Towards context-aware brain-computer interfaces

by

Andrew James Byron Myrden

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Biomedical Engineering
University of Toronto

© Copyright 2016 by Andrew James Byron Myrden
Abstract

Towards context-aware brain-computer interfaces

Andrew James Byron Myrden
Doctor of Philosophy
Graduate Department of Biomedical Engineering
University of Toronto
2016

Brain-computer interfaces (BCIs) allow individuals with disabilities to communicate and control their environment without the necessity for volitional speech or motor activation. However, most current BCIs are prone to significant performance fluctuations and incapable of adapting to their users. These shortcomings impair practical BCI usage.

This thesis investigates the feasibility of a hybrid BCI that is capable of detecting and adapting to underlying changes in user mental state. The thesis comprises four major studies, each representing a step towards this ultimate goal. The first study formulates a novel signal processing algorithm for the frequency-domain analysis of electroencephalographic (EEG) recordings. The second study examines the ability to automatically detect fluctuations in three mental states that are important to BCI usage - fatigue, frustration, and attention - based on electrical activity recorded from the surface of the scalp using EEG. The third study explores the effects of each of these mental states on the online operation of a two-class EEG-BCI. The final study investigates the efficacy of two different methods - reliability prediction and adaptive classification - by which a BCI can adapt to changes in fatigue, frustration, and attention.

In the first study, the novel algorithm, based on a clustering of spectral power features, was shown to compress EEG signals with less information loss than traditional frequency-domain analyses. In the second study, fluctuations in fatigue, frustration, and attention were detected with mean classification accuracies of 76.8%, 71.9%, and 86.1%, respectively. In the third study, a significant relationship between perceived frustration and BCI accuracy was uncovered and optimal regions for BCI performance were identified in several multi-dimensional representations of user mental state. In the final study, estimated mental state was used to predict the onset of low-accuracy BCI performance, with an 8% decrement in classification accuracy between the predicted high and low accuracy conditions. This study also demonstrated the ability
to directly adapt BCI classification, leading to statistically significant increases in classification accuracy for roughly half of participants without significantly compromising performance for the remainder of the study population.
Acknowledgements

I would like to express my gratitude to Dr. Tom Chau for his guidance and supervision throughout my time at PRISM Lab. His advice was invaluable to this work.

Many thanks also to the members of my thesis committee, Dr. Stephen Strother and Dr. Paul Yoo, and my examiners, Dr. Joseph Cafazzo and Dr. Klaus-Robert Müller. Each of them provided a different perspective on my research and their feedback improved this thesis immensely.

I would also like to acknowledge the financial support that I received from the Natural Sciences and Engineering Research Council of Canada; the Ontario Student Assistance Program; and the Institute of Biomaterials and Biomedical Engineering.

Finally, a heartfelt thank you to my family and close friends. Without your support, I certainly could not have completed this thesis.
# Contents

1 Introduction ......................................................... 1
   1.1 Access Technologies ........................................... 1
   1.2 Brain-Computer Interfaces ..................................... 2
      1.2.1 Types of BCIs ............................................. 3
      1.2.2 Types of BCI Signals ..................................... 5
      1.2.3 Electroencephalographic Brain-Computer Interfaces ..... 6
      1.2.4 Hybrid BCIs .............................................. 7
   1.3 Motivation ..................................................... 8
      1.3.1 BCI Non-Stationarity and Adaptation .................. 8
      1.3.2 Mental State and BCIs ................................... 9
      1.3.3 Mental State Detection .................................. 12
      1.3.4 User Interface Design .................................. 14
      1.3.5 Context for Brain-Computer Interfaces ................ 15
   1.4 Research Questions and Objectives ............................ 16
   1.5 Thesis Organization ........................................... 18

2 EEG Signal Processing and Machine Learning Primer ............. 19
   2.1 Introduction .................................................. 19
   2.2 Electroencephalographic Signals ............................. 19
   2.3 Signal Processing .............................................. 20
      2.3.1 Independent Component Analysis ....................... 21
   2.4 Machine Learning ............................................. 24
### 3 Feature Clustering for Robust Frequency-Domain Classification of EEG Activity

3.1 Abstract ........................................................................................................... 30
3.2 Introduction ........................................................................................................ 31
3.3 Algorithm ........................................................................................................... 34
   3.3.1 Definitions ................................................................................................... 34
   3.3.2 Procedure .................................................................................................. 36
3.4 Testing Methods .................................................................................................. 38
3.5 Results ................................................................................................................ 40
   3.5.1 Identification of Feature Clusters ............................................................... 40
   3.5.2 Mental State Classification ....................................................................... 42
   3.5.3 Mental Task Classification ....................................................................... 43
   3.5.4 Summary of Statistical Testing ................................................................. 44
3.6 Discussion ............................................................................................................ 44
3.7 Conclusions ......................................................................................................... 47

### 4 A Passive EEG-BCI for Single-Trial Detection of Changes in Mental State

4.1 Abstract ............................................................................................................ 48
4.2 Background ........................................................................................................ 49
4.3 Methods ............................................................................................................. 51
   4.3.1 Protocol ..................................................................................................... 51
   4.3.2 Signal Acquisition .................................................................................... 53
   4.3.3 Pre-Processing .......................................................................................... 53
   4.3.4 Feature Extraction .................................................................................... 53
   4.3.5 Feature Selection ....................................................................................... 54
   4.3.6 Data-Sensitive Class Balancing ................................................................. 54
   4.3.7 Classification ............................................................................................. 56
4.4 Results ............................................................................................................... 56
   4.4.1 Mental State Detection .............................................................................. 56
   4.4.2 Feature Analysis ....................................................................................... 58
4.4.3 Effects of Electrode Removal ........................................... 60
4.4.4 Generalization Between Participants .............................. 61
4.5 Discussion ................................................................. 63
  4.5.1 Feasibility of Detecting Changes in Fatigue, Frustration, and Attention ............... 63
  4.5.2 The Continuous Nature of Mental State .......................... 64
  4.5.3 Cortical Correlates of Mental State .............................. 65
  4.5.4 Electrode Removal .................................................. 66
  4.5.5 Generalizability of Feature Selection ........................... 67
  4.5.6 Comparison to Previous Literature ............................... 68
4.6 Conclusions ............................................................. 69

5 Effects of User Mental State on EEG-BCI Performance 70
  5.1 Abstract ..................................................................... 70
  5.2 Introduction ............................................................... 71
  5.3 Material & Methods ..................................................... 73
    5.3.1 Population .......................................................... 73
    5.3.2 Instrumentation .................................................... 73
    5.3.3 Training Sessions .................................................. 74
    5.3.4 BCI Development .................................................. 74
    5.3.5 Testing Sessions ................................................... 75
  5.4 Results ..................................................................... 77
    5.4.1 Choice of BCI Task ................................................ 77
    5.4.2 Online BCI Performance ......................................... 78
    5.4.3 Effects of Fatigue, Frustration, and Attention on BCI Performance .................... 79
    5.4.4 Multivariate Analysis ............................................. 83
    5.4.5 Effects of Mental State on Class Distributions .................. 85
  5.5 Discussion ................................................................. 87
    5.5.1 Optimal Mental State for BCI Control ........................ 87
    5.5.2 Towards Psychologically Adaptive BCIs ....................... 89
  5.6 Conclusions ............................................................. 90

6 Practical Considerations for EEG Classification of a Word Generation Task 92
6.1 Background .......................................................... 92
6.2 Methods ............................................................. 93
6.3 Results ............................................................... 94
6.4 Significance .......................................................... 96
6.5 Conclusions .......................................................... 97

7 Towards Psychologically Adaptive Brain-Computer Interfaces 98
7.1 Abstract ............................................................... 98
7.2 Introduction .......................................................... 99
7.3 Methods ............................................................... 101
   7.3.1 Protocol ......................................................... 101
   7.3.2 Training Sessions .............................................. 101
   7.3.3 Testing Sessions .............................................. 102
   7.3.4 Data Acquisition, Signal Processing, and Feature Extraction .............................................. 102
   7.3.5 Mental State Prediction ..................................... 103
   7.3.6 Adaptive Brain-Computer Interface ...................... 103
7.4 Results ............................................................... 109
   7.4.1 Mental State Prediction ..................................... 109
   7.4.2 Reliability Prediction ....................................... 110
   7.4.3 Adaptive Classification .................................... 111
7.5 Discussion ........................................................ 116
   7.5.1 Predictive Value of Mental State for BCIs .............. 116
   7.5.2 Contrast with Error Correction Using the Error-Related Potential .............................................. 117
   7.5.3 Variation Between Reliability Prediction and Classifier Adaptation .............................................. 117
   7.5.4 Limitations and Future Directions ....................... 118
7.6 Conclusions ........................................................ 119

8 Conclusions .......................................................... 120
8.1 Summary of Contributions ...................................... 120
8.2 Future Work ....................................................... 122
   8.2.1 Online Implementation ..................................... 123
   8.2.2 Combining Tonic and Phasic Adaptation ................ 123
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Rand indices</td>
<td>42</td>
</tr>
<tr>
<td>3.2</td>
<td>Summary of statistical testing</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison of classification algorithms</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>List of electrode subsets</td>
<td>60</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison between participant-dependent and participant-independent algorithms</td>
<td>62</td>
</tr>
<tr>
<td>5.1</td>
<td>Offline classification accuracies</td>
<td>77</td>
</tr>
<tr>
<td>5.2</td>
<td>Effects of mental state on BCI performance</td>
<td>79</td>
</tr>
<tr>
<td>5.3</td>
<td>Effects of mental state and task</td>
<td>80</td>
</tr>
<tr>
<td>7.1</td>
<td>Prediction of mental state within continuous range</td>
<td>109</td>
</tr>
<tr>
<td>7.2</td>
<td>Normalized frequency of mental state selection</td>
<td>114</td>
</tr>
<tr>
<td>7.3</td>
<td>Adaptive and non-adaptive BCI performance</td>
<td>114</td>
</tr>
</tbody>
</table>
List of Figures

1.1 An access solution .................................................. 2
1.2 The BCI cycle ....................................................... 4
1.3 P300 BCI .......................................................... 5

