chosen mental commands or their corresponding neurophysiological correlates. This type of design, and the more generalized BCI transducer required to implement it, can be utilized in further developments of a generalized BCI.

4.2   Open-Ended Brain-Computer Interfaces

Dhindsa, K., Carcone, D., & Becker, S. (2017). Towards and Open-Ended Brain-Computer Interface: A User-Centred Co-Adaptive Design Approach. Neural Computation, 29:10.
Accepted pre-published manuscript reprinted with permission.

Towards an Open-Ended BCI: A User-Centred Co-Adaptive Design Approach

Kiret Dhindsa: *Neurotechnology and Neuroplasticity Lab*
School of Computational Science and Engineering
McMaster University
Dean Carcone: *Neurotechnology and Neuroplasticity Lab*
Department of Psychology, Neuroscience, and Behaviour
McMaster University
Suzanna Becker: *Neurotechnology and Neuroplasticity Lab*
Department of Psychology, Neuroscience, and Behaviour
McMaster University

Abstract

Brain-computer interfaces (BCIs) allow a user to control a device by interpreting their brain activity. For simplicity, these devices are designed to be operated by purposefully modulating specific predetermined neurophysiological signals, such as the sensorimotor rhythm. However, the ability to modulate a given neurophysiological signal is highly variable across individuals, contributing to the inconsistent performance of BCIs for different users. On the other hand, these differences suggest that individuals who experience poor BCI performance with one class of brain signals might have good results with another. In order to take advantage of individual abilities as they relate to BCI control, we need to move beyond the current approaches.

In this paper we explore a new BCI design aimed at a more individualized and user-focused experience, which we call the Open-Ended BCI. Individual users were given the freedom to discover their own mental strategies as opposed to being trained to modulate a given brain signal. Users then underwent multiple co-adaptive training sessions with the BCI. Our first Open-Ended BCI performed similarly to comparable BCIs while accommodating a wider variety of mental strategies without a priori knowledge of the specific brain signals any individual might use. Post-hoc analysis revealed individual differences in terms of which sensory modality yielded optimal performance. We found a large and significant effect of individual differences in background training and expertise, such as in musical training, on BCI performance. Future research should be focused on finding more generalized solutions to user training and brain state decoding methods to fully utilize the abilities of different individuals in an Open-Ended BCI. Accounting for each individual's areas of expertise could have important implications on BCI training and BCI application design.

Keywords: mental imagery; brain-computer interface; human learning; individual factors; co-adaptation; user-centred design; auditory imagery; visual imagery

1   Introduction

Brain-computer interfaces (BCIs) enable direct communication between a brain and a computer by interpreting a user's intentions from brain activity [Wolpaw et al., 2002]. In contrast to BCIs based upon event-driven potentials, spontaneous (also called asynchronous) BCIs allow the user to generate mental commands at any point in time without external cues [Mason and Birch, 2000, Millan and Mouriño, 2003, Borisoff et al., 2004, Kus et al., 2012]. The lack of a well defined time window that encompasses the brain activity of interest makes segmentation particularly challenging. Moreover, the brain activity of interest may not follow some characteristic morphology that is consistent across users. For example, users are often instructed to imagine hand or foot movements to modulate the sensorimotor rhythm (SMR) when controlling a spontaneous BCI (e.g., [Diez et al., 2011, Thomas et al., 2013]). Over time, a user's strategy may shift and they may use very different patterns of brain activity for control.

Due to the nonstationarity of the user's brain signals when controlling a spontaneous BCI (see [Achtman et al., 2007, Sugiyama et al., 2007, Von Bünau et al., 2009, Von Bünau

et al., 2010]), co-adaptation between the user and the machine is required in order to provide reliable control. The user must learn to provide clear and consistent brain signals related to each mental command while the BCI, typically via machine learning, must learn the patterns of brain activity associated with the user's intentions. The high variance of neurophysiological signals, even for the same user engaged in the same mental activity, means that a BCI must attempt to solve a very complicated translation problem.

In order to simplify the problem, BCI designs traditionally require that the user issue commands by modulating a predefined brain signal which is well understood from both physiological and computational perspectives. Even with this simplification, many users have difficulty achieving a satisfactory level of control. Useful control of a binary-output BCI is often heuristically defined as above 70% classification accuracy [Kübler et al., 2001]. By this definition, 15%-25% of users are unable achieve useful control using traditional BCI approaches, which has been referred to as BCI illiteracy [Blankertz et al., 2010, Kübler and Müller, 2007]. This persistent problem prevents wider adoption of BCI in the real world [Neuper and Pfurtscheller, 2010, Allison and Neuper, 2010].

BCI illiteracy rates can be reduced by allowing for some individualization in the mental commands employed by the user. One such study incrementally trained novice users to operate a binary motor imagery BCI [Vidaurre et al., 2011a, Vidaurre et al., 2011b]. In this co-adaptive approach, users performed their first three training sessions with a set of three classes of motor imagery (left hand, right hand, and foot) which all modulate the SMR. In subsequent training sessions, users controlled a binary BCI using the pair of commands which gave the best classification accuracy during the earlier sessions. This resulted in a 50% reduction in the BCI illiteracy rate.

While the above approach is a promising first step towards more individualized BCIs, it still restricts users to a single predefined sensory modality and dictates to the user which specific mental commands to employ. Further individualization might be achieved by allowing users to employ mental strategies involving different sensory modalities. However, expanding this approach to different sensory modalities may mean that the search space of possible mental commands would become too large to efficiently find an optimal subset of commands for each user.

Although it is a challenging problem, there is a pressing need for individualized BCIs controllable with non-conventional mental commands. A significant contributing factor for this need is the fact that different individuals are able to control different brain signals to different degrees. For example, the ability to modulate SMR is affected by an individual's background and past experience [Hammer et al., 2012, Randolph et al., 2010, Randolph, 2011], as well as their cognitive profile [Burde and Blankertz, 2006, Allison et al., 2010, Vuckovic and Osuagwu, 2013], even after automatization (see [Wolpaw and McFarland, 2004, Pineda et al., 2003]) of the mental commands has taken place [Scherer et al., 2015]. Notably, stroke and spinal cord injury patients, who stand to benefit significantly from advances in BCI technology, tend to perform with significantly less accuracy when using a motor imagery BCI than do healthy individuals (e.g., [Scherer et al., 2015]), possibly due to degradation in the sensorimotor cortices [Wolpaw et al., 2002].

Rather than tackling the problem of unreliable performance for some users by providing different ways of modulating the same neurophysiological signal, a better long-term solution may lie in allowing different users to use different neurophysiological signals altogether [Allison and Neuper, 2010] (see also [del R. Millàn et al., 2002, Nai-Jen and Palaniappan, 2004] for examples of BCI studies using a variety of mental tasks). This

has also been suggested in a study in which different individuals had more classifiable brain activity with different kinds of mental imagery, illustrating that individual choice in imagery type might be beneficial for BCI performance [Friedrich et al., 2012]. In fact, it has been shown before that individual factors including personality and cognitive profile can affect how well someone can perform with a BCI using different kinds of mental imagery [Jeunet et al., 2015], but the relationship between individual factors and performance with a mental imagery BCI has not been thoroughly explored, nor has domain expertise specifically been linked to abstract user-defined visual and auditory imagery in an online BCI.

Though restricting users to controlling a BCI with a specific predetermined neurophysiological signal was a design decision that helped make modern BCIs possible, we may be required to lift this restriction in order for BCIs to provide more consistent performance across a variety of users. In order to achieve this new level of individualization, we propose a BCI which allows the user to explore and define their own mental commands, and we call this an Open-Ended BCI. Though the traditional BCI approach may be important for many specific applications meant for specific kinds of users, the development of an Open-Ended BCI may lead to a wider variety of BCI applications which can be used by the general population.

Study Purpose

In this paper, we aim to evaluate the feasibility of a simple Open-Ended BCI using existing BCI methodology and identify whether expertise in different areas relevant to certain types of mental imagery has an effect on BCI performance. In order to do this, we trained participants to control a BCI for three tasks, each by using mental imagery in a different sensory modality, but each using the same user training and computational methods. Users were free to use any mental commands which they saw fit within each sensory modality, and were not asked to reveal their choices of mental commands until after each training session.

2 Methods

Fourteen undergraduate and graduate participants (six females) trained to control an Open-Ended BCI for three tasks. For one task, the pitch of a tone was controlled using auditory imagery, for another, the size of an on-screen object was controlled using visual imagery, and for the third, the position of an object on the screen was controlled using motor imagery. Each participant completed three 30-minute sessions with each task over the course of one week, totaling nine sessions over three weeks. The order in which the three tasks were trained was counter-balanced across participants. Participants received monetary compensation for each session and the experiment was approved by the McMaster Research Ethics Board.

Participants were shown the outputs associated with each task immediately before their first training session with that particular task. For example, before a participant began their first training session with the auditory task, the lowest and highest tones were played to illustrate the range of possible tones. Participants were then given time to privately create two distinct mental commands within the appropriate sensory modality which they would use to produce each output. Since participants found it difficult to produce mental imagery involving only one sensory modality, they were instructed to

create mental commands such that the correct sensory modality was the main focus of the command. For example, while auditory imagery may have induced associated visual imagery during the auditory task, we asked the participant to create mental commands where the most salient features would be auditory rather than visual. Participants were asked to keep their mental commands consistent throughout the session, but were able to change their commands before the second session only if they did not achieve above chance control during the first session.

2.1 EEG Apparatus

EEG (electroencephalogram) was recorded using the Emotiv Epoc [Emotiv Systems, 2011]. This is a low cost consumer grade EEG headset that has previously been evaluated for its signal quality [Badcock et al., 2013, Duvinage et al., 2013, Lievesley et al., 2011] and has been used in BCI experiments [Liu et al., 2012, Carrino et al., 2012]. Though it provides significantly lower SNR (signal-to-noise ratio) than research grade EEG devices, it still provides useful EEG information. Since we are interested in the study and design of practical general-user BCI applications which might eventually move beyond the laboratory or clinic we selected this device over a research grade EEG system.

The Emotiv Epoc consists of 14 saline-based sensors (located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 according to the international 10-20 system) and two reference electrodes located at P3 and P4 (see Figure 1). The headset records at a sampling rate of 128Hz and implements a 0.2-45Hz bandpass filter and 50 Hz and 60 Hz notch filters in hardware.
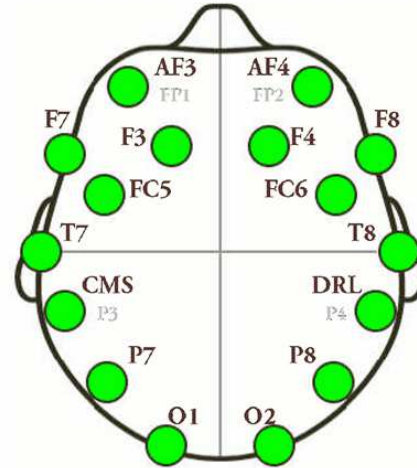
Figure 1: Emotiv Epoc electrode layout.

2.2 Experimental Procedure

The same experimental protocol was used for all tasks and sessions were held in a private room. Data acquisition and processing was done in Matlab R2013b [MATLAB, 2013] using Simulink and Psychtoolbox [Brainard, 1997].

After fitting the headset to the participant, impedance was assessed for each electrode using the proprietary Emotiv Epoc Control Panel software, because impedance measures are not given directly with this device (instead impedance is shown using a colour-coded visual display). The electrodes were adjusted and the felt tips covering the electrodes were re-moistened with saline solution as needed in order to maximize signal quality. Due to the quality of the hardware, many sessions began with at least one electrode showing poor impedance. Furthermore, the quality of the signal decayed significantly throughout most sessions. We also noticed difficulty achieving and acceptable signal quality for certain participants, as all sessions for participants P4, P5, and P6 were conducted with at least four electrodes showing either poor impedance or no signal at all.

Each session for each task followed the same structure. Text was displayed at the start of each session explaining the basic instructions of the experiment. Instructions to avoid blinking, eye movements, head movements, jaw clenches, and other muscular actions during mental imagery were given both textually and verbally by the experimenter. The participant was left alone in the experiment room and was free to begin the session when ready by pressing a key on the keyboard. Every session included ten blocks of twenty trials, each with balanced classes in a randomized order. The first block of every session was a pretraining block intended to acquire data with which to establish an initial classifier. Therefore, no online classification was performed and no feedback was presented during the pretraining block. Participants were able to take a break at the end of each block and continued when they were ready. Figure 2 illustrates the structure of each trial.

Participants completed a brief questionnaire at the end of each session in which they were asked to describe the mental imagery they used and rate the usability of the system. Participants were also asked to complete a short questionnaire on their history of athletic training, visual arts training and musical training (for each, participants were asked how many hours per week they practice, how many years they have practiced in total, and to report their overall level of expertise in the area). Note that participants P1 through P6 completed a slightly different questionnaire, and so only their self-rated expertise in each area is available for analysis.

2.3   The Three Tasks

Each task was designed to be as similar to each other task as possible, but was differentiated by the outputs of the BCI, the feedback stimulus, and the type of mental imagery participants were instructed to use. The experimenters and the BCI itself were blind to the actual mental commands used until after the training session.

During the auditory imagery task, participants were cued with text on each trial to produce either high (A5 at 880Hz) or low (A3 at 220Hz) pure tones. These tones were played at the start of each block of trials to remind the participant of the minimum and maximum pitch. The wide range of tones was chosen in order to allow participants with untrained pitch perception to more easily hear differences between the generated output and the target pitch. The participant was instructed to use auditory imagery to produce a tone as high or low as possible according to the trial cue. A tone between the low and high tones was played back as feedback.

In the visual imagery task, the goal was to change the size of a white circle displayed at the center of the screen. The circle had a minimum diameter of 50 pixels, a neutral diameter of 150 pixels, and a maximum diameter of 250 pixels. A circle of variable size was displayed as feedback.

The motor imagery paradigm involved the use of kinaesthetic motor imagery to move a white circle with a diameter of 100 pixels from the
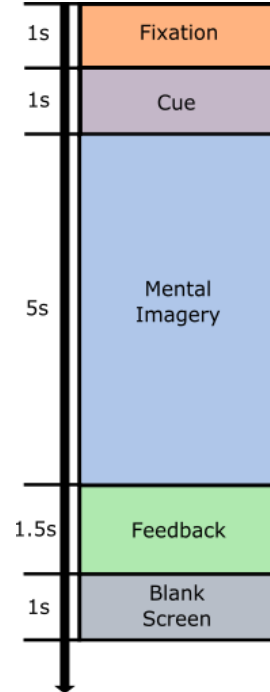
Figure 2: Trial structure

center towards the left or the right of the screen. The circle could be displaced by a maximum of 500 pixels (approximately 25% of the screen width) in either direction. A variably displaced circle was displayed for feedback.

2.4   Feedback Approach

Feedback was proportional to classifier confidence. This was measured as the estimated posterior probability that the issued mental command corresponded to the trial cue (see [Chang and Lin, 2011, Platt et al., 1999, Wu et al., 2004, Lin et al., 2004, Lin et al., 2007]). For example, in the motor imagery task, a probability of 0.7 would move the circle $2(0.7 - 0.5) = 40\%$ of the distance in the correct direction. A probability of 0.5 resulted in feedback exactly halfway between either extreme (no change for motor or visual imagery, and the tone exactly halfway between the low and high tones in audio imagery). Probabilities below 0.2 and above 0.8 were rounded to 0 or 1. Real time feedback was not presenting during the mental imagery period.

We instructed participants to try to produce outputs which were as close as possible to the target rather than instructing them to produce outputs which were simply in the correct direction. Thus, participants were training to increase classifier confidence, which implies that they were training to produce brain activity farther away from the classification boundary. We chose this approach in order to promote training towards increasingly separable mental commands in feature space over time in a more direct way compared to simple binary feedback.

2.5   Data Analysis

The experimental data was analyzed both online during each session, and offline after the entire experiment had been completed. Each session was treated independently during online analysis. A new feature space and classifier were created for each session after the pretraining block, and the feature space and classifier were updated at the end of each training block using all trials completed within the session. The updated model was used to classify trials and provide feedback during the following block.

The Open-Ended BCI approach requires generalizable signal processing and machine learning methods, because the brain activity used to issue commands is not known beforehand. A simple approach to finding a useful feature space when limited a priori knowledge is available is to select an optimal subset of features from a larger set of candidate features. Common Spatial Patterns (CSP) [Müller-Gerking et al., 1999, Ramoser et al., 2000] with power spectral density (PSD) and Filter-Bank Common Spatial Patterns (FBCSP) [Ang et al., 2008] were used for feature extraction in online and offline analysis respectively. Minimum Redundancy Maximum Relevance (MRMR) [Peng, 2005] was used for feature selection, and a linear Support Vector Machine (SVM) [Cortes and Vapnik, 1995] was used for classification with the libSVM Matlab toolbox [Chang and Lin, 2011]. It is important to note that while participants were trained with non-binary BCI outputs, the underlying system was set up to solve a binary classification problem.

2.5.1   Common Spatial Patterns

Common Spatial Patterns (CSP) is a PCA-based supervised spatial filter and feature extraction algorithm originally developed for motor imagery classification with EEG [Müller-Gerking et al., 1999, Ramoser et al., 2000]. CSP aims to learn a spatial filter

which is optimal for discriminating between different mental commands; several variations have been developed that allow for classifying a wider variety of mental imagery tasks [Blankertz et al., 2008, Kothe et al., 2013, Bobrov et al., 2011, DaSalla et al., 2009]. For online analysis, a CSP filter with four components was constructed after bandpass filtering the EEG from 8-30 Hz with a fourth order Butterworth filter. This yielded four CSP features corresponding to the largest two and smallest two eigenvalues from the eigendecomposition of the whitened class-specific spatial covariance matrices. Since CSP is a standard algorithm used in BCI, the details are omitted here (see [Blankertz et al., 2008] for a detailed description of the CSP algorithm). PSD was computed in 1 Hz bins for each CSP component, providing an additional 88 features. Features were selected using MRMR via 8-fold cross-validation.

2.5.2   FBCSP

In Filter-Bank CSP (FBCSP) [Ang et al., 2008], the EEG signals to be classified are first bandpass filtered into a set of L distinct frequency bands. The CSP algorithm is then applied to each filtered signal to obtain CSP components specific to each frequency range. The components are then concatenated together to obtain $L \times M$ CSP components in total, each of which contributes a candidate feature for feature selection. For offline analysis, FBCSP filters were constructed over 4 Hz wide passbands with 2 Hz overlap in the 6-30 Hz range, resulting in $L = 11$ filters with $M = 4$, totalling 44 features. The offline classification procedure is described in further detail below.

2.5.3   MRMR Feature Selection

The MRMR (minimum Redundancy Maximum Relevance) feature selection method developed by Peng et al. [Peng, 2005] is an information theoretic approach to feature selection for supervised classification problems. It aims to maximize the joint mutual information between a selected subset of features and the true class labels while minimizing the mutual information among the selected features themselves. The method is effective for feature sets which include a large number of highly correlated features where only a relatively small subset contains independent discriminative information.

Features are selected sequentially until K features are selected, where K is a chosen value. The first feature to be selected, z_1, is the feature which has maximum mutual information with the class labels based in a training set. If $\mathbf{X}$ is the matrix of values of all M candidate features x_i for each sample in a training set and Y are the training labels, then

$$z_1 = \max_i I(\mathbf{X} = \{x_i, i = 1, \ldots, M\}; Y),$$

where I is the mutual information function. Each kth included feature for $k = 2, \ldots K$ is selected by maximizing $D - R$. D is considered a measure of relevance and is defined as

$$D = I(\mathbf{Z}_k = \{z_i, i = 1, \ldots, k\}; Y),$$

which is the joint mutual information between all $k - 1$ selected features, including the candidate feature z_k, and Y. Similarly, R is taken as a measure of redundancy and defined as

$$R = I(\mathbf{Z}_k, \mathbf{Z}_k),$$

which is the mutual information among all selected features, including the candidate

feature z_k.

In practice, D is approximated by

$$\bar{D} = \frac{1}{k} \sum_{z_i \in \mathbf{Z}_k} I(z_i; Y)$$

and R is approximated by

$$\bar{R} = \frac{1}{k^2} \sum_{z_i, z_j \in \mathbf{Z}_k} I(z_i; z_j).$$

The number of selected features could take on the values $K = 5, 10, \ldots, 40$ in both online and offline analysis depending on which value of K yielded the greatest cross-validation accuracy.

2.6 Offline Classification Procedure

Offline analyses were conducted in order to confirm the results of online classification and to explore more computationally intensive methods of data analysis. Since CSP is more robust to overfitting when a large number of trials [Reuderink and Poel, 2008] and a small number of channels [Sannelli et al., 2010] are used, as was the case in our experiment, we do not expect that overfitting was a significant problem in this study. However, the offline results further help to ensure that the BCI performance metrics obtained in online analysis were not due to overfitting. The following procedure was used in offline analysis in order to classify trials from the same session:

1. *The Fieldtrip toolbox [Oostenveld et al., 2011] was used to reject trials containing artifacts:*

   (a) *Trials containing EOG (ocular) artifacts were removed by applying a fourth order Butterworth bandpass filter with a passband of 1-15 Hz to the de-trended and z-scored signal from each channel. Z-value thresholding was applied after averaging all channels.*

   (b) *Of the remaining trials, those containing EMG (muscular) artifacts were removed by applying a ninth order Butterworth bandpass filter with a passband of 30-60 Hz to the de-trended and z-scored signal from each channel. Z-value thresholding over was applied after averaging all channels.*

2. *The remaining data were then partitioned into non-overlapping training and test sets, with the test set containing approximately 25% of the trials.*

3. *The 0.5s to 4s time window was extracted from each trial.*

4. *FBCSP filters were learned using training data only and then used to extract features for both the training set and test set.*

5. *Each FBCSP feature was z-transformed using the sample mean and sample standard deviation in the training set.*

6. *The MRMR method was used to select features from the list of candidate FBCSP features.*

7. *A linear SVM was trained to perform classification on the test set using the selected features.*

Steps 2 through 7 were repeated 15 times for each combination of parameters. Regarding artifact rejection, all z-value thresholds were chosen per session based on visual inspection of every trial. These thresholds were chosen conservatively so that some clean trials were discarded in order to avoid including other trials containing small artifacts in the final analysis.

3   Results

3.1   Online Experiment

3.1.1   Mental Commands

Participants were asked to describe the specific mental commands used for each task in post-session questionnaires. The mental commands are summarized in Table 1. The questionnaire also asked participants to rate their level of interest in the task from one to ten. Interest in each task was correlated with BCI performance ($r = 0.25$, $p < 0.05$). Task order did not have a significant effect on performance ($F_{2,120} = 0.21$, $p = 0.81$).

3.1.2   Classification Accuracy

Table 2 summarizes the online results by session for each of the 14 participants. The average classification accuracy for the entire session is reported. Since each session began with new CSP and classification models, the average accuracy over just the last five blocks are also reported in order to mitigate the effect of suboptimal models early in each session on BCI performance estimates. The following analysis is based on the BCI performance during just the final five blocks of each session.

Nine participants surpassed the 70% heuristic for BCI control using at least one form of mental imagery. There was also a small effect of sensory modality ($F_{2,120} = 4.26$, $p < 0.05$, variance explained: $\omega^2 = \frac{SS_{between} - (k-1)MSE}{SS_{total} + MSE} = 0.05$), which was driven by the comparatively high performance with auditory imagery. The distribution of online performance results for each sensory modality is shown in Figure 3.

3.1.3   Human Learning

Since classification models were independently constructed for each session (i.e., no data from previous sessions were used in training the models used in any given session), changes in performance across sessions could not be attributed to the models themselves. Since the subject is the only part of the system which is common across sessions, improvements in BCI performance across sessions within a sensory modality could be taken as a measure of human learning. We observed that the rate of change in accuracy across sessions was significantly greater than zero ($t_{40} = 2.55$, $p < 0.01$; $t_{12} = 2.93$, $p < 0.01$ when the participant had at least one above chance session for that task). Additionally, there was a small but significant effect of session number on BCI performance ($F_{2,120} = 3.44$, $p < 0.05$, variance explained: $\omega^2 = 0.03$).

Imagery	Motor	Visual	Auditory
P1	Sweeping right arm / Sweeping left leg	Feedback stimulus itself	Piercing high note / Muffled low note
P2	N/A	N/A	N/A
P3	N/A	N/A	Opera singer / Chanting monks
P4	Raising hands / Bowling	Explosion / Black hole	Opera / Foghorn
P5	N/A	N/A	N/A
P6	N/A	Feedback stimulus itself	Hammering / Kicking a ball
P7	Guitar chord with left hand / Slapping with right hand	Growing blue circle / Shrinking marble	Buzzy kazoo / Leonard Cohen singing
P8	Boxing with right hand / Guarding with left hand	Self expanding / Shrinking ball in hands	Jazz trumpet / Heavy metal
P9	Retracting hand from hot stove / Painting with brush	Moon getting closer / Car driving away	Opera singer / Chanting om
P10	Punching with right hand / Stretching right arm to the left	*Withdrew from study*	Singing or playing high notes / Singing or playing low notes
P11	Left and right hand actions (not described)	Feedback stimulus itself	Feedback stimulus
P12	Lifting Dumbell / Dribbling basketball	Car driving away/ Balloon expanding	Bell / Foghorn
P13	Singing a high pitch / Singing a low pitch	Feedback stimulus	Right dumbell curl / Left leg extension
P14	Turning a car right / Turning a car left	Inflating a balloon / Deflating a balloon	Screeching chalkboard / Growling lion

Table 1: Shortened mental commands used by each user for each task, as reported on the post-session questionnaires. Since participants did not always answer all questions on the questionnaire or write their responses legibly, missing data was marked with N/A.

3.1.4 Classifier Confidence

Classifier confidence was highly correlated with classification accuracy across all above chance sessions according to the average of the last five blocks ($r = 0.79 \pm 0.14$, $p < 0.0001$). In addition, linear regression over the standard deviation of accuracy on each
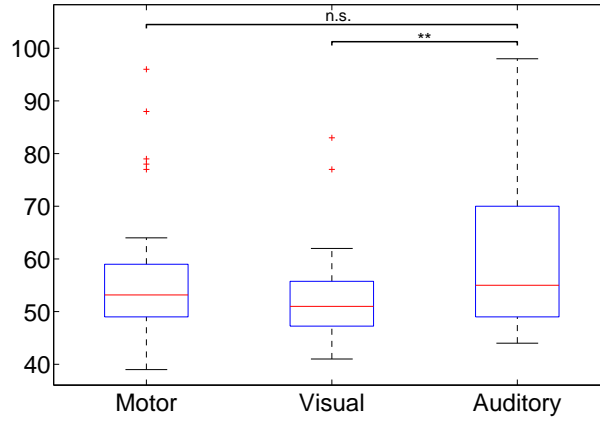
Figure 3: Boxplots showing the distribution of BCI performance for each sensory modality. **$p < 0.01$

block for sessions that yielded above chance classification reveals that classification became more stable over time ($\beta_2 = -0.4$, $p < 0.005$). Figure 4 shows online classification accuracy and classifier confidence for a high performer in each sensory modality.

3.2   Offline Results

The standard CSP approach as defined in [Blankertz et al., 2008], the methods used in our online analysis, and FBCSP were tested here. In addition, artifact rejection was employed. An average of 41.7 ± 25.2 trials per session were omitted from analysis due to artifact rejection. We found that FBCSP yielded the best results, and therefore we present only those results here. Offline classification results with FBCSP are shown in Figure 5.

Above chance classification accuracy was obtained for more sessions offline (50 of 123) than online (28 of 123). There was no difference between offline and online results for motor imagery sessions ($t_{82} = 1.17$, $p = 0.12$), but there was a trend towards higher offline accuracies compared to online accuracies for visual imagery sessions ($t_{76} = 1.60$, $p = 0.06$) and for auditory imagery ($t_{82} = 1.58$, $p = 0.06$).

Figure 6 shows the discriminatory spatial patterns for the best session of each imagery paradigm. These are the first and last common spatial patterns. Specifically, these are estimates of the first and last columns of the inverse of the CSP filter W as discussed in [Ramoser et al., 2000], [Blankertz et al., 2008], and [Blankertz et al., 2011]). However, because the number of components used in analysis was less than the number of EEG channels, the vectors used to produce the patterns shown in Figure 6 are computed using the method given in [Haufe et al., 2014]. CSP components were computed after artifact rejection and bandpass filtering from 8-30 Hz, corresponding to the bandpass filter used in the online experiment.
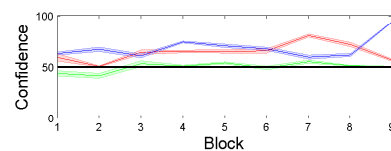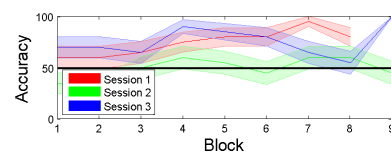
(a) P13 Motor Imagery
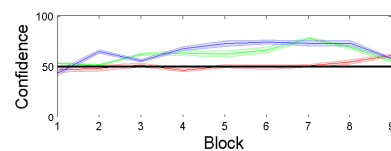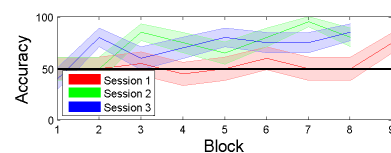
(b) P9 Visual Imagery
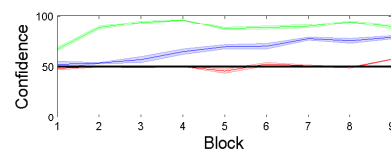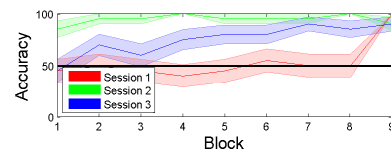
(c) P3 Auditory Imagery

Figure 4: Classification accuracy and classifier confidence for a successful subject in each sensory modality.
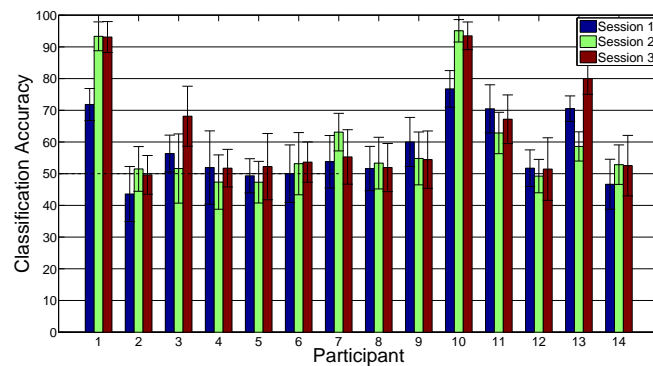
(a) Motor Imagery
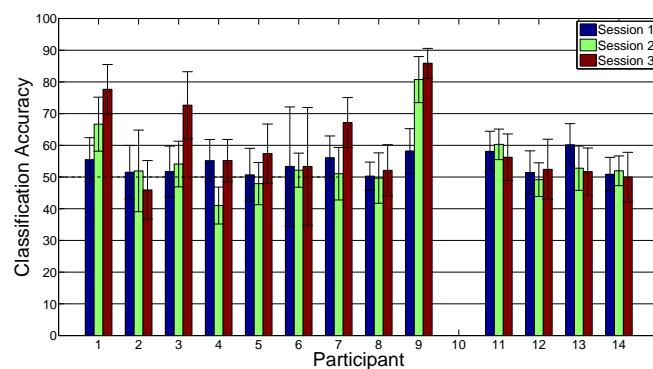
(b) Visual Imagery
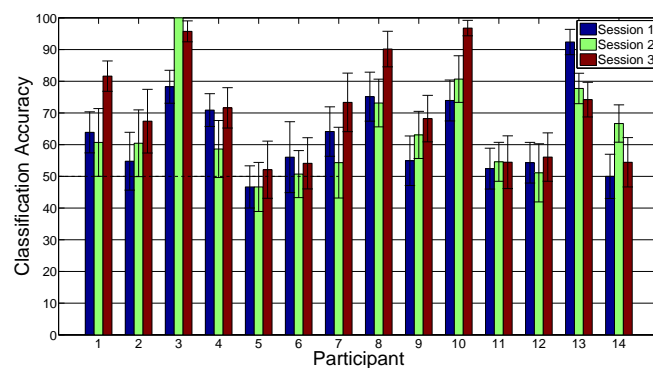
(c) Auditory Imagery

Figure 5: Offline classification results for all sessions using FBCSP
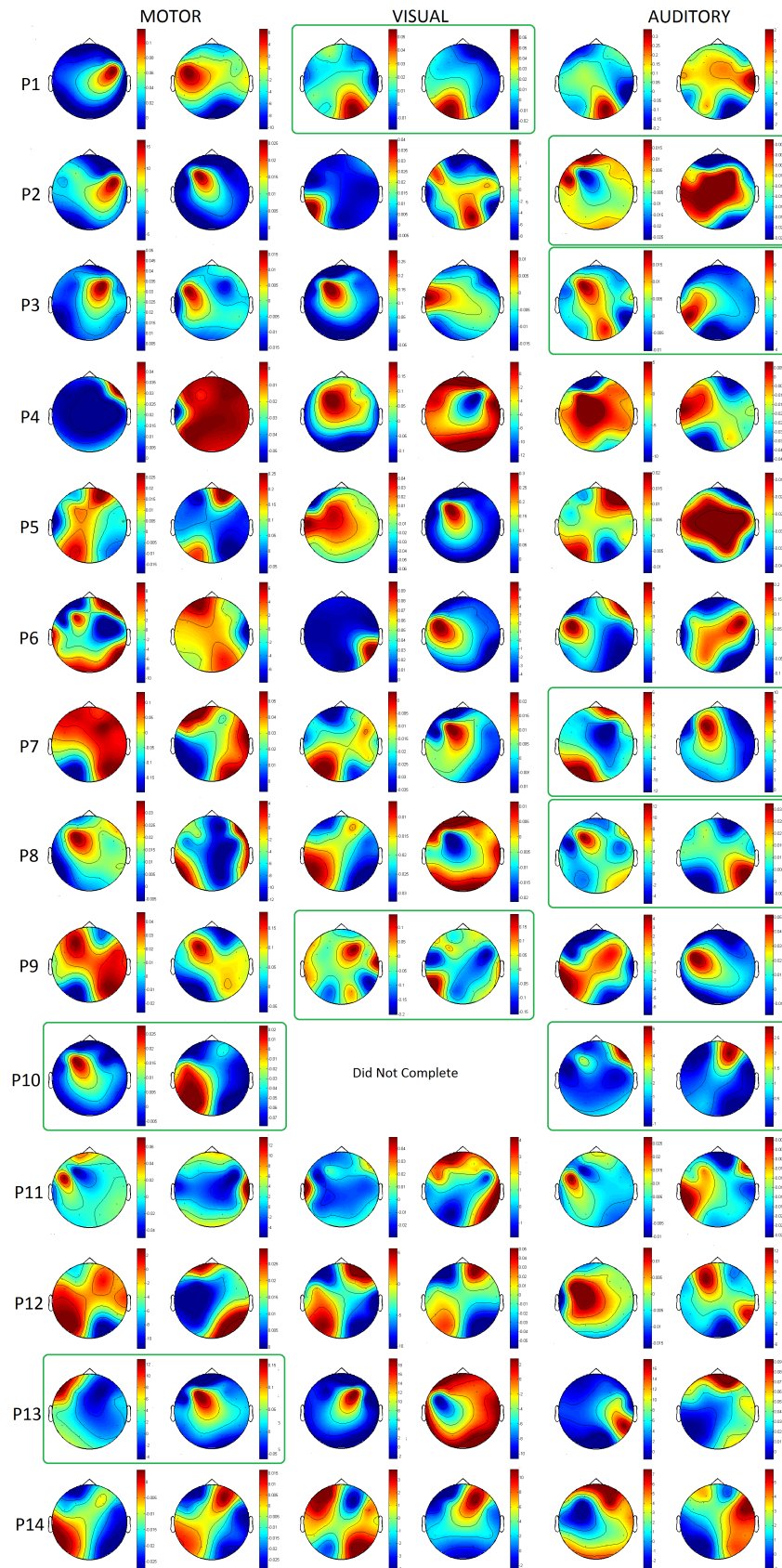
Subject	Imagery	Motor			Visual			Auditory		
	Session	1	2	3	1	2	3	1	2	3
P1	All	56	58	57	57	53	71**	49	49	61*
	Last 5	53	59	61	54	53	83**	47	49	66**
P2	All	44	52	54	50	49	52	46	53	64**
	Last 5	45	52	58	48	52	49	46	54	72**
P3	All	48	53	51	43	46	63**	53	94**	75**
	Last 5	48	49	45	41	44	63*	60	95**	85**
P4	All	46	55	47	54	52	46	53	51	49
	Last 5	49	58	39	53	48	41	55	53	52
P5	All	48	53	53	45	47	50	50	46	51
	Last 5	52	51	53	51	44	54	49	44	54
P6	All	54	44	48	51	50	55	51	50	57
	Last 5	50	41	49	50	48	55	50	50	61
P7	All	57	51	54	57	51	54	66**	51	65**
	Last 5	67**	56	56	50	52	53	70**	49	71**
P8	All	48	54	49	49	46	59*	59	60*	82**
	Last 5	50	57	47	45	46	58	63*	64*	98**
P9	All	57	57	51	54	72**	71**	49	50	45
	Last 5	63*	55	55	58	77**	78**	46	52	45
P10	All	68**	83**	81**	n/a	n/a	n/a	70**	79**	97**
	Last 5	79**	96**	88**	n/a	n/a	n/a	84**	92**	98**
P11	All	56	51	64**	61*	45	48	52	53	54
	Last 5	62	56	64*	65*	46	51	50	49	58
P12	All	47	56	45	58	47	48	47	48	52
	Last 5	53	52	58	59	50	46	47	44	55
P13	All	72**	49	76**	53	48	48	78**	70**	61*
	Last 5	78**	53	77**	54	47	51	76**	74**	61
P14	All	49	53	53	52	45	58	54	63**	52
	Last 5	53	52	62*	49	43	58	70**	68*	52

Table 2: Online classification accuracy across all sessions for each of the 14 participants. Average across all blocks given in the top row, and the average across just the last five blocks given in the bottom row. Significance is calculated based on a binomial test using the classification accuracy and the total number of trials in the session.
* denotes significance at $p < 0.01$.
** denotes significance at $p < 0.001$.

Figure 6: Common spatial patterns for the last session of each task for each participant. These are obtained using an approximation to the inverse of the CSP filter W given by computing $A = Cov(X) * W * Cov(S)^{-1}$, where $Cov(X)$ is spatial covariance matrix of the original EEG signal and S is the spatial covariance matrix of the EEG signal after being filtered by the CSP filter W [Haufe et al., 2014]. The first and last columns of A are shown here. Patterns corresponding to sessions with above 70% classification accuracy for the last five blocks of online training are marked by a green box.

3.3 Background Experience Questionnaire

Before beginning their first session, participants completed a basic questionnaire asking about any training they may have had in athletics, visual arts, or music. This questionnaire asked participants to approximate the number of hours per week spent training on athletics, visual arts, or music, the number of years spent training in total, and to give a self-rating of their level of performance/expertise (i.e., do not participate, novice, university/varsity level, or professional). In this study we used only their self-reported level of expertise in each area. These data are summarized in Table 3.

We found that self-reported level of performance or expertise had a significant effect on BCI performance ($F_{2,120} = 22.4$, $p \approx 0$, variance explained: $\omega^2 = 0.24$) and correlated significantly with BCI performance ($r = 0.46$, $p < 10^{-6}$). The difference in BCI performance at each self-reported level of expertise is shown in Figure 7.