2.1 10-10 System ....................................................... 20
2.2 EEG Signals ......................................................... 21

3.1 Sample clusterings .................................................. 41
3.2 Mental state classification ........................................... 43
3.3 Mental task classification .......................................... 44

4.1 Easy and difficult mental tasks .................................... 52
4.2 Data relabeling algorithm .......................................... 55
4.3 Mental state classification accuracies .............................. 57
4.4 Individual mental state classification accuracies ................. 58
4.5 Participant-dependent feature selection frequencies ............ 59
4.6 Electrode subset classification accuracies ........................ 61
4.7 Comparison between participant-dependent and participant-independent algorithms .............. 62
4.8 Participant-independent feature selection frequencies ........ 63

5.1 Maze Navigation .................................................... 76
5.2 Individual BCI accuracies .......................................... 78
5.3 Collective BCI accuracies .......................................... 79
5.4 Univariate performance variability .................................. 82
5.5 Multivariate performance variability - fatigue and frustration ........................................ 83
5.6 Multivariate performance variability - fatigue and attention ........................................ 84
5.7 Multivariate performance variability - frustration and attention ....................................... 85
5.8 Fisher score variability ........................................................................................................ 86
5.9 Variation in class distributions with mental state ............................................................... 87

6.1 Word generation classification accuracies ............................................................................. 94
6.2 Feature selection frequencies for classification of word generation ....................................... 95
6.3 Electrode frequencies for word generation classification ....................................................... 96

7.1 BCI reliability prediction ......................................................................................................... 106
7.2 Adaptive BCI classification .................................................................................................... 108
7.3 Training set resampling .......................................................................................................... 109
7.4 Threshold values for reliability prediction ............................................................................. 110
7.5 Accuracy of BCI reliability prediction .................................................................................. 111
7.6 Comparison of reliability results with chance performance ............................................... 111
7.7 Adaptive and non-adaptive BCI performance ....................................................................... 112
7.8 Optimal adaptive BCI performance ...................................................................................... 113
7.9 Optimal mental states and training set size for all participants ............................................ 115
List of Acronyms

AAC  Augmentative and Alternative Communication

BCA  Balanced Classification Accuracy

BCI  Brain-Computer Interface

ECoG  Electrocorticography

EEG  Electroencephalography

EOG  Electrooculography

ERD  Event Related Desynchronization

ERP  Error-Related Potential

ERS  Event Related Synchronization

FCBF  Fast Correlation-Based Filter

FFT  Fast Fourier Transform

fMRI  Functional Magnetic Resonance Imaging

GDSF  Generic Discontinuity Spatial Feature

ICA  Independent Component Analysis

LDA  Linear Discriminant Analysis
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFP</td>
<td>Local Field Potential</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetoencephalography</td>
</tr>
<tr>
<td>MEV</td>
<td>Maximum Epoch Variance</td>
</tr>
<tr>
<td>MMG</td>
<td>Mechanomyography</td>
</tr>
<tr>
<td>NIRS</td>
<td>Near-Infrared Spectroscopy</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative Predictive Value</td>
</tr>
<tr>
<td>OPV</td>
<td>Overall Predictive Value</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive Predictive Value</td>
</tr>
<tr>
<td>RBF</td>
<td>Radial Basis Function</td>
</tr>
<tr>
<td>SAD</td>
<td>Spatial Average Difference</td>
</tr>
<tr>
<td>SCP</td>
<td>Slow Cortical Potential</td>
</tr>
<tr>
<td>SED</td>
<td>Spatial Eye Difference</td>
</tr>
<tr>
<td>SSVEP</td>
<td>Steady-State Visually Evoked Potential</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>TCD</td>
<td>Transcranial Doppler Ultrasound</td>
</tr>
<tr>
<td>TK</td>
<td>Temporal Kurtosis</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Access Technologies

Communication is a pivotal part of human life. Every day, we use a combination of gestures, body language, facial expressions, writing, and speech to express ourselves and interact with others. However, there is a growing population for whom these channels of communication are not available. More than 80,000 Canadian children currently live with severe disabilities [214]. These disabilities range from trauma-induced injuries to the spinal cord or brain stem to chronic disorders such as spastic quadriplegic cerebral palsy. Impairment to communication ability is a common consequence of these conditions [45].

The identification and development of alternative means of communication for these individuals is an active research field. Within this thesis, we use the terminology defined by Tai et al. in [219]. An access pathway can be described as a sensor or input device that allows functional intent to be translated into an electrical signal. When this signal is combined with methods of signal processing and classification, an access technology is produced. The output from an access technology can be used to drive a user interface which allows control of an augmentative and alternative communication (AAC) aid, an environmental control unit, or a computer. This definition of an access solution is depicted in Figure 1.1, from [219].

Access technologies are designed to cater to the abilities of their users. Individuals who retain some level of volitional muscle control can employ access technologies that capitalize on this motion. For example,
mechanomyographic (MMG) signals can measure small muscle vibrations [8], infrared cameras can detect the opening of the mouth [146], computer vision can detect tongue protrusion in patients with spastic cerebral palsy [127], and mechanical switches can be operated using the tongue [26]. Each of these technologies thus allows a small physical movement to be used for communication.

However, such access technologies are not suitable for all individuals. For example, individuals who present as locked-in do not possess the ability to communicate through conventional means such as speech or limb motion [210]. This immediately rules out these simpler, motion-based access technologies, creating a need for more complex access technologies that rely on covert physiological variables. One access technology often considered for these individuals is a brain-computer interface (BCI) that capitalizes on their intact cognition [23].

### 1.2 Brain-Computer Interfaces

A brain-computer interface can be defined as a system that measures and analyzes brain signals, converting them in real time into outputs that do not depend on peripheral nerves and muscles [140, 240]. BCIs can therefore translate user intent into a desired output that allows control of a computer or external device, consistent with the definition of an access technology. Moreover, since BCIs bypass traditional physical access pathways, they are a suitable access technology for individuals with severe disabilities [219].
Four major application areas have been proposed for BCIs – communication and control, motor substitution, entertainment, and motor recovery [148]. Communication and control applications include BCI spellers and programs that are used to enable BCI control of internet browsers [16, 100, 169, 237]. Motor substitution focuses on restoring functional activity for individuals with disabilities. This includes the development of neuroprostheses for grasping and also assistive mobility aids such as BCI-controlled wheelchairs [147, 180]. For entertainment, it has been suggested that BCIs could be used as part of video game and virtual reality applications [120, 122]. Finally, BCIs have been investigated as a means of promoting motor recovery after stroke, with promising results [9, 52, 81, 211]. This thesis is primarily concerned with the usage of BCIs for communication and control.

BCI operation itself has been described as a cycle [227]. First, the BCI user engages in some cognitive task, typically in response to a presented stimulus. A wide variety of cognitive tasks have been used in previous BCI research, including but not limited to motor imagery, mental arithmetic, and selective attention tasks [227]. These tasks can be referred to as BCI control tasks. Performance of each of these control tasks produces a unique physiological signature in electrical and metabolic brain signals. These signals can be recorded using measurement modalities such as electroencephalography (EEG) and near-infrared spectroscopy (NIRS). The physiological signature of the control task is then detected and used to estimate user intent by the decoding module, which includes pre-processing, feature extraction, and classification stages. Based on this prediction of intent, an output signal is generated to control an external device. This output provides the user with feedback, closing the cycle [227]. A visual representation of the BCI cycle is given in Figure 1.2. More complete discussions of the individual stages of this cycle are available in other reviews [20, 30, 33, 49, 126, 139, 143, 144, 155].

Direct communication between the brain and a computer was first proposed by Vidal in 1973 [228]. In recent years, BCI research has made great strides and become increasingly widespread [140]. We will discuss some prominent considerations for modern BCI design in the following sections.

### 1.2.1 Types of BCIs

One important property of a brain-computer interface is the means by which it infers user intent. In this respect, BCIs can be divided into three categories – active, reactive, and passive BCIs [247].

---

3
Active BCIs derive their outputs from brain activity that is directly and consciously controlled by the user [247]. These BCIs require the user to intentionally modify their cognitive activity in order to make a selection or send a command. Active BCIs are often based on the performance of cognitive tasks such as mental arithmetic or motor imagery [178, 188]. For example, a BCI user may be presented with an onscreen keyboard that scans through rows and columns of letters. The user then performs a motor imagery task to select a highlighted row and column, allowing communication based solely on neural activity.

Reactive BCIs derive their outputs from brain activity that arises in reaction to external stimulation [247]. This brain activity is therefore only indirectly modulated by the user. The most common example of a reactive BCI is the P300 speller, which presents a grid of letters to the user. The user focuses on the letter that they wish to select while rows and columns of the grid are flashed randomly. The desired letter is then determined through detection of the P300 potential [60], a positive deflection in an EEG signal that occurs when a target stimulus is detected in a train of non-target stimuli. A P300 spelling interface and the P300 potential are depicted in Figure 1.3.

Passive BCIs differ from active and reactive BCIs in that they do not require any direct or indirect effort by the user. Rather, passive BCIs derive their outputs from arbitrary brain activity without voluntary control [247]. They therefore provide implicit information regarding user state. A wide variety of potential...
applications are available for passive BCIs, including but not limited to monitoring fatigue during long-term tasks such as monotonous driving [155, 225]; detecting system interaction errors using the EEG error-related potential (ERP) [70]; estimating working memory load [155, 198]; detecting the perceived loss of control over a BCI [247]; monitoring affective state [74]; and tracking user attention levels [32, 37].

1.2.2 Types of BCI Signals

BCIs depend upon the availability of a physiological signal from the brain that is correlated with cognitive activity. Both invasive and non-invasive signals have been used for BCI research, forming the basis for invasive and non-invasive BCIs, respectively.

Invasive BCIs use electrodes implanted on or within the cortex to record extracellular local field potentials and multi-unit activity [98, 132, 233]. Non-invasive BCIs use a variety of measurement modalities to record physiological signals that are correlated with cognitive activity [21]. Notably, electroencephalography (EEG) is used to record electrical activity on the surface of the scalp [186], magnetoencephalography (MEG) is used to record changes in magnetic fields induced by electrical currents within the brain [145], near-infrared spectroscopy (NIRS) is used to measure changes in cerebral blood oxygenation [188], functional magnetic resonance imaging (fMRI) is used to measure blood-oxygen-level dependent contrast [235], and transcranial Doppler ultrasound (TCD) is used to record changes in cerebral blood flow velocity [163]. This thesis focuses on non-invasive BCIs.
One important consideration for BCI design is temporal resolution – the time required for changes in cognitive activity to manifest in the monitored physiological signal [150]. This varies between signals due to the different physiological mechanisms by which each signal is modulated. Cognitive activation induces changes in neural activity within the brain, producing electrical fluctuations. This neural activity also triggers several hemodynamic processes. It causes the demand for glucose and oxygen to increase, leading to a rise in the concentration of oxygenated hemoglobin within the brain. Neural activity also sets off an astrocytic signalling cascade that results in vasodilation of the cerebral arteriolar smooth muscle walls, causing increases in cerebral blood flow velocity [94]. Due to the delay between cognitive activation and the hemodynamic response, measurement modalities that record signals based on the initial electrical activation (i.e. EEG and MEG) have superior temporal resolution to those that measure hemodynamic signals (i.e. NIRS, fMRI, and TCD) [150].

Other practical concerns also affect the choice of BCI modality [150]. MEG and fMRI require complex instrumentation which is too expensive for practical BCI usage, but boast higher spatial resolutions than the other available modalities. TCD has extremely low spatial resolution since only two signals can be acquired simultaneously. NIRS has a spatial resolution slightly superior to that of EEG but limited depth penetration of the cortex. Due to its combination of high temporal resolution, moderate spatial resolution, portability, and low cost, EEG is the most common BCI modality and is the focus of this thesis.

1.2.3 Electroencephalographic Brain-Computer Interfaces

Electrical activity within the cerebral cortex elicits voltage changes on the surface of the scalp. EEG recordings from any point on the scalp represent a temporal and spatial summation of the synaptic potentials generated by pyramidal cells within the cerebral cortex [111]. Traditionally, EEG was used to investigate brain function and evaluate neurological disorders [240]. It was first proposed by Vidal that EEG activity could also be used for direct brain-computer communication [228, 229]. Subsequently, many EEG-BCIs have been designed to harness cortical activity for communication. These include active BCIs that rely on conscious modulation of spontaneous rhythms (e.g. slow cortical potentials and sensorimotor rhythms) and reactive BCIs that rely on evoked potentials (e.g. steady-state visually-evoked potentials and the P300 response) [63].

Slow cortical potentials (SCPs) are slow voltage changes generated in the cortex. It has been shown that
individuals can learn to control SCP amplitude through operant conditioning and use this control to operate a BCI. One notable application of SCPs is one-dimensional cursor control [22]. However, extensive training is often required before individuals can successfully use BCIs that are based on SCPs [24].

The performance of certain mental tasks is known to cause event-related desynchronization (ERD) and synchronization (ERS) [227]. Most research within this area has focused on sensorimotor rhythms [63]. These rhythms occur within the mu (8-12 Hz) and beta (18-26 Hz) frequency bands over primary sensorimotor cortical areas. ERD can be observed within these regions during movement, preparation for movement, or imagined movement. Conversely, ERS can be observed within these areas after movement and with relaxation. Since these phenomena can be voluntarily induced by imagined movement (i.e. motor imagery), they have frequently been used for BCI applications [28, 29, 181, 204, 224].

Steady-state visually-evoked potentials (SSVEPs) occur when individuals focus their gaze on a periodic visual stimulus with a frequency greater than 6 Hz [63]. This causes frequency-dependent evoked potentials within the visual cortex in the occipital lobe of the brain. When BCI users are presented with multiple visual stimuli with different flicker frequencies, they can focus their gaze on the option that they wish to choose. The resultant evoked potential can be detected, allowing for BCI control through selective visual attention [42, 157]. Another type of evoked potential, the P300 potential, was discussed previously in Section 1.2.2.

1.2.4 Hybrid BCIs

Considerable attention has recently been devoted to hybrid BCIs, which consist of a typical active or reactive BCI that is combined with another device [177]. Some researchers have claimed that such BCIs may allow improved performance without increasing the burden on the user [177]. Hybrid BCIs can be characterized by the identity of their secondary system (i.e. pure or physiological) [3] and also by their mode of operation (i.e. simultaneous or sequential) [177]. In a pure hybrid BCI, the secondary system is also a BCI, while in a physiological hybrid BCI the secondary system may be a device based on physiological signals or a physical communication device. In a simultaneous hybrid BCI, both systems are used at the same time, while in a sequential hybrid BCI the secondary system may only be engaged part of the time in order to supplement the primary system (for example, by switching it on and off) [4]. The permutations of these two variables create four classifications for hybrid BCIs: simultaneous pure hybrid BCIs (e.g. a BCI speller that combines both
SSVEP and P300 classification [243]), simultaneous physiological hybrid BCIs (e.g. using both EEG and EMG to classify motor imagery for wheelchair control [124]), sequential pure hybrid BCIs (e.g. using motor imagery to activate an SSVEP BCI for prosthetic control [182]), and sequential physiological hybrid BCIs (e.g. using EOG and EEG in sequence to highlight and select an item on a computer screen [244]).

Of these classifications, only simultaneous pure hybrid BCIs are of relevance to this thesis. The most common configurations for pure hybrid BCIs are two reactive BCIs, two active BCIs, or one reactive and one active BCI. It is less common, although not unheard of, for a passive BCI to be included alongside an active or reactive BCI [51]. Simultaneous pure hybrid BCIs have been designed using two EEG-BCIs, as in the case of the dual P300-SSVEP speller [243], or using one EEG-BCI and one NIRS-BCI [68]. A comprehensive survey of hybrid BCI research is beyond the scope of this thesis, but can be found in [156].

1.3 Motivation

Despite the great strides made by BCI researchers, numerous hurdles remain before BCIs can achieve widespread practical usage. This section discusses some of the major challenges confronting BCI researchers.

1.3.1 BCI Non-Stationarity and Adaptation

BCIs are unreliable devices that are prone to significant performance fluctuations. This is partially due to user-specific differences; so-called ‘BCI illiteracy’ is common [27, 231] and individuals often show significant variance in their ability to reliably perform different BCI control tasks [148]. However, variability in performance for a single participant is also common due to the nature of the EEG signal. Like many biomedical signals, EEG is not wide-sense stationary. This means that the temporal and spectral characteristics of EEG signals vary over time [39]. Potential sources of this non-stationarity include instrumentation issues (e.g. small electrode movements and impedance variability caused by drying electrode gel) [34] and changes in psychological states such as fatigue and attention [207]. This non-stationarity can negatively affect the classification of cognitive activity.

Furthermore, this non-stationarity is exacerbated by several characteristics of practical BCI training and usage [34]. First, the data used to train a BCI often originate from sessions that may have taken place
days or weeks previously. In addition to the obvious issues with signal non-stationarity associated with this delay, the signal acquisition process itself is very sensitive. Slight deviations in electrode positioning and electrical impedance between sessions may affect the recorded signals and thus have a negative impact on the BCI itself [221, 223]. Second, BCI training sessions rarely mimic the context within which the BCI is used online. Differences between the experimental protocols used during training and online sessions can affect brain activity, further impairing the BCI during online operation [34]. Third, long-term BCI usage requires the user to repetitively complete the BCI control task or tasks, producing, over time, changes due to habituation and brain plasticity [34].

These causes of non-stationarity can be divided into those that affect BCI performance within a single session and those that affect BCI performance between multiple sessions [69, 201]. Adaptation to both types of non-stationarity has been identified as a critical goal for BCI researchers [116, 207].

A wide variety of methods have been investigated for mitigating between-session non-stationarity. Sun and Zhang adaptively updated feature extractors based on new data from each session [217]. Shenoy et al. investigated several methods of rebiasing and retraining classifiers based on online data [207]. Sugiyama et al. derived an importance-weighted covariate shift method to compensate for shifts in the feature distributions between training and testing data [216], a method replicated by numerous other research groups [130, 192, 201]. Vidaurre et al. used each sample from a testing session to update an adaptive linear discriminant analysis (LDA) classifier [230], and Llera et al. applied similar algorithms to multi-class BCIs (i.e. those that use more than two control tasks) [138]. Arvaneh et al. used both supervised and unsupervised data space adaptation to transform EEG data from the target space (i.e. the testing session) to the source space (i.e. the training session) [10]. Samek et al. investigated the development of a stationarity common spatial patterns algorithm to extract features that are invariant to signal non-stationarity [199]. However, these approaches mainly focus on the macroscopic differences in EEG signals recorded on different days and using different experimental protocols. Comparatively little attention has been devoted to within-session adaptation [230].

1.3.2 Mental State and BCIs

One common hypothesis for within-session BCI non-stationarity is that it reflects short-term changes in user mental state [80, 207, 218, 238]. It is useful here to define what constitutes user mental state. Zander
and Jatzev asserted that user mental state is comprised of two main elements - cognitive conditions and cognitive processes [246]. Cognitive conditions are defined as latent mental states such as fatigue and arousal, while cognitive processes are time-bounded cognitive events such as the error-related potential [246]. While cognitive processes can be detected and used to improve BCI performance [70, 71, 202], they contain little information regarding the short-term modulations in mental state that allegedly cause within-session non-stationarity. For this reason, we use the term ‘user mental state’ and derivatives thereof within this thesis to refer solely to cognitive conditions.

There is some evidence that mental state may affect BCI performance. Several studies have shown that cognitive conditions such as sleepiness and motivation levels are related to BCI performance on an inter-session or inter-participant basis [82, 88, 167]. On a single-trial basis, some studies have shown that EEG activity within certain cortical regions and frequency ranges may be predictive of BCI performance (i.e. fronto-parietal gamma band power in [79]). This finding may be indicative of a link between mental state and BCI performance, but these observations have several practical limitations. First, they apply solely to motor imagery BCIs, thus excluding any BCI users who can not successfully perform motor imagery tasks. Second, they are unintuitive and do not address the psychological sources of performance variability. Third, due to the abstract nature of these observations, it is difficult to target mental-state based interventions to improve BCI performance.