Imagery	Motor	Visual	Auditory
P1	2	1	1
P2	1	1	1
P3	1	2	2
P4	2	1	1
P5	1	2	1
P6	1	1	2
P7	2	1	2
P8	2	1	3
P9	1	3	1
P10	2	2	3
P11	2	2	2
P12	2	2	2
P13	3	1	1
P14	2	2	2

Table 3: Self-rated level of expertise as determined by the background experience questionnaire.
Coding: *1 = Do not practice/perform, 2 = Amateur, 3 = Varsity/University level, 4 = Professional*)

4 Discussion

This study makes contributions to BCI reserach in three major areas: first by assessing the feasibility of an Open-Ended BCI, second by measuring the contribution of human learning to performance with mental imagery BCIs, and third, by examining how domain expertise can affect BCI performance.

4.1 Feasibility of an Open-Ended BCI

The classification accuracies obtained with our participants are consistent with other studies exploring abstract non-motor imagery with low resolution EEG (e.g., [Bobrov et al., 2011, Cabrera and Dremstrup, 2008]). Nine out of 14 participants were able to
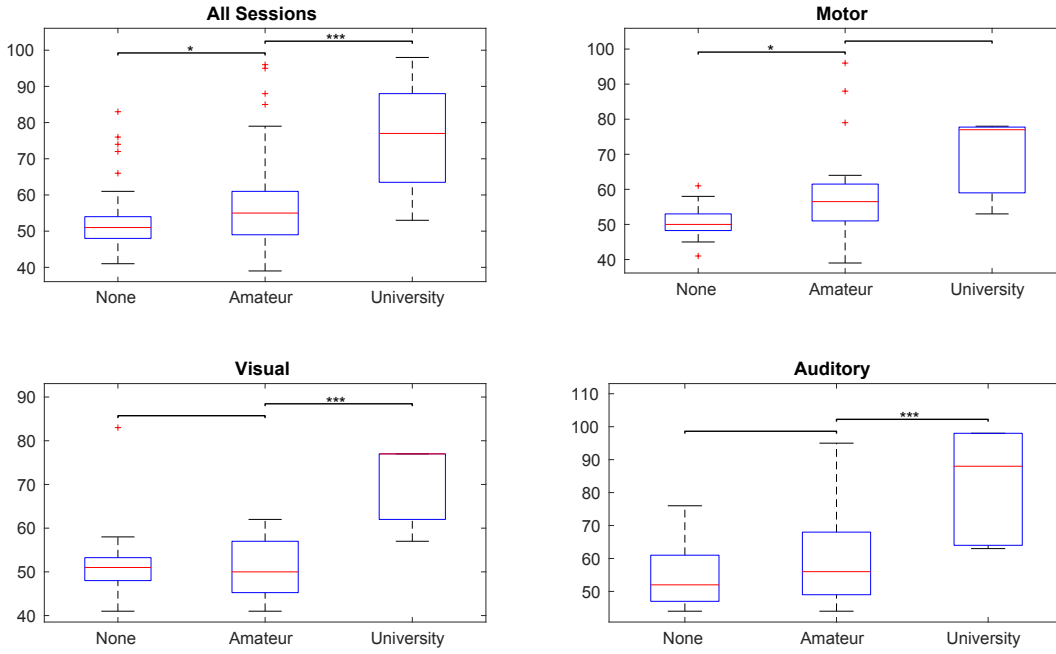
Figure 7: Boxplots showing the distribution of BCI performance at varying levels of self-reported expertise.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

achieve useful BCI control (at least 70% classification accuracy) with at least one sensory modality during the experiment, and 9 participants had data resulting in at least this level of accuracy in offline analysis. We anticipate that performance will further improve with feature extraction methods that allow for better generalization across different kinds of mental imagery. Nonetheless, these results suggest that there is promise in the concept of an Open-Ended BCI and it is worthwhile to develop this approach further.

The limitations on performance due to EEG hardware were apparent in this study. Previous studies have shown that the Emotiv Epoc yielded poorer BCI performance compared to a research grade EEG device [Bobrov et al., 2011, Duvinage et al., 2013]. However, previous studies had not reported the significant loss in SNR over the course of a study which we observed. Furthermore, the sparse and non-uniform sensor configuration is not optimal for an Open-Ended BCI, which must be able to measure a wide variety of neurophysiological signals. For example, the greater sensor coverage over temporal areas and poor coverage over motor areas may be the main reason why participants performed better with auditory imagery than motor imagery on average (motor imagery typically requires at least C3 and C4 to detect the sensorimotor rhythm). With the Emotiv Epoc's electrode configuration, the CSP algorithm would rely mainly on frontal and temporo-parietal sensors, and it is possible that participants who chose mental commands which were more distinguishable in these brain regions were to some extent inadvertently more successful in controlling the BCI. As such, this device would not be ideal for studies focused on comparing performance across tasks using different sensory modalities. However, given that our study was focused on showing that an Open-Ended BCI design is possible, the limitations attributable to this EEG hardware do not detract from the purpose or

conclusions of this study.

Another limitation was the lack of continuous real time neurofeedback during the mental imagery period of each trial. It is possible that BCI performance would have been improved with the inclusion of this neurofeedback approach. However, our results still support the conclusions that an Open-Ended BCI is a feasible approach for future BCI applications and that the domain expertise of the individual is likely to impact BCI performance with mental imagery in a related sensory modality. If continuous neurofeedback, or any other improvements to user training, were implemented in this study, we expect that it would only strengthen these conclusions due to the potentially improved BCI performance.

The presence of undetected artifacts may have also been a limitation in evaluating BCI performance in this study. Although offline analysis used a standard artifact rejection method and resulted in a slight increase in classification accuracy, it has been suggested previously that even more robust methods may be needed for BCIs [Brunner et al., 2015]. Because we did not use electromyography (EMG) electrodes on the face, larynx, and appendages, it is possible that subtle muscle activity or subvocal laryngeal activation was present during mental imagery, and thus influenced the BCI models.

4.2   The Contribution of Human Learning

There was a significant effect of human learning observed in this study. However, the current study cannot precisely determine the cause of improvements in performance across sessions within a sensory modality, even though, since computational models were independent between sessions, we could attribute those changes to the subject. However, we cannot determine for sure whether these improvements are due to subjects improving in their ability to produce reliable brain activity (although for some subjects, this may have been the case, as P10 noted that they had practiced their mental imagery strategies between sessions) or due to other factors (e.g., changes in fatigue levels, hunger, distraction, or motivation).

Though the precise causes of changes in performance across sessions is difficult to determine, the results presented here adds to previous arguments for the need for improved user training in BCI [Lotte et al., 2013], it is possible that overall classification accuracy could be enhanced with better user training. However, no user training methods have been discussed in the literature so far which appear to be particularly suitable for an Open-Ended BCI. While further sessions may have given participants enough time to find and practice more optimal personal mental strategies, a suitable neurofeedback approach which more directly promotes reliable mental commands without prior knowledge of those mental commands must also be developed.

4.3   Background Experience in BCI Performance

The effect of self-reported level of expertise in athletics, visual arts, and music on BCI performance found in this study may have important implications for BCI training. Since BCI performance was not correlated with task order, we can state with confidence that this was not a confounding factor in this result. Interest in the task was, however, correlated with performance, though to a smaller extent than expertise was. Moreover, some participants commented on their questionnaires that the task was much more interesting when the BCI responded accurately, suggesting that interest was not the cause of hig-

her BCI performance. Therefore, it is likely that the correlation between task interest and BCI performance was not a significant confounding factor affecting the impact of expertise on BCI performance.

If it is the case that expertise in an area relevant to the sensory modality used in a mental imagery BCI is predictive of subsequent BCI performance, then the BCI community could take advantage of this fact and cater BCI training to the particular background of each individual. Furthermore, BCI applications targeting specific groups of individuals, e.g., artists, might be able to suggest or specify mental commands using appropriate sensory modalities.

These results fall very much in line with previous work on personality and cognitive profiles as predictors of mental imagery BCIs [Jeunet et al., 2015, Hammer et al., 2012, Randolph, 2011, Scherer et al., 2015]. However the results presented in this study extend the idea that such individual factors can impact BCI performance to the cases of abstract user-defined visual imagery and abstract user-defined auditory imagery and suggest specifically that one's area of expertise can affect BCI performance. Furthermore, these results show that the idea of giving free and open choice of mental commands [Friedrich et al., 2012] has some merit, since on average participants performed best with mental commands within a sensory modality they found more interesting and which corresponded to their prior expertise. In future work, standardized questionnaires assessing kinaesthetic, visual, and auditory/musical aptitude may provide a clearer indication of the relationship between BCI performance and domain expertise, and whether specific variables associated with domain expertise are particularly important.

5   Conclusions and Future Work

As this is the first study on an Open-Ended design to our knowledge, we aimed to discover key areas where we should focus our attention while developing new methodology specificallly catered to Open-Ended BCIs. We identified two key areas for future work.

First, more generalizable signal processing and machine learning methods are necessary, as the naive approach of compiling a large set of candidate features from which to find an optimal subset may not be sufficiently powerful. FBCSP and related variations are useful because in principle they provide a data-driven approach to differentiate signals which are separable by their spatio-spectral distributions. However, it is possible that any combination of specific feature extraction methods may not be consistently successful in an Open-Ended BCI, because the user may choose mental commands which are not separable via any possible feature space derived from those feature extraction methods. Others have discussed the need for generalizable feature extraction methods for BCI due to the need to accommodate a greater variety of mental strategies [Nurse et al., 2015]. Notably, feature learning has recently been successfully applied in the BCI context [Stober et al., 2015]. This approach may have the flexibility needed in order to construct a custom feature space for each user and each task, providing the generalizability needed for an Open-Ended BCI.

The second key area for future work is the need for improved methods for user training. While this is true not only for open-ended BCIs (see [Lotte et al., 2013]), standard training methods may not be suitable for an Open-Ended BCI. In particular, an appropriate training method should assist the user in discovering a strong personalized mental strategy and then help the user differentiate their mental commands over time without

knowing what those commands will be in advance. We propose a data-driven approach which models the class distributions over time and uses neuorfeedback targeting areas where these distributions overlap in feature space in order to train the user to increase separation, irrespective of their choice of mental commands. In addition, we have shown that allowing and encouraging a choice of mental commands which are dominated by a sensory modality corresponding to an individual's expertise can improve BCI performance. Hence training should include a stage of discovering good personal mental commands within an appropriate sensory modality, and focus on using networks in the brain which are already trained by the user in other contexts. Whether this effect persists outside of the forms of mental imagery tested here, and what precise mechanisms explain this effect are questions that should be probed more deeply, as they may substantially improve BCI performance by enabling BCIs to better take advantage of individual factors during training.

We have learned from this study than an Open-Ended BCI is feasible. However, new methods may be required in order make better use of individual differences in the ability to modulate different neurophysiological signals. In particular, we must overcome the unique challenges involved with developing a BCI which can accommodate different mental strategies and different neurophysiological signals for different users. Given that individual factors appear to significantly impact BCI performance, developing the Open-Ended BCI further is a worthwhile endeavour.

Acknowledgements

References

[Achtman et al., 2007] Achtman, N., Afshar, A., Santhanam, G., Byron, M. Y., Ryu, S. I., and Shenoy, K. V. (2007). Free-paced high-performance brain–computer interfaces. *Journal of neural engineering*, 4(3):336.

[Allison et al., 2010] Allison, B., Luth, T., Valbuena, D., Teymourian, A., Volosyak, I., and Graser, A. (2010). BCI demographics: How many (and what kinds of) people can use an SSVEP BCI? *IEEE transactions on neural systems and rehabilitation engineering*, 18(2):107–116.

[Allison and Neuper, 2010] Allison, B. Z. and Neuper, C. (2010). Could anyone use a BCI? In *Brain-computer interfaces*, pages 35–54. Springer.

[Ang et al., 2008] Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2391–2398.

[Badcock et al., 2013] Badcock, N. A., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., and McArthur, G. (2013). Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs. *PeerJ*, 1.

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

[Blankertz et al., 2010] Blankertz, B., Sannelli, C., Halder, S., Hammer, E. M., Kübler, A., Müller, K.-R., Curio, G., and Dickhaus, T. (2010). Neurophysiological predictor of SMR-based BCI performance. *NeuroImage*, 51(4):1303–9.

[Blankertz et al., 2008] Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K.-R. (2008). Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, 25(1):41–56.

[Bobrov et al., 2011] Bobrov, P., Frolov, A., Cantor, C., Fedulova, I., Bakhnyan, M., and Zhavoronkov, A. (2011). Brain-computer interface based on generation of visual images. *PloS one*, 6(6):e20674.

[Borisoff et al., 2004] Borisoff, J. F., Mason, S. G., Bashashati, A., and Birch, G. E. (2004). Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch. *IEEE Transactions on Biomedical Engineering*, 51(6):985–992.

[Brainard, 1997] Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10:433–436.

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

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

[Cabrera and Dremstrup, 2008] Cabrera, A. F. and Dremstrup, K. (2008). Auditory and spatial navigation imagery in Brain-Computer Interface using optimized wavelets. *Journal of Neuroscience Methods*, 174(1):135–146.

[Carrino et al., 2012] Carrino, F., Dumoulin, J., Mugellini, E., Khaled, O., and Ingold, R. (2012). A self-paced BCI system to control an electric wheelchair: Evaluation of a commercial, low-cost EEG device. In *Biosignals and Biorobotics Conference (BRC), 2012 ISSNIP*, pages 1–6.

[Chang and Lin, 2011] Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

[DaSalla et al., 2009] DaSalla, C. S., Kambara, H., Sato, M., and Koike, Y. (2009). Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*, 22(9):1334–1339.

[del R. Millàn et al., 2002] del R. Millàn, J., Mouriño, J., Franzé, M., Cincotti, F., Varsta, M., Heikkonen, J., and Babiloni, F. (2002). A local neural classifier for the recognition of EEG patterns associated to mental tasks. *IEEE transactions on neural networks*, 13(3):678–686.

[Dhindsa et al., 2017] Dhindsa, K., Carcone, D., and Becker, S. (2017). A brain-computer interface based on abstract visual and auditory imagery: Evidence for an effect of artistic training. In Schmorrow, D. D. and Fidopiastis, C. M., editors, *International Conference on Augmented Cognition. Springer, Cham, 2017*.

[Diez et al., 2011] Diez, P. F., Mut, V. a., Avila Perona, E. M., and Laciar Leber, E. (2011). Asynchronous BCI control using high-frequency SSVEP. *Journal of neuroengineering and rehabilitation*, 8(July):39.

[Duvinage et al., 2013] Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., and Dutoit, T. (2013). Performance of the Emotiv Epoc headset for P300-based applications. *Biomedical engineering online*, 12:56.

[Emotiv Systems, 2011] Emotiv Systems (2011). Emotiv - brain computer interface technology. `http://www.emotiv.com`.

[Friedrich et al., 2012] Friedrich, E. V., Scherer, R., and Neuper, C. (2012). The effect of distinct mental strategies on classification performance for brain–computer interfaces. *International Journal of Psychophysiology*, 84(1):86–94.

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

[Haufe et al., 2014] Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87:96–110.

[Jeunet et al., 2015] Jeunet, C., NKaoua, B., Subramanian, S., Hachet, M., and Lotte, F. (2015). Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns. *PloS one*, 10(12):e0143962.

[Kothe et al., 2013] Kothe, C. A., Makeig, S., and Onton, J. A. (2013). Emotion recognition from EEG during self-paced emotional imagery. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, pages 855–858.

[Kübler and Müller, 2007] Kübler, A. and Müller, K. R. (2007). An introduction to brain computer interfacing. In Dornhege, G., del R. Millán, J., Hinterberger, T., McFarland, D., and Müller, K. R., editors, *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA.

[Kübler et al., 2001] Kübler, A., Neumann, N., Kaiser, J., Kotchoubey, B., Hinterberger, T., and Birbaumer, N. P. (2001). Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication. *Archives of Physical Medicine and Rehabilitation*, 82(11):1533–1539.

[Kus et al., 2012] Kus, R., Valbuena, D., Zygierewicz, J., Malechka, T., Graeser, A., and Durka, P. (2012). Asynchronous BCI Based on Motor Imagery With Automated Calibration and Neurofeedback Training. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 20(6):823–35.

[Lievesley et al., 2011] Lievesley, R., Wozencroft, M., Ewins, D., Lievesley, M., and Wozencroft, R. (2011). The Emotiv EPOC neuroheadset: an inexpensive method of controlling assistive technologies using facial expressions and thoughts? *Journal of Assistive Technologies*, 5(2):67–82.

[Lin et al., 2004] Lin, C.-J., Weng, R. C., et al. (2004). Simple probabilistic predictions for support vector regression. *National Taiwan University, Taipei*.

[Lin et al., 2007] Lin, H.-T., Lin, C.-J., and Weng, R. C. (2007). A note on Platt's probabilistic outputs for support vector machines. *Machine learning*, 68(3):267–276.

[Liu et al., 2012] Liu, Y., Jiang, X., Cao, T., Wan, F., Mak, P. U., Mak, P. I., and Vai, M. I. (2012). Implementation of SSVEP based BCI with Emotiv EPOC. *Proceedings of IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems,VECIMS*, pages 34–37.

[Lotte et al., 2013] Lotte, F., Larrue, F., and Mühl, C. (2013). Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design. *Frontiers in human neuroscience*, 7(September):568.

[Mason and Birch, 2000] Mason, S. G. and Birch, G. E. (2000). A brain-controlled switch for asynchronous control applications. *IEEE Transactions on Biomedical Engineering*, 47(10):1297–1307.

[MATLAB, 2013] MATLAB (2013). *version 8.2.0 (R2013b)*. The MathWorks Inc., Natick, Massachusetts.

[Millan and Mouriño, 2003] Millan, J. R. and Mouriño, J. (2003). Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):159–161.

[Müller-Gerking et al., 1999] Müller-Gerking, J., Pfurtscheller, G., and Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical neurophysiology*, 110(5):787–798.

[Nai-Jen and Palaniappan, 2004] Nai-Jen, H. and Palaniappan, R. (2004). Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 1, pages 507–510. IEEE.

[Neuper and Pfurtscheller, 2010] Neuper, C. and Pfurtscheller, G. (2010). Neurofeedback training for BCI control. In *Brain-Computer Interfaces*, pages 65–78. Springer.

[Nurse et al., 2015] Nurse, E. S., Karoly, P. J., Grayden, D. B., and Freestone, D. R. (2015). A generalizable brain-computer interface (BCI) using machine learning for feature discovery. *PLoS ONE*, 10(6).

[Oostenveld et al., 2011] Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011.

[Peng, 2005] Peng, H. C. (2005). Feature Selection Based on Mutual Information Criteria of Max-dependency, Max-relevance, and Min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1226–1238.

[Pineda et al., 2003] Pineda, J., Silverman, D. S., Vankov, A., Hestenes, J., et al. (2003). Learning to control brain rhythms: making a brain-computer interface possible. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):181–184.

[Platt et al., 1999] Platt, J. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74.

[Ramoser et al., 2000] Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE transactions on rehabilitation engineering*, 8(4):441–446.

[Randolph, 2011] Randolph, A. B. (2011). Not all created equal: Individual-technology fit of brain-computer interfaces. *Proceedings of the Annual Hawaii International Conference on System Sciences*, pages 572–578.

[Randolph et al., 2010] Randolph, A. B., Jackson, M. M., and Karmakar, S. (2010). Individual characteristics and their effect on predicting mu rhythm modulation. *Intl. Journal of Human–Computer Interaction*, 27(1):24–37.

[Reuderink and Poel, 2008] Reuderink, B. and Poel, M. (2008). Robustness of the common spatial patterns algorithm in the BCI-pipeline. *Centre for Telematics and Information Technology University of Twente Technical Report*.

[Sannelli et al., 2010] Sannelli, C., Vidaurre, C., Müller, K.-R., and Blankertz, B. (2010). Common spatial pattern patches-an optimized filter ensemble for adaptive brain-computer interfaces. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 4351–4354. IEEE.

[Scherer et al., 2015] Scherer, R., Faller, J., Friedrich, E. V., Opisso, E., Costa, U., Kübler, A., and Müller-Putz, G. R. (2015). Individually adapted imagery improves brain-computer interface performance in end-users with disability. *PloS one*, 10(5).

[Stober et al., 2015] Stober, S., Sternin, A., Owen, A. M., and Grahn, J. A. (2015). Deep feature learning for EEG recordings. *arXiv preprint arXiv:1511.04306*.

[Sugiyama et al., 2007] Sugiyama, M., Krauledat, M., and MÃžller, K.-R. (2007). Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8(May):985–1005.

[Thomas et al., 2013] Thomas, E., Dyson, M., and Clerc, M. (2013). An analysis of performance evaluation for motor-imagery based BCI. *Journal of neural engineering*, 10(3):031001.

[Vidaurre et al., 2011a] Vidaurre, C., Sannelli, C., Müller, K.-R., and Blankertz, B. (2011a). Co-adaptive calibration to improve BCI efficiency. *Journal of Neural Engineering*, 8(2).

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

[Von Bünau et al., 2009] Von Bünau, P., Meinecke, F. C., Király, F. C., and Müller, K.-R. (2009). Finding stationary subspaces in multivariate time series. *Physical review letters*, 103(21):214101.

[Von Bünau et al., 2010] Von Bünau, P., Meinecke, F. C., Scholler, S., and Müller, K.-R. (2010). Finding stationary brain sources in EEG data. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 2810–2813. IEEE.

[Vuckovic and Osuagwu, 2013] Vuckovic, A. and Osuagwu, B. A. (2013). Using a motor imagery questionnaire to estimate the performance of a brain–computer interface based on object oriented motor imagery. *Clinical Neurophysiology*, 124(8):1586–1595.

[Wolpaw et al., 2002] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 113(6):767–791.

[Wolpaw and McFarland, 2004] Wolpaw, J. R. and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17849–17854.

[Wu et al., 2004] Wu, T.-F., Lin, C.-J., and Weng, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5(Aug):975–1005.

4.3   Discussion

This chapter introduced the idea of an open-ended BCI and demonstrated the feasibility of a more generalized BCI transducer. The main advancement shown here is an approach that does not rely on specific knowledge about the mental commands employed by the user, thus allowing each user to choose their own mental commands. This was achieved simply by using a combination of existing methods in a way that would allow for generalization across different neurophysioloigical signals. More specifically, features were extracted in a way that could accommodate a variety of mental commands, and specific discriminative features were selected for each case.

The open-ended BCI approach can be contextualized more broadly as a feature discovery approach to generalized brain-computer interfacing. Concurrent to this work, a neural network approach for addressing the lack of individualization in BCI was proposed for motor imagery classification [20]. Similarly, a deep learning approach was proposed for music imagery classification [21]. However, whether these approaches can be generalized to mental imagery more broadly has not yet been tested, and thus cannot be said to achieve the same level of generalization as the open-ended BCI presented here. Furthermore, while none of these approaches are especially new in the machine learning literature, they are all novel in the context of brain-computer interfacing, where a commonly held view is that extensive feature engineering based on *a priori* knowledge is necessary.

Rather than representing a technological breakthrough, this study was about validating a paradigm shift in BCI design. The use of predefined mental commands has enabled much progress in the field, but it is a simplification that is not without its limitations. The transition from restricting users to specified mental commands to giving users the freedom to experiment with different kinds of mental commands introduces the significant challenge of developing a generalized BCI transducer. While this challenge may have seemed insurmountable in the past, the study presented here shows that it is indeed possible to make progress in this direction and move towards a generalized BCI with current technology. In doing so, different users can be made free to make use of mental commands which lead to optimal BCI performance for them as individuals. Though this study did not result in better BCI performance than the standard approach due to several limitations in the experimental design and suboptimal hardware, that comparable classification accuracies were achieved using a variety of novel mental commands which were only disclosed after each training session demonstrates the promise of this approach to generalized brain-computer interfacing.

One important question about the methods for generalized brain-computer interfacing presented in this thesis is how broadly they can be applied. In Chapter

5, a study using the alternative blind approach to EEG classification summarized in Chapter 2, the SF approach, is used to predict emotional responses to videos. This adds confidence to the notion that these methods open up a viable path towards a truly generalized user-centered BCI.

Bibliography

[1] M. Ahn, M. Lee, J. Choi, and S. C. Jun, "A review of brain-computer interface games and an opinion survey from researchers, developers and users," *Sensors*, vol. 14, no. 8, pp. 14601–14633, 2014.

[2] F. Miralles, E. Vargiu, X. Rafael-Palou, M. Solà, S. Dauwalder, C. Guger, C. Hintermüller, A. Espinosa, H. Lowish, S. Martin, *et al.*, "Brain–computer interfaces on track to home: Results of the evaluation at disabled end-users homes and lessons learnt," *Frontiers in ICT*, vol. 2, p. 25, 2015.

[3] L. Galway, P. McCullagh, G. Lightbody, C. Brennan, and D. Trainor, "The potential of the brain-computer interface for learning: A technology review," in *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*, pp. 1554–1559, IEEE, 2015.

[4] A. Remsik, B. Young, R. Vermilyea, L. Kiekhoefer, J. Abrams, S. Evander Elmore, P. Schultz, V. Nair, D. Edwards, J. Williams, *et al.*, "A review of the progression and future implications of brain-computer interface therapies for restoration of distal upper extremity motor function after stroke," *Expert review of medical devices*, vol. 13, no. 5, pp. 445–454, 2016.

[5] E. V. Friedrich, R. Scherer, and C. Neuper, "The effect of distinct mental strategies on classification performance for brain–computer interfaces," *International Journal of Psychophysiology*, vol. 84, no. 1, pp. 86–94, 2012.

[6] C. Jeunet, B. NKaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns," *PloS one*, vol. 10, no. 12, p. e0143962, 2015.

[7] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *Journal of Neural engineering*, vol. 4, no. 2, p. R32, 2007.

[8] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing.," *Journal of neural engineering*, vol. 9, p. 026013, Apr. 2012.

[9] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain–computer interfaces," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1567–1584, 2012.

[10] S. Sun and J. Zhou, "A review of adaptive feature extraction and classification methods for EEG-based brain-computer interfaces," in *Neural Networks (IJCNN), 2014 International Joint Conference on*, pp. 1746–1753, IEEE, 2014.

[11] C. Brunner, N. Birbaumer, B. Blankertz, C. Guger, A. Kübler, D. Mattia, J. d. R. Millán, F. Miralles, A. Nijholt, E. Opisso, *et al.*, "BNCI horizon 2020: towards a roadmap for the BCI community," *Brain-computer interfaces*, vol. 2, no. 1, pp. 1–10, 2015.

[12] B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser, "BCI demographics: How many (and what kinds of) people can use an SSVEP BCI?," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 18, no. 2, pp. 107–116, 2010.

[13] B. Z. Allison and C. Neuper, "Could anyone use a BCI?," in *Brain-computer interfaces*, pp. 35–54, Springer, 2010.

[14] C. Neuper and G. Pfurtscheller, "Neurofeedback training for BCI control," in *Brain-Computer Interfaces*, pp. 65–78, Springer, 2010.

[15] F. Lotte, F. Larrue, and C. Mühl, "Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design.," *Frontiers in human neuroscience*, vol. 7, no. September, p. 568, 2013.

[16] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Machine-learning-based coadaptive calibration for brain-computer interfaces," *Neural computation*, vol. 23, no. 3, pp. 791–816, 2011.

[17] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Co-adaptive calibration to improve BCI efficiency," *Journal of Neural Engineering*, vol. 8, no. 2, 2011.

[18] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance.," *NeuroImage*, vol. 51, no. 4, pp. 1303–9, 2010.

[19] S. Kober, M. Witte, and M. Ninaus, "Learning to modulate one's own brain activity: the effect of spontaneous mental strategies," *Frontiers in human neuroscience*, vol. 7, no. October, p. 695, 2013.

[20] E. S. Nurse, P. J. Karoly, D. B. Grayden, and D. R. Freestone, "A generalizable brain-computer interface (BCI) using machine learning for feature discovery," *PLoS ONE*, vol. 10, no. 6, 2015.

[21] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Deep feature learning for EEG recordings," *arXiv preprint arXiv:1511.04306*, 2015.

Chapter 5

Detecting Emotional Reactions from EEG Using a Generalized BCI Approach

5.1   Introduction

Methods that generalize well across individuals and a variety of mental commands do not necessarily generalize well to different types of BCI and EEG classification problems. In particular, BCI transducers based on CSP or any spatial filtering method should perform well if different mental states or mental commands are spatially separable, but only if there are enough EEG electrodes with which to compute a sufficiently precise spatial filter. Recognizing this limitation, the SF approach to generalized BCI transducers based on channel-wise spectral analysis was given in Chapter 2. Since the study presented in this chapter involved only four EEG channels, the SF approach was used.

The study presented in this chapter is on detecting emotional reactions to videos from EEG. Even though the same method that was used to develop the open-ended BCI was not used here, this study still demonstrates that generalized methods for BCI can be applied to widely different contexts. While the number of EEG channels was small, there was a sufficient amount of data available to make the SF approach feasible despite its reliance on a large number of features. Moreover, certain functions of the power spectrum, such as alpha asymmetries, are known to be correlated with certain emotional states [1, 2] (therefore approximately 1% of the features used in this study were included based on *a priori* knowledge about emotional responses in EEG). Thus, the spectral features approach was considered most appropriate for this particular study.

Note on artifact detection: The studies presented in this chapter and in Chapter 6 used the Muse EEG headband[1]. It was found that standard artifact removal algorithms did not perform adequately for this hardware, and no algorithms had been developed for real-time applications using such sparsely distributed electrodes (real-time artifact detection was especially needed for the study presented in Chapter 6). A new artifact rejection method, Filter Bank Artifact Rejection (FBAR), was developed in order to satisfy this need. FBAR is an algorithm which can detect even small and subtle artifacts in single-channel EEG in real time with a high degree of accuracy. See Appendix A for a detailed explanation of the FBAR algorithm.

---

[1]Interaxon Muse. www.choosemuse.com,

5.2 Emotional Reaction Recognition from EEG

Dhindsa, K. & Becker, S. (In Press: 2017). Emotional Reaction Recognition from EEG. *7th International Workshop on Pattern Recognition in Neuroimaging 2017*. IEEExplore.
Article reprinted with permission.

Emotional Reaction Recognition from EEG

Kiret Dhindsa
School of Scientific Computing and Engineering
McMaster University
Hamilton, Ontario, Canada, L8S 4L8
Email: dhindsj@mcmaster.ca

Suzanna Becker
Department of Psychology, Neuroscience, and Behaviour
McMaster University
Hamilton, Ontario, Canada, L8S 4L8
Email: becker@mcmaster.ca

Abstract—**In this study we explore the application of pattern recognition models for recognizing emotional reactions elicited by videos from electroencephalography (EEG). We show that both the presence and magnitude of each emotion can be predicted above chance levels with up to 88% accuracy. Furthermore, we show that there are differences in classifiability for different emotions and participants, but whether a participant's data can be classified with respect to different emotions can itself be predicted from their EEG. Index Terms– Emotion recognition, electroenecephalography (EEG), pattern recognition, classification, regression, individual differences, affective computing**

I. Introduction

The ability to recognize emotional reactions based on biological data has a wide variety of applications [1], including in therapy for mood disorders and in prosthetic devices for those with communication disorders. In particular, the rapidly growing field of affective computing is focused on integrating the emotional states of users into computer applications [2], [3]. However, more accurate pattern recognition is needed to make such applications reliable.

Machine learning is a quickly expanding field that trains data-driven pattern recognition models [4]. In recent years, machine learning approaches have been used to classify emotional imagination [5], and emotional reactions to pictures [6], audio samples [7], and videos [8] from electroencephalography (EEG). However, these studies have typically been limited to classifying only positive versus negative emotions, and use research-grade EEG hardware which is costly and impractical for most real-world applications.

In this study, we extend previous results on classification of emotional reactions and classify high versus low emotional experience with respect to 11 different emotions using a consumer-grade EEG headband. In addition, we show that we can predict from the same EEG signals whether inter-subject classification will result in above-chance classification at all. Finally, we extend our results beyond simple binary classification and employ machine learning regression analysis to predict the magnitude of the emotional reaction as well.

II. Data Acquisition

A. Participants and Materials

Forty undergraduate students from McMaster University (24 female) participated in the study. No exclusion criteria were applied. Participants provided informed consent and the study was approved by the McMaster Research Ethics Board.

A total of 102 videos acquired through Youtube were used. The videos varied in length from 51s to 300s with an average length of 138.7s. Approximately half of the video clips were from Hollywood movies, and the other half were taken in real life situations using video cameras.

EEG was recorded with the Muse headband [9]. The Muse headband has four EEG electrodes located at T9, Fp1, Fp2, and T10 according to the International 10-20 system. The reference electrode is situated at Fpz with DRLs (driven right legs) on both sides. EEG was sampled at 220 Hz with 50 Hz and 60 Hz notch filters implemented in hardware.

B. Experimental Protocol

Data were collected using Matlab R2013b [10] and stimuli were presented using the Psychophysics Toolbox [11]. Participants were set up with an EEG headband and seated in front of a laptop in a private room. The participants watched approximately 60 minutes of videos, amounting to 17.2 ± 3.5 videos watched per subject. After each video, participants rated on a scale of one to ten the extent to which they experienced the following emotions in response to the video they just watched: 'Interest', 'Amusement', 'Happiness', 'Sadness', 'Fear', 'Disgust', 'Anger', 'Hope', 'Relief', 'Surprise', and 'Sympathy'. Videos were pre-labelled as positive or negative and an approximately equal number of positive versus negative videos were played for each participant.

III. Data Analysis

A. Preprocessing

EEG corresponding to each video was de-meaned and cleaned of artifacts using the Filter-Bank Artifact Rejection (FBAR) toolbox [12]. For classification analysis, self-reported emotional ratings were transformed into binary labels by grouping ratings from 1 to 5 in one class and ratings from 6 to 10 in another class. Emotional ratings were kept in their original form for regression analysis.

B. Feature Extraction

Fourth order Butterworth bandpass filters were used to extract the theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (31-45 Hz) bands from each of the four EEG channels.

From each of these frequency bands, a wide variety of features were computed in order to increase the likelihood that a subset of features would be found to be useful for each emotion and for each task, including spectral and cross-spectral density, coherence, cross-frequency coupling, bispectrum, bicoherence, quadratic phase coupling, and alpha asymmetries. In total, 279 features were computed from the EEG recording during each video viewing.

1) Power Spectal Density: For each of the four channels, the power spectral density (PSD) was estimated for the theta, alpha, and beta bands using the definition

$$P_{XX}(f) = \lim_{T \to \infty} \mathbf{E}\left[|\hat{X}_T(f)|^2\right] \tag{1}$$

where $\hat{X}_T(\omega)$ is the finite-time Fourier transform of a signal X. This yielded 12 features.

2) Cross-Spectrum: The cross-spectral density (CSD) extends the PSD to the Fourier transform of two signals, X and Y and is defined as

$$P_{XY}(f) = \lim_{T \to \infty} \mathbf{E}\left[\hat{X}_T(f)\hat{Y}_T(f)\right]. \tag{2}$$

The CSD was computed for the theta, alpha, beta, and gamma bands for all pairs of channels, resulting in 24 features.

3) Coherence: Coherence between all pairs of electrodes within the theta, alpha, beta, and gamma bands was computed. Coherence between two signals X and Y is defined as

$$C_{XY}(f) = \frac{|P_{XY}(f)|^2}{P_{XX}(f)P_{YY}(f)}, \tag{3}$$

where P_{XX} is the PSD as defined in Eq. 1, and P_{XY} is the CSD as defined in Eq. 2. In total, 24 coherence features were computed.

4) Cross-Frequency Coupling: The weighted phase locking factor (WPLF) [13] was used as a measure of cross-frequency coupling. WPLF is a measure of coupling strength and the preferred phase angle for a pair of signals. WPLF is given by

$$WPLF = \frac{1}{T}\sum_{t=1}^{T} e^{i\theta(t)}, \tag{4}$$

where T is time and θ is the phase difference between two signals. WPLF magnitude and phase angle was computed for each pair of channels and for each pair of frequency bands, resulting in 96 features.

5) Bispectrum, Bicoherence, and Quadratic Phase Coupling: The bispectrum of a signal is the 2D Fourier Transform of the third order cumulant generating function and is given by

$$B(f_1, f_2) = \mathcal{F}^*(f_1 + f_2)\mathcal{F}(f_1)\mathcal{F}(f_2), \tag{5}$$

where $\mathcal{F}$ is the Fourier Transform and $\mathcal{F}^*$ is its complex conjugate [14].

Bicoherence is the normalized bispectrum for a signal in n bins (32 bins were used here):

$$B_c(f_1, f_2) = \frac{|\sum_n \mathcal{F}_n(f_1)\mathcal{F}_n(f_2)\mathcal{F}_n^*(f_1 + f_2)|}{\sum_n |\mathcal{F}_n(f_1)\mathcal{F}_n(f_2)\mathcal{F}_n^*(f_1 + f_2)|}. \tag{6}$$

Quadratic Phase Coupling features are obtained from the autoregressive parameters of the bicoherence of a signal. These parameters provide information about the change in bicoherence over the frequency landscape. A 6th order autoregressive analysis was performed here.

The bispectrum, bicoherence and QPC are used to capture non-linear interactions between frequency pairs in a signal with respect to magnitude, frequency-normalized magnitude, and phase [14]. The sum and sum-of-squares of the bispectrum and bicoherence for the theta, alpha, and beta bands of each channel, and their corresponding QPC autoregressive coefficients were extracted as features, for a total of 120 features.

6) Alpha Asymmetries: Asymmetric alpha power in the frontal and temporal cortices is associated with emotional responses to video [15]. Three alpha asymmetry features were computed: frontal alpha asymmetry (FAA) using only the two frontal channels, temporal alpha asymmetry (TAA) using only the two temporal channels, and global alpha asymmetry (GAA) using all four channels. Alpha asymmetries were computed using the formula

$$AA = \frac{L - R}{L + R} \tag{7}$$

where L and R are left and right alpha power respectively.

C. Feature Selection

From the 279 features extracted, small subsets were selected for classification using the minimum-Redundancy Maximum-Relevance (mRMR) feature selection method [16] to reduce the risk of overfitting. The mRMR method is an information theoretic approach to feature selection which aims to maximize the mutual information between the subset of selected features and the true training labels while simultaneously minimizing the mutual information among selected features. The number of features to select is a parameter chosen via cross-validation from the set $K \in \{4, 8, 16, 32\}$. Features were selected using training data only.

D. Classification and Regression

Two classifiers were used in this study: a support vector machine with a radial basis function kernel ($C = 1$, $\sigma_{RBF} = 1$) [17] and logistic regression (LR) with elastic net regularization ($\alpha = \frac{1}{4}$) [18]. The parameter λ was chosen with a nested validation set comprising 10% of the training data. Regression analyses were performed using a boosted decision tree with 500 nodes [19].

E. Classification and Regression Tasks

The EEG data recorded during each video were used in five different tasks. In Task 1, leave-one-subject-out cross-validation (LOSO-CV) was performed by using each participant's data as a test set once while the remaining participants for training. In Task 2, within-subject analysis was performed with leave-one-video-out cross-validation (LOVO-CV). In Task 3, LOSO-CV was performed again but this time using only the participants who had classifiable data (greater than 60% classification accuracy or statistically greater than
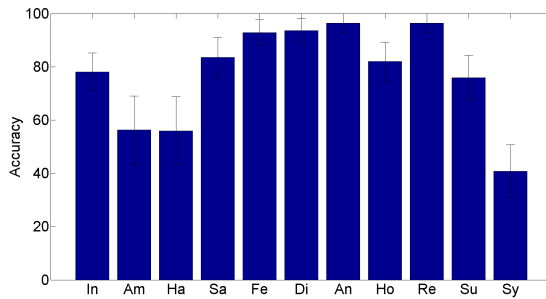
Fig. 1: Prediction rate of classifiability of participants with respect to each emotion. In = Interest, Am = Amusement, Ha = Happiness, Sa = Sadness, Fe = Fear, Di = Disgust, An = Anger, Ho = Hope, Re = Relief, Su = Surprise, Sy = Sympathy.

Fig. 2: Prediction rate of Classifiability of each participant averaged across all emotions.

zero correlation between predicted and observed emotional response) in Task 2. In Task 4, data from all participants were used together to perform regular 10-fold cross-validation with randomized partitioning of training and test sets with a 75%-25% split. Finally, in Task 5, 10-fold cross-validation was performed using only the participants who had classifiable data in Task 2.

IV. RESULTS

Table I shows the classification accuracies for each task and each emotion. The dominant relevant features were alpha asymmetries, frontal alpha and beta bicoherence, temporal theta coupling, and alpha and beta fronto-temporal WPLF. Similarly, the results of regression analysis are shown in Table II. Regression results were driven almost exclusively by beta band QPC features in the frontal channels.

There was a high degree of inter-subject variance in classification accuracy for each emotion. This was in part driven by the fact that some participants' EEG could not be classified beyond chance levels, while the EEG from others could be classified with over 90% accuracy. We used the results from Task 1 to label those for whom greater than 60% classification accuracy was achieved as 'Classifiable' and those for whom less than 50% accuracy was achieved as 'Non-Classifiable' and trained new classifiers based on these labels. The LOSO-CV classification accuracy for predicting whether participants would be classifiable with respect to each emotion is shown in Figure 1, and the average classifiability across emotions for each participant is shown in Figure 2. These results were driven by temporal theta and beta bicoherence, temporal theta and alpha cross-spectral features, and frontal beta power.

V. DISCUSSION AND CONCLUSIONS

In this study we showed that it is possible to predict both the presence and magnitude of several different types of emotional experiences to videos with a consumer-grade EEG headband. As such, the method developed here may be more directly applicable to real-world affective computing
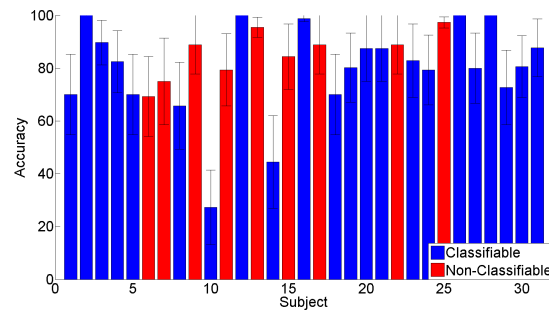
applications. We achieved average classification accuracies of over 70% when performing LOSO-CV with and without non-classifiable participants (Tasks 1 and 3), LOVO-CV within participants (Task 2), and regular 10-fold CV with and without non-classifiable participants (Tasks 4 and 5). We achieved over 80% classification accuracy with some emotions, such as 'Interest' and 'Anger'. Furthermore, the regression analyses resulted in models which could predict the degree of the reported emotional experience with, in some cases, $R > 0.4$.

For most emotions, we were also able to predict from the EEG which participants would be classifiable or non-classifiable using LOSO-CV. Although the same data were used in the previous classification analyses, a unique set of features were selected when determining whether a participant would be classifiable. This result suggests that it may be possible in the case of emotion recognition or potentially in brain-computer interfacing to determine a priori whether or not a given participants will be easily classifiable. Interestingly, non-classifiability was less of a problem in regression analysis, where the amount of training data seemed to be the main factor in learning accurate models. Further analysis is required to determine why certain participants were non-classifiable. One possibility is that these participants represent a subset of individuals whose emotional reactions produce a different pattern of brain activity from the majority.

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal processing magazine*, vol. 18, no. 1, pp. 32–80, 2001.
[2] J. Tao and T. Tan, "Affective computing: A review," in *International Conference on Affective computing and intelligent interaction*. Springer, 2005, pp. 981–995.

TABLE I: CLASSIFICATION ACCURACY (STD)

	Task 1		Task 2		Task 3		Task 4		Task 5	
	SVM	LR	SVM	LR	SVM	LR	SVM	LR	SVM	LR
Interest	82 (17.6)	73 (27.3)	83 (12.3)	81 (19.2)	80 (14.6)	79 (11.4)	76 (3.0)	76 (3.8)	84 (6.1)	88 (5.6)
Amusement	61 (16.2)	50 (18.3)	71 (13.1)	66 (17.8)	59 (15.5)	47 (14.0)	54 (3.7)	53 (4.5)	59 (6.6)	58 (7.6)
Happiness	61 (13.7)	58 (14.1)	66 (15.0)	59 (12.6)	47 (6.2)	56 (22.3)	55 (2.4)	60 (2.1)	60 (11.5)	58 (15.1)
Sadness	73 (15.6)	70 (19.0)	69 (14.9)	68 (15.8)	76 (5.8)	77 (11.1)	68 (5.8)	70 (3.3)	73 (7.0)	78 (8.1)
Fear	81 (14.9)	79 (17.2)	72 (12.0)	72 (12.0)	80 (5.6)	79 (8.3)	75 (2.8)	77 (2.3)	79 (6.5)	79 (8.0)
Disgust	80 (13.4)	80 (15.3)	75 (8.6)	74 (10.8)	78 (7.9)	77 (10.1)	76 (3.0)	80 (2.5)	78 (1.6)	77 (5.2)
Anger	82 (14.4)	81 (16.7)	78 (10.3)	74 (11.5)	83 (6.8)	83 (9.8)	79 (3.0)	80 (3.1)	84 (2.8)	83 (7.7)
Hope	78 (16.1)	74 (21.9)	78 (12.5)	70 (15.5)	79 (14.2)	79 (11.8)	76 (3.9)	75 (3.6)	80 (6.2)	82 (8.0)
Relief	83 (15.0)	80 (19.1)	76 (12.5)	75 (13.8)	81 (10.6)	75 (22.5)	79 (3.8)	81 (3.3)	76 (4.5)	78 (5.3)
Surprise	73 (19.7)	67 (26.6)	76 (13.2)	71 (16.8)	81 (11.5)	62 (23.9)	68 (3.3)	69 (4.4)	62 (4.6)	61 (6.7)
Sympathy	66 (17.4)	58 (20.5)	73 (13.7)	69 (15.0)	70 (14.6)	66 (12.1)	59 (5.4)	62 (4.1)	71 (6.2)	68 (8.0)
Average	75 (8.4)	70 (10.6)	74 (4.7)	71 (5.6)	74 (11.3)	71 (11.5)	70 (9.5)	71 (9.3)	73 (9.2)	74 (10.6)

TABLE II: REGRESSION MSE and CORRELATION

	Task 1		Task 2		Task 3		Task 4		Task 5	
	MSE	R	MSE	R	MSE	R	MSE	R	MSE	R
Interest	5.17**	0.13*	5.67**	N/A	4.97**	0.28**	4.75**	0.27**	4.89**	0.31**
Amusement	11.31**	0.08	11.50**	0.01	10.53**	0.27**	11.15**	0.12**	11.22**	0.09*
Happiness	13.47**	0.07	21.45	N/A	14.01**	0.25**	13.97**	0.07*	14.95**	0.07
Sadness	12.11**	0.13*	14.88*	0.00	12.68**	0.34**	11.33**	0.13**	11.07**	0.25**
Fear	10.06**	0.09*	8.39**	N/A	7.79**	0.30**	9.34**	0.06*	10.09**	0.15*
Disgust	11.33**	0.06	21.99	N/A	10.26**	0.22**	10.78**	0.08*	11.26**	0.05
Anger	9.29**	0.06	15.37*	0.04	8.65**	0.43**	9.62**	0.02	8.23**	0.34**
Hope	10.02**	0.09	15.09*	0.00	8.44**	0.31**	9.48**	0.11**	9.83**	0.21**
Relief	9.50**	0.03	12.64*	0.04	11.34**	0.15*	9.00**	0.03	5.73**	0.36**
Surprise	11.95**	0.02	11.90*	0.01	6.03*	0.41**	8.72**	0.18**	4.76**	0.38**
Sympathy	11.78**	0.10	13.61*	0.01	11.43*	0.37**	10.26**	0.27**	12.70**	0.25**

* denotes significance from chance level accuracy at $p < 0.05$, and ** denotes significance at $p < 0.005$

[3] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98–125, 2017.

[4] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[5] C. A. Kothe, S. Makeig, and J. A. Onton, "Emotion recognition from EEG during self-paced emotional imagery," *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, pp. 855–858, 2013.

[6] D. J. McFarland, M. A. Parvaz, W. A. Sarnacki, R. Z. Goldstein, and J. R. Wolpaw, "Prediction of subjective ratings of emotional pictures by eeg features," *Journal of Neural Engineering*, vol. 14, no. 1, p. 016009, 2016.

[7] S. Makeig, G. Leslie, T. Mullen, D. Sarma, N. Bigdely-Shamlo, and C. Kothe, "First demonstration of a musical emotion BCI," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6975 LNCS, no. PART 2, pp. 487–496, 2011.

[8] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "Eeg-based emotion recognition during watching movies," in *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*. IEEE, 2011, pp. 667–670.

[9] Interaxon, "Muse," www.choosemuse.com, 2014.

[10] MATLAB, *version 8.2.0 (R2013b)*. Natick, Massachusetts: The MathWorks Inc., 2013.

[11] D. H. Brainard, "The psychophysics toolbox," *Spatial vision*, vol. 10, pp. 433–436, 1997.

[12] K. Dhindsa, "Filter-Bank Artifact Rejection: High Performance Real-Time Single-Channel Artifact Detection for EEG," *Biomedical Signal Processing and Control*, In Press: 2017.

[13] J.-P. Lachaux, E. Rodriguez, J. Martinerie, F. J. Varela *et al.*, "Measuring phase synchrony in brain signals," *Human brain mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[14] U. Greb and M. Rusbridge, "The interpretation of the bispectrum and bicoherence for non-linear interactions of continuous spectra," *Plasma physics and controlled fusion*, vol. 30, no. 5, p. 537, 1988.

[15] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V. Friesen, "Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I." *Journal of personality and social psychology*, vol. 58, no. 2, p. 330, 1990.

[16] H. C. Peng, "Feature Selection Based on Mutual Information Criteria of Max-dependency, Max-relevance, and Min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.

[17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[18] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[19] T. Hastie, R. Tibshirani, and J. Friedman, "Boosting and additive trees," in *The Elements of Statistical Learning*. Springer, 2009, pp. 337–387.

5.3   Discussion

Accurate detection of emotional states and reactions has interesting applications in affective computing, psychotherapy, and BCIs (particularly passive BCIs, which rely on detection of changes in mental states rather than willful mental commands [3]). For example, one can imagine that the ability to detect fear and disgust can be used to monitor the individual progress of a patient undergoing treatment for a phobia. The approach laid out in this work has not yet been tested for real-time classification, so further research would need to be done in order to evaluate its efficacy in that context. However, this work does demonstrate that a generalized approach can be used for the complex task of emotion recognition from EEG. Moreover, this work shows that generalized methods can be used in the analysis of a variety of emotions, rather than being fine-tuned for just one or two emotions.

The generalized approach based on SF performed well on many of the emotions studied even though the approach was not specifically designed to detect emotions at all. The neural correlates as expressed in EEG of each of these emotions have not yet been fully characterized in the literature either, so designing a specialized feature space specifically for each emotion would require further neuroscientific study similar to the study presented in Chapter 3. Given this fact, this study serves to demonstrate an additional benefit gained by generalized BCI methods, which is that they can enable applications even if there is insufficient *a priori* knowledge with which to develop handcrafted models. Further analysis on the specific features which are discriminative for each emotion can add new information to the literature regarding which features of the EEG are relevant to the classification of various emotional states.

It must be noted that because each emotion was analyzed independently (this was done because the emotional ratings are highly correlated across emotions given that multiple emotions could be rated highly per video), it is possible that the classifiers were actually identifying general arousal levels or video-specific visual processing rather than the emotions themselves. This explanation seems unlikely, because upon further analysis the specific discriminative features were different for each emotion and a variety of videos were associated with each emotion, suggesting that general arousal or visual aspects of the videos cannot account for the results that were obtained. However, deeper feature-level analysis is required to confirm that the classification and regression models truly reflected emotional states and not some other signal present in the EEG.

One critical limitation in this study is that no method was used to try and identify where in the EEG time series different emotions were experienced. The videos were quite long with respect to the usual time windows used in EEG signal processing. Computing features across these long windows can be unreliable due

to nonstationarity in the EEG, which can be seen at timescales as short as several seconds. A standard approach to EEG analysis would have been to subdivide the EEG time series recorded during each video into many short (*e.g.*, 2s) epochs which share the same class label. However, it would not be reasonable to expect that each epoch would actually reflect the labelled emotion, which would be very detrimental to the success of any purely supervised method. Therefore, computing features per video, while reducing the discriminative power of those features due to smearing across discriminative and non-discriminative time windows within the time series, was a conservative approach to the problem that at least avoided introducing even more noise into the class labels.

In order to maximize discriminative power, it would be necessary to use a semi-supervised approach, which would first identify time windows in which the emotional experience was present. One idea, which will be experimented with upon revisiting these data, is to track changes in the features across the video compared to a baseline measure. More specifically, spectral features could be computed for short overlapping time windows in order to approximate a continuous time series of feature values. These feature values could be compared to the average of all feature values across all videos for the given participant, which would serve as a baseline measure, or centroid, of the feature space. The total deviation of the feature values over time with respect to the baseline (*e.g.*, the mean-squared difference) might be useful in determining where the discriminative information resides in each video, particularly if the feature values deviate from baseline in a relatively smooth fashion for at least several seconds. Using a statistical threshold, a segment of the EEG time series corresponding to where the feature values deviate from the baseline may provide a means to retain the discimrinative data separate from the extraneous portions of the time series. However, it may be necessary to introduce some supervision in this proposed method, since it is also possible that the extracted time segments will be more reflective of neural processing related to non-emotional features of the video, *e.g.*, audiovisual features. One option for introducing supervision is to test whether the direction of the deviation is predictive of the class labels themselves, and to choose time windows accordingly.

Bibliography

[1] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V. Friesen, "Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I.," *Journal of personality and social psychology*, vol. 58, no. 2, p. 330, 1990.

[2] R. J. Davidson, "Anterior cerebral asymmetry and the nature of emotion," *Brain and cognition*, vol. 20, no. 1, pp. 125–151, 1992.

[3] T. O. Zander and C. Kothe, "Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general," *Journal of neural engineering*, vol. 8, no. 2, p. 025005, 2011.

Chapter 6

Improving Human Self-Control Over Neural Activity with Adaptive Neurofeedback

6.1   Introduction

So far the focus of this thesis has been on developing an appropriate machine learning approach for a generalized BCI. However, in Chapter 2 it was noted that redesigning the way the human user is trained to control a BCI is just as important as making advancements in machine learning. The need for improved methods for user training may be even greater for generalized BCIs and open-ended BCIs than for standard BCIs, because there is no guarantee that the user will choose mental commands which have easily separable neural correlates to begin with. This chapter, along with the next, address the question of how to advance human training protocols for BCI.

Humans are trained to use a BCI via neurofeedback (NFB) [1, 2, 3, 4]. NFB was originally developed as a way to train individuals to modulate or regulate their own brain activity with respect to a single simple feature (*e.g.*, power in a specific frequency band) or a ratio of simple features (*e.g.*, the ratio of power between two bands) usually at a single electrode site [5, 6, 7]. The applicability of NFB to BCI is clear due to the need for the user to learn to modulate their own brain activity in order to produce consistent and distinct patterns of brain activity for BCI control. However, standard NFB methodology may need to be modified for the BCI context, where the user must learn to willfully control their brain activity with respect to several features simultaneously.

In current practice NFB is applied in a variety of ways for BCI user training. However, these methods do not vary widely from the standard methods used in NFB outside of BCI, with one notable difference. In a traditional NFB protocol the target brain state is defined either as a specific value for the feature being trained, or as simply increasing or decreasing that feature. In the BCI context, NFB targets are instead defined by the machine learning model. Therefore, the NFB target in a BCI is somewhat amorphous; there is a different target for each mental command and the targets change as the system adapts to the user. The targets depend on whatever patterns of brain activity the model recognizes as belonging to each mental command at the time.

There are several problems with current implementations of classifier-based NFB (see [4] for a more detailed discussion). First, because each class is usually not completely separable in feature space unless perfect classification accuracy is achieved, feedback can be ambiguous near the classification boundary. Near this boundary, the classifier itself would exhibit low confidence of its classification prediction and can easily be wrong. Furthermore, since most individuals only achieve low to moderate degrees of BCI control, classifier-based NFB is not usually highly reliable. For both of these reasons, there is only a low degree of confidence that any feedback derived from the classifier would promote changes in brain

activity in the correct direction.

This problem is even worse in a real-time BCI application with live classifier updates. In this scenario, the classifier generally begins with very poor accuracy and improves over time. If a user begins training with NFB derived from a poor classifier which has not yet established clear patterns for each mental command, it may be difficult to learn useful strategies during the early, and arguably most important, stages of training. Moreover, the user may learn counter-productive strategies which can limit performance in the long term. This problem could be partially mitigated if there were a clear way to inform the user about which features of brain activity need to be adjusted and in which direction. Unfortunately, and particularly in the case of non-linear classifiers, it is very difficult to determine exactly why a certain class is selected in the first place. Therefore, simply determining which features violate the pattern expected by the classifier can be a very difficult problem which prevents the implementation of feature-specific NFB training.

Due to the problems in directly applying standard NFB methods to BCI, there is a growing consensus that user training for BCIs is itself one of the more serious barriers to improving BCI usability [8, 9, 3, 4]. Recent empirical work using BCI training methods for non-BCI tasks strongly supports this notion [10, 11]. However, alternative training methods or adaptations of traditional NFB methodology to better fit the BCI context have not yet been proposed.

Since NFB is central to user training in BCI, the NFB literature was studied in order to gain insight into how NFB protocols could be modified to better serve as a tool in brain-computer interfacing. What was discovered was that even in traditional NFB training, the problem of training someone to change their brain activity in the correct direction remained unsolved. This could be seen as an opportunity, because if this problem could be solved in the simpler traditional NFB setting, then the solution might be adaptable to the more complex BCI setting.

Current automated NFB methodology (*i.e.*, algorithm-controlled NFB as opposed to manual clinician-controlled NFB) lacks a clear and reliable method for shaping, even though it is in essence based on an operant conditioning paradigm [7, 12]. Shaping, in Learning Theory, is the use of incremental reinforcement for progressive approximations or improvements with respect to a task, skill, or behaviour to be learned [13]. When implemented well, shaping promotes learning in a manageable way with specificity, or with a reduced risk of producing extraneous and erroneous behaviour. Due to a lack of shaping in automated NFB, users often learn to modulate their brain activity in incorrect or irrelevant ways [3, 14, 15]. The failure rate is surprisingly similar to the BCI illiteracy rate, a parallel which may or may not be coincidental.

This chapter presents the results of a new NFB training protocol called Progressive Thresholding. The version presented here is a specific and simple implementation of the Progressive Thresholding algorithm designed for traditional and clinical neurofeedback applications. This new user training approach was first validated in a standard NFB protocol in order to compare it more directly to current NFB methods. Moreover, extending Progressive Thresholding for BCI user training involves a more complex algorithm (see Chapter 7), and BCI studies involve added layers of technical complication which can influence the results. Moreover, BCI user training involves NFB training with respect to several features simultaneously, as opposed to traditional NFB training which involves learning to modulate only one or two features. These factors together suggested that it would be easier to assess the impact of Progressive Thresholding in a standard NFB protocol before moving to a BCI study. There is also a clear and present need for a such a technique for clinical NFB. This gives the following study much more importance than just as a precursor to a BCI study based on the same principles, even if improved BCI user training was the original intention behind developing this new approach.

6.2   Adaptive Neurofeedback: Training Individuals to Control their Brain Activity

Dhindsa, K., Gauder, K. D., Marszalek, K. A., Terpou, B., & Becker, S. (Revision under review). Progressive Thresholding: Incorporating Shaping and Specificity into Automated Neurofeedback Training.

Progressive Thresholding: Incorporating Shaping and Specificity into Automated Neurofeedback Training

Kiret Dhindsa: *Neurotechnology and Neuroplasticity Lab*

School of Computational Science and Engineering, McMaster University

Kyle D. Gauder: *Neurotechnology and Neuroplasticity Lab*

Department of Psychology, Neuroscience, and Behaviour, McMaster University

Kristen A. Marszalek: *Neurotechnology and Neuroplasticity Lab*

Department of Psychology, Neuroscience, and Behaviour, McMaster University

Braeden Terpou: *Neurotechnology and Neuroplasticity Lab*

Department of Psychology, Neuroscience, and Behaviour, McMaster University

Suzanna Becker: *Neurotechnology and Neuroplasticity Lab*

Department of Psychology, Neuroscience, and Behaviour, McMaster University

1280 Main Street West, Hamilton, Ontario, Canada, L8S 4L8

Phone: 1-905-525-9140 ext. 23030

Email: becker@mcmaster.ca

Abstract

Neurofeedback, a type of biofeedback which trains individuals to modify their own brain activity, has long been discussed as a promising form of adjunctive non-pharmaceutical treatment for many neurophychological disorders. However, there is an active debate over how efficacious and specific neurofeedback treatments are, and how to best design and implement them. One of the central issues being debated is whether reward thresholds, which are meant to provide an essential component of the training signal by indicating to the patient or trainee when they are changing their brain activity correctly, can be set automatically by an algorithm or must be set manually by a specially trained clinician. The current debate has largely settled on the conclusion that because automatic reward thresholding, as it is currently implemented, does not involve shaping the trainee's progress towards the goal, effective and reliable neurofeedback must be done with manual thresholding.

In this study we weigh in on this debate by suggesting a third option. It is possible to achieve both the convenience and affordability of automatic thresholding as well as the efficacy and reliability of manual thresholding simultaneously. The problem with automatic thresholding does not lie in the fact that it is automatic, but in the fact that current algorithms utilize a flat reward rate and therefore do not incorporate shaping. Here, we show that automatic reward thresholding is not synonymous with a flat reward rate algorithm and present a new automatic thresholding algorithm, Progressive Thresholding, which does away with flat reward rates and instead uses difficulty tuning and inter-session progress models for each individual in order to simulate the kind of shaping a clinician might perform when setting reward thresholds manually. We show that Progressive Thresholding leads to far superior learning outcomes compared to the current standards in automatic reward thresholding.

Keywords: Neurofeedback, Electroencephalography (EEG), Automatic Reward Thresholding, Frontal Alpha Asymmetry, Learning Theory, Algorithm Design

Introduction

Neurofeedback is the use of classical or operant conditioning to train an individual to regulate or modify their own brain activity (Sherlin et al., 2011; Thatcher, 2000). This is accomplished by pairing a real-time display of an individual's brain activity with some indicator of performance with respect to a defined target or goal. The combination of these key ingredients enables an individual to associate their own mental states with changes in the feedback stimulus and then to learn how to generate brain activity in order to maximize performance.

A neurofeedback protocol typically begins with a feature of brain activity which is to be trained, such as the sensorimotor rhythm (Wyrwicka & Sterman, 1968) or the difference in frontal alpha activity between the left and right hemispheres, referred to as frontal alpha asymmetry (FAA) (Allen, Harmon-Jones, & Cavender, 2001). The goal of training with respect to this neurofeedback target can be defined as increasing or decreasing the specified activity, or tuning the activity to a certain level, depending on the application. The feedback stimulus shows trainees their current brain activity with respect to the neurofeedback target,

and usually also indicates where this feature of brain activity must move in order for trainees to be rewarded. Setting and updating this threshold is referred to as reward thresholding, and is typically done based on the intra-session statistics of the target feature of brain activity.

Neuroefeedback is most commonly employed as a neuropsychiatric tool, where the rationale is that if a patient can learn to change the patterns of brain activity which are associated with, and perhaps causally related to, their psychiatric condition, their behaviour and cognition may change in a corresponding way (Angelakis et al., 2007; Heinrich, Gevensleben, & Strehl, 2007). Many studies have reported positive findings using neurofeedback as a therapeutic agent. *e.g.*, for attention deficit hyperactivity disorder (ADHD) (Arns, de Ridder, Strehl, Breteler, & Coenen, 2009). Furthermore, studies seeking to shed light on the mechanisms underlying such improvements seem to indicate that neurofeedback can lead to both short-term (Ros, Munneke, Ruge, Gruzelier, & Rothwell, 2010; Ghaziri et al., 2013; Kluetsch et al., 2014) and long-term neuroplastic changes involving the targeted brain areas and deeper brain structures with which they communicate (Lévesque, Beauregard, & Mensour, 2006; Scheinost et al., 2013; Simkin, Thatcher, & Lubar, 2014). However, other studies have reported mixed or negative results using similar protocols (Lofthouse, Arnold, & Hurt, 2012; van Dongen-Boomsma, Vollebregt, Slaats-Willemse, & Buitelaar, 2013; Zuberer, Brandeis, & Drechsler, 2015). While the exact reasons for this serious discrepancy in the literature are still actively debated (for example, problems with methodology and implementation could account for the results of some studies (Pigott & Cannon, 2014)), even prominent and optimistic neurofeedback researchers write that there are some important limitations to current neurofeedback technologies (Sherlin et al., 2011; Lofthouse et al., 2012; Arns & Kenemans, 2014).

The main limitation preventing clear validation of the efficacy of neurofeedback therapies is the difficulty in designing a double-blind randomized controlled study, and the debate over proper neurofeedback methodology often centers around the question of how to appropriately design a double-blind controlled study (Sherlin et al., 2011; Arns & Kenemans, 2014). The reason double-blinding is difficult to implement is that in order to incorporate shaping, that is, rewarding incremental progress towards the goal, reward thresholds are typically updated manually by an experimenter or a clinician who monitors progress on a separate screen (Arns, Heinrich, & Strehl, 2014). Recently, double-blind designs have been proposed using manual thresholding and multiple testing sites in an attempt to maintain blinding (Kerson, 2013). However, this is an expensive solution which still leaves patients who might benefit from neurofeedback reliant on clinicians to dedicate a significant amount of time to overseeing their training. An automated neurofeedback approach would allow many more to benefit from neurofeedback in a more simple and cost-effective way.

Researchers have attempted to conduct double-blind randomized controlled trials by employing automatic reward thresholding algorithms. However, it is with these algorithms that results with neurofeedback studies appear to be particularly mixed or even negative (DeBeus & Kaiser, 2011; Perreau-Linck, Lessard, Lévesque, & Beauregard, 2010; Lansbergen, van Dongen-Boomsma, Buitelaar, & Slaats-Willemse, 2011; Arnold et al., 2012; van Dongen-Boomsma et al., 2013; Arns et al., 2014). Prominent neurofeedback researchers have

argued that the reason for these mixed results is that automatic reward thresholding algorithms currently do not incorporate shaping (Sherlin et al., 2011; Arns et al., 2014; Strehl, 2014). Instead, automatic reward thresholding is typically implemented with a flat reward rate. For example, the reward threshold might be updated in order to maintain a 75% reward rate at all times, regardless of whether the trainee performs better or worse with respect to the actual goal.

Neurofeedback training with current automatic reward thresholding may actually contribute to an additional problem seen in the neurofeedback literature, where approximately 15% to 25% of participants are either not successful in changing their brain activity, or erroneously change their brain activity in the incorrect direction (Vollebregt, Dongen-Boomsma, Buitelaar, & Slaats-Willemse, 2014; Zuberer et al., 2015). This can happen because current automatic reward thresholding algorithms take no account of whether a participant moves towards or away from the goal and do not consider whether a participant has improved with respect to their baseline brain activity.

Automatic reward thresholding with a fixed reward rate is still used in the literature (*e.g.*, (van Dongen-Boomsma et al., 2013; Vollebregt et al., 2014)) because performing manual reward thresholding with a specially trained clinician is expensive, laborious, and highly inconvenient in practice. Unfortunately, the unreliability of the current approach to automatic reward thresholding thresholding has led to an association between automatic reward thresholding and an incorrect application of learning theory (Sherlin et al., 2011; Arns et al., 2014; Strehl, 2014). As such, it is commonly stated in the neurofeedback literature that automatic reward thresholding is an ineffective approach to neurofeedback training and is not methodologically sound.

In this paper we aim to disambiguate automatic reward thresholding from reward thresholding with a fixed reward rate. Here we propose that it is entirely feasible to incorporate both shaping and specificity in an automatic reward thresholding algorithm by incorporating more sophisticated statistical analysis and intelligently implementing variable reward rates. We present a new automatic reward thresholding algorithm called Progressive Thresholding (PT), which incorporates shaping into automatic neurofeedback protocols by simulating what a clinician might aim to do. We compare PT to the standard automatic reward thresholding with a flat rate of reward (ST) and show that individuals who were trained with PT were more successful in learning to control their brain activity.

Experiment Overview

A double-blind randomized controlled trial was conducted in order to compare the learning outcomes of participants trained to regulate frontal alpha asymmetry (FAA) with either ST or PT. Each participant underwent a 30min prescreening session in order to determine eligibility to participate in neurofeedback training. Participants who met the eligibility criteria during prescreening were invited to participate in four to six weeks of neurofeedback training. When possible, three sessions were completed per week, with at least one day in between sessions. Each session took approximately 35 minutes and participants were compensated with money at the end of their participation.

PT can be implemented with virtually any neurofeedback protocol. In order to focus on the question of how PT compares to ST in training individuals to modulate some feature of brain activity towards a target state, it was important to choose an already established neurofeedback protocol for which ST is already used. Moreover, the use of non-established neurofeedback protocols involves a degree of risk, as undergoing neurofeedback training can result in changes in behaviour, mood, or cognitive functioning which are not necessarily positive (Hammond & Kirk, 2007).

The established neurofeedback protocol of reducing FAA (Rosenfeld, Cha, Blair, & Gotlib, 1995; Allen et al., 2001; Hammond, 2005; Choi et al., 2011; Peeters, Ronner, Bodar, van Os, & Lousberg, 2014; Baehr, Rosenfeld, & Baehr, 1997, 2001) was chosen primarily for two reasons. First, FAA balancing protocols are used in the treatment of depression, anxiety and stress because of the established associations between left-dominant FAA (*i.e.*, greater right frontal activation), mood, motivation, and predisposition to mood disorders (Henriques & Davidson, 1991; Davidson, 1998; Harmon-Jones, 2003; Harmon-Jones, Gable, & Peterson, 2010; Lewis, Weekes, & Wang, 2007; Stewart, Coan, Towers, & Allen, 2014). Though our main goal was to evaluate the efficacy of PT relative to ST, we were also aware that many undergraduate students, who make up our most reliable pool of potential participants, experience symptoms of, or are at risk of, clinically significant levels of depression, anxiety, and stress. Therefore, to the extent that FAA neurofeedback protocols are effective in helping such individuals, we aimed to provide some potential benefit to our participants in addition to empirically probing our research question.

The second reason for choosing the FAA balancing protocol was that we aimed to make the the neurofeedback sessions as simple and convenient for our participants as possible. In addition, we aimed to test our algorithm in a setting which would be compatible with low-cost and accessible options for neurofeedback treatments. For these reasons we used low-cost commercial EEG hardware which did not require electrode gel or long set up times (see the section titled, "EEG Apparatus" below). This device had electrodes positioned conveniently for measuring FAA, making the FAA protocol the best available option for a first test of PT versus ST.

Methods

Participants

Participants were recruited through McMaster University's online experiment recruitment service and with poster advertisements. A total of 102 participants (80 females) took part in a prescreening session to determine eligibility for neurofeedback training based on handedness, mood, and baseline resting EEG. Of the 102 prescreened participants, 21 (16 females) were deemed eligible for neurofeedback training. Of the 21 participants invited for neurofeedback training, 13 completed at least five sessions. One participant's data were discarded due to poor signal quality (less than 10% of data from each session survived artifact rejection).

Participants were randomly assigned to ST or PT, and completed all of their neurofeedback training sessions with the same thresholding algorithm. Six (three females) were trained using ST, and six (five females) were trained using PT. In order to ensure data privacy and blinding, all data were stored in an anonymized form and all data processing required during neurofeedback training was done automatically. Analysis of prescreening data was also done automatically, and the acceptance or rejection decision was passed onto the experimenter without any additional information. Participants were not informed that there were two training algorithms, and both experimenters and participants were blind to group assignment.

EEG Apparatus

EEG was recorded using the Muse EEG headband (Interaxon, 2014). The Muse Headband is a commercially available dry-sensor EEG headband with four EEG sensors located at Fp1, Fp2, TP9 and TP10 and with a reference at Fpz and DRLs (driven right legs) one inch from the reference on either side. This hardware configuration was used throughout this study. Data was collected from this headset at a 220 Hz sampling rate with a 0.5 Hz highpass filter and 50 Hz and 60 Hz notch filters applied automatically by the hardware in order to remove noise from surrounding electrical wires. The EEG sensors used in this device have been shown to acquire signals which are comparable to those acquired by more traditional research devices (Lee, Chin, Yi, Lee, & McKeown, 2015; Interaxon, 2015), lending confidence to its use in a research setting.

EEG Processing

All data processing, including FAA computation and neurofeedback thresholding, was performed in MAT-LAB R2013b (MATLAB, 2013) and all stimulus presentation and experimental instructions were displayed using Psychtoolbox (Brainard, 1997) in MATLAB. EEG was streamed into MATLAB in real time using the MuLES toolbox (Cassani, Banville, & Falk, 2015).

On an ongoing basis, the most recent 1s of EEG data from channels Fp1 and Fp2 was used to compute FAA. EEG from each channel was transformed to have a mean of zero and checked for artifacts using the FBAR (Filter-Bank Artifact Rejection) toolbox (Dhindsa, 2017). If both frontal channels were free of artifacts, the signal from each channel was filtered to be between 0.5 Hz and 40 Hz using a fourth-order Butterworth filter and alpha power (8-13 Hz) was computed for each channel. Segments of EEG which were marked as contaminated with artifacts by FBAR were not included in analysis or FAA computations.

FAA was computed as follows:

$$\text{FAA} = \frac{L - R}{L + R},$$

where L is alpha power calculated from the left frontal electrode (Fp1) and R is the alpha power calculated from the right frontal electrode (Fp2). Note that this formula is simply the negative of the formula commonly used in the literature, as first defined by (Rosenfeld et al., 1995), meaning that greater left alpha power would drive feedback to the right side of the feedback stimulus in Figure 1 rather than the left side.

FAA is a highly variable measure. In order to present interpretable feedback to participants, smoothing was required. In this study the average of the past 100 FAA computations was presented as feedback. As FAA was computed approximately 15-20 times per second, the past 100 FAA computations accounted for the past 5-7s of data. Thus, the feedback signal, while immediately responsive to changes in FAA, moved smoothly and was influenced by a time frame within which an individual could be expected to associate their mental activity with the feedback stimulus.

It has been suggested that percent-time with asymmetric alpha (PTAA) is a more reliable measure than FAA (Baehr et al., 1997). We computed PTAA in addition to FAA in order to determine whether there would be any significant differences in their correlations to our mood measures or learning outcomes. PTAA was computed as the proportion of time FAA was below zero.

Prescreening

Participants undergoing prescreening completed the Waterloo Handedness Inventory (WHI) (Steenhuis & Bryden, 1989), the Oxford Happiness Questionnaire (OHQ) (Hills & Argyle, 2002), and a 10min EEG baseline recording. Participants were eligible for neurofeedback training if they scored above 20 on the WHI (indicating that they were strongly right-handed and therefore unlikely to have reversed brain lateralization), less than four on the OHQ (indicating that they were experiencing below typical levels of happiness) and if their baseline FAA was greater than 0.05.

We allowed for one exception to these criteria. Individuals scoring above 20 on the WHI and below three on the OHQ were invited into the study even if their baseline FAA score was between 0 and 0.05. This exception was made because FAA tends to be an unstable measure, especially for individuals who are depressed (Debener et al., 2000), suggesting that their level of FAA could not be reliability determined on the basis of one baseline recording (*e.g.*, their FAA could be high at other times). Moreover, individuals exhibiting low OHQ scores are also more likely to be depressed and thus are more likely to benefit from FAA neurofeedback if FAA neurofeedback is indeed beneficial. The OHQ was used instead of a clinical inventory so that the many participants who underwent prescreening could provide an indication of their happiness levels through a less invasive questionnaire.

Together, these eligibility requirements were meant to reduce the risk of training individuals who may have reversed hemispheric functionality, who would not potentially benefit from mood enhancement which might be derived from neurofeedback training with an FAA protocol, and whose baseline FAA was not left-dominated. With the exception of those who scored very low on the OHQ (for the reasons mentioned above), those with right-dominant FAA were not included because depression, anxiety, and clinically significant degrees of stress are particularly associated with left-dominant FAA, as mentioned earlier. In contrast, right-dominant FAA is usually associated with positive affect.

Neurofeedback Sessions

Each neurofeedback session included a 3min pre-session resting baseline EEG recording and a 3min post-session resting baseline EEG recording. In between the baseline recordings, participants underwent approximately 20mins of neurofeedback training, split into five 4min blocks with 1min breaks in between. On the first and last sessions, participants completed the Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the Beck Anxiety Inventory (BAI) (Beck, Epstein, Brown, & Steer, 1988), and Cohen's Perceived Stress Scale (PSS) (Cohen, Kamarck, & Mermelstein, 1983). At the end of their last session, participants were asked to fill out a short questionnaire in which they were asked to rate the difficulty of the neurofeedback training task and their subjective assessment of how well they were able to perform on the task, as well as to describe the strategy they eventually settled on in order to perform the task.

At the end of every session, participants also completed a three-item mood inventory (TIM) based on depression screening work by Henkel et al. (Henkel et al., 2004). We included the following items in the questionnaire, which ask the participant how they have felt since their last session: "I have felt cheerful and in good spirits", "I have felt active and vigorous", and "I have felt calm and relaxed". For each of these items, participants could respond with six possible responses ranging from "All of the time", to "At no time". Finally, participants were also asked whether any events took place since their last training session which affected their mood, and whether they were affected positively or negatively. Altogether, these questions were used to identify whether changes in baseline FAA scores were affected by external circumstances without probing participants for details about their personal lives.

Neurofeedback Stimulus

Feedback was presented using the colour-coded bar shown in Figure 1. Perfectly balanced FAA was represented as the yellow line at the middle of the bar, and thresholds were marked symmetrically on either side by the green target area. Under both PT and ST, a decreased threshold was represented by a narrowing of the green area, and an increasing threshold was represented by a broadening of the green area. The white line moved along the bar and represented the participant's current FAA score. Participants were only instructed to learn how to control the white line and to keep the white line in the green area as much as possible.

Participants accumulated points while the white line remained in the green target area as a secondary reinforcer. The rate of point accumulation was increased over 15s from a multiplier of $1\times$ to a multiplier of $4\times$ in a linear fashion if the participant could maintain the white line in the target area without leaving it. This means that after 15 consecutive seconds of being in the target (green) area, participants accumulated points at four times the default rate. This multiplier dropped linearly back to $1\times$ over five seconds when the participant left the target area.

144

Figure 1: The feedback stimulus used to present real-time feedback for training.

Standard Thresholding

Standard thresholding was implemented as has been described throughout the neurofeedback literature (*e.g.*, (Lansbergen et al., 2011; van Dongen-Boomsma et al., 2013; Vollebregt et al., 2014)). Specifically, ST was implemented according to the pseudocode given in Algorithm 1. For clarity, the pre-session FAA distribution refers to the histogram of FAA values observed during the 3min pre-session baseline, and the block FAA distribution is the histogram of FAA values observed during the given neurofeedback block.