We propose that adaptation to user mental state may be more effective if the relationships between BCI performance and more intuitive cognitive conditions can be deciphered. In particular, we are interested in quantifiable mental states that can be realistically expected to fluctuate continuously during BCI usage. We have identified three such states that are of particular interest - mental fatigue, attention, and frustration. Although these mental states represent only a subset of all the mental states that may affect BCI performance, we have limited the scope of this thesis to these conditions. Each of these mental states is intriguing because both previous literature and our subjective experiences indicate that modulations in each of these states may naturally arise from BCI usage while also affecting BCI performance.

Mental fatigue can be defined as a feeling of tiredness and exhaustion which impairs both ability and willingness to perform a task [40]. Fatigue is often an unavoidable consequence of BCI usage. For example, SSVEP and P300 BCIs are dependent on rapidly flashing visual stimuli that can induce significant fatigue in users [5, 40, 142]. Even in mental task-based BCIs, repetitive performance of tasks such as mental
arithmetic and motor imagery during a long period of BCI usage may lead to fatigue. This may be linked
to the nature of BCI interaction. Most BCIs employ one of two control paradigms - either process control
or goal selection [239]. A process control BCI requires the user to generate every necessary command signal.
On the other hand, a goal selection BCI requires the user only to communicate intent, at which point an
intelligent execution unit intervenes to perform the necessary actions to achieve this goal. The difference
between these two paradigms can be understood through the example of wheelchair navigation. Controlling
a wheelchair using process control requires the user to specify every movement and rotation required to
navigate from the current location to the intended destination. Using goal selection, the BCI user need
only specify the destination and the wheelchair will autonomously navigate to that location. In essence, the
process control BCI forces its user to generate a high-speed series of interactions that would normally be
handled by reflex pathways in healthy individuals [239]. This is very taxing and places a heavy burden on
the BCI user [4, 140, 239]. Despite their tiring requirements, process control BCIs are much more common
than goal selection BCIs in current research, potentially because they are easier to implement and require
less specialized knowledge of the environment in which they are used [239]. However, the fatigue caused by
these BCIs may be an impediment to BCI training and control [49].

Although many definitions of attention are possible depending on context, one pertinent definition is “the
capacity to allocate processing resources selectively to stimuli or classes of stimuli” [195]. In BCI usage,
this equates to the ability to remain focused on the BCI control task while disregarding other stimuli. In
some respects, attention is linked to fatigue; when BCI users become tired, they may find it difficult to
continue performing the control task [40]. However, other variables may also cause lapses in attention,
including motivation, ambient noise, and the presence of distractors. No well-designed EEG-BCI studies
appear to have been conducted to analyze the effects of attention level on BCI performance. However, Falk
et al. showed that an NIRS-BCI devolved to chance accuracies in a noisy environment with distractions [66].
Other researchers have noted the potential for stimulus-rich feedback to distract users from the BCI control
task, thereby decreasing performance [168, 184].

Frustration can be defined as “an interference with the occurrence of an instigated goal-oriented response
at its proper time in the behavior sequence” [59]. Since perfect performance is rare in a BCI, frustration is
commonly induced by the loss of control when a BCI fails to recognize user intent [36, 83, 92, 193]. Frustration
can also be induced by technical issues that may occur when BCIs are used without the oversight of an
experienced researcher [101], and by the awkwardness of the system-paced configuration that is most common
in BCI research [75]. Frustration has been identified as a potential risk factor for poor BCI performance [193]. It is also possible that the neural correlates of frustration may overlap with those of BCI control tasks like motor imagery, leading to dramatically reduced performance [170]. Frustration may also cause abandonment of the BCI itself, thus posing an existential threat to BCI usage [149].

The causality between BCI performance and each of these mental states is murky. Fluctuations in BCI performance could ostensibly be either a cause or an effect of changes in mental state. This raises the possibility of a feedback loop leading to complete BCI failure. The integration of mental state detection and an active or reactive BCI may help mitigate this risk [50, 148]. However, the development of this type of hybrid BCI depends upon the ability to reliably detect changes in mental state.

### 1.3.3 Mental State Detection

We propose five requirements that mental state detection must meet in order to be combined with a BCI. First, it must be reliable between sessions, as BCIs are generally used on many different days. Second, it must be sensitive to a range of mental states that includes those most likely to be induced by BCI usage. Third, it must permit accurate single-trial detection of changes in mental state, rather than globally averaging across many trials. Fourth, it must be capable of performing this single-trial detection online in real-time. Finally, to maximize the practicality of detection, it must be achieved using a modality like EEG or NIRS that can also be used for a BCI. Current efforts to detect fluctuations in fatigue, attention, and frustration fall short of these requirements.

Most EEG fatigue detection research has focused on driver fatigue. For example, Jap et al. compared slow and fast wave EEG activity during long and short driving tasks, but did not attempt fatigue detection [105]. Yeo et al. claimed to classify mental fatigue during driving with 99% accuracy but all data that did not exhibit the expected patterns in frequency-domain activity during manual examination were discarded prior to classification, rendering the results dubious [242]. Regardless, results achieved using long-term simulated driving protocols are unlikely to apply to a BCI setting. Some other studies have used protocols that translate to a BCI setting more readily. Liu et al. classified pre and post-task EEG data from vigilance, switch, and arithmetic tasks with 84% accuracy but ignored all transitional states between unfatigued and exhausted [134]. Shen et al. claimed to have classified five levels of mental fatigue with 91% accuracy but in actuality classified five different levels of vigilance task performance within a 25-hour sleep deprivation
study [206]. Murata et al. showed that event-related potential delays increase with fatigue, but they only investigated values before and after a three-hour mental arithmetic task and they did not attempt single-trial detection [158]. Trejo et al. attained 97% classification between low and high fatigue during mental arithmetic but again achieved this by discarding all data from states other than ‘alert’ and ‘exhausted’ [225]. We conclude that although extremely accurate detection has been achieved, no study has successfully demonstrated that fatigue can be detected within a range of intensities likely to occur during BCI usage. It is also concerning that we have not located any studies that used regression to predict fatigue on a continuous scale. Moreover, the aforementioned studies generally ignored the potential for fatigue to fluctuate, assuming a monotonic trend in fatigue over time, as assumption that may not hold on a fine time scale.

Less research has been devoted to EEG classification of attention. Hamadicharef et al. differentiated attention and non-attention conditions with 89% accuracy using spectral-spatial patterns [87]. However, they did not attempt to differentiate varying levels of attention for the same task, as would be expected during BCI usage. Li et al. classified three levels of self-reported attention during four mental tasks but with only 57% accuracy, which may not be sufficiently reliable for BCI adaptation [131]. Liu et al. differentiated attentive and inattentive states with 76% accuracy using a support vector machine (SVM) classifier [135]. However, they discarded any data that could not easily be labeled as one of these states. Müller et al. accurately predicted attention levels for one participant during a vigilance task, but these results must be replicated for more participants within a protocol that more closely resembles BCI operation [155]. Outside of single-trial classification, some studies have identified spectral components of attention and engagement that may be relevant to attention detection [185], while others have professed to show measures of engagement but do not demonstrate any compelling reason to believe that these measures are accurate [17, 18].

Although frustration has been cited as an important topic in BCI research and human-computer interaction in general [50, 65, 77, 193], we could find no peer-reviewed articles that demonstrated the ability to detect changes in frustration using EEG. This constitutes a clear gap in the scientific literature. We conclude that further development of automatic detection of changes in fatigue, attention, and frustration is necessary before mental state detection can be combined with an active or reactive BCI. However, should this goal be achieved, the resultant hybrid BCI may help address some of the inherent shortcomings of BCI interaction.
1.3.4 User Interface Design

Computer interfaces are typically characterized by their bandwidth, dimensionality, and delay [237]. In these respects, BCIs are fundamentally different from conventional interaction devices such as a mouse and keyboard. These conventional devices provide high-bandwidth, high-dimensional input signals with very little delay. BCI communication, on the other hand, has extremely limited bandwidth, generally low dimensionality, and long and variable delays introduced by both classification algorithms and the user himself or herself [237].

BCI communication is also characterized by severe asymmetry between the input and feedback channels [237]. Most BCIs provide feedback on a computer monitor. Consequently, the feedback channel has a high bandwidth, in contrast to the extremely low-bandwidth input channel. This asymmetry is much less pronounced when using conventional interaction devices [237]. Despite this, many BCI studies have simply mapped BCI signals onto existing control metaphors such as cursor control and button-pressing [64]. The vast difference in parameters between BCIs and conventional interaction devices makes this approach inefficient.

The limited bandwidth of most BCIs is one reason why unreliable performance can induce frustration, as a single error can take a great deal of time and effort to correct. There are at least two ways to cope with this. The first is to carefully design BCI user interfaces such that they take into account the unconventional control properties of BCI communication rather than morphing existing control metaphors to fit BCI signal properties [29, 237]. The second is to increase the amount of information that can be transmitted by the BCI by increasing its efficiency or by adding additional channels, as in a hybrid BCI.

However, hybrid BCIs that consist of any combination of active and reactive BCIs may increase bandwidth at the cost of additional fatigue for the user. Those BCIs that require the user to master two control tasks increase the burden on the user rather than decreasing it, and those that use two measurement modalities to detect the same task increase the complexity and setup time for the BCI. In contrast, a hybrid BCI that includes a passive BCI can make use of the principles of complementarity, composability, and controlled cost [247]. The passive BCI can work in concert with an active or reactive BCI without interfering with it; multiple passive BCIs can be used in parallel, to, for instance, measure different mental states; and the passive BCI requires no conscious effort (or cognitive cost) from the user. The information from a passive BCI can also be used to adapt the BCI interface [148], although it is unclear how this information can best
be integrated with a BCI. To address this, we must explore the concept of context awareness.

### 1.3.5 Context for Brain-Computer Interfaces

Within the field of computing, context has been defined by Dey et al. [56] as “any information that can be used to characterize the situation of an entity.” An entity has been defined as any person, place, or object relevant to the interaction between a user and an application, including the user and application themselves [56]. Entities that are relevant to BCI interaction include the BCI user, the environment surrounding the user, and the BCI itself. Information regarding the mental state of the user thus constitutes contextual information. Schilit et al. have proposed four ways in which such contextual information can be integrated with an application: automatic contextual reconfiguration, context-triggered actions, proximate selection, and contextual commands [203].