Algorithm 1 Standard Thresholding

1: **procedure**
2:     $D_{\mathrm{pre}}(n) \leftarrow \mathrm{n^{th}}$ percentile of the pre-session FAA distribution
3:     $D_{\mathrm{block}}(n) \leftarrow \mathrm{n^{th}}$ percentile of the current neurofeedback block's FAA distribution
4:     $T \leftarrow$ The reward threshold
5:     $\delta \leftarrow$ Time elapsed since last threshold update, in seconds
6: *Setting the initial threshold for the session*:
7:     **if** New Session **then**
8:         $T = D_{\mathrm{pre}}(75)$
9: *Updating the threshold during training*:
10:     **if** $\delta \geq 15$ **then**
11:         $T = D_{\mathrm{block}}(75)$
12:     **if** $T < 0.05$ **then**
13:         $T = 0.05$

Progressive Thresholding

Progressive thresholding incorporates two novel features which aim to facilitate specificity and shaping within sessions and across sessions. Within-session progress is promoted because the algorithm gradually increases the difficulty of training by decreasing the threshold as the participant succeeds in changing their FAA to become more symmetric. The difficulty only decreases marginally when a participant fails to reduce the threshold for an entire block, and is limited to decreasing only as far as the 75th percentile of the pre-session baseline FAA distribution. In contrast to ST, where the threshold area can theoretically change by any amount in any direction, PT uses much smoother and gradual transitions of the threshold and the change in threshold is asymmetric towards increasing difficulty by design. Furthermore, participants are not able to achieve a 75% reward rate if their FAA scores move farther away from zero compared to when they began the session.

Progress across sessions is promoted for participants by using their session history to seed the initial threshold of each new session. A linear regression analysis of the difference between the pre-session and

post-session baseline FAA scores is performed using data from all of that participant's previous training sessions. For participants who have completed more than two sessions, once a new pre-session baseline is complete, the results of this regression analysis are used to estimate the expected post-session FAA score for the current session, assuming that their rate of progress will be maintained. The initial threshold is set to halfway between the current pre-session baseline FAA score and the expected post-session baseline FAA score, requiring the participant to move closer to the goal compared to previous sessions in order to be rewarded. Note that this usually corresponds to an initial threshold which is less than the 75th percentile of the pre-session baseline FAA distribution. For the first and second neurofeeedback sessions, there is insufficient data with which to compute the regression model, so the initial threshold is set the same that it is set in ST (using the 75th percentile of the pre-session baseline FAA distribution). An idealized illustration of the initial threshold setting in PT is shown in Figure 2. The PT algorithm is described in pseudocode in Algorithm 2. Note that for our experiment, $\hat{\beta}_1$ is usually less than zero.

Algorithm 2 Progressive Thresholding

1: **procedure**
2:      $D_{\text{pre}}(n) \leftarrow n^{\text{th}}$ percentile of the pre-session FAA distribution for the current session
3:      $D_{\text{block}}(n) \leftarrow n^{\text{th}}$ percentile of the current neurofeedback block's FAA distribution
4:      $S \leftarrow$ Session Number
5:      $FAA_{\text{pre},i} \leftarrow$ Average FAA score of the ith pre-session baseline, $i = 1, \dots, S-1$
6:      $FAA_{\text{post},i} \leftarrow$ Average FAA score of the ith post-session baseline, $i = 1, \dots, S-1$
7:      $T \leftarrow$ The reward threshold
8:      $\delta \leftarrow$ Time elapsed since last threshold update, in seconds
9:      $flag \leftarrow$ set to *true* if the entire previous block has elapsed without a reduction in T
10: *Setting the initial threshold for the session*:
11:      **if** $S \leq 2$ **then**
12:          $T = D_{\text{pre}}(75)$
13:      **if** $S > 2$ **then**
14:          $\hat{\beta}_1 = \frac{\sum(S - mean(S))(FAA_{post,i} - FAA_{pre,i} - AVG(FAA_{post,i} - FAA_{pre,i}))}{\sum(S - mean(S))^2}$
15:          $T = mean(D_{pre}) + \hat{\beta}_1/2$
16: *Updating the threshold during training*:
17:      **if** $\delta \geq 15$ **then**
18:          **if** $\text{Prop}(-T < D_{\text{block}} < T) > 0.75$ **then**
19:              $T = 0.9 \times T$
20:          **if** $\text{Prop}(-T < D_{\text{block}} < T) < 0.75$ & $flag$ **then**
21:              $T = (1/0.9) \times T$
22:      **if** $T < 0.05$ **then**
23:          $T = 0.05$
24:      **if** $T > D_{\text{pre}}(75)$ **then**
25:          $T = D_{\text{pre}}(75)$

In both ST and PT procedures, T was set to a minimum of 0.05 in order to prevent the threshold area from becoming too small visually and to prevent the target from becoming too difficult given the normal moment to moment variability of FAA.
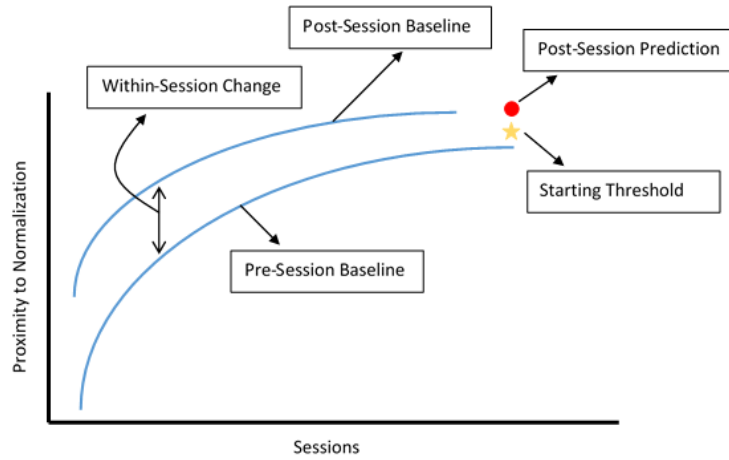
Figure 2: An illustration of how Progressive Thresholding promotes progress across sessions in the idealized case where a participant's target brain activity moves smoothly in the intended direction.

Participant Instructions

We provided participants with no indication of potential strategies which they could use to perform the neurofeedback task. Such information was deliberately withheld in accordance with previous research indicating that a lack of a prescribed mental strategy best leads to automaticity, meaning that FAA control was more likely to become a natural separate willful action (Hardman et al., 1997; Kober, Witte, & Ninaus, 2013; Strehl, 2014). Instead, participants were told that they would have to discover their own strategy for controlling the feedback signal.

Data Analysis

Offline data analysis was done in MATLAB R2015a (MATLAB, 2015). Before conducting statistical tests, baseline or session data were removed if less than 25% of the EEG signal survived artifact rejection. Each statistical test was performed after checking whether the assumptions of that test were valid for the data being analyzed. For tests of differences in means and variances, the Kolmogorov-Smirnov test of normality (Stephens, 1974) was used before using tests which assume normality. All p-values are reported after corrections for multiple comparisons using Holmes-Bonferroni corrections (Holm, 1979) where required.

Results

Prescreening Results

Of the 102 prescreening participants, 78 completed the baseline EEG recording (the remaining 24 were excluded before reaching the EEG recording stage). The EEG hardware used in this study is a novel

apparatus and the OHQ is a novel questionnaire for neurofeedback experiments. For this reason we checked to see whether the OHQ was correlated with FAA and could therefore be used in place of more invasive clinical questionnaires for prescreening purposes. Sixteen data points were removed from analysis because less than two full minutes of clean EEG data were available with which to compute FAA, leaving data from 62 participants. We found that OHQ scores correlated with FAA computed using the two frontal channels used during neurofeedback training ($r(60) = 0.28$, $p = 0.03$) as well as PTAA computing these same channels ($r(60) = -0.28$, $p < 0.03$). Scatter plots of these data are shown in Figures 3 and 4. OHQ scores did not correlate with temporal alpha asymmetry (TAA) or PTAA scores when they were computed using the temporal channels, for which only 45 participants passed the signal quality requirement (TAA: $r(43) = -0.18$, $p = 0.24$, PTAA: $r(43) = 0.19$, $p = 0.21$) or all channels together (global alpha asymmetry: $r(34) = -0.14$, $p = 0.38$, PTAA: $r(34) = 0.14$, $p = 0.37$), suggesting that the correlation is specific to frontal EEG.
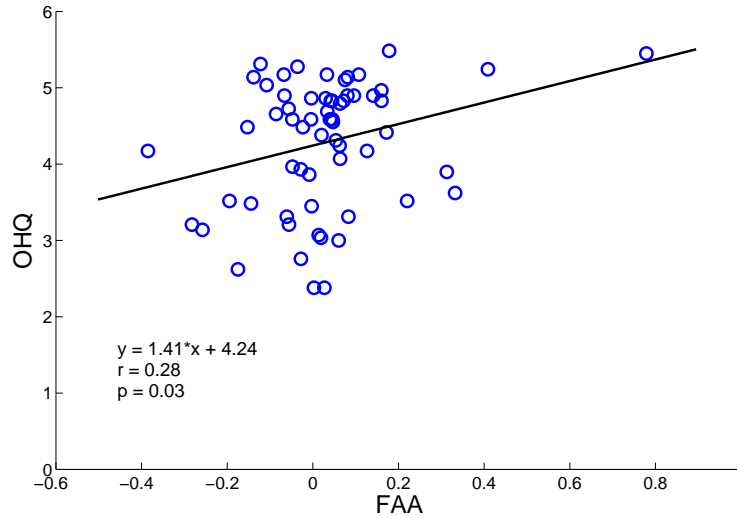
Figure 3: Scatterplot of Frontal FAA vs. OHQ score.

FAA versus PTAA

Statistics computed using PTAA were equivalent in interpretation to statistics computed using FAA and provide no additional information (*i.e.*, hypothesis tests using PTAA always led to the same conclusion, though the exact statistic or p-value might vary slightly). Since participants were trained using FAA rather than PTAA, we omit reporting statistics with PTAA for the remainder of the paper.
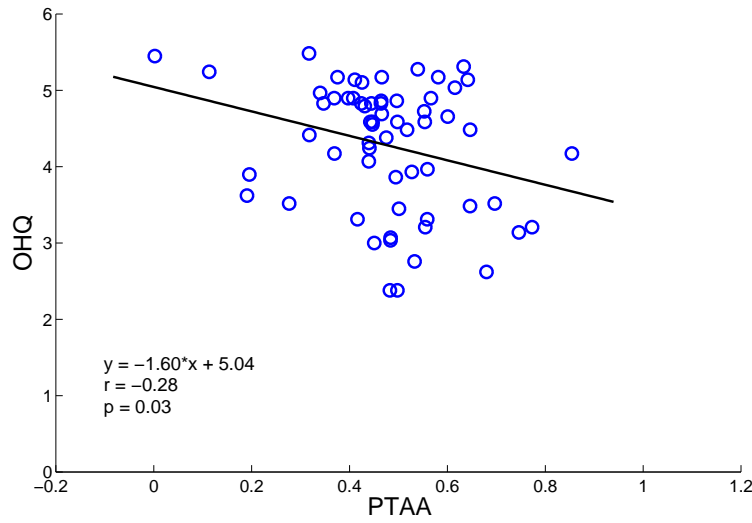
Figure 4: Scatterplot of Frontal PTAA vs. OHQ score.

Neurofeedback Learning Outcomes

Within-Session Changes

We measured the change in FAA scores between each participant's pre-session and post-session baselines. Kolmogorov-Smirnov tests were used to assess whether each variable deviated significantly from normality. The results of these tests are given in Table 1 and show that we fail to reject the hypothesis that each variable is normally distributed. Therefore we used paired samples T-tests in order to assess whether there were significant differences between pre-session and post-session mean FAA within the same group, and we used independent samples T-tests in order to compare means across the two groups .

We performed a two-factor ANOVA in order to assess the independent and interactive effects of time and group on the within-session changes in FAA. This analysis revealed that there was an overall effect of time (pre- versus post-session baseline) on FAA (pre-session baselines: $M = 0.059$, $SD = 0.125$; post-session baselines: $M = 0.029$, $SD = 0.068$; $F(1, 265) = 6.06$, $p = 0.0144$). There was also a significant group effect (ST: $M = 0.058$, $SD = 0.108$; PT: $M = 0.029$, $SD = 0.093$; $F(1, 265) = 5.64$, $p = 0.0183$). However, there was no interaction effect between time and group when analyzing within-session changes ($F(1, 265) = 1.33$, $p = 0.2492$). These data are summarized in Figure 5.

Given the significant effects of time and group, we also performed T-tests in order to further explore those differences. We used one-sided paired samples T-tests to compare pre-session FAA to post-session FAA because successful neurofeedback should lead to lower post-session FAA compared to pre-session FAA, taking into consideration that participants began training with baseline FAA above zero. The ST group showed no difference between pre-session FAA scores and post-session FAA scores (pre-session baseline:

$M = 0.066$, $SD = 0.134$; post-session baseline: $M = 0.050$, $SD = 0.074$; $t(66) = 1.08$, $p = 0.28$). However, a significant reduction in post-baseline FAA scores compared to pre-baseline FAA scores was found for the PT group (pre-session baseline: $M = 0.051$, $SD = 0.116$; post-session baseline: $M = 0.007$, $SD = 0.053$; $t(66) = 2.99$, $p = 0.004$). A two-sided independent samples T-test revealed that the post-session FAA scores were also significantly lower for the PT group versus the ST group ($t(133) = 3.84$, $p = 0.0006$). Since there was no difference in pre-session baselines between the groups ($t(132) = 0.68$, $p = 0.49$), the lower post-session FAA scores in the PT group cannot be attributed to differences in starting FAA scores. All p-values obtained through tests which use at least one variable in common were corrected with Holmes-Bonferroni corrections for multiple comparisons.

Measure	KS Statistic	p
ST Pre-Session FAA	0.188	$p = 0.06$
ST Post-Session FAA	0.120	$p = 0.78$
PT Pre-Session FAA	0.103	$p = 0.56$
PT Post-Session FAA	0.119	$p = 0.78$

Table 1: Kolmogorov-Smirnov test results for the distribution of FAA scores for pre-session and post-session baselines for both ST and PT groups.

As a measure of algorithm reliability, we compared the proportion of sessions for which FAA scores moved towards symmetry using a chi-square test of proportions. The PT group had a greater proportion of successful sessions, or sessions in which FAA moved towards symmetry rather than away from symmetry (ST: 0.582; PT: 0.821; $\chi^2(1) = 10.16$, $p = 0.001$). The proportion of successful sessions in the ST group was not significantly higher than the chance level of 0.5 ($\chi^2(1) = 1.21$, $p = 0.27$), but the PT group did have a greater proportion of successful sessions compared to the chance level of 0.5 ($\chi^2(1) = 27.60$, $p < 10^{-6}$).

We also tested whether there were reductions in the variance of FAA scores using one-sided two-sample F-tests for equal variances. Post-baseline FAA scores had a significantly lower variance compared to pre-session FAA scores in both the ST group ($F(66, 67) = 3.23$, $p < 10^{-5}$). and the PT group ($F(66, 66) = 4.74$, $p < 10^{-9}$). Two-sided F-tests were used to compare ST and PT groups in order to avoid the assumption that one group should have a reduced variance compared to the other. There was no difference in the variance of pre-session FAA scores between the ST and PT groups ($F(66, 66) = 1.32$, $p = 0.26$). However, the PT group had a significantly reduced variance in post-session FAA scores compared to the ST group ($F(67, 66) = 1.94$, $p = 0.008$). This can easily be seen in Figure 5.

There were no differences in FAA scores between sessions where participants reported that external factors led to a more positive mood versus a more negative mood (pre-session FAA scores: $p = 0.40$, post-session FAA scores: $p = 0.78$). There was also no correlation between the score on the TIM and pre-session FAA scores ($p = 0.78$) or post-session FAA scores ($p = 0.25$). However, TIM scores were significantly higher for the PT group ($M = 13.3$, $SD = 1.9$) than the ST group ($M = 11.4$, $SD = 2.2$) ($t(131) = -4.96$, $p < 10^{-5}$). No difference was found in the proportion of sessions for which positive (ST = 0.03; PT = 0.08; $\chi^2(1) = 0.16$, $p = 0.69$) or negative (ST: 0.15; PT: 0.17; $\chi^2(1) = 0.076$, $p = 0.78$) moods was reported.
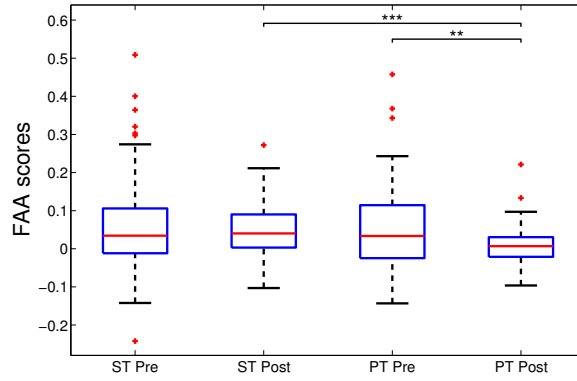
Figure 5: Distribution of FAA scores for pre-session baselines and post-session baselines for both groups. Distributions include all sessions belonging to each category.
$**$ denotes a significant difference in the distribution means with $p < 0.005$.
$***$ denotes a significant difference in the distribution means with $p < 0.0005$.

Changes Across Training

An average of 11.4 ($SD = 4.7$) sessions were completed (PT: $M = 11.50$, $SD = 3.6$; ST: $M = 11.33$, $SD = 6.0$). In order to rule out the above results being influenced by differences in the amount of training, we assessed the correlation between the number of completed sessions and the change in FAA scores across training. No significant correlations were found between change in FAA scores across training and the number of sessions completed ($p = 0.59$ for pre-session baselines, and $p = 0.89$ for post-session baselines). This was true for the ST group ($p = 0.65$ for pre-session baselines, and $p = 0.92$ for post-session baselines) and the PT group ($p = 0.56$ for pre-session baselines, and $p = 0.66$ for post-session baselines) when their data were observed separately.

In order to measure the overall effect of neurofeedback training, we compared FAA scores between the first and last sessions. The results of Kolmogorov-Smirnov tests performed on each of these variables are reported in Table 2 and justify the use of T-tests when comparing their means. Pre-session and post-session baselines were compared separately in order to minimize the influence of within-session effects on these analyses.

A two-factor ANOVA comparing the first pre-session baseline to the last pre-session baseline across ST and PT groups revealed no significant group difference ($F(1, 20) = 2.60$, $p = 0.12$) and no time difference ($F(1, 20) = 1.77$, $p = 0.20$). However, the interaction between group and time was significant ($F(1, 20) = 16.48$, $p = 0.0006$). Comparing post-session baselines revealed no significant group difference ($F(1, 20) = 1.21$, $p = 0.28$), time difference ($F(1, 20) = 0.26$, $p = 0.61$), or interaction effect ($F(1, 20) = 0.01$, $p = 1.0$). These data are summarized in Figure 6.

A one-tailed paired samples T-test revealed no change between the first pre-session baseline FAA scores

Measure	KS Statistic	p
ST First Pre-Session FAA	0.29	$p = 1.0$
ST First Post-Session FAA	0.20	$p = 1.0$
ST Last Pre-Session FAA	0.35	$p = 1.0$
ST Last Post-Session FAA	0.27	$p = 1.0$
PT First Pre-Session FAA	0.33	$p = 1.0$
PT First Post-Session FAA	0.24	$p = 1.0$
PT Last Pre-Session FAA	0.28	$p = 1.0$
PT Last Post-Session FAA	0.26	$p = 1.0$

Table 2: Kolmogorov-Smirnov test results for the distribution of FAA scores for the first and last pre-session and post-session baselines for both ST and PT groups.

and the last pre-session baseline FAA scores in the ST group ($t(5) = -1.43$, $p = 0.89$). Likewise, there was no significant change in the ST group when comparing first and last post-session baselines ($t(5) = 0.42$, $p = 0.35$). In contrast, there was a significant reduction in FAA scores from the first pre-session baseline to the last pre-session baseline for the PT group ($t(5) = 4.47$, $p = 0.0033$), though no change was found in the post-session baselines ($t(5) = 0.30$, $p = 0.39$). Two-tailed Independent-samples T-tests revealed that the PT group had significantly lower FAA scores in the last pre-session baseline compared to the ST group ($t(10) = 2.43$, $p = 0.035$), but there was no difference in the last post-session baseline ($t(10) = 0.80$, $p = 0.44$). All p-values obtained through tests which use at least one variable in common were corrected with Holmes-Bonferroni corrections for multiple comparisons.

We also compared the variance in FAA scores between the first and last session as well as between groups using the two-sample F-test for equal variance. There was no change in variance between the first and last session FAA scores for the ST group when comparing pre-session baselines ($F(5,5) = 1.57$, $p = 0.63$) and a trend towards a difference when comparing post-session baselines ($F(5,5) = 0.15$, $p = 0.059$). A significant difference in variance from the first to the last session was found in the PT group for both pre-session FAA scores ($F(5,5) = 7.84$, $p = 0.041$) and post-session FAA scores ($F(5,5) = 66.42$, $p = 0.0003$). A trend towards a difference in variance was observed between the ST group and the PT group when comparing the last pre-session baselines ($F(5,5) = 5.53$, $p = 0.084$), and the difference was significant for the last post-session baselines ($F(5,5) = 7.70$, $p = 0.043$).

All six participants from the PT group successfully moved their pre-session baseline FAA scores towards zero versus only four participants from the ST group. While the difference in proportion only tended towards significance, ($\chi^2(1) = 3.00$, $p = 0.08$), additional data may lead to a significant result.

Changes in Mood Inventory Scores

Two participants discontinued participation in the study without notification. As a result, mood inventory data for all twelve participants were available for the first neurofeedback session, but data from only ten participants were available for the last session. The changes in FAA scores between the first and last training sessions and the changes in mood inventory scores are summarized below. No correlations were found between
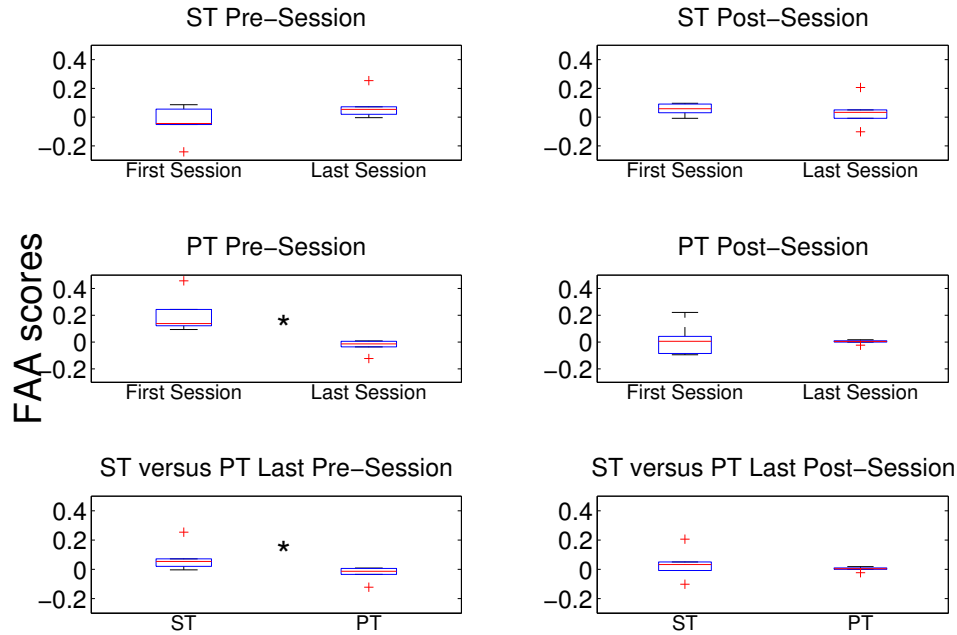
Figure 6: Comparison between FAA scores from the first neurofeedback training session and the last neurofeedback training session. a) Pre-session baseline FAA during the first training session versus the last training session for the ST group. b) Post-session baseline FAA during the first training session versus the last training session for the ST group. c) Pre-session baseline FAA during the first training session versus the last training session for the PT group. d) Post-session baseline FAA during the first training session versus the last training session for the PT group. e) Pre-baseline FAA from the last session for the ST group and the PT group. f) Post-baseline FAA from the last session for the ST group and the PT group. ∗ denotes a significant difference in the distributions ($p < 0.05$).

changes in pre-session or post-session FAA scores and scores on any of the mood inventories.

We observed the changes in mood inventory scores before and after neurofeedback training for those participants who had elevated stress, depression, or anxiety scores at baseline. These data are shown in Figure 7. Pre and post scores are shown for the BDI (including only participants who initially scored 20 or higher, indicating moderate or severe depression), the BAI (including only participants who initially scored 19 or higher, indicating moderate or severe anxiety) and the PSS (including only participants who initially scored 20 or higher, indicating high stress). For these participants, scores either decreased or remained consistent depending on the participant. No difference was observed between ST and PT participants. These data suggest that studies testing FAA neurofeedback with clinically depressed or anxious participants, or participants with high stress, may show changes in mood inventory scores related to neurofeedback training. However, our study, which aimed only to test differences in learning outcomes between ST and PT irrespective of clinical status, did not find statistically significant differences.
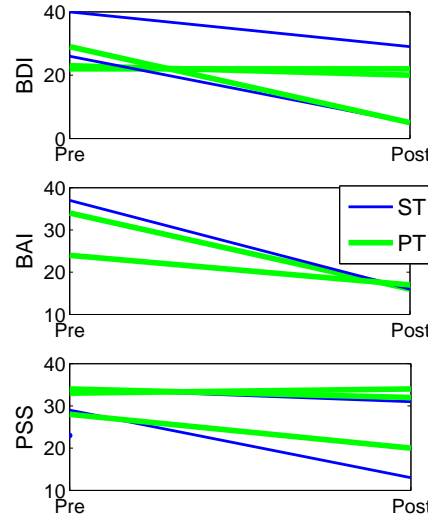
Figure 7: Changes in mood inventory scores for only those participants who had elevated stress, depression, or anxiety scores at baseline.

We conducted an exploratory analysis to determine whether any items in the three mood inventories used were correlated with changes in FAA scores. After correction for multiple comparisons, only one item of the BDI, related to poor sleep, was found to correlate with change in FAA scores with nearly statistical significance ($r(10) = 0.83$, $p = 0.06$).

Training Questionnaire

Subjective reports of task difficulty were correlated with how much participants were able to reduce their FAA scores between their first and last sessions ($r(9) = 0.78$, $p = 0.006$), with participants who rated the task as more difficult performing better. The PT group also trended towards reporting higher task difficulty despite performing better compared to the ST group (ST task difficulty: $M = 4.6$, $SD = 1.9$; PT task difficulty: $M = 6.5$, $SD = 1.6$; $t(9) = -1.76$, $p = 0.056$). This may have driven the relationship between perceived task difficulty and performance, since the PT algorithm is designed to provide more difficult training, but also led to better learning outcomes in this study. Interestingly, change in FAA scores did not correlate with subjective self-evaluations of how well participants were able to control the feedback cursor ($r(10) = 0.19$, $p = 0.91$), suggesting that participants were not able to evaluate their own performance accurately.

Discussion

Neurofeedback Training and the Impact of Progressive Thresholding

The goal of this study was to test whether the PT algorithm would lead to better learning outcomes compared to the ST algorithm, which is the current standard in automatic thresholding. In order to do this, we trained 12 participants to balance the alpha power recorded at two prefrontal EEG electrodes (FAA) over several sessions using a commercial EEG headband.

In the current study, the PT group completed training with significantly lower FAA scores compared to the ST group despite the PT group beginning with a higher average FAA score (see Figure 6) and despite the PT group reporting higher subjective task difficulty compared to the ST group. There was a significant decrease in the average FAA scores for the PT group, resulting in FAA scores slightly below zero. Though the change in FAA scores was non-significant in the ST group, this could have been due to the ST group beginning training closer to an average FAA score of zero rather than the ST algorithm failing to promote correct change. The PT group also trended towards a smaller variance in FAA scores when comparing their final pre-session and post-session baselines to their first pre-session and post-session baselines. Thus we can conclude that on the basis of our sample, neurofeedback with PT was more effective in training participants to reduce their FAA scores than was neurofeedback with ST. Combined with results of the two-factor ANOVA, the data suggest that the effect of PT is reflected as improved training over time, which is expected and desired for a neurofeedback training algorithm.

A commonly cited limitation in automatic neurofeedback training is that the lack of shaping and the fixed flat reward rate is believed to be associated with the observation that trainees often fail to change their brain activity correctly, or even erroneously change their brain activity in the wrong direction (Vollebregt et al., 2014; Pigott & Cannon, 2014). It is extremely important to develop methods which rectify this issue because incorrect neurofeedback training can lead to detrimental side-effects (Hammond & Kirk, 2007). In the current study the PT group showed a greater proportion of sessions in which FAA scores were reduced compared to ST (0.821 for PT versus 0.582 for ST), though we note that this statistic is limited in its interpretability because baseline FAA scores are meant to change across sessions. However, we also compared the groups with respect to their changes in FAA scores across training. Given the clear trend for a greater proportion of PT participants completing training with FAA scores having moved in the desired direction (6/6 PT participants versus 4/6 ST participants), we expect that with more participants the difference will be statistically significant. These results suggests that PT more reliably results in correct neurofeedback training, thereby reducing the risk of erroneous changes in brain activity and unintended effects.

There has been extensive debate over automatic thresholding, with many prominent neurofeedback experts arguing that automatic thresholding should not be used because it does not apply any shaping mechanism (Sherlin et al., 2011; Arns et al., 2014; Strehl, 2014). After Lansbergen's group introduced automatic thresholding (Lansbergen et al., 2011), they reverted back to manual thresholding methods but noted that

in a small sample, there was no differnece between manual thresholding and automatic thresholding (van Dongen-Boomsma et al., 2013). However, as Pigott explains (Pigott & Cannon, 2014), there may have been important methodological flaws which led to this conclusion, because instead of using a manual thresholding method with shaping, they simulated automatic thresholding with manual thresholding (*i.e.*, the clinician aimed to provide the same flat 80% reward rate that was used in the automatic thresholding algorithm). The PT algorithm is an example of a new kind of automatic thresholding algorithm which incorporates shaping by introducing elements of statistical modelling and individualization. This study suggests that the problem with automatic thresholding is not that it is automatic, as one might assume from reading the neurofeedback literature, but that a flat reward rate, however it is administered, fails to convey any useful training signal that could specifically promote progress towards the goal.

The current study should change the debate over automatic reward thresholding in neurofeedback, as we have demonstrated with Progressive Thresholding that an automatic algorithm can in principle simulate what a clinician might do manually. We note that the PT algorithm presented here is not the only way PT could be implemented, and that it is not necessarily the best way to implement PT. Instead, PT is really a class of neurofeedback reward thresholding algorithms defined by its use of individualized adaptive statistical models of training progress. The version presented here is one of the most simple implementations of PT, using only very basic modeling techniques. In addition, PT, like ST, requires no *a priori* information about the specific neurofeedback protocol or target, other than whether the goal is to increase the measure provided, decrease it, or move it to some specific level. For the purpose of this study, we sought only to demonstrate that it was possible to design an automatic reward thresholding algorithm which incorporated a shaping mechanism and individualization, which we hope contributes to the development of safe, automatic, affordable, reliable, and effective neurofeedback-based treatment options in neuropsychiatry. In future work, we aim to validate the Progressive Thresholding algorithm on other neurofeedback protocols with larger sample sizes, and in particular, compare it to clinician-controlled manual thresholding. In addition, we intend to experiment with more advanced machine learning modeling methods in order to improve the efficacy of PT.

Measuring FAA

An issue with FAA neurofeedback which does not often receive sufficient attention in the literature is the high degree of instability of the FAA score. In our study, the average standard deviation of the FAA score across participants was $M = 0.82$, $SD = 0.02$, combining pre-session and post-session baselines. Additionally, the standard deviation of the average FAA scores across sessions per participant was $M = 0.07$, $SD = 0.04$. These values are very high considering the range of FAA scores shown in Figure 6 and when considering the amount of change in FAA scores observed between the first and last sessions. An alternative method of computing FAA may be required to address the instability of the current FAA formulation.

Baehr, Rosenfeld and Baehr (Baehr et al., 2001) report this variability as their reason for introducing PTAA as an alternative measure for FAA. However, in our study, analyzing PTAA led to the same results

as analyzing FAA. More importantly, participants are usually trained with FAA rather than PTAA because of the loss of amplitude-related information when computing PTAA from FAA. Reporting results using PTAA, while perhaps preserving the conclusions being drawn, is inconsistent with respect to the measure with which participants are actually trained. Furthermore, PTAA may not be a reliable replacement for FAA because the loss of degree of asymmetry means it is possible to score low on PTAA with an average FAA score suggesting that one is extremely asymmetric (*e.g.*, if a patient's FAA score followed a positively skewed normal distribution with a median less than zero). One possible alternative to the FAA computation used here is to compute FAA using the individual alpha peak frequencies (IAF) (Doppelmayr, Klimesch, Pachinger, & Ripper, 1998; Klimesch, 1999; Quaedflieg et al., 2015). The IAF is considered to be a relatively stable measure over time (van Boxtel et al., 2012), suggesting that computing FAA using the IAF may reduce noise in the FAA signal and may result in more impactful neurofeedback training. This option should be explored further in future work.

The instability of FAA is perhaps most problematic when considering moment to moment changes in the FAA score. As we noted in the Methods section, extensive smoothing was required in order to present a smoothly varying FAA score for feedback. Without such smoothing, the FAA score can appear indistinguishable from pure noise. Even with the extensive smoothing done in this study, the FAA score can vary enormously. We took a random neurofeedback session from our study and computed the FAA score in one-second time windows. The FAA score as it is typically reported in the literature is displayed in the top plot of Figure 8, while the same FAA score with our smoothing function applied to it is displayed in the bottom plot. It is clear that without extensive smoothing, this participant would have not have been presented an information-conveying feedback signal. However, details explaining whether similar smoothing was used for other studies or whether FAA was calculated over wider time windows is often omitted from other studies (*e.g.*, (Rosenfeld et al., 1995; Choi et al., 2010; Kerson, Sherman, & Kozlowski, 2009), though such information is included in some studies, *e.g.*, (Peeters et al., 2014)). Therefore, we cannot say whether our feedback presentation method, and thus our results, are easily comparable to previous studies using FAA neurofeedback with EEG. More importantly, if similar smoothing was not done in some FAA neurofeedback studies, we might expect that participants in those studies would have had poor learning outcomes. We recommend that precise methods of computing FAA and any smoothing operations applied to generate feedback signals be reported in full detail in future reports.

Pre-screening Data

A moderate correlation was found between OHQ scores and FAA scores (see Figure 3), suggesting that the OHQ might be useful as a prescreening tool or as a less invasive substitute for the BDI in some situations. However, the correlation was in the opposite of the expected direction given what is known about the association between left-dominant FAA and depression. We suspect that this may be caused by the placement of the reference electrode on the Muse headband, which is half way between Fp1 and Fp2. Given the close
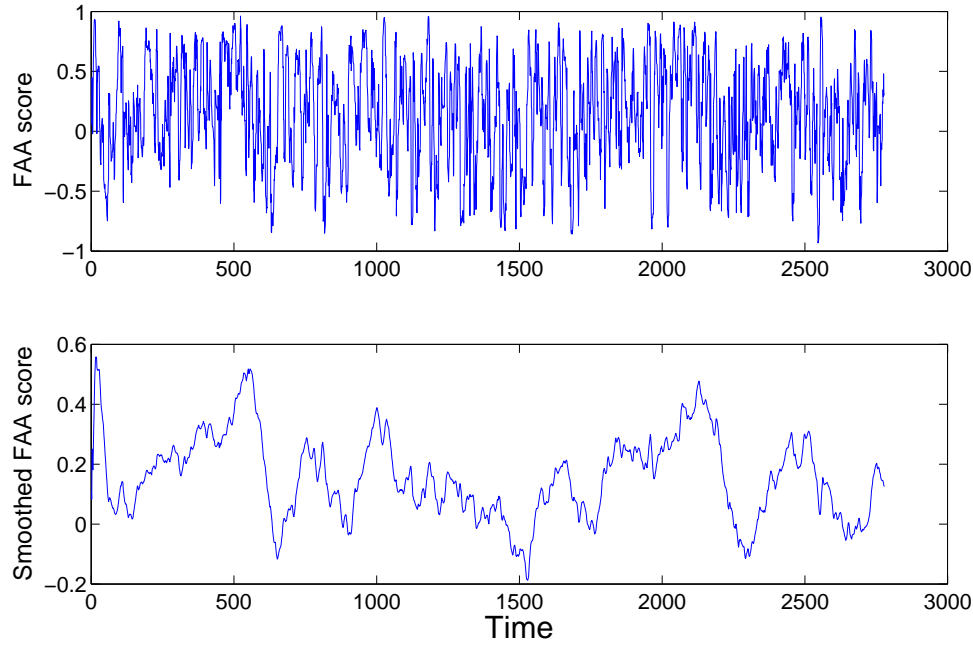
Figure 8: Plots of the FAA score as it is typically reported in the literature (top) versus the smoothed FAA score we used to provide feedback (bottom).

proximity of the reference to these electrodes, it is possible that the reference picks up some of the field potentials centered at Fp1 and Fp2. If this is the case, when the reference signal is subtracted from each electrode some fraction of the signals from Fp1 and Fp2 may be swapped, potentially resulting in sign reversal of our FAA score.

Regardless of the precise reason for the reversal of correlation between OHQ scores and FAA scores, the fact that the correlation is in the opposite from the expected direction implies that we should not necessarily expect a reduction in FAA to correspond to a reduction in BDI, BAI, or PSS scores in this study. Further study on the properties of the Muse headband may be required in order to determine how best to translate current neurofeedback protocols for this device. However, with more recent software compatible with the Muse headband, the signals can be re-referenced, for example to the average of TP9 and TP10. This may resolve the unexpected result of the reversed correlation between OHQ scores and FAA scores.

Mood Inventories

Our study did not find meaningful relationships between FAA scores and the scores of mood inventories. Only one result, the relationship between FAA change in the post-session baselines of the PT group and the change in BAI scores, was found to be potentially significant, but after correction for multiple comparisons, the

statistic was no longer significant. While these results do not support the hypothesis that FAA neurofeedback is useful for treating depression, we also do not take them as counter-evidence for several reasons.

Since our sample was not restricted to a clinical population, it is not surprising that no significant changes in mood inventory scores were observed. In Figure 7, we plotted pre-post changes in mood inventory scores for just those individuals who score high enough to indicate potential clinical pathology. An overall pattern of stagnant or decreasing mood inventory scores for these individuals suggests that a clinical study might result in significant clinical improvements with FAA neurofeedback. While these changes could also be due to external factors in each participants' life, the fact that there was no worsening of mood scores is encouraging. Whether the changes can be attributed to neurofeedback can not be determined in our study, as we did not aim to test FAA neurofeedback in a clinical population.

Measuring changes in mood scores across time points in undergraduate students may be particularly unreliable due to fluctuations in stress levels as students enter periods of higher or lower work loads. In the current study, we did not ask students about their work loads, school-related deadlines, exam schedules, or personal lives, and different participants began and finished the study at different periods in the school term. In previous work, it has been shown that BDI, BAI, and PSS levels are sensitive to school-related stressors. Therefore, it is possible that school-related stressors influenced the results of the BDI, BAI, and PSS scores.

Finally, we used a commercial EEG headband due to the growing interest in commercial and low-cost neurofeedback treatment options. However, the commercial headband we used did not have electrodes at F3 or F4, which are commonly used to measure FAA. Furthermore, the headband did not have an EEG electrode at Cz and did not have enough electrodes to compute a reliable average reference. FAA is thought to be best measured using an average reference or a reference at Cz (Davidson, 1998; Baehr et al., 1997; Rosenfeld et al., 1995). Instead we used the default reference provided by the Muse headband, which is between Fp1 and Fp2. Given that our electrode configuration differed significantly from what is typically used in the literature, it is possible that our measure of FAA did not carry the same mood-related information as FAA measures used in other studies.

Given the above, the lack of clinically significant effects in the present study does not rule out the utility of our approach for clinical populations. Determining whether PT is helpful in a clinical context would require further study.

Conclusions and Future Work

Here we presented Progressive Thresholding, a new automatic reward thresholding algorithm which incorporates shaping in automated neurofeedback training in order to improve learning outcomes. We showed that participants trained with Progressive Thresholding had greatly improved learning outcomes compared to participants trained with the standard automatic reward thresholding method. In particular, participants trained with Progressive Thresholding were more successful in balancing their frontal alpha activity and

showed a greater proportion of successful neurofeedback sessions. Additionally, only those trained with Progressive Thresholding showed across-session training effects whereby participants came to their final sessions already having balanced frontal alpha activity.

In future work, Progressive Thresholding should be validated for different neurofeedback protocols and with different EEG hardware to ensure that the benefits of this new automatic thresholding method is robust across different neurofeedback contexts. Progressive Thresholding should also be tested against manual reward thresholding in order to determine how it compares to neurofeedback with thresholding controlled by a clinical expert. Finally, Progressive Thresholding should be validated in a clinical population in order to assess whether the improvements in learning outcomes translate into improvements in clinical outcomes.

Compliance with Ethical Standards

References

Allen, J. J., Harmon-Jones, E., & Cavender, J. H. (2001). Manipulation of frontal EEG asymmetry through biofeedback alters self-reported emotional responses and facial EMG. *Psychophysiology*, *38*(4), 685–693. doi: 10.1111/1469-8986.3840685

Angelakis, E., Stathopoulou, S., Frymiare, J. L., Green, D. L., Lubar, J. F., & Kounios, J. (2007). EEG neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly. *The Clinical neuropsychologist*, *21*(1), 110–129. doi: 10.1080/13854040600744839

Arnold, L. E., Lofthouse, N., Hersch, S., Pan, X., Hurt, E., Bates, B., ... Grantier, C. (2012). Eeg neurofeedback for adhd: double-blind sham-controlled randomized pilot feasibility trial. *Journal of attention disorders*, 1087054712446173.

Arns, M., de Ridder, S., Strehl, U., Breteler, M., & Coenen, A. (2009). Efficacy of neurofeedback treatment in ADHD: the effects on inattention, impulsivity and hyperactivity: a meta-analysis. *Clinical EEG and neuroscience : official journal of the EEG and Clinical Neuroscience Society (ENCS)*, *40*, 180–189. doi: 10.1177/155005940904000311

Arns, M., Heinrich, H., & Strehl, U. (2014). Evaluation of neurofeedback in adhd: the long and winding road. *Biological psychology*, *95*, 108–115.

Arns, M., & Kenemans, J. L. (2014). Neurofeedback in ADHD and insomnia: Vigilance stabilization through sleep spindles and circadian networks. *Neuroscience and Biobehavioral Reviews*, *44*, 183–194. Retrieved from `http://dx.doi.org/10.1016/j.neubiorev.2012.10.006` doi: 10.1016/j.neubiorev .2012.10.006

Baehr, E., Rosenfeld, J. P., & Baehr, R. (1997). The clinical use of an alpha asymmetry protocol in the neurofeedback treatment of depression: Two case studies. *Journal of neurotherapy*, *2*(3), 10–23.

Baehr, E., Rosenfeld, J. P., & Baehr, R. (2001). Clinical Use of an Alpha Asymmetry Neurofeedback Protocol in the Treatment of Mood Disorders. *Journal of Neurotherapy*, *4*(4), 11–18. doi: 10.1300/ J184v04n04_03

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of consulting and clinical psychology*, *56*(6), 893.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuing depression. *Archives of General Psychiatry*, *4*, 561–571.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Cassani, R., Banville, H., & Falk, T. H. (2015). Mules: An open source eeg acquisition and streaming server for quick and simple prototyping and recording. In *Proceedings of the 20th international conference on intelligent user interfaces companion* (pp. 9–12).

Choi, S. W., Chi, S. E., Chung, S. Y., Kim, J. W., Ahn, C. Y., & Kim, H. T. (2010). Is alpha wave neurofeedback effective with randomized clinical trials in depression? A pilot study. *Neuropsychobiology*, *63*, 43–51. doi: 10.1159/000322290

Choi, S. W., Chi, S. E., Chung, S. Y., Kim, J. W., Ahn, C. Y., & Kim, H. T. (2011). Is alpha wave neurofeedback effective with randomized clinical trials in depression? a pilot study. *Neuropsychobiology*, *63*(1), 43–51.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of health and social behavior*, 385–396.

Davidson, R. J. (1998). Anterior electrophysiological asymmetries, emotion, and depression: conceptual and methodological conundrums. *Psychophysiology*, *35*(5), 607–614. doi: 10.1017/S0048577298000134

Debener, S., Beauducel, A., Nessler, D., Brocke, B., Heilemann, H., & Kayser, J. r. (2000). Is resting anterior eeg alpha asymmetry a trait marker for depression? *Neuropsychobiology*, *41*(1), 31–37.

DeBeus, R., & Kaiser, D. (2011). Neurofeedback with children with attention deficit hyperactivity disorder: A randomized doubleblind placebo-controlled study. *Neurofeedback and neuromodulation techniques and applications*, 127–152.

Dhindsa, K. (2017). Filter-Bank Artifact Rejection: High Performance Real-Time Single-Channel Artifact Detection for EEG. *In Press: Biomedical Signal Processing and Control*.

Doppelmayr, M., Klimesch, W., Pachinger, T., & Ripper, B. (1998). Individual differences in brain dynamics: important implications for the calculation of event-related band power. *Biological cybernetics*, *79*(1), 49–57.

Ghaziri, J., Tucholka, A., Larue, V., Blanchette-Sylvestre, M., Reyburn, G., Gilbert, G., ... Beauregard, M. (2013). Neurofeedback training induces changes in white and gray matter. *Clinical EEG and neuroscience*, 1550059413476031.

Hammond, D. C. (2005). Neurofeedback treatment of depression and anxiety. *Journal of Adult Development*, *12*(2), 131–137.

Hammond, D. C., & Kirk, L. (2007). Negative effects and the need for standards of practice in neurofeedback. *Biofeedback*, *35*(4).

Hardman, E., Gruzelier, J., Cheesman, K., Jones, C., Liddiard, D., Schleichert, H., & Birbaumer, N. (1997). Frontal interhemispheric asymmetry: self regulation and individual differences in humans. *Neuroscience Letters*, *221*(2), 117–120.

Harmon-Jones, E. (2003). Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology*, *40*(6), 838–848.

Harmon-Jones, E., Gable, P. A., & Peterson, C. K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological psychology*, *84*(3), 451–462.

Heinrich, H., Gevensleben, H., & Strehl, U. (2007). Annotation: Neurofeedback - Train your brain to train behaviour. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*, 3–16. doi: 10.1111/j.1469-7610.2006.01665.x

Henkel, V., Mergl, R., Coyne, J. C., Kohnen, R., Möller, H. J., & Hegerl, U. (2004). Screening for depression in primary care: Will one or two items suffice? *European Archives of Psychiatry and Clinical Neuroscience*, *254*(4), 215–223. doi: 10.1007/s00406-004-0476-3

161

Henriques, J. B., & Davidson, R. J. (1991). Left frontal hypoactivation in depression. *Journal of abnormal psychology*, *100*(4), 535.

Hills, P., & Argyle, M. (2002). The oxford happiness questionnaire: A compact scale for the measurement of psychological well-being. *Personality and individual differences*, *33*(7), 1073–1082.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Interaxon. (2014). *Muse.* www.choosemuse.com.

Interaxon. (2015). Muse: Technical specifications, validation, and research use [Computer software manual]. Retrieved from \url{https://storage.googleapis.com/ix_choosemuse/uploads/2015/07/tech_specs.pdf}

Kerson, C. (2013). A proposed multisite double-blind randomized clinical trial of neurofeedback for ADHD: need, rationale, and strategy. *Journal of attention disorders*, *17*(5), 420–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23590978 doi: 10.1177/1087054713482580

Kerson, C., Sherman, R. a., & Kozlowski, G. P. (2009). Alpha Suppression and Symmetry Training for Generalized Anxiety Symptoms. *Journal of Neurotherapy*, *13*(3), 146–155. doi: 10.1080/10874200903107405

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

Kluetsch, R. C., Ros, T., Théberge, J., Frewen, P. A., Calhoun, V. D., Schmahl, C., ... Lanius, R. A. (2014). Plastic modulation of PTSD resting-state networks and subjective wellbeing by EEG neurofeedback. *Acta Psychiatrica Scandinavica*, *130*(2), 123–136.

Kober, S., Witte, M., & Ninaus, M. (2013). Learning to modulate one's own brain activity: the effect of spontaneous mental strategies. *Frontiers in Human Neuroscience*, *7*(October), 695. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3798979\&tool=pmcentrez\&rendertype=abstract$\backslash$nhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC3798979/ doi: 10.3389/fnhum.2013.00695

Lansbergen, M. M., van Dongen-Boomsma, M., Buitelaar, J. K., & Slaats-Willemse, D. (2011). Adhd and eeg-neurofeedback: a double-blind randomized placebo-controlled feasibility study. *Journal of neural transmission*, *118*(2), 275–284.

Lee, C., Chin, J., Yi, L., Lee, B.-S., & McKeown, M. J. (2015). Feasibility analysis and adaptive thresholding for mobile applications controlled by eeg signals. In *Signal processing conference (eusipco), 2015 23rd european* (pp. 2416–2420).

Lévesque, J., Beauregard, M., & Mensour, B. (2006). Effect of neurofeedback training on the neural substrates of selective attention in children with attention-deficit/hyperactivity disorder: A functional magnetic resonance imaging study. *Neuroscience Letters*, *394*(3), 216–221. doi: 10.1016/j.neulet.2005.10.100

Lewis, R. S., Weekes, N. Y., & Wang, T. H. (2007). The effect of a naturalistic stressor on frontal EEG asymmetry, stress, and health. *Biological psychology*, *75*(3), 239–247.

Lofthouse, N., Arnold, L. E., & Hurt, E. (2012). Current status of neurofeedback for attention-deficit/hyperactivity disorder. *Current Psychiatry Reports*, *14*(5), 536–542. doi: 10.1007/s11920-012-0301-z

MATLAB. (2013). *version 8.2.0 (r2013b)*. Natick, Massachusetts: The MathWorks Inc.

MATLAB. (2015). Matlab version 8.5.0.197613 (r2015a) [Computer software manual]. Natick, Massachusetts.

Peeters, F., Ronner, J., Bodar, L., van Os, J., & Lousberg, R. (2014). Validation of a neurofeedback paradigm: manipulating frontal EEG alpha-activity and its impact on mood. *International Journal of Psychophysiology*, *93*(1), 116–120.

Perreau-Linck, E., Lessard, N., Lévesque, J., & Beauregard, M. (2010). Effects of neurofeedback training on inhibitory capacities in adhd children: a single-blind, randomized, placebo-controlled study. *Journal of Neurotherapy*, *14*(3), 229–242.

Pigott, H. E., & Cannon, R. (2014). Neurofeedback requires better evidence of efficacy before it should be considered a legitimate treatment for adhd: What is the evidence for this claim? *NeuroRegulation*, *1*(1), 25.

Quaedflieg, C., Smulders, F., Meyer, T., Peeters, F., Merckelbach, H., & Smeets, T. (2015). The validity of individual frontal alpha asymmetry eeg neurofeedback. *Social cognitive and affective neuroscience*, nsv090.

Ros, T., Munneke, M. a. M., Ruge, D., Gruzelier, J. H., & Rothwell, J. C. (2010). Endogenous control of waking brain rhythms induces neuroplasticity in humans. *European Journal of Neuroscience*, *31*(4), 770–778. doi: 10.1111/j.1460-9568.2010.07100.x

Rosenfeld, J. P., Cha, G., Blair, T., & Gotlib, I. H. (1995). Operant (biofeedback) control of left-right frontal alpha power differences: Potential neurotherapy for affective disorders. *Biofeedback and Self-Regulation*, *20*(3), 241–258.

Scheinost, D., Stoica, T., Saksa, J., Papademetris, X., Constable, R. T., Pittenger, C., & Hampson, M. (2013). Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity. *Transl Psychiatry*, *3*(4), e250. Retrieved from `http://dx.doi.org/10.1038/tp.2013.24` doi: 10.1038/tp.2013.24

Sherlin, L. H., Arns, M., Lubar, J., Heinrich, H., Kerson, C., Strehl, U., & Sterman, M. B. (2011, October). Neurofeedback and Basic Learning Theory: Implications for Research and Practice. *Journal of Neurotherapy*, *15*(4), 292–304. Retrieved from `http://www.tandfonline.com/doi/abs/10.1080/10874208.2011.623089` doi: 10.1080/10874208.2011.623089

Simkin, D. R., Thatcher, R. W., & Lubar, J. (2014). Quantitative EEG and Neurofeedback in Children and Adolescents. Anxiety Disorders, Depressive Disorders, Comorbid Addiction and Attention-deficit/Hyperactivity Disorder, and Brain Injury. *Child and Adolescent Psychiatric Clinics of North America*, *23*(3), 427–464. Retrieved from `http://dx.doi.org/10.1016/j.chc.2014.03.001` doi: 10.1016/j.chc.2014.03.001

Steenhuis, R. E., & Bryden, M. (1989). Different dimensions of hand preference that relate to skilled and unskilled activities. *Cortex*, *25*(2), 289–304.

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, *69*(347), 730–737.

Stewart, J. L., Coan, J. A., Towers, D. N., & Allen, J. J. (2014). Resting and task-elicited prefrontal eeg alpha asymmetry in depression: Support for the capability model. *Psychophysiology*, *51*(5), 446–455.

Strehl, U. (2014). What learning theories can teach us in designing neurofeedback treatments. *Frontiers in Human Neuroscience*, *8*(November), 1–8. Retrieved from `http://journal.frontiersin.org/article/10.3389/fnhum.2014.00894/abstract` doi: 10.3389/fnhum.2014.00894

Thatcher, R. W. (2000). Eeg operant conditioning (biofeedback) and traumatic brain injury. *Clinical EEG and Neuroscience*, *31*(1), 38–44.

van Boxtel, G. J., Denissen, A. J., Jäger, M., Vernon, D., Dekker, M. K., Mihajlović, V., & Sitskoorn, M. M. (2012). A novel self-guided approach to alpha activity training. *International journal of psychophysiology*, *83*(3), 282–294.

van Dongen-Boomsma, M., Vollebregt, M. A., Slaats-Willemse, D., & Buitelaar, J. K. (2013). A randomized placebo-controlled trial of electroencephalographic (eeg) neurofeedback in children with attention-deficit/hyperactivity disorder. *The Journal of clinical psychiatry*, *74*(8), 821–827.

Vollebregt, M. A., Dongen-Boomsma, M., Buitelaar, J. K., & Slaats-Willemse, D. (2014). Does eeg-neurofeedback improve neurocognitive functioning in children with attention-deficit/hyperactivity disorder? a systematic review and a double-blind placebo-controlled study. *Journal of Child Psychology and Psychiatry*, *55*(5), 460–472.

Wyrwicka, W., & Sterman, M. B. (1968). Instrumental conditioning of sensorimotor cortex EEG spindles in the waking cat. *Physiology & Behavior*, *3*(5), 703–707. doi: 10.1016/0031-9384(68)90139-X

Zuberer, A., Brandeis, D., & Drechsler, R. (2015). Are treatment effects of neurofeedback training in children with ADHD related to the successful regulation of brain activity? A review on the learning of regulation of brain activity and a contribution to the discussion on specificity. *Frontiers in Human Neuroscience*, *9*(March), 1–15. Retrieved from `http://journal.frontiersin.org/article/10.3389/fnhum.2015.00135` doi: 10.3389/fnhum.2015.00135

6.3   Discussion

The work presented here clearly demonstrates that even a simple algorithmic implementation of the core principle of shaping can have a significant impact on a traditional NFB protocol. However, the core idea behind this particular algorithm, that progressive, incremental reinforcement and difficulty scaling can be used to guide a trainee to modulate their brain activity more correctly and optimally, does not apply trivially to the BCI context. Some modifications are required, and a new algorithm must be developed.

For a generalized user-centered BCI, shaping is extremely important. Rather than being prescribed mental commands which, on the basis of *a priori* neuro-scientific knowledge, can be expected to be classifiable for many individuals, the user must be trained to make their chosen mental commands separable and consistent over time. Individually-adapted shaping is what Progressive Thresholding provided over Standard Thresholding in the traditional neurofeedback context. Applying this to the BCI context requires an extension of Progressive Thresholding which takes into account the very notion of class separability in an arbitrary multidimensional feature space with a moving NFB target.

Bibliography

[1] E. Curran, "Learning to control brain activity: A review of the production and control of EEG components for driving braincomputer interface (BCI) systems," *Brain and Cognition*, vol. 51, pp. 326–336, Apr. 2003.

[2] J. Pineda, D. S. Silverman, A. Vankov, J. Hestenes, *et al.*, "Learning to control brain rhythms: making a brain-computer interface possible," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, no. 2, pp. 181–184, 2003.

[3] C. Neuper and G. Pfurtscheller, "Neurofeedback training for BCI control," in *Brain-Computer Interfaces*, pp. 65–78, Springer, 2010.

[4] F. Lotte, F. Larrue, and C. Mühl, "Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design.," *Frontiers in human neuroscience*, vol. 7, no. September, p. 568, 2013.

[5] E. Angelakis, S. Stathopoulou, J. L. Frymiare, D. L. Green, J. F. Lubar, and J. Kounios, "EEG neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly.," *The Clinical neuropsychologist*, vol. 21, no. 1, pp. 110–129, 2007.

[6] H. Heinrich, H. Gevensleben, and U. Strehl, "Annotation: Neurofeedback - Train your brain to train behaviour," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 48, pp. 3–16, 2007.

[7] L. H. Sherlin, M. Arns, J. Lubar, H. Heinrich, C. Kerson, U. Strehl, and M. B. Sterman, "Neurofeedback and basic learning theory: Implications for research and practice," *Journal of Neurotherapy*, vol. 15, pp. 292–304, Oct. 2011.

[8] B. Z. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser, "BCI demographics: How many (and what kinds of) people can use an SSVEP BCI?," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 18, no. 2, pp. 107–116, 2010.

[9] B. Z. Allison and C. Neuper, "Could anyone use a BCI?," in *Brain-computer interfaces*, pp. 35–54, Springer, 2010.

[10] C. Jeunet, A. Cellard, S. Subramanian, M. Hachet, B. N'Kaoua, and F. Lotte, "How well can we learn with standard BCI training approaches? A pilot study," in *6th International Brain-Computer Interface Conference*, 2014.

[11] C. Jeunet, E. Jahanpour, and F. Lotte, "Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study," *Journal of neural engineering*, vol. 13, no. 3, p. 036024, 2016.

[12] U. Strehl, "What learning theories can teach us in designing neurofeedback treatments," *Frontiers in Human Neuroscience*, vol. 8, no. November, pp. 1–8, 2014.

[13] R. G. Miltenberger, *Behavior modification: Principles and procedures.* Cengage Learning, 2011.

[14] M. A. Vollebregt, M. Dongen-Boomsma, J. K. Buitelaar, and D. Slaats-Willemse, "Does eeg-neurofeedback improve neurocognitive functioning in children with attention-deficit/hyperactivity disorder? a systematic review and a double-blind placebo-controlled study," *Journal of Child Psychology and Psychiatry*, vol. 55, no. 5, pp. 460–472, 2014.

[15] A. Zuberer, D. Brandeis, and R. Drechsler, "Are treatment effects of neurofeedback training in children with ADHD related to the successful regulation of brain activity? A review on the learning of regulation of brain activity and a contribution to the discussion on specificity," *Frontiers in Human Neuroscience*, vol. 9, no. March, pp. 1–15, 2015.

Chapter 7

A Generalized BCI with Progressive Neurofeedback for Answering Binary Questions

7.1   Introduction

The work presented in this chapter combines the generalized BCI transducer presented in Chapter 4 with an extension to the Progressive Thresholding neurofeedback algorithm presented in Chapter 6. The goal of the study is to develop a BCI that can be used to answer "Yes" or "No" questions for severely disabled persons, as well as to test the usefulness of the new approach to NFB in the BCI context. Though this study is ongoing, pilot data collected from healthy volunteers are presented below after further contextualization of this work with respect to the broader scope of this thesis.

7.1.1   The Need for Generalization and Improved User Training

The need for a generalized BCI which can accommodate a broader variety of users and mental tasks has been discussed several times throughout this thesis. However, generalization serves an additional purpose for severely disabled persons, especially those who have experienced traumatic brain injuries. Accidents causing traumatic brain injuries may also involve injuries leaving scarring on the scalp and loss of brain function that is unique for different individuals. Such scarring, in addition to injuries and the presence of medical equipment can all interfere with the placement of EEG electrodes. These factors combine to make it difficult to construct a BCI based on predetermined kinds of mental imagery for this special population because such a BCI would require a similar electrode configuration for all patients. Generalized methods like the ones which have been presented in this thesis can make it easier to accommodate the needs of each individual patient since they allow for a variety of patterns of brain signals from different regions of the brain to serve as control signals.

The need for improved user training has also been discussed throughout this thesis. NFB is the primary means of training users to control a BCI [1, 2, 3, 4]. Since the goal of user training is to improve the classifiability of mental commands, the purpose of NFB in the BCI context is to train users to produce a more consistent pattern in the EEG for each mental command which is distinct from the patterns associated with other mental commands and other brain activity not at all associated with BCI control [5]. NFB as it is currently implemented in BCIs has been shown to be useful in this respect [6, 7, 8, 9]. However, more recent empirical results have demonstrated that current NFB protocols are far from optimal for BCI user training [10]. While there is an emerging consensus regarding the inadequacies of current NFB methodology [4, 11, 10, 12], substantive proposals for addressing this problem are still lacking [8, 13, 14, 15, 16].

Almost all user training methods in brain-computer interfacing can be described as variations of the classic Graz Protocol [17, 3, 10]. The Graz Protocol trains BCI users over a series of structured trials. In each trial, the user is cued to perform one of their mental commands. Neurofeedback in the Graz Protocol is entirely machine learning driven. Whether feedback is provided while the user focuses on the mental command or after the mental command period, some visual or auditory stimulus is used to inform the user about which mental command was predicted by the BCI. This tells the user whether or not they produced brain activity in line with the machine's expectations, and is therefore useful feedback for the user. However, it does not tell the user anything about how to adjust their brain activity. Moreover, this approach does not involve any explicit mechanism for specifically training users to produce more consistent and distinct mental commands over time, *i.e.*, shaping.

The Graz Protocol is foundational with respect to user training in the field of brain-computer interfacing. Trial-based user training and NFB derived from a machine learning model are perhaps unavoidable when training a person to use a BCI with intentionally generated mental commands. However, it is becoming increasingly clear that advances to the Graz Protocol are needed to improve BCI usability in preparation for real-world applications, and that there is much to be gained by applying principles of learning theory and instructional design [4, 11, 10]. Some advances in this direction have already been made.

Recently proposed improvements to the standard Graz Protocol have focused mainly on either improving the instructions and feedback for the user or environment in which training takes place, rather than the NFB algorithm or protocol. In line with the argument for generalization presented in this thesis, one study showed that providing mental strategies that were not overly specific, and thus allowing for some flexibility in the exact mental commands used by each user, was beneficial in a traditional NFB setting (*e.g.*, in an experiment similar to the one presented in Chapter 6) [18]. With regards to implementing shaping, one approach has been developed that trains users with a progressively more complex feature space [19]. Specifically, users were trained by first applying the Graz Protocol to simple alpha and beta band activity. Users were then trained based on CSP features from those same frequency bands in subsequent training runs. This approach led to shorter training times and led to one user overcoming their previous BCI illiteracy.

While the above approach addresses the problem of providing manageable step-wise BCI user training, it can only be implemented in a BCI which is designed for a specific neurophysiological signal. This is because the multiple hierarchical feature spaces are predefined based on knowledge of the mental commands which users are instructed to use. In particular, finding a simplified feature space for the initial training phase when the mental commands are unknown could be a

very difficult problem to solve. Without a solution to this problem, this approach is not yet compatible with a generalized BCI, though an adaptation based on a hierarchy of selected features from a pretraining dataset may serve as a close approximation. More importantly, this approach is still missing a mechanism with which to explicitly train users to increase the consistency and distinctness of their mental commands over time within a feature space.

Without shaping mental commands during BCI user training, improvement in mental command generation over time is based mainly on guesswork on the part of the user. As discussed in Chapter 6, the same problem exists in traditional NFB training. However, the question of how to introduce shaping in BCI user training is somewhat different than the question of how to introduce shaping into traditional NFB. Instead of training a user to produce brain activity closer to some target, shaping for brain-computer interfacing must train users to improve class (*i.e.*, mental command) separability in a feature space with respect to a changing classifier. This is even more difficult in a generalized BCI, where the feature space and mental commands must be treated as arbitrary and potentially unknown. However, an analogue to Progressive Thresholding might prove to be useful in the BCI context as well.

Here, a new formulation of NFB specifically designed for BCIs is proposed. This method, called Progressive Neurofeedback (PNFB) is an extension of the Graz Protocol for the case of improving class separability in an arbitrary multi-dimensional feature space inhabited by two or more classes. The formulation of PNFB is given below.

7.2   Progressive Neurofeedback

The purpose of preprocessing, feature extraction, and feature selection in machine learning is to maximize class separability by manipulating the feature space. In contrast, NFB, properly applied to BCI, should support classification by training the user to produce more separable classes, or mental commands. In other words, NFB manipulates the data generation process itself rather than just the way those data are processed. PNFB represents an attempt to improve this function of NFB by explicitly incorporating a mechanism for shaping.

Restricting the discussion to the usual machine learning problem, where the features are represented in a space in which Euclidean distance is defined, class separability depends on a combination of the intraclass variance for each class and interclass distances in this feature space (note that these two measures together determine the degree of interclass overlap, which is inversely proportional to class

separability). Therefore, NFB should train the user to reduce the intraclass variances relative to interclass distances of their mental commands in feature space over time in order to maximize class separability. Note that this is a description of class separability in the linear sense, which, in comparison to non-linear separability, is a simpler basis upon which to formulate and test the core idea behind PNFB.

Class Separability in a Feature Space

The degree of linear class separability in a Euclidean space in statistics and machine learning is measurable as the degree of class overlap, which is related to the ratio of interclass distance (here interclass distance refers to the Euclidean distance between class means, but this concept generalizes to many other definitions of interclass distance) to intraclass variances (see for example the Davies-Bouldin Index for evaluating clustering solutions [20]). Maximizing classification accuracy in the linear sense can be interpreted as maximizing this ratio.

Given an $M \times N_c$ data matrix $X^{(c)}$ containing the N_c samples for class $c \in \{1, \ldots, C\}$ represented as a set of $M \times 1$ feature vectors $\{x_1^{(c)}, \ldots, x_{N_c}^{(c)}\}$, the interclass distance between any pair of classes c_1 and c_2 can be simply defined as the Euclidean distance between class means $\mu^{(c_1)}$ and $\mu^{(c_2)}$:

$$D^{(c_1,c_2)} = \|\mu^{(c_2)} - \mu^{(c_1)}\|_2. \tag{7.1}$$

It is also possible to define interclass distance as the shortest vector from one class boundary to another class boundary, but for applications in statistics and machine learning, this definition is difficult to apply. The difficulty arises from the fact that the boundary of a random variable's distribution is typically not precisely defined. In contrast, the sample mean tends to be a more stable estimate than the class boundary, including for non-ergodic signals, due to the Central Limit Theorem [21].

The intraclass variance $R^{(c)}$ is related to the total distance between points within the class and the class centroid. For example, if $\hat{R}^{(c)}$ is some estimator for $R^{(c)}$ that needs to be chosen, then one possibility is to use

$$\hat{R}^{(c)} = \sum_{i=1}^{N_c} \|x_i^{(c)} - \mu^{(c)}\|_2. \tag{7.2}$$

Note that there is a clear similarity in form between this choice of $\hat{R}^{(c)}$ and the variance of a random variable. However, intraclass variance can also be defined as the volume of the class' bounding region, which is mathematically represented

as an alternative choice for $\hat{R}^{(c)}$:

$$\hat{R}^{(c)} \approx \int_{D_1} \cdots \int_{D_M} F^{(c)}(a_1, \ldots, a_M) da_1 \ldots da_M, \tag{7.3}$$

where F approximates the boundary of class c. Though a data-driven estimate of F can be obtained by computing a smoothed alpha hull [22], this was found to be too computationally intensive for a real-time BCI implementation.

After selecting estimators of $D^{(c_1,c_2)}$, $R^{(c_1)}$, and $R^{(c_2)}$, the choices for which are used in PNFB will be defined below, the linear separability between two classes can be defined as

$$S^{(c_1,c_2)} = \frac{D^{(c_1,c_2)}}{R^{(c_1)} + R^{(c_2)}}. \tag{7.4}$$

For linear classification problems in Euclidean spaces, maximizing binary classification accuracy and maximizing S are commensurate. Given this formulation, higher values of $S^{(c_1,c_2)}$ mean greater class separability, and vice versa. If the volumes of the class models can be estimated, then a more precise measure of class separability is the inverse of the overlapping volume between the class models. However, even with other measures of intraclass variance, the goal of PNFB is framed as training a user to maximize S in order to simplify the task of the BCI transducer.

The Setup for PNFB

Given the definition of class separability above, PNFB can be implemented in a variety of ways. Here, a simple implementation is proposed in order to test the basic concept of PNFB. First, a more specific measure of class separability is needed which can be efficiently computed. An approach based on first and second order statistics is proposed.

The distance between class centroids is used to measure interclass distance D as in Equation 7.1 due to its simplicity and reliability for changing class models. A common measure of intraclass variance is the class covariance matrix (here we simplify the notation and just use $R^{(c)}$ for the measure actually used in PNFB) [23]

$$R^{(c)} = v \cdot \text{cov}(X^{(c)}). \tag{7.5}$$

Here v is used as a scaling factor which controls the proportion of the samples contained by the ellipsoid defined by $R^{(c)}$ and centered on the estimated class centroid of $X^{(c)}$, $\bar{X}^{(c)}$. $R^{(c)}$ is a confidence region which is used in place of a true bounding region, and the proportion of the probability mass of $X^{(c)}$ that it contains is controlled by v (*e.g.*, for normally distributed $X^{(c)}$, $v = 2$ produces

172

an ellipsoid which contains approximately 95% of the samples in $X^{(c)}$). Note, however, that $X^{(c)}$ can be non-Gaussian and non-convex. Therefore, $R^{(c)}$ can sometimes be a poor representation of the boundary of $X^{(c)}$ under non-Gaussian distributions, high skewness, and for large deviations from convexity. This may not be highly detrimental because PNFB does not rely heavily on a precise estimate of class boundaries. For the purpose of this study, class models based on second order statistics are sufficient, but the inclusion of skewness estimators and non-Gaussian shape estimators can be used in future extensions to PNFB.

Given these choices of D and R, the goal is to design an NFB protocol that trains the user to maximize D and minimize R, thus maximizing S. This is an operation that must be performed over time while the BCI is being operated. Thus PNFB must operate with respect to changing class models, and it becomes necessary to define $X_t^{(c)}$, $D_t^{(c)}$, $R_t^{(c)}$, and $S_t^{(c_1, c_2)}$ as the class data, intraclass variance, interclass distance, and interclass separability, respectively, at time t. Here t increments when the models are updated, which depends on the specific design of the BCI.

Training BCI Users with PNFB

PNFB extends simple classifier-based NFB by presenting NFB based on surrogate class models that have slightly greater D and slightly lower R compared to the true class models. Thus the NFB provided to the user with PNFB is based on a classifier that has been given slightly idealized versions of the class models. Although the classifier provides feedback to the user in the usual way, because it is trained with the surrogate models, the user is required to learn to perform their mental commands as well as the surrogate models would suggest in order to receive feedback which is as positive as they would normally receive. If the user is able to learn to generate mental commands which, over a number of trials, produce similar D and R to the surrogate models, then PNFB generates new surrogate models with even further improved values for D and R.

The above suggests that PNFB may make training more difficult in comparison to training with the standard Graz protocol. However, it also creates a series of progressive training steps that are automatically adapted to the individual user and that move in the direction of increasing class separability. Furthermore, the increase in difficulty can be controlled by how different the surrogate models are compared to the true class models. Even if these differences are kept small in order to avoid making training too difficult, over several iterations, the effect on learning and BCI performance are potentially large. The exact implementation of PNFB proposed for an initial study of this approach is given below.

Assume a user is being trained to control a two-class BCI. Assume further that at time $t = 0$ the first set of models has been trained using a set of pretraining trials which are represented as feature matrices $X_0^{(1)}$ and $X_0^{(2)}$, and that these models will be used to provide feedback for the next block of trials. With PNFB, surrogate models which have $\hat{D}_0^{(1,2)} > D_0^{(1,2)}$, $\hat{R}_0^{(1)} < R_0^{(1)}$, and $\hat{R}_0^{(2)} < R_0^{(2)}$ are needed. First, the intraclass variances are reduced relative to the true current class variances using the simple transformation of the original data given by

$$\hat{X}_t^{(c)} = \sqrt{\alpha}\left(X_t^{(c)} - \bar{X}_t^{(c)}\right) + \bar{X}_t^{(c)}, \quad \forall c \in \{1, \dots, C\}. \tag{7.6}$$

Here α is a scaling factor which controls the degree to which intraclass variance is reduced ($0 < \alpha < 1$, but values between 0.85 and 0.95 are recommended so that the surrogate class models remain similar to the true class models). Note that this transformation simply reduces the covariance structure of the original class data proportionally such that $\text{cov}(\hat{X}_t^{(c)}) = \alpha \cdot \text{cov}(X_t^{(c)})$, since $\alpha \cdot \text{cov}(X) = \text{cov}(\sqrt{\alpha}X)$. Hence α is used to control the degree to which the user is conditioned to make their mental commands more consistent.

Increasing D can also be done using a simple transform. A vector $V^{(c)}$ whose initial point is the centroid of a given class model and whose direction points away from all other class centroids can be obtained by computing

$$\mathbf{V}_t^{(c)} = \sum_{\forall (i \neq c) \in \{1, \dots, C\}} \bar{X}_t^{(c)} - \bar{X}_t^{(i)}. \tag{7.7}$$

Taking $U_t^{(c)}$ to be the unit vector of $V_t^{(c)}$, a shift along U that is proportional to the standard deviation of $\hat{X}_t^{(c)}$ is given by

$$\frac{\sigma(\hat{X}_t^{(c)})}{m} U_t^{(c)}, \tag{7.8}$$

where σ denotes the standard deviation and m is a parameter which determines the magnitude of the shift along $U_t^{(c)}$ as a factor of $\sigma(\hat{X}_t^{(c)})$. Thus m can be used to control the degree to which the user is conditioned to make their different mental commands unique from one another in feature space.

Combining the two transforms, the surrogate data for class c at time t can be obtained by transforming $X_t^{(c)}$ by

$$\hat{X}_t^{(c)} = \sqrt{\alpha}\left(X_t^{(c)} - \bar{X}_t^{(c)}\right) + \bar{X}_t^{(c)} + \frac{\sigma(\hat{X}_t^{(c)})}{m} U_t^{(c)}. \tag{7.9}$$

The core concept of this transformation is illustrated in Figure 7.1.

The use of surrogate class models is unconventional in NFB, but it is the
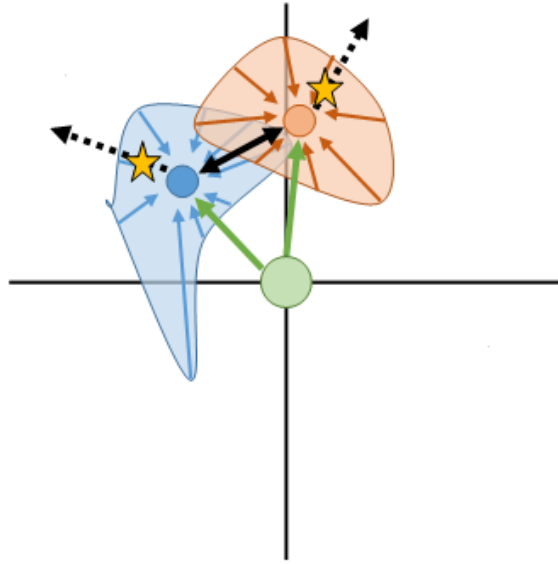
Figure 7.1: An illustration of the transformation which is used to generate the surrogate class models in Progressive Neurofeedback. Note that this illustration is not limited to the specific implementation of PNFB used here, as non-Gaussian class models are displayed.

basis of implementing shaping in PNFB. The transformed data $\hat{X}_t^{(c)}$, $\forall c \in C$ are used to construct a surrogate feature space and train a surrogate classification model. Since the surrogate class models are slightly more separable, classifier-based NFB using the surrogate classification model assumes the user being trained is able to produce slightly more reliable mental commands than they can in reality. Therefore, it should be slightly more difficult for the user to achieve the same degree of control over the BCI than training without the surrogate models. In order to achieve the same degree of success as would be achieved with the standard Graz protocol, the user must learn to produce more reliable mental commands, because the classifier is "tricked", or has been "lied to", about the true reliability of the user's ability to generate mental commands. Hence, the user is required to adapt in the direction of improved class separability. Note that the way shaping is implemented algorithmically is similar conceptually to how it is implemented in Progressive Thresholding; progress is promoted with difficulty tuning and an imposed directionality to the feedback.

Recent evidence suggests that moderate to high difficulty in NFB is beneficial to learning self-modulation of brain activity, including in a BCI setting, though extreme levels of difficulty can be detrimental to training [24, 25]. The parameters v and m control the difficulty in PNFB and therefore should be set so that the surrogate class models do not deviate so far from the true class models that the user cannot adapt, but also so that there is enough of an increase in difficulty to enhance learning. In order to ensure the user is learning correctly, the surrogate

models are not updated until the user is able to produce true class models which are very close to the surrogate models in a statistical sense, *i.e.*, their statistics can no longer be distinguished from one another. Once the user achieves this, new surrogate models can be generated in order to further train the user towards increased class separability. In order to train the user to maximize class separability, the process of generating surrogate models and tracking the user's progress over a number of trials is repeated until either the classes become fully separable, or more likely, until progress plateaus. Once the user is no longer able to improve their performance, their true data can be used to train the classifier with which they will actually use the BCI.

7.2.1   PNFB Simulation

A simple simulation with two features can be performed in order to illustrate how surrogate models change on each iteration under PNFB. Figure 7.2 shows how PNFB transforms the data over successive iterations. Since new surrogate models are only generated when the user is able to adjust their mental commands to approximate the surrogate class models, this simulation assumes the user is successful in this regard and does not consider how long it takes the user to accomplish this.

The simulation follows a similar pattern as long as the data are classifiable beyond chance to begin with, irrespective of along which set of directions the classes are separable. The particular simulation given here for illustration was performed with class one (shown in magenta) sampled from a distribution having mean $\mu^{(c1)} = (0.9, -0.3)$ and diagonal covariance matrix with standard deviation $\sigma^{(c1)} = (0.8, 1.2)$, class two (shown in green) sampled from a distribution having mean $\mu^{(c2)} = (1.2, 1)$ and diagonal covariance matrix with standard deviation $\sigma^{(c2)} = (1.5, 0.8)$, and baseline EEG (shown in blue) sampled from a standard normal distribution with mean $\mu^{(B)} = (0, 0)$ and standard deviation $\sigma^{(B)} = (1, 1)$.

A divide and conquer multiclass classification scheme, described in Chapter 2 Section 2.6.5, was performed before PNFB and at each iteration, similarly to the classification scheme used in the BCI experiment described below. In this classification scheme, baseline EEG versus either mental command class was classified (referred to as Null classification throughout this chapter), and the data were classified as belonging to one or the other of the two mental commands in an independent analysis (referred to as Command classification throughout this chapter). The change in classification error over PNFB iterations in this simulation is given in Figure 7.3.