Researchers have recently used some of these strategies to enhance BCIs with contextual information. Navarro et al. have proposed BrainAble, an ongoing project focused on using contextual information about the environment to adapt a BCI using contextual commands [164]. Faller et al. also used environmental context to provide contextual commands to BCI users [67]. Bos et al. proposed several methods by which context-triggered actions and contextual commands can be used in BCIs but focused primarily on environmental and application context rather than user context [35]. Zander and Jatzev stopped short of integrating contextual information of user state with a BCI but hypothesized that a passive BCI may make it possible to reliably infer information about ambiguous brain activity in uncontrolled environments [246]. Given the scant evidence for BCI adaptation to user context, we must look elsewhere for guidance.

Within the broader field of human-computer interaction, Fairclough has hypothesized that computer systems can adapt to physiological information in two modes - explicit and implicit adaptation [65]. For the user, explicit adaptation is overt and immediately apparent while implicit adaptation is covert. A BCI could implement explicit adaptation through context-triggered actions (e.g., shutting down when distraction is detected) or contextual commands (e.g., providing different options to the user based on their mental state). However, explicit adaptation is inherently risky. There is a high probability of false alarms that may induce frustration in the BCI user and potentially push them towards abandonment [65]. Therefore, this thesis focuses on implicit adaptation. Statistical changes in EEG feature distributions during BCI usage may be driven by changes in user mental state [207, 246]. Automatic contextual reconfiguration through classifier
adaptation in response to user mental state may provide a way to covertly mitigate the effects of fluctuations in fatigue, attention, and frustration.

1.4 Research Questions and Objectives

The primary objective of this thesis was the development of a novel hybrid BCI that combines an active mental task-based BCI with mental state detection. In pursuit of this objective, and in light of the aforementioned gaps in the scientific literature, the following secondary objectives were identified:

Objective 1
Investigate the feasibility of EEG-based single-trial detection of fluctuations in fatigue, attention, and frustration in able-bodied participants. It is necessary to develop an algorithm capable of detecting short-term fluctuations in each mental state, rather than those between pre-task and post-task conditions separated by multiple hours of cognitive effort. For ecological validity, these fluctuations should occur during the performance of cognitive tasks similar to those used in previous BCI research. To allow comparison to previous literature, binary classification of each mental state should be performed. This objective is the focus of Chapter 4.

Objective 2
Investigate the effects of fatigue, attention, and frustration on BCI performance. No statistically significant quantifiable relationship between BCI performance and short-term changes in mental state has been previously demonstrated. Changes in fatigue, attention, and frustration during online usage of a mental task-based EEG-BCI must be recorded and statistically compared with BCI performance. This objective is the focus of Chapter 5.

Objective 3
Adapt the mental state detection algorithm developed during Objective 1 to allow prediction of fatigue, attention, and frustration on a continuous scale. Moving away from the multi-level mental state classification that is currently prevalent in the scientific literature may allow for more accurate and more sensitive measurement of fluctuations in mental state. This objective is one focus of Chapter 7.

Objective 4
Investigate covert methods of combining user mental state detection with a mental-task based BCI. Such methods may include strategies such as error detection and classifier adaptation. The classification accuracy attained by this adaptive hybrid BCI must be compared to a traditional non-adaptive BCI. This objective is the second focus of Chapter 7.

Based on these objectives, the following research questions were identified (note that each enumerated research question is linked to the objective of the same number):

**Research Question 1a**
With what single-trial accuracy can fluctuations in fatigue, attention, and frustration be classified based on EEG data?

**Research Question 1b**
What cortical areas and what frequencies of EEG activity are most useful for classification of fatigue, attention, and frustration?

**Research Question 2**
How is BCI performance affected by changes in fatigue, attention, and frustration?

**Research Question 3**
Can fluctuations of fatigue, attention, and frustration be predicted on a continuous scale using a regression algorithm? If so, what strength of correlation between self-reported and predicted values can be attained?

**Research Question 4a**
How accurately can BCI errors be predicted based solely on estimated user mental state?

**Research Question 4b**
Can a BCI that automatically adapts to user mental state classify cognitive activity more accurately than a BCI that does not?
1.5 Thesis Organization

Chapter 2 introduces some of the signal processing and machine learning concepts used throughout this thesis. Chapters 3, 4, 5, and 7 each consist of an independent manuscript that has been published, or is currently under review, in a peer-reviewed journal. Each manuscript is self-contained, with the consequence that certain material - particularly relating to experimental methods and a review of the scientific literature - is repeated in multiple chapters. Chapter 6 presents an addendum to the study detailed within Chapter 5. Chapter 8 summarizes the contributions of this thesis to the scientific literature and proposes several avenues for future research.
Chapter 2

EEG Signal Processing and Machine Learning Primer

2.1 Introduction

This chapter reviews some of the signal processing and machine learning algorithms used within by this thesis.

2.2 Electroencephalographic Signals

EEG signals recorded from the surface of the scalp reflect both tonic changes in cognitive activation and phasic evoked potentials [76]. The frequency-domain structure of EEG signals encodes significant information regarding cognitive activity. Signal bandwidth is approximately 50 Hz, with significant cortical activity taking place in the delta (0.5-4 Hz), theta (4-7.5 Hz), alpha (8-13 Hz), and beta (14-26 Hz) frequency bands [200]. We note here that these definitions of the four major frequency bands are far from universal. For example, some researchers define the alpha band as the range from 8 to 15 Hz, and some even treat the so-called ‘upper alpha’ and ‘lower alpha’ bands separately [113].
The placement of EEG electrodes on the scalp is governed by the International 10-20 and 10-10 systems, which separate adjacent electrodes by either 20% or 10% of the front-back or left-right length of the skull [43, 102]. Within these systems, each electrode location is denoted by an alphanumeric code (e.g. F<sub>3</sub>). The alphabetical portion of this code represents the lobe of the brain on which the electrode resides, with F, T, C, P, and O signifying the frontal, temporal, central, parietal, and occipital lobes, respectively. The numerical portion of the code represents the lateral position on the skull, with odd numbers referring to the left hemisphere and even numbers to the right hemisphere. Electrode locations on the midline are represented with the letter ‘z’ (e.g. F<sub>z</sub>) rather than a number. An example of the 10-10 system is depicted in Figure 2.1.

![Figure 2.1](image.jpg)

Figure 2.1: International 10-10 system for electrode placement.

The maximum amplitude of spontaneous EEG signals on the surface of the scalp is quite low, requiring signal amplification for most analyses. Evoked potential amplitudes are typically lower than the signal-to-noise threshold and must be averaged across multiple trials to achieve detection [179]. Figure 2.2 depicts a collage of EEG signals in both the time and frequency-domains.

### 2.3 Signal Processing

Within this thesis, a B-Alert X24 headset (Advanced Brain Monitoring, Carlsbad, CA) with a sampling frequency of 256 Hz was used to record cortical activity. Signals were acquired from 15 electrodes placed over the F<sub>3</sub>, F<sub>1</sub>, F<sub>z</sub>, F<sub>2</sub>, F<sub>4</sub>, C<sub>3</sub>, C<sub>1</sub>, C<sub>2</sub>, C<sub>4</sub>, CP<sub>z</sub>, P<sub>1</sub>, P<sub>z</sub>, P<sub>2</sub>, and PO<sub>z</sub> cortical locations specified by the 10-20 system. The resultant signals were band-pass filtered (2-30 Hz) to isolate the frequencies of interest using an equiripple finite impulse response filter. This section details the additional filtering methods necessary to remove electrical artefacts resulting from eye movement.
2.3.1 Independent Component Analysis

Signal artefacts caused by eye movement, muscle contractions, high electrode impedance, and electrical interference are common in EEG signals [152]. Techniques to remove these artefacts include epoch rejection and regression based on reference signals near the eyes (i.e., EOG signals) [46, 78], but these approaches either discard significant amounts of data or require additional instrumentation. Recently, independent component analysis (ICA) has been adopted as another alternative for EEG artefact removal [152].

ICA was initially proposed as a means of solving the blind source separation problem [109]. In this problem, N independent source signals \( s = \{s_1(t), s_2(t), \ldots, s_N(t)\} \) are linearly mixed by an unknown matrix \( A \), referred to as the mixing matrix. Considering each potential source of artefacts as an independent source contributing to a montage of EEG recordings, ICA can be applied to estimate each underlying source signal. Some of the resultant source signals represent cognitive activity while others represent external artefacts, allowing EEG signals to be reconstructed using only the source signals of interest. However, most work with ICA artefact removal for EEG signals relies on manual identification of the source signals that represent artefacts [152]. Practical usage of ICA for artefact removal requires automatic detection of which source
signals should be removed.

Within this thesis, an adapted version of the ADJUST algorithm developed by Mognon et al. was used to remove signal artefacts [152]. We use the following notation. The recorded EEG signals \( g(t) \) are a linear combination of the source signal components \( s(t) \) as follows:

\[
g(t) = A \cdot s(t) + n(t)
\]  

(2.1)

where \( A \) represents the mixing matrix and \( n(t) \) represents the contributions of stochastic noise, which were ignored. The matrix \( g(t) \) is of size \( 15 \times N_s \), where \( N_s \) represents the number of samples within a segment of EEG data. Each row of \( g(t) \) represents the EEG signal from one electrode and each column represents the instantaneous value of the signals from each electrode at one time point. \( A \) is a square matrix of size \( 15 \times 15 \) and \( s(t) \) is of the same dimensions as \( g(t) \). Each row \( a_n, 1 \leq n \leq 15 \), represents the spatial loading of source signal \( s_n(t) \) on each electrode. For each source signal, temporal features were computed from \( s_n(t) \) and spatial features from \( a_n \). A total of six features were computed for each source signal based on [152]:

**Maximum epoch variance (MEV)**

Each source signal \( s_n(t) \) was divided into 2.5-second non-overlapping epochs. The variance of \( s_n(t) \) within each epoch was computed and the ratio between the maximum epoch variance and the mean epoch variance was computed. To make this measure more robust, the top and bottom 2.5% of values were removed before computing the maximum and mean epoch variance.

\[
MEV_n = \frac{\text{trimmax}(\text{var}_e[p[s_n(t)]])}{\text{trimmean}(\text{var}_e[p[s_n(t)]])}
\]  

(2.2)

**Generic discontinuity spatial feature (GDSF)**

The generic discontinuity spatial feature is sensitive to local spatial discontinuities. It used the proximity of each pair of electrodes to identify the element of the weight vector \( a_n \) that was most dissimilar to its neighbours. GDSF can be expressed as:

\[
GDSF_n = \max(|a_n - \langle k_{mn} a_m \rangle_n|)
\]  

(2.3)
where $k_{mn}$ represents an exponential weight vector that decreases as the distance between electrodes $n$ and $m$ increases and $\langle \ldots \rangle_m$ represents the average over all electrodes $m \neq n$.