The simulation shows that PNFB can gradually improve classification accuracy
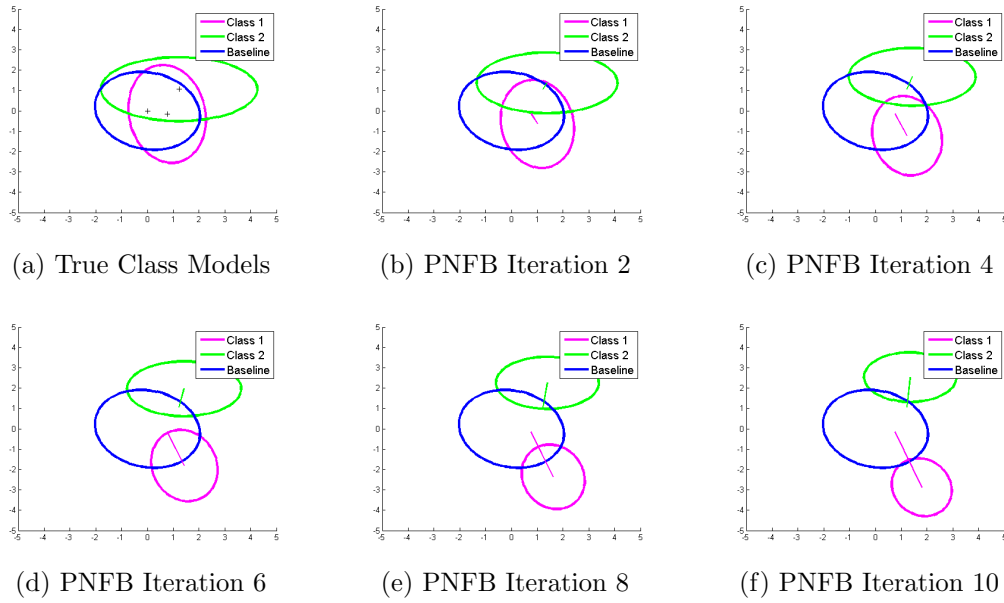
Figure 7.2: The change in class models over successive simulated PNFB iterations. Displayed are the 2σ ellipsoids of each class model with their means marked with crosses in subfigure a). A line from the original class means to their new class means are shown to illustrate the direction in which the surrogate models are shifted. Note that the class models also become more homoscedastic at higher PNFB iterations.

by reducing intraclass variance and increasing interclass distance. In addition, it promotes greater homoscedasticity in the class models. This simulation, however, only serves to illustrate that the surrogate models used in PNFB become gradually more ideal for classification. The success in PNFB in a BCI depends primarily on the ability of the user to learn to adjust their mental commands in order to approximate those surrogate models. Thus the success of PNFB depends on whether NFB adjusted by the surrogate models helps the user to appropriately adjust their mental commands towards improved separability. This requires empirical study with a BCI.

7.3   BCI Experiment

A BCI was constructed for the purpose of enabling users to answer "Yes" or "No" questions using personalized mental commands. This system was piloted with healthy volunteers using a standard experimental protocol used in BCI research.

The experiment was organized into multiple sessions. The first session was reserved for pretraining. During pretraining, participants simply answered ques-
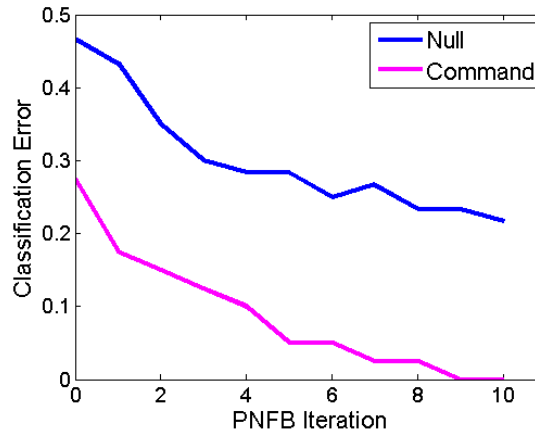
Figure 7.3: The change in Null classification error and Command classification error across simulated PNFB iterations. Iteration zero refers to the classification errors of the original true class models.

tions one at a time using their chosen mental commands without feedback. EEG was recorded during this time and used to train an initial classification model for each individual participant. In subsequent sessions, denoted training sessions, participants actively attempted to answer questions through the BCI, with feedback, in order to improve the performance of both the user and the computer algorithms.

7.3.1   Participants and Data Acquisition

Data for the pretraining session of the experiment were collected from three healthy right-handed volunteers (all three female, ages 20-25). One of these participants continued to donate their time to data collection for this experiment by participating in three further pretraining sessions using different mental commands as well as three training sessions. All sessions using the same type of mental commands were conducted on different days and the data collected across different days were combined when training models. This was done in order to ensure that machine learning models would be required to adapt to day-to-day changes in the EEG.

EEG was recorded using a 64-channel BioSemi headcap [26] with five external electrodes used to record EOG and EMG. One external electrode was placed just lateral to the right eye, and another was placed just superior to the same eye. These were used to capture the waveforms related to horizontal and vertical eye movements. One external electrode was placed over each masseter for acquisition of jaw clench EMG. Finally, one electrode was placed over the larynx in order to

measure subvocal laryngeal (LSV) activations.

Impedance was checked after equipping the participant with the EEG cap. Electrodes were adjusted and electrode gel was reapplied as needed until impedance was below 20 kOhms for as many electrode sites as possible. For each session, between two and four electrodes had impedance values above this threshold because the issue could not be resolved in a timely manner. EEG was sampled at 2 kHz and downsampled to 256 Hz prior to processing.

During EEG recordings the participant was seated in a private room in front of a standard computer display. Data was collected and processed using Matlab R2016a [27] and Psychtoolbox [28]. Instructions and visual stimuli were displayed on the computer display and auditory stimuli were played through standard computer speakers.

7.3.2   BCI Task

Since the motivation behind this study is to develop a BCI as a communication aid for severely disabled individuals, primarily so they can answer questions posed by healthcare staff regarding their current condition, we asked participants to provide a list of 130 questions about themselves (65 to which the answer was "Yes", and 65 to which the answer was "No") with the corresponding answers. The first 100 questions were used for training only, and the remaining 30 questions were provided for testing (*i.e.*, data related to the test questions were never used to train algorithms used by the BCI, but were instead intended for use as a completely independent test set). Though the questions were self-selected, participants were made aware that the experimenters would have access to the questions and answers, and so they should avoid including any personal information which they did not wish to be revealed. Most of the questions were neutral and generic in nature (*e.g.*, "Do you have a brother?", or, "Are you Canadian?"). The questions were recorded as audio files by a lab member who did not take part in the experiment in any other capacity.

Both pretraining and training sessions were organized into ten blocks of ten trials. During training the experiment paused after every block in order to update any models used in classification. Once these updates were completed, the participant was able to continue training when ready by pressing any key on the keyboard in front of them. The specific timings of different components of each trial are given in Table 7.1. In order, each trial was divided into the question presentation period (QP), the response period (RSP), which was also the NFB period during training sessions, the predicted response display period (DISP), and the intertrial interval (ITI).

Session	QP	RSP/NFB	DISP	ITI
Pretraining	≈3s	5s	2s	4s
Training	≈3s	Max. 12s	2s	4s
Usage	≈3s	Max. 12s	2s	4s

Table 7.1: The amount of time allocated to each period within each trial for the BCI in pretraining sessions, training sessions, and proposed timings for the final BCI application (referred to as "Usage"). QP refers to the question presentation period, RSP refers to the response period, NFB refers to the response period with neurofeedback, DISP refers to the predicted response display period, and the ITI refers to the intertrial interval.
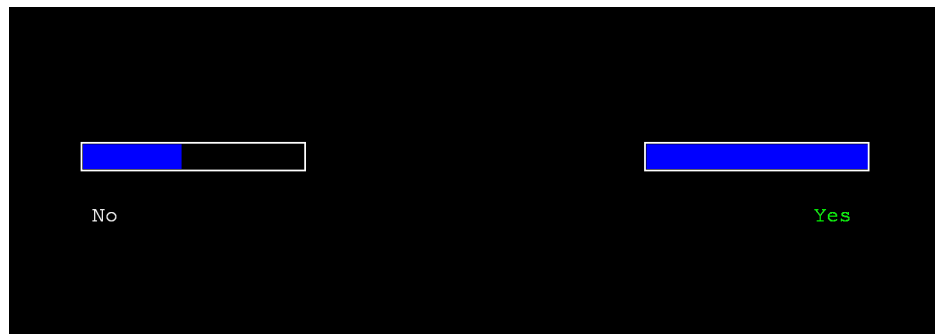
Figure 7.4: The NFB display used during the training sessions. In this example, the "Yes" response was selected, but the classifier did have moderate confidence in the "No" response as well.

Two kinds of feedback were provided during the training sessions. First, NFB was provided while the participant attempted to answer the question in order to facilitate participant training. NFB was presented in the form of two bars underneath the words "Yes" and "No", as shown in Figure 7.4. The classifier made a prediction of the participant's answer approximately once every 0.5s. The bar corresponding to the prediction filled by $1/12^{th}$ of its total length in order to indicate to the user how the BCI interpreted their most recent brain activity, and the speed at which the bar filled reflected the classifier's confidence in the corresponding response.

The second kind of feedback was displayed after the response period in order to indicate the final decision of the BCI. This feedback was based on the state of the NFB bars at the end of the response period. If one of the bars filled all the way, the response period was terminated, the font colour for the word displayed on the computer screen was changed to green, and an audio recording of the corresponding answer, "Yes" or "No", was played. If, by the end of the 12s response period, one of the bars was more than 70% filled and the other was less than 50% filled, the font colour corresponding to the predicted answer changed to yellow to indicate low confidence in the predicted response, and no audio response

was played. Finally, if neither bar was filled past the 70% threshold, then the screen displayed the text "Unclear Response".

7.3.3   Mental Commands

The BCI was piloted using four categories of mental imagery. Motor imagery was used both because it is know to be an effective means of BCI control with many individuals and because it is the most common strategy for BCI control. Thus motor imagery serves as a point of comparison with more standard BCI designs. In accordance with the generalized BCI framework and the broader goal of bringing personalized mental commands to bear in brain-computer interfacing, three additional control strategies were used. One strategy was to directly focus on the answer to the question by mentally repeating the answer to the question as clearly as possible. This response approach was used in a similar study on a BCI designed for answering "Yes" or "No" questions, but using fNIRS [29]. A second alternative to motor imagery involved musical imagery, whereby the participant chose to imagine two familiar tunes as mental commands. Finally, the third alternative strategy was based on defining complex lateralized mental commands based on personal preferences (for example, imagining reaching excitedly for a favourite food using the right hand to answer "Yes" versus pushing away a disliked food using the left hand to answer "No"). Participants were given time to choose their own mental commands within the relevant category at the start of each pretraining session.

7.4   Data Analysis

7.4.1   Preprocessing

EEG recorded during the QP, RSP/NFB, DISP, and ITI periods of each trial were stored and preprocessed separately. Denoising and artifact rejection were performed after each of these recording periods for each trial independently. Each EEG window was zero-meaned by channel and re-referenced to an average reference. A 6th order 60 Hz Butterworth notch filter and a 4th order 4-40 Hz Butterworth bandpass filter were applied to the signals.

Artifact rejection was performed with the ADJUST toolbox [30] in EEGLAB [31]. EOG artifacts were suppressed using an adaptive filter based on conventional recursive least squares (CRLS) regression [32] with respect to the two EOG external electrodes [33]. EMG artifacts were removed using an automated blind

source separation approach seeded with the two EMG external electrodes [34]. Finally, CRLS regression was also used with the LSV external electrode in order to suppress the intrusion of EMG due to LSV activation. The remaining cleaned 64-channel EEG signals were used for further analysis.

7.4.2   Feature Extraction

The purpose of the pilot experiment is to determine a set of methods which can be used in a full study focused on the application of the BCI and the utility of PNFB. Therefore, both of the generalized feature extraction approaches described in Section 2.6 of this thesis were used to analyze the pilot data. While Filter-Bank Common Spatial Patterns (FBCSP) [35] is appropriate given the large number of EEG channels, spectral features (SF) were also tested because the neural activity pertaining to direct "Yes" or "No" mental responses is very likely to arise from neural populations which are highly overlapping and very close together (SF features were extracted from the top 6 PCA components in order to reduce the number of features). However, during online analysis that was conducted during data acquisition, only FBCSP was used due to its speed in online feature extraction.

7.4.3   Classification

Classification was performed using logistic regression with elastic-net regularization (LRE) [36] to predict the binary class labels. This method was used for its efficiency in model training and parameter selection during real time experiments. In addition, elastic-net regularization behaves as an embedded feature selection method with built-in sparsity constraints that help to avoid overfitting and redundancy. In addition, the number of features chosen for classification can be restricted. Moreover, elastic-net regularization allows for feature selection by taking into consideration a feature's discriminative power in linear combinations with other features, as opposed to non-embedded sequential feature selection methods, such as MRMR [37], which only observes the added discriminative power of each feature on its own. Details of LRE are given below.

Logistic Regression with Elastic Net Regularization

Logistic regression uses a generalized linear model with the logistic link function in order to estimate the probability that class $c \in C$ is the correct class given the

trial feature vector $\mathbf{x}$:

$$P(Y = y|\mathbf{x}) = \frac{1}{1 + e^{-c(b+\beta\mathbf{x})}}, \qquad (7.10)$$

where b and β are the linear coefficients. The parameters (b, β) were learned using LRE in the typical fashion using its convex optimization formulation

$$\min_{b,\beta} \frac{1}{N} \sum_i \log(1 + e^{y_i(\beta\mathbf{x_t}+b)}) + \alpha\lambda\|\beta\|_1 + \alpha(1 - \lambda)\|\beta\|^2, \qquad (7.11)$$

where α is a parameter which influences model sparsity and λ is the elastic-net mixing parameter. Note that when $\lambda = 1$, the elastic-net regularizer becomes equivalent to the lasso regularizer. Both parameters were optimized using nested cross-validation.

The elastic-net penalty accommodates for the fact that adjacent CSP features and spectral features are likely to be correlated. In particular, the regularized $l2$ penalty given by $\alpha(1 - \lambda)\|\beta\|^2$ introduces a sparsity constraint at the feature level which encourages removal or reduced weighting for correlated features. For additional sparsity, LRE models were constrained to keep a maximum of $N/4$ features, where N was the number of samples in the training set.

Cross-Validation Scheme

During offline analysis, cross-validation was performed in order to choose values for λ and α. For each value of $\lambda \in \{1/8, 1/4, 1/2, 3/4\}$, 10-fold cross-validation was executed with 25% of the training data randomly partitioned for model testing. For each model, α was estimated via gradient descent starting from a value of 0.5 using a nested 10-fold cross-validation loop with an additional 10% of the training data separated for use as a validation set. Additional offline analyses were performed treating each session as independent in order to assess whether within-session analyses and between-session analyses produced substantially different results.

During the online analyses which took place during training sessions, cross-validation was performed using all previous trials associated with the participant and the specific mental strategy, including trials from the associated pretraining session. This was done in an attempt to construct models which could accommodate for the day-to-day nonstationarity in EEG and thus generalize across training days, which is an important feature for BCIs designed for medium-term or long-term use. In order to reduce the amount of processing time required during online classifier updates, the values of α and λ for the LRE model and the learned FBCSP filters were stored from the previous session and not relearned

from scratch. The LRE parameters and the FBCSP filters were only updated again after the session was completed.

The classification accuracies reported below are based on the best model's average performance on the withheld test data taken from the training data for each cross-validation run (not the 30 questions which were withheld for later testing, as these have not been used in the pilot experiment so far). Since these test data were used in order to choose λ, the results reported below should be taken as cross-validation accuracies rather than as true test accuracies. Results are reported this way because this is the cross-validation scheme used to choose a model during real time BCI training, where the true test data are the subsequent trials for which the participant attempts to answer their "Yes" or "No" questions. Hence the offline classification accuracies are analogous to the metrics which are used for model selection in the online experiment.

7.5 Pilot Experiment Results

7.5.1 Online Analysis

The results of online classification are given in Table 7.2. Since Participant 3 (referred to as P3) was the only participant whom had completed any training sessions, online analysis was only ever performed with this participant. P3 completed two training sessions with the Direct mental command strategy and one training session with Musical mental imagery.

Participant	Strategy	Class	Train 1	Train 2
P3	Direct	Null	66.5	79.0
		Command	49.0	49.5
	Music	Null	69.5	–
		Command	51.0	–

Table 7.2: Classification accuracy for each session in the pilot experiment using FBCSP.

7.5.2 Offline Analysis

Results of offline analyses using the session data just as they were used during the online analysis, *i.e.*, with data from previous sessions using the same mental strategy combined with data from the given session, are given in Table 7.3 for

analysis using FBCSP and in Table 7.4 for analysis using SF features. Similarly, results of offline analyses with each session treated independently are given in Tables 7.5 and 7.6 for the FBCSP and SF approaches respectively.

Participant	Strategy	Class	Train 1	Train 2
P3	Direct	Null	68.8 (2.9)	73.2 (2.3)
		Command	50.5 (10.1)	49.8 (4.6)
	Music	Null	72.2 (6.6)	–
		Command	51.4 (9.4)	–

Table 7.3: Classification accuracy for each session in the pilot experiment using FBCSP. Here, dependencies across sessions are taken into account, so each result is based on a machine learning analysis using the data belonging to that particular session and all previous sessions.

Participant	Strategy	Class	Train 1	Train 2
P3	Direct	Null	61.3 (4.9)	62.9 (4.1)
		Command	49.7 (7.4)	50.9 (6.5)
	Music	Null	60.9 (7.2)	–
		Command	49.2 (8.6)	–

Table 7.4: Classification accuracy for each session in the pilot experiment using SF. Here, dependencies across sessions are taken into account, so each result is based on a machine learning analysis using the data belonging to that particular session and all previous sessions.

7.6 Discussion

7.6.1 BCI Performance in the Pilot Experiment

Above chance mental command classification was only achieved when each session was treated as separate. This suggests that the approaches that have been tried so far fail to generalize across sessions that take place on separate days, even though they yield low to moderate rates of correct classification in within-session analyses (with FBCSP outperforming SF overall). This drop in accuracy was not as pronounced for Null classification as it was for Command classification. Command classification involving data from multiple sessions, in contrast, never resulted in above chance classification. Even in the case of within-session classification, the results never surpassed the 70% threshold used as a heuristic for assessing whether or not a binary BCI is useful [38]. Null classification rates were higher in general, which suggests that one option for answering "Yes" or "No" questions could be to

Participant	Strategy	Class	Pre-Train	Train 1	Train 2
P1	Motor	Null	67.8 (9.6)	–	–
		Command	58.5 (6.3)	–	–
	Direct	Null	83.3 (4.3)	–	–
		Command	65.0 (7.8)	–	–
P2	Direct	Null	65.0 (4.8)	–	–
		Command	61.0 (4.4)	–	–
P3	Direct	Null	63.0 (7.4)	89.0 (5.3)	75.5 (6.0)
		Command	64.5 (8.7)	61.5 (7.5)	65.0 (9.4)
	Motor	Null	52.5 (10.3)	–	–
		Command	56.0 (5.7)	–	–
	Music	Null	70.5 (6.9)	71.5 (4.1)	–
		Command	50.5 (5.5)	58.5 (4.6)	–
	Emotive	Null	66.5 (5.1)	–	–
		Command	51.5 (10.2)	–	–

Table 7.5: Classification accuracy (and standard deviation) for each session in the pilot experiment using FBCSP. Here, each session is treated as independent, so each result is based on a machine learning analysis using only the data belonging to that particular session.

use an active command for "Yes" and a rest state for "No". However, this would require a highly controlled environment in which the patient's caregiver would need to turn on the BCI before asking questions, and thus limiting the desired degree of freedom provided by the BCI.

Since only one participant underwent any actual training to use the BCI, and BCI performance varies widely across individuals, the reliability of the current methods cannot be determined. It could be the case that a substantial proportion of individuals could achieve greater than 70% classification accuracy with at least one mental strategy with one of the current feature extraction approaches, or it could be the case that most individuals would have less success than P3. However, it does appear to be the case that a new strategy needs to be developed in order to overcome the challenge of generalizing across sessions.

One potential solution is to replace the base CSP algorithm in the FBCSP scheme with stationary CSP [39] and to include further regularization if needed (*e.g.*, by removing trials which may be outliers prior to learning the CSP filters [40]). Since stationary CSP seeks to find a stationary subspace within the EEG for which to construct a CSP spatial filter, its use may allow for improved generalization across sessions. Another approach would be to compute SF features from CSP components, as was done in Chapter 4. A more detailed analysis of various candidate feature spaces will help to determine which method in par-

Participant	Strategy	Class	Pre-Train	Train 1	Train 2
P1	Motor	Null	64.3 (9.9)	–	–
		Command	52.5 (5.2)	–	–
	Direct	Null	69.0 (6.1)	–	–
		Command	58.5 (4.5)	–	–
P2	Direct	Null	56.8 (4.2)	–	–
		Command	50.5 (8.9)	–	–
P3	Direct	Null	61.4 (6.2)	82.5 (4.7)	69.5 (6.1)
		Command	57.5 (8.3)	55.0 (7.2)	55.5 (9.8)
	Motor	Null	48.8 (9.7)	–	–
		Command	50.5 (8.1)	–	–
	Music	Null	66.8 (5.6)	67.0 (4.9)	–
		Command	51.0 (6.4)	54.0 (5.1)	–
	Emotive	Null	59.3 (6.2)	–	–
		Command	50.0 (8.4)	–	–

Table 7.6: Classification accuracy for each session in the pilot experiment using SF. Here, each session is treated as independent, so each result is based on a machine learning analysis using only the data belonging to that particular session.

ticular should be used for future sessions. In particular, it would be useful to check whether the Emotive strategy produced lateralized brain activity for P3 as was expected, because lateralized mental commands should be classifiable with FBCSP.

A great deal of work remains to be done before the BCI presented here develops into a BCI with which users can reliably answer "Yes" or "No" questions. The pilot data collected so far have provided valuable information and have allowed for the systems design to be tested. The pretraining data was paramount in informing the current implementation of the BCI, especially with respect to the choice of LRE over the SVM used in Chapter 4. In addition, the few training sessions conducted so far were indispensable with regards to identifying the need to craft an approach which is robust to the nonstationary changes in the EEG across sessions. Unfortunately, the small number of training sessions conducted, all of which were performed by one participant without above chance results in the online setting, mean that it is impossible at this time to infer anything about the utility of PNFB for BCI user training.

7.6.2   The Choice of Mental Commands

The choice of mental commands remains a difficult problem to address. Given the evidence presented in this thesis and in the literature discussed in previous

chapters, it would seem helpful to have each participant complete aptitude and background questionnaires in order to make informed choices about potential personalized mental commands. For example, the Kinaesthetic and Visual Imagery Questionnaire [41] was found to be correlated with BCI performance when using motor imagery in one study [42]. Adding to this evidence, the results presented in Chapter 4 suggest that a questionnaire probing a user's past experience in a variety of fields might help predict the kind of mental strategy with which they might be successful. How reliable the relationship is between aptitude in different domains and BCI performance, as well as how broadly this relationship can be applied, require further study, but what is known so far can potentially be exploited in this work.

With the current pilot study, not enough data are available with which to determine whether certain mental strategies are more promising for further exploration compared to other strategies. However, the pretraining data acquired so far suggest that direct mental imagery can be a viable mental strategy for answering "Yes" versus "No" questions with a BCI. This has been done successfully with fNIRS [29], but no reports of attempting this with EEG have yet appeared in the literature. With P3, the Direct mental response strategy was, unexpectedly, the most successful strategy of the four attempted. This may be a positive finding, since direct mental "Yes" or "No" responses would usually be the most natural, intuitive, and desirable set of mental commands for users. However, additional data might reveal other mental strategies to ultimately be more useful.

7.6.3   Plans for a Full Study Including Clinical Patients

The results presented in this chapter are based on a pilot study which was used to help develop and test the technical details of the of the BCI and PNFB, as well as the user interface. Additional pilot data may be needed in order to find an optimal machine learning strategy for this BCI. After the methodology is finalized, a full study must be undertaken in order to properly test PNFB and the usability of the BCI. This will be done in two stages: one experiment with healthy participants, and one with clinical patients.

Approximately 15-20 healthy participants split into two groups will be needed in order to evaluate the efficacy of PNFB compared to a standard Graz Protocol (implemented in the exact same way but with standard classifier-based NFB instead of PNFB). Participants will need to complete at least four training sessions plus one pretraining session on different days so that both within-session and between-session progress can be assessed. If PNFB shows improvements over standard training methods, an additional follow-up session one month after completing training would be conducted if possible in order to determine whether

PNFB also helps participants to better retain the skill of using the BCI in the long term. Addressing the problem of the long-term reliability of a BCI without requiring the user to undergo tedious amounts of training and retraining is another critical issue in developing viable BCIs for real world applications [43, 44].

Once the healthy participant pool has completed training, a deeper analysis of the online results and exploratory offline analysis will be used to make additional improvements or modifications to the BCI in order to maximize its expected utility and reliability. Once this is complete, the system will be deployed as an assistive communications prosthetic for severely disabled patients in a clinical setting. With help from caregivers and clinicians, the BCI will hopefully allow patients to answer queries about the usability, practicality, and reliability of the system. These data, along with ongoing analysis of the data collected during active BCI use, would be leveraged in order to continually update the system. Once the BCI becomes sufficiently reliable, it could be used to aid caregivers and clinical staff in interacting with severely disabled patients. Most importantly, the BCI will hopefully provide a significant improvement in the lives of severely disabled individuals.

Bibliography

[1] E. Curran, "Learning to control brain activity: A review of the production and control of EEG components for driving braincomputer interface (BCI) systems," *Brain and Cognition*, vol. 51, pp. 326–336, Apr. 2003.

[2] J. Pineda, D. S. Silverman, A. Vankov, J. Hestenes, *et al.*, "Learning to control brain rhythms: making a brain-computer interface possible," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, no. 2, pp. 181–184, 2003.

[3] C. Neuper and G. Pfurtscheller, "Neurofeedback training for BCI control," in *Brain-Computer Interfaces*, pp. 65–78, Springer, 2010.

[4] F. Lotte, F. Larrue, and C. Mühl, "Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design.," *Frontiers in human neuroscience*, vol. 7, no. September, p. 568, 2013.

[5] B. Z. Allison and C. Neuper, "Could anyone use a BCI?," in *Brain-computer interfaces*, pp. 35–54, Springer, 2010.

[6] D. J. McFarland, W. A. Sarnacki, T. M. Vaughan, and J. R. Wolpaw, "Brain-computer interface (BCI) operation: signal and noise during early training sessions," *Clinical Neurophysiology*, vol. 116, no. 1, pp. 56–62, 2005.

[7] C. Neuper and G. Pfurtscheller, "Neurofeedback training for BCI control," in *Brain-Computer Interfaces*, pp. 65–78, Springer, 2009.

[8] H.-J. Hwang, K. Kwon, and C.-H. Im, "Neurofeedback-based motor imagery training for brain–computer interface (BCI)," *Journal of neuroscience methods*, vol. 179, no. 1, pp. 150–156, 2009.

[9] K. Shindo, K. Kawashima, J. Ushiba, N. Ota, M. Ito, T. Ota, A. Kimura, and M. Liu, "Effects of neurofeedback training with an electroencephalogram-based brain–computer interface for hand paralysis in patients with chronic stroke: a preliminary case series study," *Journal of rehabilitation medicine*, vol. 43, no. 10, pp. 951–957, 2011.

[10] C. Jeunet, E. Jahanpour, and F. Lotte, "Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study," *Journal of neural engineering*, vol. 13, no. 3, p. 036024, 2016.

[11] F. Lotte and C. Jeunet, "Towards improved BCI based on human learning principles," in *Brain-Computer Interface (BCI), 2015 3rd International Winter Conference on*, pp. 1–4, IEEE, 2015.

[12] R. Chavarriaga, M. Fried-Oken, S. Kleih, F. Lotte, and R. Scherer, "Heading for new shores! Overcoming pitfalls in BCI design," *Brain-Computer Interfaces*, pp. 1–14, 2017.

[13] J. Schumacher, C. Jeunet, and F. Lotte, "Towards explanatory feedback for user training in brain-computer interfaces," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, pp. 3169–3174, IEEE, 2015.

[14] S. Teillet, F. Lotte, B. N'Kaoua, and C. Jeunet, "Towards a spatial ability training to improve mental imagery based brain-computer interface (MI-BCI) performance: a pilot study," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016)*, p. 6, 2016.

[15] C. Jeunet, B. NKaoua, and F. Lotte, "Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates," *Progress in brain research*, vol. 228, pp. 3–35, 2016.

[16] E. Loup-Escande, F. Lotte, G. Loup, and A. Lécuyer, "User-centered BCI videogame design," *Handbook of Digital Games and Entertainment Technologies*, pp. 225–250, 2017.

[17] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[18] S. Kober, M. Witte, and M. Ninaus, "Learning to modulate one's own brain activity: the effect of spontaneous mental strategies," *Frontiers in human neuroscience*, vol. 7, no. October, p. 695, 2013.

[19] C. Vidaurre and B. Blankertz, "Towards a cure for BCI illiteracy," *Brain topography*, vol. 23, no. 2, pp. 194–198, 2010.

[20] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 224–227, 1979.

[21] J. Rice, *Mathematical statistics and data analysis*. Nelson Education, 2006.

[22] B. Guo, J. Menon, and B. Willette, "Surface reconstruction using alpha shapes," in *Computer Graphics Forum*, vol. 16, pp. 177–190, Wiley Online Library, 1997.

[23] C. R. Rao, "Familial correlations or the multivariate generalisations of the intraclass correlations," *Current Science*, vol. 14, no. 3, pp. P66–67, 1945.

[24] R. Bauer, M. Vukelić, and A. Gharabaghi, "What is the optimal task difficulty for reinforcement learning of brain self-regulation?," *Clinical Neurophysiology*, vol. 127, no. 9, pp. 3033–3041, 2016.

[25] R. Bauer, M. Fels, V. Royter, V. Raco, and A. Gharabaghi, "Closed-loop adaptation of neurofeedback based on mental effort facilitates reinforcement learning of brain self-regulation," *Clinical Neurophysiology*, vol. 127, no. 9, pp. 3156–3164, 2016.

[26] BioSemi, "BioSemi Headcap." `https://www.biosemi.com/headcap.htm`. Accessed: 2017-07-10.

[27] MATLAB, *version 9.0.0 (R2016a)*. Natick, Massachusetts: The MathWorks Inc., 2016.

[28] D. H. Brainard, "The psychophysics toolbox," *Spatial vision*, vol. 10, pp. 433–436, 1997.

[29] U. Chaudhary, B. Xia, S. Silvoni, L. G. Cohen, and N. Birbaumer, "Brain–computer interface–based communication in the completely locked-in state," *PLoS biology*, vol. 15, no. 1, p. e1002593, 2017.

[30] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011.

[31] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[32] D. Lee, M. Morf, and B. Friedlander, "Recursive least squares ladder estimation algorithms," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 3, pp. 627–641, 1981.

[33] P. He, G. Wilson, and C. Russell, "Removal of ocular artifacts from electro-encephalogram by adaptive filtering," *Medical and biological engineering and computing*, vol. 42, no. 3, pp. 407–412, 2004.

[34] W. De Clercq, A. Vergult, B. Vanrumste, J. Van Hees, A. Palmini, W. Van Paesschen, and S. Van Huffel, "A new muscle artifact removal technique to improve the interpretation of the ictal scalp electroencephalogram,"

in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pp. 944–947, IEEE, 2006.

[35] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2391–2398, 2008.

[36] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[37] H. C. Peng, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.

[38] A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. P. Birbaumer, "Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication," *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 11, pp. 1533–1539, 2001.

[39] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing.," *Journal of neural engineering*, vol. 9, p. 26013, Apr. 2012.

[40] X. Yong, R. K. Ward, and G. E. Birch, "Robust common spatial patterns for EEG signal preprocessing," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 2087–2090, IEEE, 2008.

[41] F. Malouin, C. L. Richards, P. L. Jackson, M. F. Lafleur, A. Durand, and J. Doyon, "The kinesthetic and visual imagery questionnaire (KVIQ) for assessing motor imagery in persons with physical disabilities: a reliability and construct validity study," *Journal of Neurologic Physical Therapy*, vol. 31, no. 1, pp. 20–29, 2007.

[42] A. Vuckovic and B. A. Osuagwu, "Using a motor imagery questionnaire to estimate the performance of a brain–computer interface based on object oriented motor imagery," *Clinical Neurophysiology*, vol. 124, no. 8, pp. 1586–1595, 2013.

[43] 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, 2010.

[44] E. Holz, L. Botrel, and A. Kübler, "Bridging gaps: long-term independent BCI home-use by a locked-in end-user," in *TOBI Workshop IV. Sion, Switzerland*, 2013.

Chapter 8

Discussion and Conclusions

8.1 Contribution to Brain-Computer Interfacing

Brain-computer interfacing has the potential to redefine the relationship between humans and computers. A great deal of progress has been made, and remains to be made, using the traditional approach of designing BCIs around mental commands with known neural correlates. However, BCIs are best controlled via types of mental imagery and mental processes which are highly developed within the individual. Therefore, in the long-term, BCIs will need to accommodate a wide variety of mental commands in order to enable reliable control for the general population. Until such time that BCIs are able to do this, it is unlikely that BCIs will become widely adopted as useful human-computer interfaces.

In order to achieve sufficient usability for the general population by the traditional approach, a separate BCI transducer would need to be developed for each possible type of mental command. Then the correct BCI transducer, or set of BCI transducers, would have to be selected for each individual and adapted to the unique way in which their brain produces those mental commands. Given the open-ended nature of mental imagery, not only would this approach require an impractically large number of specialized BCI transducers, but the extent of specialization required for each BCI transducer risks precluding the flexibility needed in order to adapt to variations in mental commands over time and across individuals. That is why it is proposed in this thesis that generalized methods for brain-computer interfacing are required if BCIs are to become usable and impactful for the majority of people.

A successful generalized BCI transducer is one which can accommodate a wide range of mental commands. Though this requires a solution to the more difficult problem of classifying different kinds mental imagery without knowing beforehand what they are or what their neural correlates are, overcoming this challenge is necessary for the long-term success of BCIs, especially in moving them outside of the laboratory and into the clinic and the homes of the general public. Given that widespread usability, broader applicability, and highly variable performance across individuals have become central points of discussion in the literature [1, 2, 3, 4], the need for a substantial shift in methodology is becoming increasingly recognized. Generalized methods for brain-computer interfacing should be developed in order to evaluate whether this strategy will serve as part of the solution.

8.1.1   Progress in Generalized BCI Transducers

This thesis presents some of the first generalized methods for EEG brain-computer interfacing which have been shown to accommodate a wide variety of mental commands across individuals. Two approaches to a generalized BCI transducer were introduced and empirically validated. In Chapter 4, the open-ended BCI was implemented using a feature learning approach based on Common Spatial Patterns (CSP). This work showed that with a generalized approach users were more free to choose their own mental commands compared to a standard approach. This included several types of mental commands that have never been used to control BCIs before, yet the generalized BCI transducer acheived comparable performance to specialized BCI transducers. In addition, new evidence was obtained showing that the choice of mental commands, including their primary sensory modality, should take into account the user's domain expertise and experience outside of the BCI context (*e.g.*, musicians may perform much better using relevant auditory imagery than visual or motor imagery). While there were clear advantages to using this generalized approach, there did not appear to be any significant downsides compared to using non-generalized methods that were revealed in the studies conducted so far.

In Chapter 5, a second generalized BCI transducer based on spectral features (SF) and feature selection was implemented. This approach, which can be combined with the CSP approach by extracting SF features from CSP components, was developed for BCIs which do not have sufficient spatial resolution with which to effectively learn spatial filters, and for mental commands that are not spatially separable. This method was used to determine whether self-reported emotional reaction to videos was high or low on eleven different emotions with up to 88% accuracy, despite taking into account very little about what is known regarding the neural processing of various emotions. Interestingly, this method was also useful in predicting which participants would have classifiable reactions to the videos in the first place, suggesting that there is significant heterogeneity in neural processing of emotional stimuli. The nature of this difference and its implications require further study.

The results obtained through the work in Chapter 5 show that a generalized approach is applicable in a wide variety of applications and opens the door to new BCI applications without the need for extensive research into the neural correlates of all mental activity of interest. In this case, a method that was not designed to detect emotional reactions in particular performed well for that problem, and could be used to develop affective BCIs for diagnosis, monitoring, and treatment of certain affective disorders. For example, real-time detection of fearful reactions to stimuli can help determine which stimuli evoke such reactions for a patient suffering from post-traumatic stress disorder or a phobia and whether prescribed

treatments are effective in mitigating those reactions.

Whether or not the methods for generalized BCI transducers presented here will form the basis of future generalized BCIs remains to be seen. What this thesis has acheived, however, is to show that the field may need not continue to restrict itself to the traditional approach of designing a brand new highly specialized BCI transducer for each variation on mental imagery in many applications. Moreover, this body of work substantiates the argument that the traditional approach is not optimal in the long term and that a generalized approach can be expected to become increasingly necessary as BCIs expand beyond the laboratory.

8.1.2   Improved Neurofeedback for BCI User Training

Taking steps towards a generalized BCI, because it is a fundamentally different design paradigm, invites one to revisit the very definition of a BCI, its function, and its purpose. Brain-computer interfacing is a highly multidisciplinary field which is closely related to the broader field of human-computer interaction. Unsurprisingly, BCIs are seen through the lens of human-computer interfaces (HCIs). However, the broader HCI paradigm treats the user as an operator of a system which must respond to user input. This paradigm is appropriate for most HCIs, but does not fit the basic concept of a BCI. A BCI is a fundamentally new kind of HCI.

The current paradigm used in BCI research is borrowed from research on other types of HCIs. This paradigm encourages BCI researchers to focus on the BCI transducer in order to improve usability, applicability, and reliability. This does not contradict the goal of developing a generalized BCI transducer, and should still be seen as an important aspect of BCI research. However, the field has focused almost exclusively on developing new BCI transducers until just the last few years [3, 2] and has neglected to consider the other half of the problem in a serious way. What must be recognized is that the human is part of a BCI in a unique way compared to other HCIs. Where in other HCIs, like a keyboard, the system is static and the human must learn to use it effectively, a BCI is a dynamic system which must train the user to adapt to it while simultaneously training itself to respond to the user. This makes a BCI is unique in that it depends on co-adaptation between the user and the system, which is equipped with its own adaptive artificial intelligence. This is not to say that elements of HCI research are not beneficial to BCI research. However, the unique qualities of a BCI must be taken into consideration.

As noted previously, some researchers have begun to recognize the importance of this symbiotic relationship [3, 4, 5, 2]. Some approaches to improving

co-adaptation recognize that the best choice of mental commands depends on the user, but still focus heavily on the BCI transducer and mental commands which are chosen by the designer and given to users [6, 7, 8, 9]. Some degree of generalization is achieved by collecting EEG as the user performs a few different mental commands and using cross-validation techniques to determine the best pair of mental commands for further BCI training, but this is far from the level of generalization achieved in the work presented in this thesis. Moreover, these approaches to co-adaptation do not aim address one of the primary barrier to optimal co-adaptation, which is the poor approach to neurofeedback (NFB) training used in BCIs [3, 4, 5, 2].

A proposal for improved NFB for BCIs was given in Chapter 7 based on an algorithmic formulation of shaping from learning theory. The founding principles behind this approach was validated in the classical NFB setting in Chapter 6, where participants were much better able to balance their frontal alpha activity using the proposed Progressive Thresholding algorithm than with a standard automated NFB implementation. Moreover, participants in the Progressive Thresholding group showed balanced frontal alpha asymmetry in their pre-session baselines (*i.e.*, in their resting EEG prior to beginning training sessions) in their later sessions, suggesting a greater inter-session effect. With additional data, a similar effect might be found with Progressive NFB (PNFB) in BCI user training. PNFB is one of the first proposals for directly tackling the problem of suboptimal and inappropriate NFB training for BCIs. However, the question of how to redesign NFB protocols for BCI user training remains an open question. Nonetheless, the need to further incorporate principles of learning theory, such as shaping, is recognized.