**Spatial average difference (SAD)**

For a weight vector $a_n$, the spatial average difference represents the degree to which a source signal is biased towards the frontal electrodes and away from the posterior electrodes:

$$SAD_n = \text{mean}(a_{n, FA}) - \text{mean}(a_{n, PA})$$

(2.4)

where $a_{n, FA}$ represents the average weighting on electrodes $F_z, F_1, F_2, F_3, \text{ and } F_4$ and $a_{n, PA}$ represents the average weighting on electrodes $CP_z, P_2, P_1, P_2, \text{ and } P0_z$.

**Spatial eye difference (SED)**

For each weight vector $a_n$, the spatial eye difference was defined similarly to $SAD_n$ but compared the average weighting on $F_1$ and $F_3$ (i.e. the left hemispheric frontal electrodes) to the average weighting on $F_2$ and $F_4$ (i.e. the right hemispheric frontal electrodes).

**Temporal kurtosis (TK)**

As in the definition of MEV, each component was divided into non-overlapping 2.5-second intervals. The kurtosis was computed within each interval and the 5% trimmed mean of the kurtosis from all epochs was computed.

These features were used to remove four types of artefacts: eye blinks, vertical eye movements, horizontal eye movements, and generic discontinuities [152]. For each feature, k-means clustering was used to divide the values observed for all source signals into two groups, implicitly defining an upper threshold. Each type of artefact was characterized by two features, and source signals were flagged as artefact-contaminated if they exhibited values exceeding this threshold for both features. Additional constraints were applied for most types of artefacts to prevent false positives.

**Eye Blinks**

Source signals containing eye blinks were identified through higher values of SAD and TK. An additional constraint was applied by mandating that the spatial loading on all frontal electrodes must have the same sign.
Vertical Eye Movements
Source signals containing vertical eye movements were identified through higher values of SAD and MEV. The same constraint was applied as for eye blinks.

Horizontal Eye Movements
Source signals containing horizontal eye movements were identified through higher values of SED and MEV. An additional constraint was applied by mandating that the spatial loading on the left frontal electrodes have a different sign than that on the right frontal electrodes.

Generic Discontinuities
Generic discontinuities were identified through higher values of MEV and GDSF. No additional constraints were applied.

Source signals classified as containing any type of artefacts were subtracted from the recorded EEG signals to remove the effects of eye movement.

2.4 Machine Learning

We formulate the general machine learning problem as follows based on [25]. Suppose that we have a data set of N observations \( X = (x_1 \cdots x_N) \) along with corresponding target values \( t = (t_1 \cdots t_N) \). Each input vector \( x \) consists of D values, or features, such that \( x = (x_1 \cdots x_D) \). When referring to individual features, we use the notation that the entire feature set \( F = (F_1 F_2 \cdots F_D) \). Each target value \( t_i \) may be either discrete or continuous. Based on these known observations and labels, the goal of a machine learning algorithm is to predict the expected value of \( \hat{t} \) for any new input \( \hat{x} \). Within this paradigm, \( X \) and \( t \) are said to represent the training data while \( \hat{x} \) and \( \hat{t} \) represent the testing data.

2.4.1 Dimensionality Reduction

In many cases, it is necessary to transform the set of D-dimensional input vectors \( X \) into a space of lower dimensionality. There are several reasons for this. First, machine learning problems suffer from the so-called ‘curse of dimensionality’. In a very high-dimensional space, the amount of training data necessary to
fully characterize a classification problem becomes very large. When \( D \) is increased without a commensurate increase in the amount of available training data, predictive performance on new testing data is compromised. Moreover, as \( D \) increases, more and more complex methods are necessary to capture the dependencies within the data, causing computation to become unwieldy [25]. Second, in a high-dimensional feature space, irrelevant features (i.e. those that have no predictive value for the targets) and redundant features (i.e. those that contain the same predictive information as other features) are commonplace. Training learning algorithms on irrelevant and redundant features often has deleterious effects on the ability to classify new testing samples [84].

Several approaches can be used for dimensionality reduction, and a full discussion of these approaches is beyond the scope of this thesis. However, there are two major categories of dimensionality reduction methods - feature selection methods, which identify a subset of the original variables that is optimal in some way for classification, and feature projection methods, which transform the original data to a lower-dimensional space [48]. This section provides an overview of feature selection and describes one feature selection algorithm used frequently within this thesis.

**Feature Selection and the Fast Correlation-Based Filter**

Feature selection methods can be split into two categories: wrapper methods and filter methods [84]. Wrapper methods involve the usage of an induction algorithm on an inner cross-validation of the training set to estimate some measure of model fit. They are computationally expensive and sensitive to the choice of classifier. Filter methods evaluate features based on some general measure of similarity to the target vector. They are less computationally expensive than wrapper methods and generate reduced representations that often generalize well to various classifiers [84]. The feature selection algorithm used most frequently within this thesis is the fast correlation-based filter (FCBF).

The FCBF, as the name implies, is an example of a filter method [245]. The goal of the FCBF is to transform the original feature set \( F = (F_1, F_2, \ldots, F_D) \) to a reduced feature set \( F' = (F'_1, F'_2, \ldots, F'_M) \), where \( M < D \), \( M, D \in \mathbb{N} \). This is accomplished by selecting individual features from \( F \) rather than by combining features from \( F \). For a set of \( N \) observations \( \mathbf{X} \), each of which comprises \( D \) feature values, and a corresponding set of target values \( \mathbf{t} \), we used an adapted version of the FCBF algorithm from [245] to construct a reduced feature set \( F' \) of dimensionality \( M \) from the original feature set \( F \):
1. For every feature $F_i$, $1 < i < D$, compute a similarity measure $S_i$ between $F_i$ and $t$.

   (a) If $S_i$ is greater than a feature-target similarity threshold $\delta_{ft}$, add $F_i$ to a candidate feature set $F_p$.

2. Select the feature $F_n$ from the candidate feature set $F_p$ that has the maximum value of $S_i$ and add it to the reduced feature set $F'$.

   (a) Remove $F_n$ from the candidate feature set $F_p$.

   (b) For every remaining feature $F_j$ within the candidate feature set, compute a similarity measure $S_j$ between $F_j$ and the recently selected feature $F_n$.

   (c) If $S_j$ is greater than the inter-feature threshold $\delta_{ff}$, remove $F_j$ from the candidate feature set $F_p$.

   (d) If the dimensionality of the new feature set $F'$ is now equal to the desired dimensionality $M$, terminate the algorithm.

   (e) If the dimensionality of the new feature set $F'$ is still less than $M$, return to the beginning of Step 2 and reiterate.

This adaptation of the original FCBF algorithm allows for the desired number of features to be specified ahead of time. The employed similarity measure can be any function that returns a scalar measure between two variables. While the symmetrical uncertainty was recommended in [245], we used the absolute value of the linear correlation coefficient. $\delta_{ft}$ and $\delta_{ff}$ control the degree to which each selected feature must be predictive of the target values $t$ and uncorrelated with the other selected features. We used values of $\delta_{ft} = 0$ and $\delta_{ff} = 0.65$ throughout this thesis. In theory, the algorithm can empty the candidate feature set $F_p$ without meeting the required dimensionality $M$ for $F'$, but the risk of this occurring was limited by maintaining a low value for $\delta_{ft}$. In the event that this measure was insufficient and the candidate feature set was reduced to the empty set before the required dimensionality for $F'$ was reached, the non-redundancy requirement was relaxed by incrementally increasing $\delta_{ff}$ until the algorithm could reach the required dimensionality $M$. 

26
2.4.2 Classification

Classification algorithms are used when the target vector \( t \) contains only discrete values such as \( \pm 1 \). There are two types of classifiers. Generative classifiers learn a model of the joint probability distribution \( p(X, t) \) and render predictions by using Bayes rule to compute \( p(t|X) \) and selecting the most likely class label. Discriminative classifiers learn a direct mapping from the input data \( X \) to the target vector \( t \) [108]. Three classification algorithms were used within this thesis - linear discriminant analysis (LDA), support vector machines (SVM), and naive Bayes. This section provides a brief overview of each of these algorithms.

Linear Discriminant Analysis

Linear discriminant analysis is, as the name implies, a discriminative classifier that aims to find a linear function that best separates the different classes in \( X \) [25]. LDA projects features vectors into a one-dimensional space that maximizes the separation between classes:

\[
f(X) = w_0 + w^T \cdot X \tag{2.5}
\]

where \( w_0 \) represents the bias value and \( w \) is a weight vector. Classification decisions are rendered based on the sign of this projection as follows:

\[
t = sgn(f(X)) = sgn(w_0 + w^T \cdot X) \tag{2.6}
\]

The mathematical formulation of LDA depends upon how the separation between classes is defined. Fisher’s LDA defines this separation \( S \) as the ratio of between-class variance and within-class variance [73]:

\[
S = \frac{w \cdot \mu_1 - w \cdot \mu_2}{w \Sigma_1 w^T + w \Sigma_2 w^T} \tag{2.7}
\]

where \( \mu_1 \) and \( \Sigma_1 \) represent the mean vector and covariance matrix for the first class and \( \mu_2 \) and \( \Sigma_2 \) represent the same for the second class. It can be shown that maximizing \( S \) with respect to \( w \) leads to [25]:

27
\[ w = S_w^{-1} \cdot (\mu_1 - \mu_2) \]  \hspace{1cm} (2.8)

where \( S_w \) is the within-class scatter matrix. This choice of projection is referred to as Fisher’s linear discriminant, and maximizes the difference between the class means in the projected space while minimizing the within-class variances. We note that this formulation omits any regularization, assuming that the number of samples in the training set is sufficient to support the usage of the entire feature set. A discussion of shrinkage LDA, a regularized form of LDA that is useful when this assumption is false, may be found in [30].

For a set of D-dimensional feature vectors \( \mathbf{X} \), LDA can be interpreted as a specification for a (D-1)-dimensional hyperplane that optimally separates the two target classes. The weight vector \( \mathbf{w} \) is normal to this hyperplane while the bias \( w_0 \) stipulates the distance of this hyperplane from the origin [25].

**Support Vector Machines**

Like LDA, SVM is a linear discriminative classifier that identifies a (D-1)-dimensional hyperplane that separates a set of D-dimensional feature vectors \( \mathbf{x} \). However, for SVM, this hyperplane is optimal in the sense that it provides maximal margin, defined as the distance between the hyperplane and the nearest training points from either class [25]. The boundary of this hyperplane can be calculated by maximizing the margin subject to a constraint that no feature vectors fall within the margin [95].

One important characteristic of SVM is that this optimization depends only on dot products between the input feature vectors [95]. This allows SVM to be used even when the input data are not separable in the feature space. A kernel function can be used to map the input feature vectors into a higher-dimensional space in which they are linearly separable. A linear decision boundary in this higher-dimensional space can thus be transformed into a highly complex decision boundary within the original input space. This adaptability to data that are not linearly separable provides a significant advantage over LDA.

Various kernel functions have been used in previous SVM research. Some of the most common are the linear, polynomial, and Gaussian, or radial basis function (RBF), kernels [103, 154]. Within this thesis, we employed the RBF kernel, defined as:
\[ \phi(x_i, x_j) = exp\left(\frac{-||x_i - x_j||^2}{2\sigma^2}\right) \] (2.9)

**Naive Bayes**

Naive Bayes is a generative classifier based on Bayes’ theorem. Suppose that a sample feature vector \( x_i \in \mathbb{R}^D \) belongs to one of two classes such that \( t_i = \pm 1 \). The posterior probability of membership in each class can be written by Bayes’ theorem as [25]:

\[
\begin{align*}
p(t_i = 1|x_i) &= \frac{p(t_i = 1) p(x_i|t_i = 1)}{p(x_i)} \quad (2.10) \\
p(t_i = -1|x_i) &= \frac{p(t_i = -1) p(x_i|t_i = -1)}{p(x_i)} \quad (2.11)
\end{align*}
\]

where \( p(t_i = \pm 1) \) represents the prior probability for each class. By making a conditional independence assumption for each feature \( F = (F_1 F_2 \cdots F_D) \), each conditional probability can be expanded as [129]:

\[ p(x_i|t_i = \pm 1) = \prod_{i=1}^D p(x_i|t_i = \pm 1) \quad (2.12) \]

The conditional probability for each individual feature can be estimated using the Gaussian distribution specified by the sample mean and standard deviation for each class within the training data [107]. Ignoring the denominators of Equations 2.10 and 2.11, the class membership for a sample feature vector \( x_i \) can be predicted by estimating the posterior probability that this feature vector could have been produced by the observed distribution of each class:

\[ t_i = \arg\max_C \left[ p(C) \prod_{i=1}^D p(x_i|C) \right], \ C \in \pm 1 \quad (2.13) \]
Chapter 3

Feature Clustering for Robust Frequency-Domain Classification of EEG Activity

With permission from Elsevier, the majority of this chapter has been reproduced from the following journal article: Myrden A., and Chau T. Feature clustering for robust frequency-domain classification of EEG activity. *Journal of Neuroscience Methods*, in press, 2016. ©2016. This manuscript version is made available under the CC-by-NC-ND 4.0 license provided at [http://creativecommons.org/licenses/by-nc-nd/4.0/](http://creativecommons.org/licenses/by-nc-nd/4.0/).

3.1 Abstract

Background: The analysis of electroencephalograms is often performed in the frequency-domain. These analyses typically involve the computation of spectral power either over pre-defined frequency bands (e.g. theta, delta, alpha, and beta bands) or over a large number of narrow frequency ranges. However, the former technique ignores variability in the frequency bands over time and between participants while the
latter ignores the significant redundancy and correlations between these spectral powers.

New Method: This paper details an unsupervised feature extraction method for EEG data that uses a clustering of features to automatically agglomerate narrow-band spectral powers based on their similarities. This method computes a set of analogues to the traditional frequency bands that are data-driven rather than pre-defined and participant-specific rather than participant-independent.

Results: The new feature clustering algorithm was applied to the detection of changes in three mental states and the detection of the performance of three mental tasks. Balanced classification accuracies approaching or exceeding 70% were attained for each classification problem.

Comparison to Existing Methods: Classification accuracies attained by this algorithm were compared to those attained by two frequency-domain algorithms that did not employ clustering - a wide-band algorithm based on the spectral power within the theta, delta, alpha, and beta bands and a narrow-band algorithm based on the spectral power within 1-Hz ranges. Across the six classification problems considered, the feature clustering algorithm was statistically superior to both the wide-band and narrow-band algorithms.

Conclusions: The new feature clustering algorithm provides a promising alternative to conventional frequency-domain EEG analysis.

3.2 Introduction

Electroencephalography, or EEG is a clinical technique that is used to record fluctuations in electrical activity on the surface of the scalp [166]. Since these fluctuations are caused in part by changes in covert mental activity, EEG is frequently used as a means of studying cognition. This includes applications in cognitive state monitoring [105, 110] and brain-computer interfaces [60, 241]. In these fields, recorded EEG signals are used to detect patterns of brain activity that are of practical interest. Examples include the detection of underlying cognitive states (e.g. fatigue [105] or alertness [110]) and the detection of the performance of specific tasks (e.g. motor imagery) [60, 241]. Machine learning algorithms are usually required to achieve accurate and reliable single-trial detection [126, 139].

Analysis of EEG recordings is often performed in the frequency domain [139]. Cortical neuronal activity within different frequency bands (the theta, delta, alpha, and beta bands) has been correlated with different
psychological variables and conditions [90, 112]. Many analyses rely upon the computation of spectral power within each of these bands [105]. However, it can be difficult to accurately specify the endpoint of each frequency band, since these bands can vary over time and between participants [61, 85]. Consequently, small deviations in the definitions used for each frequency band may affect the performance of EEG-based analyses of cognitive activity. A data-driven approach that flexibly identifies the information-bearing frequency bands may generalize more effectively over time, between participants, and to different applications [85].

One alternative to this traditional approach is to compute the spectral power over many narrow frequency ranges. For example, rather than using estimations of the power within the four major frequency bands, spectral power can be computed in 1 Hz increments for the EEG signals recorded from each electrode [18]. Each of these computed powers represents a signal feature, and the most useful subset of features for each participant can be identified using any standard feature selection algorithm [84]. The usage of many precise features rather than a few general features from each electrode allows this approach to flexibly identify frequency ranges of interest, but it also introduces several disadvantages. The most obvious is that increasing the dimensionality of the feature set also increases the complexity and computational requirements of the analysis. More important, however, is the issue of redundancy. Spectral power features extracted from similar cortical locations at similar frequencies are very highly correlated. The presence of these highly correlated features within the feature set complicates feature selection, as redundant features, like irrelevant features, have deleterious effects on the performance of learning algorithms [245].

Feature selection methods are often divided into two main categories - wrapper and filter methods [245]. Wrapper methods identify an optimal feature subset by comparing the performance of an inductive learning algorithm over a wide array of possible subsets. This is often effective but is also computationally intensive and time-consuming. Filter methods, on the other hand, do not involve a learning algorithm. Simple examples of filter methods include ranking features by their correlation with the target labels or by using Fisher’s criterion. EEG features sets can be quite high-dimensional, particularly if spectral power is computed over many narrow frequency ranges for each electrode. This motivates the usage of filter methods to limit the required computational time. Simple filter methods like those discussed previously only remove irrelevant features from the feature set, as they assess features based only on their relation to the target variable. This ignores the redundancy between features that characterizes a spectral power-based EEG feature set. Effective feature selection within this feature set therefore requires an algorithm that is capable of both removing irrelevant features and compensating for redundant features.
A number of such algorithms have been proposed. Correlation-based Feature Selector (CFS) chooses a subset of features such that these features are highly correlated with the target and uncorrelated with each other [86]. Fast Correlation-based Filter (FCBF) pursues the same goal without an involved subset search, greatly reducing complexity and computational time for high-dimensional data [245]. Conditional Mutual Information Maximization (CMIM) differs from FCBF mainly in the usage of conditional mutual information as a metric for feature-target and feature-feature similarity rather than the linear correlation coefficient or symmetrical uncertainty [72]. Minimal-redundancy-maximal-relevance (MRMR) classification also uses mutual information to identify a subset of relevant features with low redundancy, and has been used as part of a two-stage feature selection approach that includes a wrapper algorithm as well [175]. Feature selection within highly correlated high-dimensional feature sets can also be performed using methods based on a clustering of features. Such algorithms were originally used to identify groups of correlated words for text classification [57]. This can be performed in either a supervised fashion using the empirically observed probabilities of each word belonging to each document class [57, 106] or in an unsupervised fashion using semantic properties of each word [15]. Recent work has investigated the applications of clustering algorithms to feature selection for other types of classification problems using graph-theoretic clustering (the FAST algorithm) [212].

One property unifying these algorithms is the way in which they treat redundant features. CFS, FCBF, CMIM, and MRMR discard redundant features entirely. Even in the FAST algorithm, all features are clustered but ultimately just one feature is selected as a representative for each cluster, and the rest are again discarded. However, it has been shown in the past that even presumably redundant features may have value during classification, so long as the features are not perfectly redundant [84]. Since features are rarely perfectly correlated in practice, there may be value in seeking a way to combine clustered features rather than selecting one and discarding the rest. This is particularly true if individual features are noisy and weakly predictive. In this case, simply selecting one feature from a set of correlated features may lead to poor generalizability and vulnerability to outliers.