8.2   Limitations and Future Directions

The work presented in this thesis advances brain-computer interfacing both in terms of generalized BCI transducers and improved NFB training. However, there are several limitations and shortcomings which must be addressed in future work. Each of these prongs of the generalized BCI problem represents a full program of research with applications outside of BCI, including EEG analysis and clinical NFB, but they are interconnected and support progress towards improved generalized BCIs. The next steps towards improving on both of these areas of BCI design are discussed below.

8.2.1 Further Generalization of the BCI Transducer

The approach to generalizing a BCI presented in this thesis relies heavily on generalizing the feature extraction component of the BCI transducer. A feature learning approach based on CSP was presented along with a more classical approach based on finding optimal subsets after extracting many spectral features (the SF approach). A great deal of work can be done in improving both of these approaches, including by combining them in an effective way.

With respect to the spectral feature approach, the main limitation going forward will be minimizing growth in the list of candidate features as new features are included to capture distinct kinds of information from brain activity (*e.g.*, polyspectra based on higher order statistics if enough samples are available for reliable estimation, or measures of causality, such as partial directed coherence). As more features are added, feature extraction will become less computationally efficient and the ratio of useful features to irrelevant features for any given problem may become increasingly small. Furthermore, the number of candidate features can easily become large compared to the amount of training data available, leading to problems associated with the curse of dimensionality [10]. Therefore, strategies for reducing the number of candidate features need to be developed.

Rather than the variety of features included, the primary reason the SF approach results in a very large number of candidate features is because an entire set of features is extracted for each EEG time series. Therefore, a promising way to reduce the number of features is to reduce the number of EEG time series included in the analysis. This can be done using channel selection (see for example [11]), or when enough spatial resolution is available, with spatial filtering (*e.g.*, CSP or PCA), and source separation (*e.g.*, Independent Components Analysis [12] or beamforming [13]). If the method used for reducing the number of time series is compatible with a generalized BCI, as all of the listed examples are, then spectral features can then be computed over the new smaller number of time series without significant loss of discriminative power. In some cases, for example with CSP, it may be easier to extract discriminative information from the new time series compared to the original EEG channels, thus improving the quality of the BCI transducer.

This extension of the SF method is how the CSP and SF approaches can be combined, since the CSP features can be added as candidate features to SF features computed over CSP components. However, since FBCSP, which is preferred for generalized BCIs, relies on learning many CSP filters, SF features should only be used after a few CSP components are selected in order to avoid the same problem of computing too many candidate features.

One approach, which is described here only hypothetically because the methodology has not yet been developed, is to take advantage of important discriminative information which might appear only in combinations of features and not in any of the features individually. While combinations of features are taken into account at the classification or regression stages of analysis, they are typically not taken into account during feature selection. When features are selected one at a time, as in MRMR [14] and most supervised feature selection methods [15], consideration is only given to whether an individual feature provides unique useful information on its own. Therefore, there is a mismatch between the needs of classifiers, which discriminate data on the basis of learned combinations of features, and non-embedded feature selectors (feature selectors which are independent from the classifier rather than combined with the classifier, see [15] for more details), which do not consider the discriminatory value of combinations of features.

Methods for which feature selection is embedded with classification, such as logistic regression with elastic net regularization [16] described in Chapter 7, do consider combinations of features in feature selection, but these methods often must still be preceded with a non-embedded feature selection step in order to reduce dimensionality, particularly with the SF approach. A non-embedded feature selection method that could learn the coefficients of maximally discriminative combinations of features might not only lead to better classification and regression results, but should also reduce the number of features needed during modeling because useful information from a large number of features could be represented by a relatively small number of values. Such a method could not be found in the literature, and thus developing such a method could be a useful task for future research.

With respect to the use of FBCSP, there are simple improvements that can be made. While FBCSP provides the generalization needed for mental imagery BCIs, it is not necessarily robust to noise or stable over time. Fortunately, FBCSP is simply a wrapper for the base CSP algorithm. This means that many extensions to CSP can be used in place of the base CSP algorithm in FBCSP. For example, regularized CSP [17] can be used to add robustness to noise and stability over time can be gained through the use of stationary CSP [18] within FBCSP. It is possible that the results obtained in the work which used FBCSP in this thesis could be improved by replacing one of these extensions of CSP with the base CSP algorithm used in FBCSP. An offline analysis comparing implementations of FBCSP using enhanced versions of CSP should be performed in order to determine which, if any, should be used in future generalized BCIs.

Finally, one possible path to developing a generalized BCI transducer is through deep learning classification and regression methods. Deep learning represents an entirely different feature learning approach [19] which combines learning a hierarchy of features with classification or regression in one graphical model. Its success

in other domains, such as image analysis, suggests that deep learning could improve generalization and reduce the amount of work required on the part of the designer in brain-computer interfacing. However, it has proven difficult so far to adapt existing deep learning models for EEG classification. Very few EEG and BCI-related analyses have been conducted using deep learning methods, and those that have done so have had only limited success (*e.g.*, [20]). Furthermore, tests using a multichannel Convolutional Neural Network [21] on the data presented in Chapter 4 failed to produce classification accuracies better than those achieved by FBCSP. One reason for this is that deep learning can only be expected to outperform alternative methods if hierarchical features are useful, which may not be the case for many kinds of mental imagery. It is possible that a deep learning approach could prove successful if an appropriate model architecture could be found, but this alone can be a difficult task. If a deep learning approach does prove to be successful, the question will remain whether improvements in BCI usability, if any, outweigh the challenge of finding appropriate model architectures for different individuals and different applications.

In addition to the limitations regarding the software involved in BCIs, the studies presented here were limited by the availability of EEG hardware. Commercial EEG hardware was used due to limited access to full EEG caps, and because the generalized approach to brain-computer interfacing presented here has been especially designed to enable real-world applications for the general public, which will typically make use of commercial EEG devices comparable to those used here. However, these devices are limited in their electrode coverage and often exhibit poorer signal quality than research-grade EEG hardware. As such, continued development of generalized methods would ideally take place with research-grade or higher quality EEG hardware, a process which has begun with the work presented in Chapter 7, in order verify the results obtained so far and to improve confidence in these findings and their implications.

8.2.2   The Adaptive Neurofeedback Framework

The Adaptive Neurofeedback Framework (ANFB) is an overarching framework for NFB which encompasses both the Progressive Thresholding method studied in Chapter 6 and its Progressive NFB extension for BCIs proposed in Chapter 7. However, this framework is much broader than these algorithms and includes further ideas and proposed algorithms for improving NFB in both its traditional and BCI contexts. The ANFB framework is founded on the idea that machine learning and statistical modeling can be used to develop dynamic neurofeedback protocols which adapt to the user throughout training, analogously to how a BCI adapts to its user. Selected proposed mechanisms for achieving this are outlined below as avenues for future research.

One of the fundamental challenges in NFB training is choosing appropriate features of brain activity to train. For clinical NFB, Z-score neurofeedback addresses this problem by identifying features in the resting EEG which deviate significantly from a normalized database and which also have been reported in the literature to be relevant to the particular disorder to be treated [22]. A multilayer extension of this approach might be useful in a generalized BCI. Given an initial pretraining dataset for a set of mental commands and a set of candidate features, those features which deviate from baseline or resting EEG form the first layer of selected features. Of those features, those which deviate across mental commands would form the second layer of features selected for training. Then, rather than using NFB training to normalize those features, NFB training could be applied to further differentiate those mental commands from one another and from baseline EEG using PNFB.

Suppose an optimal method for selecting features for NFB training was already available. There is still no method available at the present time which provides explanatory feedback, *i.e.*, feedback which tells the user which features of brain activity need to be adjusted and in what way [3, 4, 2]. How to instruct a user on how to control specific features of brain activity remains an open question. However, users are likely to find it easier to train to control one feature at a time rather than learning to control all relevant features simultaneously. Therefore it may be beneficial to train consistency and separation in mental commands beginning with the most discriminative feature first and then adding one feature at a time as a desired level of control is achieved. This may, in turn, reduce the number of features required for classification if precise control over a small number of features can be learned by the user. A study testing this approach should consider whether this would increase training time, since it may take some time to learn how to control each feature and then to learn to control features together. In addition, a study testing this approach should consider whether it would change the user's mental commands, since training each feature may not require the original mental commands, and once features are recombined, the original mental command could be significantly altered.

One question which has arisen multiple times throughout this thesis is the question of how to choose optimal mental commands in the first place. For a generalized or open-ended BCI, the choice of mental command is left to the user rather than determined by the design of the system. While individuals might self-select "good" mental commands on average (keeping in mind that many may not), they might not self-select the "best" mental commands. Furthermore, there is no way to choose different mental commands for a given BCI without restarting the training process from the beginning, which introduces the risk of interference from previous training attempts and loss of interest and motivation in the entire endeavor. Therefore finding a good initial set of mental commands prior to training is extremely important.

Should a BCI try to predict an optimal set of mental commands for a user, or should the BCI simply help a user to fine-tune a self-selected mental command? It may be possible to do both, to a degree. Given what is known about how cognitive profiles and aptitude in different areas influence which mental commands yield better BCI control with different individuals [23, 24], including what has been revealed in Chapter 4, it may be reasonable to measure such individual factors and suggest to users that they come up with mental commands which correspond to certain skills they might have in a particular sensory modality. Given mental commands which have been self-selected with some well-reasoned constraints, NFB methods which operate directly in the feature space, such as PNFB, might help to optimize those mental commands for BCI control.

One limitation of the BCIs presented in this thesis is that they do not take advantage of the fact that the user's brain will usually produce an error-related potential when the BCI produces an incorrect output [25]. These potentials can be detected using techniques similar to those employed in P300 BCIs. The advantage gained through detection of error-related potentials during BCI training is that the BCI transducer can avoid being trained with incorrect labels on training data. For a binary BCI, error-related potentials provide a means to automatically label incoming brain activity to continually adapt the BCI transducer, since the detection of an error-related potential indicates that the opposite of the predicted class label was the true class label.

In addition to errors made by the BCI transducer, the user may also make errors from time to time. During training, a user may misread the training cue and produce the incorrect mental command, or simply not pay attention during the trial and miss producing any mental command at all. In both cases, inclusion of these data is detrimental to successful adaptation of the BCI transducer. The first type of error can be partially mitigated with outlier detection (see for example [26]), especially if the generated mental activity is closer to some other mental command than it is to the cued mental command. The second type of error could be mitigated by real-time EEG-based attention monitoring, or mind-wandering detection [27]. However, reliable methods for doing this need to be developed before they can be applied to a BCI.

While the Progressive Thresholding improved performance on an NFB task over standard methods, and PNFB may prove to do the same in the BCI context, an important question remains. Does training with these methods result in stronger neuroplastic changes that lead to improved NFB and BCI use in the long term? The problem of enabling long-term BCI use is a challenging one due to the nonstationarity of the EEG over time. Typically, long-term use requires frequent retraining (*e.g.*, [28]). The amount of training could be reduced if methods like PNFB do lead to more stable patterns of brain activity and more stable neuroplastic changes. This could be tested by observing changes in functional

connectivity before and after training, as seen in successful clinical NFB training [29, 30, 31, 32], and by observing changes in the lasting activations resulting from transcranial magnetic stimulation (TMS), also seen after successful NFB training in the clinical setting [33].

As with the development of generalized BCI transducers, the work on Progressive Thresholding presented in Chapter 6 can only be interpreted so far due to the use of commercial non-standard EEG hardware. While commercial EEG hardware is ideal for NFB applications from a practical point of view, from a research perspective, it introduces certain limitations. In particular, the sensor locations used for measurnig frontal alpha asymmetry were not precisely the same as is usually reported in the literature. Moreover, the reference electrode was in a significantly different location. This electrode configuration may have contributed to the lack of behavioural effects seen in this study, even when NFB training was successful. However, that is mainly a question about the applicability of this particular NFB protocol for influencing mood. The main goal was to show that given a measure of brain activity Progressive Thresholding would lead to better learning outcomes compared to the current standard automatic thersholding approach, and this was clearly achieved.

8.3   Conclusions

The successes of brain-computer interfacing have resulted in much-deserved excitement and anticipation within the brain-computer interfacing and neurotechnology communities. Research in this area has led to markedly improved lives for some individuals suffering from severe cases of paralysis and ALS. In addition, recent advances in controlling software and hardware directly by thought have stimulated the imaginations of many scientists and engineers who see brain-computer interfacing as an unprecedented from of technological progress which could have significant implications for society at large.

While the successes of the brain-computer interfacing community should be celebrated, it must be recognized that the field is only in its infancy. BCIs remain unheard of to the majority of people around the world because they have not yet made a significant impact outside of research laboratories and experimental clinical cases. The reality of the current state of BCI research should serve as a reminder that there is still a long way to go before BCIs are relevant in the lives of everyday people.

The philosophical underpinning of this thesis is that brain-computer interfacing is unlikely to become a useful tool for the general population until a reliable

generalized user-centered BCI is developed. Generalization can enable BCIs to sufficiently account for the enormous variability between the brains of individuals, and only by recognizing that the role of the user is central to BCI performance can optimal co-adaptive methods be developed. The empirical research presented here contributes to the understanding of how humans and machine learning interact to attain reliable BCi performance, introduces new tools for improved user training through NFB, and illustrates why generalization is a valuable approach moving forward. Most importantly, this work brings closer the development of BCIs which can be effective and reliable tools in clinics and homes for a wide variety of applications and users.

Bibliography

[1] C. Brunner, N. Birbaumer, B. Blankertz, C. Guger, A. Kübler, D. Mattia, J. d. R. Millán, F. Miralles, A. Nijholt, E. Opisso, *et al.*, "BNCI Horizon 2020: towards a roadmap for the BCI community," *Brain-computer interfaces*, vol. 2, no. 1, pp. 1–10, 2015.

[2] R. Chavarriaga, M. Fried-Oken, S. Kleih, F. Lotte, and R. Scherer, "Heading for new shores! overcoming pitfalls in bci design," *Brain-Computer Interfaces*, pp. 1–14, 2017.

[3] F. Lotte, F. Larrue, and C. Mühl, "Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design.," *Frontiers in human neuroscience*, vol. 7, no. September, p. 568, 2013.

[4] F. Lotte and C. Jeunet, "Towards improved BCI based on human learning principles," in *Brain-Computer Interface (BCI), 2015 3rd International Winter Conference on*, pp. 1–4, IEEE, 2015.

[5] C. Jeunet, E. Jahanpour, and F. Lotte, "Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study," *Journal of neural engineering*, vol. 13, no. 3, p. 036024, 2016.

[6] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Machine-learning-based coadaptive calibration for brain-computer interfaces," *Neural computation*, vol. 23, no. 3, pp. 791–816, 2011.

[7] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Co-adaptive calibration to improve BCI efficiency," *Journal of Neural Engineering*, vol. 8, no. 2, 2011.

[8] J. Faller, R. Scherer, E. V. Friedrich, U. Costa, E. Opisso, J. Medina, and G. R. Müller-Putz, "Non-motor tasks improve adaptive brain-computer interface performance in users with severe motor impairment," *Frontiers in neuroscience*, vol. 8, p. 320, 2014.

[9] A. Schwarz, R. Scherer, D. Steyrl, J. Faller, and G. R. Müller-Putz, "A co-adaptive sensory motor rhythms brain-computer interface based on common spatial patterns and random forest," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 1049–1052, IEEE, 2015.

[10] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE transactions on information theory*, vol. 14, no. 1, pp. 55–63, 1968.

[11] G. Bin, Z. Lin, X. Gao, B. Hong, and S. Gao, "The ssvep topographic scalp maps by canonical correlation analysis," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 3759–3762, IEEE, 2008.

[12] R. Vigário, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE transactions on biomedical engineering*, vol. 47, no. 5, pp. 589–593, 2000.

[13] F. Darvas, D. Pantazis, E. Kucukaltun-Yildirim, and R. Leahy, "Mapping human brain function with meg and eeg: methods and validation," *NeuroImage*, vol. 23, pp. S289–S299, 2004.

[14] H. C. Peng, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.

[15] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[16] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[17] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized Common Spatial Pattern with aggregation for EEG classification in small-sample setting," vol. 57, pp. 2936–2946, Dec. 2010.

[18] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing.," *Journal of neural engineering*, vol. 9, p. 26013, Apr. 2012.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Deep feature learning for EEG recordings," *arXiv preprint arXiv:1511.04306*, 2015.

[21] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Time series classification using multi-channels deep convolutional neural networks," in *International Conference on Web-Age Information Management*, pp. 298–310, Springer, 2014.

[22] R. W. Thatcher, "Latest developments in live z-score training: Symptom check list, phase reset, and loreta z-score biofeedback," *Journal of Neurotherapy*, vol. 17, no. 1, pp. 69–87, 2013.

[23] E. V. Friedrich, R. Scherer, and C. Neuper, "The effect of distinct mental strategies on classification performance for brain–computer interfaces," *International Journal of Psychophysiology*, vol. 84, no. 1, pp. 86–94, 2012.

[24] C. Jeunet, B. NKaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns," *PloS one*, vol. 10, no. 12, p. e0143962, 2015.

[25] P. W. Ferrez and J. d. R. Millán, "Error-related eeg potentials generated during simulated brain–computer interaction," *IEEE transactions on biomedical engineering*, vol. 55, no. 3, pp. 923–929, 2008.

[26] M. Fatourechi, R. Ward, and G. Birch, "A self-paced brain–computer interface system with a low false positive rate," *Journal of neural engineering*, vol. 5, no. 1, p. 9, 2007.

[27] C. Braboszcz and A. Delorme, "Lost in thoughts: neural markers of low alertness during mind wandering," *Neuroimage*, vol. 54, no. 4, pp. 3040–3047, 2011.

[28] 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, 2010.

[29] J. Lévesque, M. Beauregard, and B. Mensour, "Effect of neurofeedback training on the neural substrates of selective attention in children with attention-deficit/hyperactivity disorder: A functional magnetic resonance imaging study," *Neuroscience Letters*, vol. 394, no. 3, pp. 216–221, 2006.

[30] D. Scheinost, T. Stoica, J. Saksa, X. Papademetris, R. T. Constable, C. Pittenger, and M. Hampson, "Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity," *Transl Psychiatry*, vol. 3, no. 4, p. e250, 2013.

[31] J. Ghaziri, A. Tucholka, V. Larue, M. Blanchette-Sylvestre, G. Reyburn, G. Gilbert, J. Lévesque, and M. Beauregard, "Neurofeedback training induces changes in white and gray matter," *Clinical EEG and neuroscience*, p. 1550059413476031, 2013.

[32] R. C. Kluetsch, T. Ros, J. Théberge, P. A. Frewen, V. D. Calhoun, C. Schmahl, R. Jetly, and R. A. Lanius, "Plastic modulation of PTSD resting-state networks and subjective wellbeing by EEG neurofeedback," *Acta Psychiatrica Scandinavica*, vol. 130, no. 2, pp. 123–136, 2014.

[33] T. Ros, M. a. M. Munneke, D. Ruge, J. H. Gruzelier, and J. C. Rothwell, "Endogenous control of waking brain rhythms induces neuroplasticity in humans," *European Journal of Neuroscience*, vol. 31, no. 4, pp. 770–778, 2010.

Appendices

Appendix A

Filter-Bank Artifact Rejection

Dhindsa, K. (2017). Filter-Bank Artifact Rejection: High performance real-time single-channel artifact detection for EEG. *Biomedical Signal Processing and Control*, Volume 38: 224-235.
Article reprinted with permission.

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Filter-Bank Artifact Rejection: High performance real-time single-channel artifact detection for EEG

Kiret Dhindsa

Neurotechnology and Neuroplasticity Lab, School of Computational Science and Engineering, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

ABSTRACT

Recent developments in electroencephalography (EEG) have led to a variety of consumer-grade EEG devices for brain–computer interfacing (BCI) and neurofeedback (NF) equipped with only one or a few EEG sensors. With minimal electrode coverage, most methods of detecting artifacts in the signal which arise from non-brain sources are not applicable. Furthermore, methods which can be used on single-channel EEG are typically not sufficiently accurate or fast for BCI and NF applications. In this paper a new highly accurate artifact rejection method is introduced, called Filter-Bank Artifact Rejection (FBAR), which is designed for real-time EEG applications using just a few or even a single EEG channel. FBAR is compared to a current state-of-the-art method, Fully Automated Statistical Thresholding for EEG artifact Rejection (FASTER). FBAR outperformed FASTER on all test data, due mainly to its ability to detect small artifacts in the presence of high amplitude EEG. This makes FBAR particularly useful for BCI and NF applications, which are especially dependent on achieving the highest possible signal-to-noise ratio in a real-time setting. A MATLAB toolbox allowing for use and experimentation with FBAR including several customizable options is available as a Git repository at https://bitbucket.org/kiretd/FBAR.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, low-cost consumer-grade electroencephalography (EEG) devices have entered the market with home-use brain–computer interfacing (BCI) and neurofeedback (NF) applications in a variety of areas, including gaming and therapy [1]. NF requires real-time processing of brain recordings in order to provide real-time feedback enabling users to learn how to regulate their own brain activity. The inherent difficulty of learning to control one's own brain activity means that NF applications require precise feedback in order to provide the user with a clear training signal. If this feedback is contaminated by noise, the learning process is likely to be compromised, leading to anything from longer and more tedious training periods to incorrect training and unintended outcomes [2]. While BCI often incorporates NF, BCI applications differ in that they involve the extra step of interpreting user intention from the observed EEG in order to drive a computer application. BCI typically uses highly tuned machine learning models to differentiate between patterns of brain activity, which may differ in only subtle ways. The presence of artifacts in the EEG can distort the features used in these models and therefore prevent them from precisely learning the correct patterns [3,4]. For these reasons, NF and BCI applications are particularly sensitive to noise in the EEG. In order for these devices and applications to be suc-cessful, tools are needed to ensure that the quality of the signal acquired is as high as possible.

As NF and BCI applications enter the consumer market, new processing methods will be needed to compensate for the differences between research-grade and consumer-grade hardware. Consumer EEG devices are susceptible to the usual sources of artifacts found with traditional research devices, including eye blinks and eye movements (electrooculographic, or EOG, artifacts), as well as artifacts from jaw clenches, facial expressions, and other muscle activity (electromyographic, or EMG, artifacts). For research-grade EEG devices, artifacts are typically removed using spatial filtering methods [5], such as blind source separation with independent components analysis (ICA) [6,7], usually requiring at least 14 EEG channels [8], or with Fully Online and Automated Artifact Removal for Brain–Computer Interfacing (FORCe), requiring 16 EEG channels [9]. However, these methods cannot readily be applied to many consumer devices as the requisite number of sensors are simply not available. The lack of appropriate artifact rejection methods for consumer hardware has been identified as one limitations preventing more widespread adoption of consumer BCI and NF applications [10–12].

Methods have been developed for EEG applications involving small-channel or single-channel EEG in order to make applications with these hardware configurations feasible. These methods, such

as the current state-of-the-art, Fully Automated Statistical Thresh-olding for EEG artifact Rejection (FASTER) [13], usually rely on applying thresholds to statistics of the EEG signal. For epoched single-channel data, FASTER uses the mean, variance, amplitude range, and median gradient of the input signal. The statistics for these features are computed across all epochs, and epochs are rejected if the value for any of these features exceeds three stan-dard deviations outside of the mean. However, FASTER and similar methods in the small-channel or single-channel context may not be sufficiently accurate for applications in BCI and NF. In particular, this approach makes it particularly difficult to accurately identify artifacts in the presence of high amplitude EEG.

Other methods are able to suppress, or clean, artifacts from EEG. Recently, a method for cleaning EMG artifacts in single-channel EEG has been developed [14]. However, this method does not work on EOG artifacts, which contaminate the lower frequency bands in the EEG. Similarly, methods for cleaning EOG artifacts in single-channel EEG were also recently developed, but they do not clean EMG arti-facts [15–17]. These methods are only suitable for a narrow variety of artifacts and do not handle small artifacts effectively. While it is possible to use multiple methods sequentially, it is often desirable to have just one tool for handling artifacts, particularly in a real-time application where fast processing is necessary for the success of the application.

In this paper a new method is proposed which performs well in the new and challenging context of artifact detection for single-channel or small-channel EEG. Filter-Bank Artifact Rejection (FBAR) is a fast and highly accurate machine learning based EEG artifact detection method designed specifically for real-time applications using small-channel (fewer than four or six EEG channels) or single-channel EEG and can detect even very small-amplitude artifacts in the presence of high-amplitude EEG. A MATLAB toolbox allow-ing for accessible use of FBAR is available at https://bitbucket.org/kiretd/FBAR.

1.1. Overview of the FBAR approach

FBAR uses amplitude and power statistics in multiple frequency bands in a feature selection and classification pipeline. While this is a common approach in the machine learning literature, it has not been applied to the problem of artifact detection in EEG. The few existing methods based on machine learning are limited for the same reasons as most other methods: they rely on the con-struction of spatial filters using several EEG channels [18,19]. Here we take a different machine learning approach in order to fit the context of detecting artifacts for single-channel EEG.

Since many types of artifacts can be spatially localized in a lin-ear sense, and because the statistics of EOG and EMG signals differ considerably with respect to true EEG signals, spatial filters greatly simplify the problem of identifying spatial components which are reflective of artifacts. However, in the single-channel EEG case, the EEG, EOG, and EMG signals cannot be unmixed using a spatial fil-ter. Therefore, it is more convenient in the single-channel case to frame the problem in terms of identifying segments of the EEG time-series which are contaminated by artifacts. This can be for-mulated as finding a function, which maps time-series segments to a binary variable indicating the pres-ence or absence of an artifact. Since artifacts cannot always be identified on the basis of just one or two features of the signal, especially in the case of small artifacts present among high ampli-tude EEG, a combination of features is required. However, even with an appropriate set of features, determining the optimal combina-tion of features which define the best mapping from time-series segments to artifact identification is a non-trivial problem.

Machine learning methods are particularly well-suited to solv-ing this type of mapping problem [20]. In particular, many modern machine learning classifiers are able to learn non-linear mappings from input data to output classes in high dimensional feature spaces without the need for an analytical solution. This is particularly use-ful for the case of artifact detection in single-channel EEG where the presence of a wide variety of artifacts ensure that some artifacts are indistinguishable from true EEG except for by a combination of some small subset of features, while other artifacts are indistin-guishable from true EEG by those same features but are classifiable based on an entirely different combination of features.

Broadly, there are three machine learning approaches which are appropriate for the problem at hand. The first is a purely feature engineering approach, in which a set of features are computed that are known to be discriminative with respect to clean EEG and artifacts a priori. The success of this approach depends on the degree of domain knowledge and a priori information that can be included in the handcrafted feature space. Since single-channel artifact detection has not yet been studied deeply enough for com-plete characterization of an optimal feature space on the basis of a priori knowledge alone (especially in the case of detecting small-amplitude artifacts), a purely feature engineering approach would require several assumptions and educated guesses. It can thus be expected to result in a suboptimal solution. Furthermore, because the goal of FBAR is to be applicable for as many kinds of arti-facts as possible, using purely preselected features would introduce increased risk of poor generalizability for artifacts which were not considered (or for which insufficient data were available) at the outset.

A second approach is to use deep learning [21], which in the-ory, is capable of learning a set of discriminative features from the EEG time-series itself [22,23]. While this is a promising avenue for artifact detection in EEG, deep learning is not without its down-sides. Of particular relevance to the current problem, deep learning can be said to replace laborious feature engineering with laborious architecture engineering (i.e., experimentally finding a good com-bination of network architecture and hyperparameter settings). While deep learning avoids the need for a high degree of domain knowledge, it is often the case that if sufficient domain knowledge is available, a feature engineering or a feature selection approach is more efficient and potentially as successful.

Here a third approach was used, which could be seen as a bal-ance between a purely feature engineering approach and a deep learning approach. In the feature selection approach, an engineered feature space based on available domain knowledge is constructed just as in the purely feature engineering approach. However, these are treated merely as candidate features, rather than as the final set of features used for classification. These candidate features are passed through a supervised feature selection algorithm in order to reduce the dimensionality of the feature space in a way which preserves as much discriminative information as possible. Compared to the purely feature engineering approach, the feature selection step has the added benefit of reducing the dimensionality of the feature space, and thus the classification model, by remov-ing redundant and irrelevant features. While the feature selection approach uses some a priori information about how artifacts are characterized in the EEG time-series, it is more forgiving when only partial knowledge is available and when some types of arti-facts are not well-represented in the training data. Since the feature selection approach provides an efficient means of generating a discriminative feature space given incomplete domain knowledge which is not sufficiently fine-grained, it avoids the need for the expensive model selection experiments which are required in the deep learning approach.

The feature selection approach still demands that some domain knowledge is available with which to construct a useful list of can-didate features to begin with. A discussion of which features were

included in this list and why they were chosen to be candidate features is given in Section 2.5.

2. Methods

2.1. The Muse EEG headband

The current study was conducted using Interaxon's Muse headband [24], a recently released commercial EEG headset that comes paired with an NF application. The Muse headband is a high quality dry electrode EEG headset with four channels located at Fp1, Fp2, TP9, and TP10 with reference at Fpz and DRLs one inch from the reference on both sides. EEG is sampled at 2 kHz and downsampled to 220 Hz for analysis. The headband is worn like a pair of glasses, with the frontal electrodes placed over the middle of the forehead, and the rear electrodes placed behind the ears. Previous studies using the Muse headband have shown that high quality EEG can be obtained (e.g., [25]). Furthermore, simultaneous recording with the Muse headband and a Brain Vision actiCHamp system [26] conducted by Interaxon produced a nearly identical signal, while a simultaneous recording with the Muse headband and a g.Tec g.USBamp system [27] during a battery cognitive tasks conducted by McMaster University showed comparable performance between the two systems [28].

2.2. Data and participants

Data from nine individuals using Muse EEG headbands were acquired for training, all using the four-channel electrode configuration described above in Section 2.1. Three training sets (referred to as Datasets 1, 2 and 3 throughout this paper) were collected from three right-handed male volunteers between the ages of 21 and 27. A different Muse headband was used for data collection with each participant in order to account for any variability in hardware performance across headbands. Since these were final-product manufactured models, little variability was expected. However, since the goal was to develop an artifact detection method which would generalize well across users, each potentially using a different headband, multiple Muse headbands were used for data collection. The volunteers for Datasets 1, 2, and 3 were verbally instructed to perform different artifact-generating actions on cue, including various eye movements, blinks, jaw clenches, and facial expressions.

Six of the training sets (denoted Datasets 4 through 9 throughout this paper) were obtained from Interaxon's database of anonymized EEG data. Thus their genders, ages, and handedness were all unknown. These data were collected from customers who had purchased a Muse headband and volunteered to have their data saved for research purposes. All six of these datasets were acquired from customers using an application designed to facilitate an eyes-closed mindfulness meditation exercise with neurofeedback (Datasets 4 through 9 also include some eyes-open data while participants performed a signal quality check within the meditation application). The application also provides clear instructions for wearing the headband, whose electrode configuration is unchangeable. Therefore it can be reasonably expected that the same electrode configuration was used as described above. These datasets contain very high alpha wave activity, leading several existing artifact detection methods to fail because the amplitudes of the alpha waves are much higher than typical EEG.

In total, approximately 157 min of EEG data were included as training data (average and standard deviation of dataset length: 17.5 ± 7.0 min). Since Datasets 1, 2, and 3 were obtained by having subjects specifically perform several kinds of artifact-generating actions, an especially large number and variety of artifacts is

Table 1
A summary of relevant properties of the datasets used for training and testing the artifact detection model. Length is given in minutes, and proportions are given as the average proportion of EEG samples labeled as EOG or EMG across channels for each dataset.

Dataset	Length	Prop. EOG	Prop. EMG
1	21.9	0.374	0.100
2	22.7	0.194	0.083
3	22.0	0.299	0.046
4	21.9	0.063	0.007
5	13.4	0.061	0.016
6	8.3	0.720	0.026
7	4.2	0.065	0.025
8	21.5	0.049	0.008
9	21.5	0.104	0.036

included in these training data. Table 1 summarizes the length of each dataset and proportion of artifacts contained therein. Datasets 1, 2, and 3 are available with the FBAR MATLAB toolbox, but because Datasets 4 through 9 are the property of Interaxon, they cannot be provided for download.

2.3. Data labeling

FBAR uses supervised machine learning in order to detect the presence of an artifact. Therefore, labeled training data are required in order to train an FBAR model. The process of acquiring labeled data can be laborious, especially if the data is labeled manually, either in the time-series or in previously windowed EEG. Thus it is required that someone experienced with EEG data visually scan through the available data and mark artifacts by hand. While manual labeling is not perfect and cannot be taken as a gold standard, it is the best method available given that the goal is to develop an objective method which can automatically label such data.

Manual labeling was performed by an EEG researcher with eight years of experience working with EEG, and these labels were checked and confirmed by a senior researcher with more than 20 years of EEG signal processing experience. The labelers were aware that the purpose of labeling was to develop an artifact detection method appropriate for the Muse headband and similar hardware. However, the labels were provided before beginning work on developing the method. The labelers were not aware whether the method would be for single-channel EEG or multi-channel EEG, whether or not it would utilize a machine learning approach, or what kind of information would be extracted from the signals in order to perform artifact detection. Each of these factors were determined after labeling had been completed.

Since FBAR requires epoched EEG for feature extraction, labeling methods which produce labels in the time-series must use a method to convert those labels into an epoched form. In order to avoid predominately clean epochs from being mislabeled as containing an artifact, which can occur for epoch containing edges of artifacts, a cutoff parameter γ was used. The parameter γ denotes the proportion of an epoch which must be labeled as artifactual in the time-series before that epoch as a whole is labeled as artifactual. Small values of γ risk labeling as artifactual epochs which are mostly clean, and may produce models prone to false positives (labeling clean epochs as contaminated). However, large values of γ are very liberal with labeling epochs as clean, and may lead to models prone to false negatives (labeling contaminated epochs as clean). The FBAR toolbox leaves γ as a parameter to be set by the experimenter depending on their priorities with respect to leniency or strictness in artifact detection. The value of γ, which should usually be set between 0 and 0.5, can also be chosen empirically via cross-validation. Fig. 4 shows that the FBAR models generated with the datasets used in this paper were not very sensitive to values of γ between 0 and 0.2, but that higher values resulted in worse per-
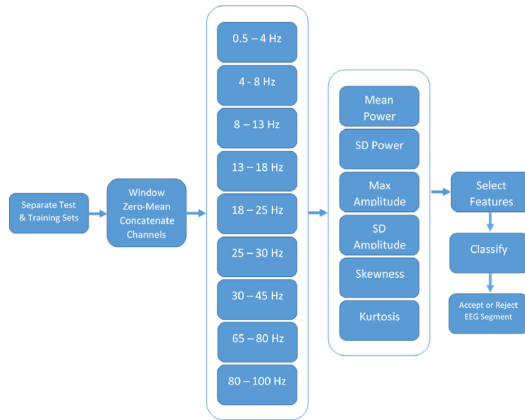
Fig. 1. A diagrammatic description of the FBAR processing pipline.

formance. For the purposes of this paper, the value of γ was set to 0.2 throughout.

In the following subsections, the FBAR processing pipeline is described. This processing pipeline is illustrated in Fig. 1.

2.4. Data preprocessing

Throughout this paper, EEG signals were windowed into one-second epochs ($n = 220$ samples per epoch) with a 0.1 s shift, which mimics many real-time processing schemes. However, the FBAR toolbox allows for both the window size and and the shift between windows to be set by the experimenter. Features were computed for each epoch after the mean of the EEG time series in that epoch was removed. After computing all of the features in the training set, each feature was normalized by that feature's mean and standard deviation in the training set. The feature statistics computed from the training set were used to normalize features in any test sets or validation sets.

2.5. Features

FBAR separates the EEG signal from each epoch into multiple frequency bands and extracts features from each band. Third-order Butterworth filters were used to filter the signals. Power-band features and amplitude statistics within each band were computed for each frequency band. While the FBAR toolbox provides experimenters with the options to select custom frequency bands and features, in this study a standard set of features were used throughout based on domain knowledge of common sources of artifacts in EEG. The following frequency bands were used: 0.5–4 Hz, 4–8 Hz, 8–13 Hz, 13–18 Hz, 18–25 Hz, 25–30 Hz, 30–45 Hz, 65–80 Hz and 80–100 Hz. Frequency bands between 45 Hz and 65 Hz were omitted because notch filters were applied at 50 Hz and 60 Hz in the Muse hardware to reduce environmental noise. The following features were extracted from the signal x in each frequency band for each epoch: mean signal power, standard deviation of signal power, maximum absolute amplitude, standard deviation of amplitude, skewness, and kurtosis. In total, 54 features were computed for each epoch (nine frequency bands with six features in each).

The frequency bands used correspond to the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (>30 Hz) bands commonly used in EEG research, with beta and gamma broken up in three bands each due to their increased width.

Since there are many subtypes of EOG and EMG artifacts which can occur across a wide range of frequencies, a wide range of frequency bands were used in feature extraction. For example, EOG artifacts typically have low frequency components which affect the delta and theta bands, and they sometimes exhibit noise which affects the alpha band. The goal was to ensure that enough features would be computed in order to capture the difference between clean EEG and any type of artifact. Since with FBAR the number of features depends on the number of bands used, three roughly evenly spaced bands which correspond to the delta, theta, and alpha bands commonly found in the EEG literature were used to address the problem of EOG artifacts.

The same reasoning was used for EMG artifacts, which typically reside in the beta and gamma bands. However, since the frequency range of EMG artifacts is much larger, wider bands were used to avoid dedicating a comparatively large number of features to EMG artifacts. Furthermore, very little EEG activity of interest occurs in the gamma range under normal circumstances and for most EEG research. Moreover, what activity does occur in the gamma band is usually of small amplitude. Therefore, EMG artifacts stand out more easily in these high frequency ranges, and broad frequency bands could be used without great risk that the features extracted from these wide bands would not be discriminative.

FBAR can be implemented with different frequency bands, and the FBAR toolbox allows researchers to define their own filter-bank so that FBAR can be adapted to different data when needed. It should be noted that while the combination of frequency bands and features used here was effective for the data available and provided a good balance of breadth versus the dimensionality of the feature space, they may require some adjustment depending on the specific application or hardware used.

The specific definition for each feature and the reason they were included are given below.

2.5.1. Mean power

Signal power $\mathcal{X}$ was calculated as the squared magnitude of the discrete Fourier Transform using the FFT (Fast Fourier Transform) algorithm over the Hamming-windowed signal. The mean power was calculated as

$$\frac{1}{K_B}\sum_{i=1}^{K_B}\mathcal{X}_B,$$

where X_B is the power spectrum within a frequency band B and K_B is the number of points in $\mathcal{X}$ which fall within B. Mean power is a commonly used measure for characterizing both EOG [29] and EMG [30] signals. These artifacts typically arise from sources closer to the EEG sensors than the brain and also produce larger electromagnetic fields. As a result, they produce noticeable spikes in the EEG power spectrum which affects its first and second order statistics.

2.5.2. Standard deviation of power

Similarly to the mean power, the standard deviation of signal power was calculated for each frequency band as

$$s_n = \sqrt{\frac{1}{K_B - 1}\sum_{i=1}^{K_B}(\mathcal{X}_{Bi} - \bar{\mathcal{X}}_B)^2}.$$

As in the case of mean power, the large fluctuations in power introduced by the presence of artifacts can result large differences in the standard deviation of the power spectrum for contaminated epochs versus clean epochs, particularly for large artifacts.

2.5.3. Max absolute amplitude

The max absolute amplitude was calculated for the filtered signal x_B as

$$\max |x_B|.$$

The increases in power that arise due to EOG and EMG artifacts are associated with large amplitude changes in the EEG time-series. For this reason, first and second order amplitude statistics are commonly used in artifact detection [4], including in FASTER [13]. However, here the max absolute amplitude was used over the mean amplitude in order to account for the possibility of large amplitude artifacts with short durations compared to the length of an epoch. Such artifacts are more easily noticed with respect to the max absolute amplitude of the signal, because their short duration may result in a negligible impact on the mean of the signal.

2.5.4. Standard deviation of amplitude

The sample standard deviation of amplitude was calculated as

$$s_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_{B,i} - \bar{x}_B)^2},$$

where $n = 220$ is the number of samples in each EEG epoch. As above, this feature was included because the amplitude changes seen with medium and large artifacts often result in large changes in the second order statistics of the EEG time-series.

2.5.5. Skewness of amplitude

Corrected sample skewness was calculated as

$$g_n = \frac{n^2}{(n-1)(n-2)} \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{B,i} - \bar{x}_B)^3}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_{B,i} - \bar{x}_B)^2\right]^{3/2}}.$$

Skewness was included in the list of candidate features because many artifacts are asymmetric and thus result in deviations from normality in the EEG signal. Deviations in skewness can be expected even when the artifact is too small to be detectable based on its first and second order statistics. In fact, skewness has been found to be useful when determining which components obtained from ICA reflect artifacts because of the asymmetry they introduce [31,19]. This is especially true for EOG artifacts.

2.5.6. Kurtosis of amplitude

Finally, corrected sample kurtosis was calculated as

$$k_n = \frac{n-1}{(n-2)(n-3)} ((n+1)k_1 - 3(n-1)) + 3,$$

where k_1 is the uncorrected sample kurtosis given by

$$k_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{B,i} - \bar{x}_B)^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_{B,i} - \bar{x}_B)^2\right]^2}.$$

Like skewness, kurtosis has previously been used as a feature to identify artifactual components after ICA because EOG artifacts tend to have high kurtosis while the kurtosis of clean EEG is usually close to zero [32]. Kurtosis was included in FBAR because small EOG artifacts should exhibit elevated levels of kurtosis due to their morphology even if they do not appear unusual in their first and second order statistics.

2.6. Feature selection

Using a robust feature selection method with FBAR serves several purposes. Feature selection reduces the number of features used in the classification model, thereby reducing computation
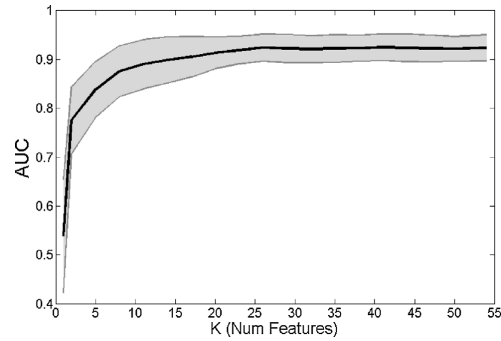

Fig. 2. Average cross-validation accuracy (and shaded standard deviation) by number of selected features, K, using $D = 8$. Accuracy is measured by testing on the labeled dataset which was not included in training.

time during real-time artifact detection. Feature selection also allows experimenters to cast a wide net by computing many features in many frequency bands when the correct features to use cannot be determined by a priori knowledge alone. Finally, feature selection often improves classification accuracy by removing redundant and noisy features. Minimum Redundancy Maximum Relevance (MRMR) was used for feature selection in this study (full details found in [33]). MRMR is an information theoretic supervised feature selection method which aims to maximize the mutual information between a subset of features in a training set with the true class labels while simultaneously minimizing the joint mutual information among the selected features. Features were selected on the basis of training data alone, and those selected features were used for testing and validation. The number of features selected, K, is a parameter that must be chosen by the experimenter or set experimentally via cross-validation. Fig. 2 shows the results of cross-validation over K.

2.7. Modeling

For simplicity, results are only presented using a support vector machine (SVM) with a radial basis function (RBF) kernel. In principle, any classifier can be used with FBAR. However, earlier analyses using only Datasets 1, 2 and 3 showed that an SVM with an RBF kernel yielded slightly better classification results compared to other methods. Nonetheless, the FBAR toolbox includes options for several classifiers, including a multi-layer perceptron (MLP), the choice of SVM kernels and parameters, and linear and quadratic discriminant analysis (LDA and QDA). In addition, no attempt was made to optimize SVM hyperparameters, as performance was already sufficiently high.

All of the FBAR models presented in this paper were trained using data from all EEG channels and were built to generalize to both low frequency and high frequency artifacts. However, the FBAR toolbox includes options to build separate models for different types of artifacts, as well as to build separate models for each EEG channel, which may be desirable in some cases. Building separate models for low frequency and high frequency artifacts did yield slightly higher classification accuracy in preliminary analysis with Datasets 1, 2, and 3 (increases between 0.5% and 2% time-series accuracy depending on the test set), but at the cost of increasing classification time since two classifiers were needed rather than one. Building separate models for each EEG channel is only necessary if different EEG channels have very different characteristics. This is typically not the case, and the cost is a reduction of training data available per model by a factor of the number of EEG channels.
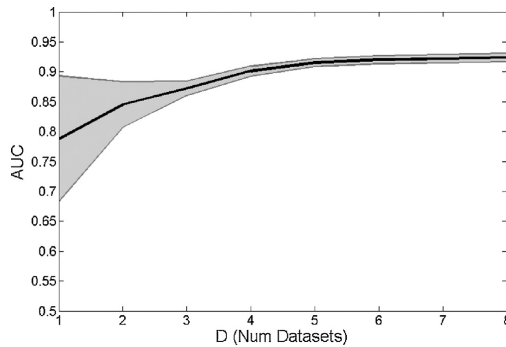
Fig. 3. Average cross-validation accuracy (and shaded standard deviation) by number of training sets used, *D*, using *K* = 25. Accuracy is measured by averaging test results on all labeled datasets which were not included in training.
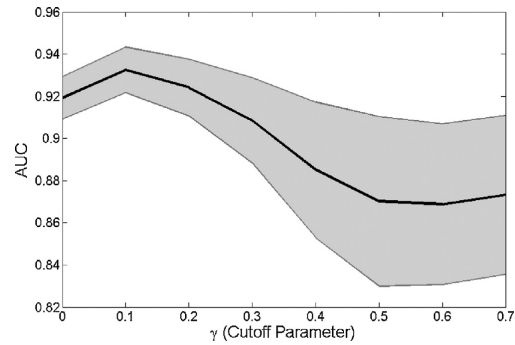

Fig. 4. Average cross-validation accuracy (and shaded standard deviation) over values of the cutoff parameter, γ, using *K* = 25. Accuracy is measured by testing on the labeled dataset which was not included in training.

The Muse EEG headset used in this is an exceptional case in that the frontal channels and the temporal channels are made of different materials, and therefore do show substantially different EEG statistics and characteristics (the frontal channels generally record signals with much smaller amplitudes and much lower power). FBAR was able to generalize across all channels regardless of this fact, suggesting that models for individual channels may not be necessary in most cases. With all channels combined, a total of 628 min of labeled data was available (157 min times four channels).

Aside from the selection of frequency bands, the most important parameters to set with FBAR are the number of features, *K*, and the number of training sets, *D*. An optimal value of *K* can be selected based on the error curve across possible values of *K*, as shown in Fig. 2. The number of training sets *D* is mainly important because a minimum *D* is desired both to reduce the computational time required to train the FBAR model, but especially to reduce the labour involved in labeling data.

Cross-validation results over *K* are presented in Fig. 2. These results were obtained using leave-one-out (LOO) cross-validation, where for each value of *K*, each dataset was used for testing once while the other eight datasets were used for training (therefore nine models were generated for each value of *K*). The same procedure was used to obtain cross-validation results over γ, shown in Fig. 4, except that *K* was fixed to a value of *K* = 25, corresponding to the value of *K* found to be close to optimal.

Cross-validation results over *D* are presented in Fig. 3 in order to provide an estimate of how much labeled data is be required to obtain a sufficiently generalizable artifact detection model with FBAR. These results were obtained by generating an FBAR model with every possible combination of *d* training sets for each value of $d \in 1 \ldots D$, and averaging the test results of every labeled dataset which was not included in the *d* training sets. Hence, for each value of *d*, $\binom{D}{d}$ models were generated and $D - d$ tests were performed with each model. Note that for applications using highly controlled conditions during EEG recording, fewer training sets may be sufficient compared to what is suggested by Fig. 4. This is because the nine training sets used here contain a mix of eyes-open and eyes-closed EEG collected in both a controlled laboratory setting and in an uncontrolled home setting while users performed a meditation exercise. Moreover, some EEG samples were selected specifically to address the challenge of exceptionally high amplitude alpha EEG exhibited by some expert meditators. Thus the data used here may be significantly more varied than what is typically seen in a research

setting, requiring a greater degree of generalization in the FBAR models than would be otherwise needed.

3. Results

3.1. Cross-validation results

The results of each cross-validation experiment are presented below. Based on the results of LOO cross-validation over *K*, shown in Fig. 2, the number of selected features was set to *K* = 25 for all subsequent analyses. The specific features selected by MRMR are shown in Table 2. An interesting symmetry emerges from this table, which could help to reveal a deeper understanding of the structure of various artifacts in future work. Power statistics were only important in this model in the lowest and highest frequency bands. Similarly, the maximum and standard deviation of the amplitude was important for the lowest two and highest three frequency bands, while not being needed in the middle frequency bands. In contrast, skewness was an important feature for every frequency band, while kurtosis was only used in the two lowest frequency bands. Further discussion regarding the importance of these features is provided in Section 4.

Results of cross-validation over the number of training sets used, *D* with *K* = 25 is presented in Fig. 3. AUC approached convergence after four or five training sets. There was a sharp decrease in the variance in AUC from *D* = 1 to *D* = 3 because models which were trained using only the non-meditation EEG datasets usually misclassified very high amplitude alpha waves seen during successful meditation as artifacts. With less than four subjects included in the training set, some combinations of subjects used for training

Table 2

Features selected for FBAR trained with *D* = 9 and *K* = 25. The same features were selected by MRMR for the LOO cross-validation results presented in Tables 4–7. Selected features are marked with a •.

Filter	Mean Pow.	SD Pow.	Max Amp.	SD Amp.	Skewness	Kurtosis
0.5–4 Hz	•	•	•	•	•	•
4–8 Hz			•	•	•	•
8–13 Hz					•	
13–18 Hz					•	
18–25 Hz					•	
25–30 Hz					•	
30–45 Hz			•	•	•	
65–80 Hz			•	•	•	
80–100 Hz	•	•	•	•	•	

included none of the data acquired during meditation task, thus resulting in models which generalized poorly to this condition.

As mentioned previously, the cutoff parameter γ was set to 0.2 for all analyses done in this paper, as it is included in the FBAR tool-box only as a heuristic. The effect of γ on AUC with $K=25$ and $D=8$ is shown in Fig. 4. FBAR models do not appear to be very sensitive to γ until γ is set to be greater than 0.2 or 0.3.

3.2. Comparison with FASTER and EEGLAB

FASTER [13] was used to detect artifacts in the nine training sets used in this study. Additionally, FASTER was used in conjunction with the remaining single-channel artifact detection methods available in EEGLAB [34] (specifically, EEGLAB's amplitude and variance thresholding algorithms, kurtosis-based algorithm, linear trend algorithm, and spectrum-based algorithm). For simplicity, the combination of these methods is referred to as FASTER+. The results are given in terms of time-series accuracy (the percentage of EEG samples correctly classified), area under the ROC curve (AUC), false negative rate (FNR) and false positive rate (FPR). These are shown in Tables 4–7 respectively. FBAR results were obtained using LOO cross-validation with $K=25$.

While FASTER performed well in terms of time-series accuracy for the eyes-closed datasets, it performed poorly on the eyes-open datasets, leading to a high variance in the classification accuracy (grand average 85.2 ± 12.1). More importantly, AUC was more consistently low across all datasets (grand average 0.713 ± 0.062) with high FNR (grand average 0.141 ± 0.146) on eyes open datasets, and high FPR (grand average 0.239 ± 0.169) on eyes closed datasets. This was especially true for those datasets characterized by high amplitude alpha EEG during meditation.

Visual inspection revealed that the high FNR was due to small artifacts tending to be left undetected by FASTER, but FASTER and FASTER+ also missed several typical eye blink artifacts in some datasets (see Fig. 8 for examples). The high FPR in Datasets 4 through 9 were due to the unusually high alpha waves being falsely marked as artifacts (see Fig. 7 for an example). Both FASTER and FASTER+ yielded similar results with a similar pattern of misclassification. Time-series classification accuracy was more consistent with FASTER+ compared to FASTER, but a slightly lower accuracy was obtained on average (grand average 82.5 ± 7.3). However, the AUC was both higher and more consistent (grand average 0.759 ± 0.038, false negative rate 0.112 ± 0.121, false positive rate 0.432 ± 0.208). Therefore, the exemplars shown in the figures below focus on FASTER+ rather than FASTER.

FBAR performed better than FASTER or FASTER+ both in terms of time-series accuracy (grand average 93.3 ± 2.3) and AUC (grand average 0.923 ± 0.016, FNR 0.024 ± 0.045, FPR 0.206 ± 0.074) for every dataset. The difference in performance was especially notable in terms of the AUC, which was due to the ability of FBAR to simultaneously detect small artifacts and recognize the high alpha activity of meditators as clean EEG. Paired-sample T-tests were used to evaluate the differences in performance between FBAR and FASTER or FASTER+. FBAR significantly outperformed FASTER and FASTER+ in terms of time-series accuracy (FASTER: $t_8 = 2.37$, $p < 0.05$; FASTER+: $t_8 = 5.35$, $p < 0.001$), AUC (FASTER: $t_8 = 10.68$, $p < 10^{-5}$; FASTER+: $t_8 = 13.45$, $p < 10^{-6}$) and FNR (FASTER: $t_8 = -3.10$, $p < 0.05$; FASTER+: $t_8 = -3.09$, $p < 0.05$). With respect to the FPR, FBAR did not outperform FASTER ($t_8 = -0.61$, $p = 0.56$), but it did outperform FASTER+ ($t_8 = -3.43$, $p < 0.01$). Therefore, FBAR only failed to outperform FASTER on one measure, FPR, and outperformed both FASTER and FASTER+ in every other case. The reason for this is given further below.

Table 3 displays the results of Kolmogorov–Smirnov (KS) tests of the normality assumption in order to justify the use of T-tests. Paired T-tests require that the difference between the paired data is
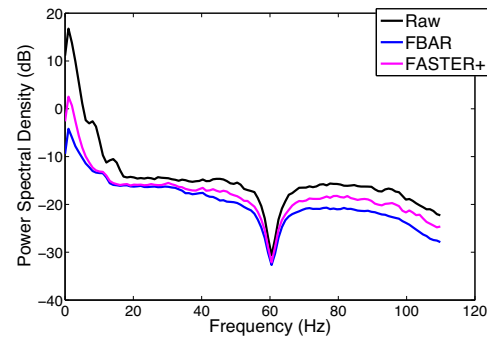

Fig. 5. Power spectra for one channel of a new dataset taken from Interaxon's research database and not used in training.

normally distributed. Therefore, KS tests of the difference between of the compared results were conducted. In each case, the test failed to reject the null hypothesis that the difference was normally distributed, thus justifying the use of paired T-tests.

In addition to the above results, FBAR was tested on 30 new datasets after training with all nine labelled datasets. However, since these new datasets were not prelabeled, visual inspection was used to confirm that FBAR results in similar changes to the power spectrum and accurately detects both small and large artifacts. Examples of these are shown in Figs. 5 and 8 .

The power spectra of the original EEG signal, the signal marked clean by FBAR, and the signal marked clean by FASTER+ were compared. Fig. 5 shows the corresponding power spectral densities (PSD) obtained with a new dataset randomly selected from Interaxon's research database that was not part of the training set. FBAR greatly reduced power in the low frequencies (particularly 0–8 Hz, with a mean difference of 4.6 dB, $p < 0.01$ in a paired-sample T-test) due to a greater reduction in EOG compared to FASTER+. FBAR also more significantly reduced power in the high frequency ranges (40 Hz and above, with a mean difference of 2.1 dB, $p < 10^{-16}$ in a paired-sample T-test), where EMG artifacts typically reside. However, FBAR and FASTER+ maintain close to the same power in the alpha (mean difference of 0.5 dB, $p = 0.02$) and beta (mean difference of 0.3 dB, $p = 0.01$) frequency bands. Note that some difference in the alpha and beta bands is expected because FBAR removes more artifacts than FASTER+, and thus removes the contribution of these artifacts to the PSD in the alpha and beta bands. These PSD results, when taken together with what can be seen from visual inspection of the EEG time-series, suggests that FBAR is able to remove EOG and EMG artifacts without significantly affecting the true EEG signal more effectively than FASTER+. Additional examples illustrating how FBAR and FASTER+ compared when viewing the EEG time-series are given below.

Visual inspection of the EEG signals after they had been marked by FBAR and FASTER+ provided insight into some of the patterns seen in the benchmarking tables above. Most notably, FBAR produced a greater proportion of false positives compared to FASTER in Datasets 1, 2, and 3 (eyes-open EEG), but a much smaller proportion of false positives in Datasets 4 through 9 (eyes-closed meditation EEG). The eyes-open datasets contained several small artifacts resulting from small eye movements and other muscle movements which were not labeled in the training data, but were marked as artifacts by FBAR nonetheless, resulting in an inflated false positive rate. These small artifacts were often missed by FASTER and FASTER+. It is important to note, however, that there are also exam-

Table 3

The results of Kolmogorov–Smirnov tests of normality for the difference in results between FBAR and FASTER and between FBAR and FASTER+ with respect to each performance measure.

Method	Accuracy	AUC	FNR	FPR
FBAR–FASTER	0.35 ($p = 0.17$)	0.28 ($p = 0.42$)	0.32 ($p = 0.26$)	0.24 ($p = 0.61$)
FBAR–FASTER+	0.17 ($p = 0.92$)	0.22 ($p = 0.72$)	0.34 ($p = 0.20$)	0.14 ($p = 0.98$)

Table 4

Time-series classification accuracy using all methods. Each value is the average across all four channels.

Method	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Avg. (SD)
FASTER	60.2	83.3	70.1	91.9	91.0	92.1	91.3	95.0	92.1	85.2 (12.1)
FASTER+	68.5	79.8	78.4	77.8	89.1	81.4	89.5	89.6	88.8	82.5 (7.3)
FBAR	**90.2**	**91.1**	**90.5**	**95.1**	**92.7**	**93.9**	**94.4**	**96.7**	**94.8**	**93.3 (2.3)**[a, b]

Bold font denotes the best performance value per column (dataset).
- [a] Denotes FBAR performed significantly better than FASTER.
- [b] Denotes FBAR performed significantly better than FASTER+.

Table 5

AUC using all methods. Each value is the average across all four channels.

Method	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Avg. (SD)
FASTER	0.603	0.743	0.632	0.686	0.775	0.711	0.734	0.750	0.784	0.713 (0.062)
FASTER+	0.685	0.753	0.748	0.734	0.808	0.761	0.755	0.788	0.801	0.759 (0.038)
FBAR	**0.914**	**0.912**	**0.906**	**0.935**	**0.916**	**0.944**	**0.909**	**0.918**	**0.950**	**0.923 (0.016)**[a, b]

Bold font denotes the best performance value per column (dataset).
- [a] Denotes FBAR performed significantly better than FASTER.
- [b] Denotes FBAR performed significantly better than FASTER+.

Table 6

False negative rate using all methods. Each value is the average across all four channels.

Method	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Avg. (SD)
FASTER	0.439	0.177	0.324	0.059	0.046	0.075	0.053	0.027	0.067	0.141 (0.146)
FASTER+	0.374	0.152	0.237	0.040	0.036	0.049	0.048	0.021	0.056	0.112 (0.121)
FBAR	**0.141**	**0.028**	**0.013**	**0.003**	**0.002**	**0.003**	**0.016**	**0.008**	**0.006**	**0.024 (0.045)**[a, b]

Bold font denotes the best performance value per column (dataset).
- [a] Denotes FBAR performed significantly better than FASTER.
- [b] Denotes FBAR performed significantly better than FASTER+.

Table 7

False positive rate using all methods. Each value is the average across all four channels.

Method	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Avg. (SD)
FASTER	**0.041**	**0.123**	**0.027**	0.397	0.420	**0.162**	0.316	0.480	**0.188**	0.239 (0.169)
FASTER+	0.135	0.289	0.154	0.728	0.545	0.516	0.475	0.653	0.396	0.432 (0.208)
FBAR	0.142	0.207	0.192	**0.230**	0.236	**0.075**	**0.191**	0.233	0.206	**0.206 (0.074)**[b]

Bold font denotes the best performance value per column (dataset).
- [a] Denotes FBAR performed significantly better than FASTER.
- [b] Denotes FBAR performed significantly better than FASTER+.

ples of genuine false positives, where portions of EEG which appear clean on visual inspection were marked as artifacts.

Fig. 6 shows an example of small EMG artifacts in Dataset 2 (produced when participants were instructed to "gently smile") which were more reliably detected by FBAR compared to FASTER+. Fig. 7 shows an example from Dataset 4 where FASTER and FASTER+ incorrectly labeled high amplitude alpha activity as artifacts. Examples of very small EOG artifacts can also be seen, which are typically generated by small eye-movements that naturally occur while the eyes are closed. The PSDs for this dataset reveal that even though FBAR and FASTER+ were equal in most frequencies, only FBAR was able to maintain the original signal in the alpha and beta frequencies. Finally, Fig. 8 shows an example of a dataset separate from the training data for which FASTER+ performs poorly, but FBAR successfully marks artifacts accurately and precisely. Also in Fig. 8, examples of small artifacts which are typically not detected by other methods are visible.

3.3. Online processing speed

A real-time application requires that artifact detection takes minimal computational resources so that as much CPU time is available for running the algorithms associated with the application itself. In order to provide an estimate of how much CPU time is required by FBAR, the amount of time taken to output the presence or absence of an artifact for one 1s sample of single-channel EEG was measured. This included all steps after data acquisition (filtering, feature extraction, normalizing features by the training set statistics, and classification). A Windows 7 desktop computer using MATLAB R2013a with an Intel Core i5 3.30 GHz CPU and 12 GB of RAM was used to run this test. The results are shown in Fig. 9, indicating that FBAR does process data with sufficient efficiency for real-time applications. However, because channels are processed independently, processing time will be multiplied for each channel of EEG, unless parallel processing is performed. Therefore it is recommended that FBAR only be used in its intended context, which
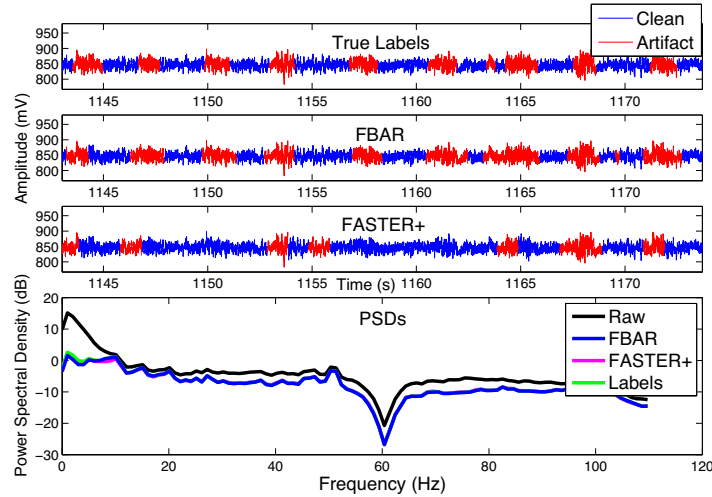
Fig. 6. Example of FBAR correctly detecting small EMG artifacts produced by gentle smiles in Dataset 2. FASTER and FASTER+ are often unable to detect artifacts of this size. PSDs of the entire signal are also shown.
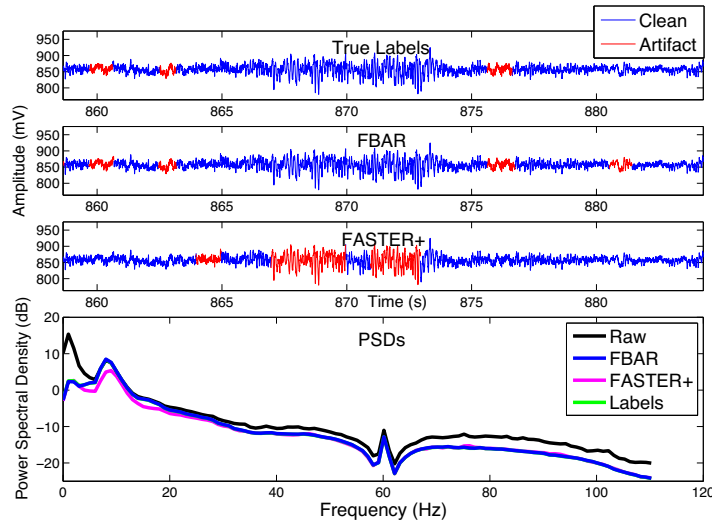

Fig. 7. An example from dataset 4, where FASTER and FASTER+ often mistake high amplitude alpha EEG for artifacts. PSDs of the entire signal are also shown.

is single-channel or small-channel EEG with fewer than five or six EEG channels.

For comparison, the same test was conducted with FASTER and FASTER+ in a simulated online setting. The processing time for FASTER was 0.176 ± 0.012 s and FASTER+ was 0.184 ± 0.009 s per epoch, which is much higher than FBAR ($p \approx 0$ for both comparisons based on two-sample T-tests). Thus FBAR is significantly more efficient for real-time artifact detection.

4. Discussion

The results of this study show FBAR to be a highly accurate artifact detection algorithm for real-time single-channel or small-channel EEG. FBAR significantly outperformed the popular FASTER algorithm, as well as the combination of FASTER and additional EEGLAB artifact detection tools both on the training data presented here and on independent test data.
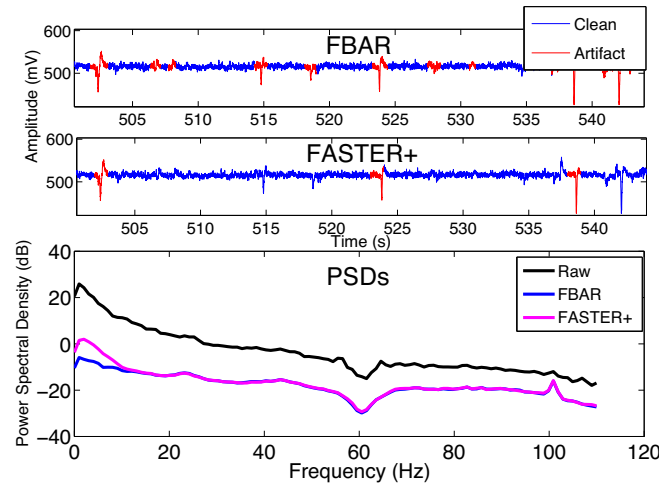
Fig. 8. Example from a new dataset taken from the Interaxon research database where FASTER+ performs particularly poorly while FBAR performs well. PSDs of the entire signal are also shown. True labels are not available because these data have not been given training labels.
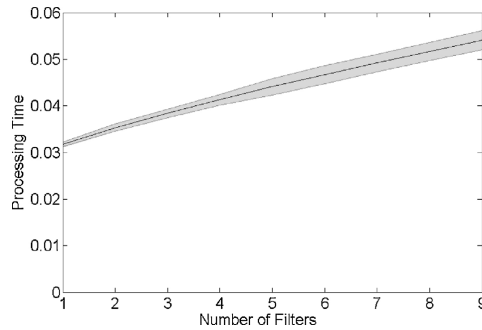

Fig. 9. Online processing time (in seconds) required for one single-channel EEG epoch per number of filters used in FBAR (each filter uses six features). Average across 1000 trials with shaded standard deviation is shown.

The reason FBAR outperforms FASTER and related methods for single-channel artifact detection is that FBAR uses a multidimensional numerically optimized model to differentiate artifacts from clean EEG. FASTER uses mainly low order statistics (though kurtosis is included with FASTER+) and looks only for statistically extreme values. Furthermore, instead of combining features in a multidimensional model, FASTER+ performs artifact rejection with respect to one feature at a time in a sequential manner. Thus FASTER can be viewed as approximating a series of one-dimensional classifiers which cannot take into account any interaction between features. This makes FASTER prone to false negatives with respect to small-amplitude artifacts, especially in the presence of high-amplitude EEG because they are not well-separable in any one dimension. In contrast, FBAR can take into account linear and non-linear interactions between features because it explicitly uses a multidimensional model. Moreover, FBAR incorporates a wider variety of features than FASTER+, allowing it to take into consideration additional information about the input signals.

According to Table 2, FBAR makes use of low order statistics only in the very low frequency and very high frequency bands. These features have been used in previous artifact detection methods to successfully classify medium-to-large-amplitude EOG and EMG artifacts, since they produce extremely large power and amplitude spikes [29,30,4,13]. However, these features are not able to distinguish small-amplitude artifacts.

Small-amplitude artifacts introduce non-zero skewness and kurtosis in what is otherwise a close to normally distributed EEG signal [31,32,19] even though they have little to no impact on the lower order statistics of the EEG time-series. Therefore, FBAR must rely on these higher order statistics for detecting these artifacts. However, even with these higher order statistics, a statistical thresholding approach would be inadequate for small-amplitude artifact detection, as evidenced by the inability of FASTER+ to detect such artifacts even with the inclusion of kurtosis as a feature. The reason is that small-amplitude artifacts do not result in statistically extreme values of skewness or kurtosis. Instead, it is the combination of slightly elevated skewness, kurtosis, and power which indicates the presence of a small-amplitude artifact. This requires a multidimensional model which can taken into consideration such interactions between features, such as FBAR.

4.1. Limitations and future work

The main disadvantages of FBAR are that it is more labour-intensive than many artifact detection methods. Given the short classification time required once a model is built, computational time will not be a significant issue for real-time applications using modern computers. However, the model building stage can involve several hours of CPU time if large amounts of training data are used and many cross-validation runs and performed. The trade-off, however, is a much more accurate and reliable artifact detection tool, especially if small-amplitude artifacts are a concern. Therefore FBAR is particularly well-suited for BCI and NF applications where computational efficiency and high accuracy may outweigh the desire for a more "plug-and-play" tool.

As with many data-driven studies, further development of FBAR would benefit from a larger training set. As mentioned in Section 2.2, subject characteristics such as age and gender were unknown for Datasets 4 through 9. Thus, it cannot be claimed that the general

population was well-represented in the training data with which the current FBAR models were trained. However, there are two main points to make regarding this limitation. First, the statistics of EOG and EMG artifacts vary much more within a subject than they do between subjects, because they most often arise from basic and simple involuntary physiological processes, such as eye blinks. While statistics of the true EEG signal may change with age or gender [35], the kinds of changes seen would not be expected to be detrimental for FBAR, which appears to be robust to EEG collected under different conditions (in this case, eyes-open, eyes-closed, and meditating, which introduce larger variability in the EEG than do age or gender). Second, because FBAR is based on a machine learning pipeline, it has an additional advantage: when a condition is found for which FBAR fails, it is always possible to add appropriate training data and further optimize its models. In addition, the FBAR toolbox allows for the inclusion of additional features in case its users discover cases where FBAR does not perform adequately even with additional training data. It remains to be seen whether or where such conditions would arise, excluding incorrectly acquired EEG data, as it was not found to fail for any of the 9 training sets or 30 test sets used for study.

Currently, FBAR simply removes segments of EEG which are contaminated by artifacts. While it is generally preferable to filter out the artifactual data and recover the underlying EEG signal, this is difficult to accomplish without spatial filtering and a high number of sensors. In contrast to recent methods which are able to clean only a certain kind of artifact [14–17], FBAR can accurately detect a wide variety of both EMG and EOG artifacts, even when they are small compared to the surrounding EEG but cannot recover the underlying EEG. For applications in which a high density of artifacts are expected (e.g., mobile EEG applications or the use of EEG in an ambulance), using FBAR to remove data segments could result in an unacceptable amount of rejected data. While the current use of FBAR may not be appropriate for these particular applications, in future work, FBAR could be combined with an artifact suppression method in such a way that FBAR could be used to detect the artifact before suppression is applied. Alternatively, FBAR could be used to provide labels for a supervised learning algorithm trained to suppress artifacts rather than remove them.

In Section 1.1 the rationale for using a more standard feature engineering and feature selection approach with an SVM classifier instead of a deep learning approach was given. While the current approach is successful and there are logical reasons for its use over a deep learning approach, the deep learning approach should not be dismissed. A deep learning approach, though it might require additional training data, may be a promising avenue for improvement over FBAR if a suitable combination of network architecture and hyperparameter settings can be found. To date, no studies on single-channel artifact detection using a deep learning approach could be found. A deep learning approach should be explored as an alternative to FBAR in future work on single-channel artifact detection.

5. Conclusions

In summary, the FBAR toolbox provides a means to achieve high performance real-time artifact detection with only one or a few EEG channels. Most notably, the very low false negative rate of FBAR suggests that FBAR is less likely to miss artifacts than FASTER or FASTER+, and FBAR is much more successful in detecting small-amplitude artifacts, even in the presence of high-amplitude EEG. While methods which are easier to implement may be preferable for certain applications, FBAR may be preferable for EEG applications which require exceptionally accurate and precise artifact detection, such as BCI, NF, or the extraction of small event-related

potentials (ERPs). The FBAR MATLAB toolbox is offered freely and made available at https://bitbucket.org/kiretd/FBAR.

References

[1] L.F. Nicolas-Alonso, J. Gomez-Gil, Brain computer interfaces, a review, Sensors (Basel, Switzerland) 12 (2) (2012) 1211–1279.
[2] L.H. Sherlin, M. Arns, J. Lubar, H. Heinrich, C. Kerson, U. Strehl, M.B. Sterman, Neurofeedback and basic learning theory: implications for research and practice, J. Neurother. 15 (October (4)) (2011) 292–304, http://dx.doi.org/10.1080/10874208.2011.623089 (Online).
[3] J.J. Shih, D.J. Krusienski, J.R. Wolpaw, Brain–computer interfaces in medicine, Mayo Clin. Proc. Mayo Clin. 87 (March (3)) (2012) 268–279.
[4] M. Fatourechi, A. Bashashati, R.K. Ward, G.E. Birch, EMG and EOG artifacts in brain computer interface systems: a survey, Clin. Neurophysiol. 118 (March (3)) (2007) 480–494, Available: http://www.ncbi.nlm.nih.gov/pubmed/17169606 (Online).
[5] J.A. Urigüen, B. Garcia-Zapirain, EEG artifact removal state-of-the-art and guidelines, J. Neural Eng. 12 (3) (2015) 031001.
[6] R.N. Vigário, Extraction of ocular artefacts from EEG using independent component analysis, Electroencephalogr. Clin. Neurophysiol. 103 (3) (1997) 395–404.
[7] R.N. Vigário, J. Särelä, V. Jousmäki, E. Oja, M. Hämäläinen, Independent component approach to the analysis of EEG and MEG recordings, IEEE Trans. Bio-med. Eng. 47 (5) (2000) 589–593.
[8] S. Makeig, A.J. Bell, T.-P. Jung, T.J. Sejnowski, et al., Independent component analysis of electroencephalographic data, Adv. Neural Inf. Process. Syst. (1996) 145–151.
[9] I. Daly, R. Scherer, M. Billinger, G. Müller-Putz, FORCe: fully online and automated artifact removal for brain–computer interfacing, IEEE Trans. Neural Syst. Rehabil. Eng. 23 (5) (2015) 725–736.
[10] V. Mihajlović, B. Grundlehner, R. Vullers, J. Penders, Wearable, wireless EEG solutions in daily life applications: what are we missing? IEEE J. Biomed. Health Inform. 19 (1) (2015) 6–21.
[11] G. Kristo, J. Höhne, R. Ortner, B. Reuderink, N. Ramsey, BNCI Horizon 2020, 2015.
[12] J. Minguillon, M.A. Lopez-Gordo, F. Pelayo, Trends in EEG-BCI for daily-life: requirements for artifact removal, Biomed. Signal Process. Control 31 (2017) 407–418.
[13] H. Nolan, R. Whelan, R.B. Reilly, FASTER: fully automated statistical thresholding for EEG artifact rejection, J. Neurosci. Methods 192 (1) (2010) 152–162.
[14] X. Chen, A. Liu, J. Chiang, Z.J. Wang, M.J. McKeown, R.K. Ward, Removing muscle artifacts from EEG data: multichannel or single-channel techniques? IEEE Sens. J. 16 (7) (2016) 1986–1997.
[15] X. Li, C. Guan, H. Zhang, K.K. Ang, Discriminative ocular artifact correction for feature learning in EEG analysis, IEEE Trans. Biomed. Eng. (2016).
[16] S. Khatun, R. Mahajan, B.I. Morshed, Comparative study of wavelet-based unsupervised ocular artifact removal techniques for single-channel EEG data, IEEE J. Transl. Eng. Health Med. 4 (2016) 1–8.
[17] R. Patel, M.P. Janawadkar, S. Sengottuvel, K. Gireesan, T.S. Radhakrishnan, Suppression of eye-blink associated artifact using single channel EEG data by combining cross-correlation with empirical mode decomposition, IEEE Sens. J. 16 (18) (2016) 6947–6954.
[18] S.-Y. Shao, K.-Q. Shen, C.J. Ong, E.P. Wilder-Smith, X.-P. Li, Automatic EEG artifact removal: a weighted support vector machine approach with error correction, IEEE Trans. Biomed. Eng. 56 (2) (2009) 336–344.
[19] I. Winkler, S. Haufe, M. Tangermann, Automatic classification of artifactual ICA-components for artifact removal in EEG signals, Behav. Brain Funct. 7 (1) (2011) 30.
[20] S.S. Haykin, S.S. Haykin, S.S. Haykin, S.S. Haykin, Neural Networks and Learning Machines, vol. 3, Pearson, Upper Saddle River, NJ, USA, 2009.
[21] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.
[22] M. Längkvist, L. Karlsson, A. Loutfi, A review of unsupervised feature learning and deep learning for time-series modeling, Pattern Recognit. Lett. 42 (2014) 11–24.
[23] Y. Ren, Y. Wu, Convolutional deep belief networks for feature extraction of EEG signal, 2014 International Joint Conference on Neural Networks (IJCNN), IEEE (2014) 2850–2853.
[24] Interaxon, "Muse", 2014 http://www.choosemuse.com.
[25] C. Lee, J. Chin, L. Yi, B.-S. Lee, M.J. McKeown, Feasibility analysis and adaptive thresholding for mobile applications controlled by eeg signals, 2015 23rd European in Signal Processing Conference (EUSIPCO), IEEE (2015) 2416–2420.

[26] "Brain Vision LLC – actiCHamp," http://www.brainvision.com/actichamp.
[27] "g.Tec Medical Engineering- g.USBamp," http://www.gtec.at/Products/Hardware-and-Accessories/g.USBamp-Specs-Features.
[28] Interaxon, Muse: Technical Specifications, Validation, and Research Use.
[29] J.L. Whitton, F. Lue, H. Moldofsky, A spectral method for removing eye movement artifacts from the EEG, Electroencephalogr. Clin. Neurophysiol. 44 (6) (1978) 735–741.
[30] I.I. Goncharova, D.J. McFarland, T.M. Vaughan, J.R. Wolpaw, EMG contamination of EEG: spectral and topographical characteristics, Clin. Neurophysiol. 114 (9) (2003) 1580–1593.
[31] L. Shoker, S. Sanei, J. Chambers, Artifact removal from electroencephalograms using a hybrid bss-svm algorithm, IEEE Signal Process. Lett. 12 (10) (2005) 721–724.

[32] A. Delorme, T. Sejnowski, S. Makeig, Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis, Neuroimage 34 (4) (2007) 1443–1449.
[33] H.C. Peng, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1226–1238.
[34] A. Delorme, S. Makeig, Eeglab: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (1) (2004) 9–21.
[35] W. Klimesch, EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis, Brain Res. Rev. 29 (2) (1999) 169–195.