This study investigates a new clustering-based feature selection method for EEG analysis that bridges the gap between the analysis of the spectral powers from wide pre-defined frequency ranges and the analysis of the highly correlated spectral powers from many narrow frequency ranges. Here, the extracted features are the spectral powers from many narrow frequency ranges for each electrode. However, by using clustering to identify highly-correlated subsets of these features, individual features can be agglomerated into groupings
that represent data-driven, participant-specific analogues of the traditional EEG frequency bands. This negates the need for global, artificially rigid definitions of these bands. Each cluster can then be independently compressed to reduce the dimensionality of the feature set without discarding features. We hypothesize that, for each of a diverse set of classification problems, the feature clustering algorithm will match or exceed the overall performance attained by using either narrow or wide frequency ranges for feature extraction without feature clustering.

3.3 Algorithm

3.3.1 Definitions

Supervised similarity metrics based on the target document classes are typically used to guide clustering during text analysis [57]. We eschewed this approach in favor of using an unsupervised clustering method with a similarity metric based directly on the relationships between features. This allowed the derivation of clusterings that may be useful for generic EEG analysis rather than specialized applications. For the $i$th sample of recorded EEG data of generic length, we define the feature vector $\Theta_i$ as:

$$\Theta_i = \{P_{i,f_1}^e, P_{i,f_2}^e, \ldots, P_{i,f_{max}}^e\} = \{\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,N_f}\}$$

(3.1)

where $P_{i,f_i}^e$ represents the spectral power of the EEG signal recorded from electrode $e_1$ within the frequency range $(f_1 - \Delta f, f_1)$, $N_e$ represents the total number of electrodes used in the recording, and $f_{max}$ represents the maximum frequency of interest at each electrode. We used non-overlapping 1 Hz frequency ranges such that $\Delta f = f_2 - f_1 = f_3 - f_2 = \cdots = 1$, although many other configurations are possible. The second, simplified notation will be used for the remainder of this section, with $N_f$ representing the total number of features per sample, equal to $(N_e \cdot f_{max})$ For a set of feature vectors $\Theta = \{\Theta_1, \Theta_2, \ldots, \Theta_{N_s}\}$, we define the values of all samples for one feature $\hat{F}_j$ as:
where $\Theta_i$, $1 < i < N_s$, and $\hat{\Theta}_j$, $1 < j < N_f$, therefore represent the rows and columns of the feature matrix. We then define a similarity matrix between features as:

$$S = \begin{bmatrix}
\rho_{1,1} & \cdots & \rho_{1,N_f} \\
\vdots & \ddots & \vdots \\
\rho_{N_f,1} & \cdots & \rho_{N_f,N_f}
\end{bmatrix}$$

(3.3)

where $\rho_{j,k}$ is the Pearson correlation coefficient between $\hat{\Theta}_j$ and $\hat{\Theta}_k$, expressed as:

$$\rho_{j,k} = \frac{\text{cov}(\hat{\Theta}_j, \hat{\Theta}_k)}{\sigma_{\hat{\Theta}_j} \sigma_{\hat{\Theta}_k}}$$

(3.4)

The similarity matrix can then be expressed as $S = \{P_1 P_2 \cdots P_{N_f}\}$, where $P_i = [\rho_{1,i} \rho_{2,i} \cdots \rho_{N_f,i}]$ is simply a column vector expressing the similarity between the $i$th feature and all other features. Now, based on $S$, we compute a distance matrix $D$ specifying the distance $D_{j,k}$ between each pair of features $(\hat{\Theta}_j, \hat{\Theta}_k)$ as:

$$D = \begin{bmatrix}
D_{1,1} & \cdots & D_{1,N_f} \\
\vdots & \ddots & \vdots \\
D_{N_f,1} & \cdots & D_{N_f,N_f}
\end{bmatrix}$$

(3.5)

$$D_{j,k} = ||P_j - P_k|| = \sqrt{(\rho_{1,j} - \rho_{1,k})^2 + (\rho_{2,j} - \rho_{2,k})^2 + \cdots + (\rho_{N_f,j} - \rho_{N_f,k})^2}$$

(3.6)

Suppose that we have a set of $N_c$ clusters $C = \{C_1 C_2 \ldots C_{N_c}\}$ and a membership vector $M = \{m_1 m_2 \cdots m_{N_f}\}$ representing the cluster membership of each feature, where $m_n \in [1, \ldots, N_f]$, $n = 1, \ldots, N_f$, and $m_n = k$ when the $n$th feature belongs to cluster $C_k$. We define the distance between a feature $\hat{\Theta}_j$ and a cluster $C_k$
in terms of the inter-feature distance as:

\[ DC_{j,k} = \frac{\sum_n D_{j,n}}{|C_k|} \quad \forall n \mid m_n = k \]  

(3.7)

\[ |C_k| = \sum_{n=1}^{N_f} \delta(m_n - k) \]  

(3.8)

where \( \delta(\cdots) \) is the Kronecker delta function. For an arbitrary clustering \( C \), the fit of each cluster \( C_k \) was computed based on the ratio of within cluster similarity \( S_{wk} \) to between cluster similarity \( S_{bk} \), defined here as:

\[ S_{wk} = \frac{\sum_n DC_{n,k}}{|C_k|} \quad \forall n \mid m_n = k \]  

(3.9)

\[ S_{bk} = \max_l \left( \frac{\sum_n DC_{n,k}}{|C_l|} \right) \quad \forall n \mid m_n = l, \forall l \in [1, N_c], l \neq k \]  

(3.10)

For any feature \( \hat{\Theta}_j \) that represents the spectral power at an electrode \( e_j \) within a 1 Hz range terminating at \( f_j \), we define the adjacent features as those located at \( \{e_j, f_j \pm 1\} \) and \( \{e_n, f_j\} \), where \( e_n \) is the electrode spatially closest to \( e_j \). Cluster expansion was restricted by requiring new additions to a cluster to be adjacent to one or more features already belonging to that cluster. This ensured that clusters expanded only in the local topographic area where high correlations were expected.

### 3.3.2 Procedure

The inputs to the clustering algorithm are:

- A distance matrix \( D \) specifying the distance between each pair of features \( \hat{\Theta}_i \) and \( \hat{\Theta}_j \)
- A threshold value \( T \) for adding unclustered features to existing clusters
- A maximum number of iterations without improvement \( I_{limit} \)
- A membership vector \( M \) initialized with all values equal to zero
• Measures to track clustering performance - a maximum cluster quality score $Q_{\text{max}}$, initialized to zero, and corresponding membership vector $M_{\text{max}}$, likewise initialized with all values equal to zero.

The algorithm proceeds as follow:

1. Check proportion of currently unclustered features. If less than 90% of features are clustered:

   (a) Randomly select a seed feature $\hat{\Theta}_s$ for a new cluster $C_k$ from the set of all unclustered features.

   (b) Initialize a test set comprising the features adjacent to $\hat{\Theta}_s$.

   (c) Compute $DC_{j,k}$ for all features $\hat{\Theta}_j$ in the test set.

   (d) Identify the candidate feature, $\hat{\Theta}_c$ with the lowest value of $DC_{c,k}$ of all features in the test set.

   (e) If $DC_{c,k} < T$ and $\hat{\Theta}_c$ either does not belong to another cluster ($m_c = 0$) or belongs to another cluster $C_l$ ($m_c = l$) such that $DC_{c,k} < DC_{c,l}$:

      i. Add $\hat{\Theta}_c$ to $C_k$ (set $m_c = k$).

      ii. Add all features adjacent to the newly added feature $\hat{\Theta}_c$ to the test set.

      iii. Return to (c).

   (f) If $DC_{c,k} < T$ but belongs to another cluster $C_l$ ($m_c = l$) such that $DC_{c,l} < DC_{c,k}$:

      i. Remove $\hat{\Theta}_c$ from the test set.

      ii. Return to (c).

   (g) If $DC_{c,k} > T$:

      i. Remove all features from the test set.

      ii. Return to 1 to form a new cluster $C_{k+1}$.

2. If more than 90% of features have been added to clusters:

   (a) Disband any cluster $C_k$ for which $|C_k| < 10$, and, for all $\hat{\Theta}_j$ in $C_k$, set $m_j = 0$. 

37
(b) For each remaining feature $\hat{\Theta}_j$ for which $m_j = 0$, find all pre-existing clusters $C_k$ to which features adjacent to $\hat{\Theta}_j$ belong. From this set of clusters, find the cluster $C_k$ for which $D_{j,k}$ is minimized and add $\hat{\Theta}_j$ to that cluster.

(c) Locate any features $\hat{\Theta}_j$ that belong to a cluster $C_k$ to which none of their adjacent features belong. Find the cluster $C_l$ which minimizes $D_{j,l}$ while containing at least one feature adjacent to $\hat{\Theta}_j$ and reallocate $\hat{\Theta}_j$ to this cluster.

(d) Compute $\frac{S_w}{S_b}$ for each cluster and $Q = \frac{\sum_k S_{wk}}{N_e}$. The inversion of the typical ratio prevents the algorithm from arriving at only trivial clusterings (e.g. clusters containing features from only one electrode or frequency).

(e) If $Q > Q_{max}$, set $Q_{max} = Q$ and $M_{max} = M$.

(f) If $Q \leq Q_{max}$ and has been for $I_{limit}$ iterations, terminate clustering with $M_{max}$ as the output. Otherwise, disband any clusters for which $\frac{S_{wk}}{S_{bk}} < Q$ and return to 1.

The threshold $T$ for cluster membership affects the number and compactness of clusters, and can be tuned for specific applications. Once $Q_{max}$ had been unchanged for $I_{limit}$ iterations, the clustering that yielded $Q_{max}$ was used for dimensionality reduction. Features from each electrode were compressed individually. For every cluster to which features from an electrode belonged, the mean value of all features at that cortical location belonging to that cluster was computed. Since most clusters included features from all electrodes, the dimensionality of the resultant feature set was equal to the product of the number of clusters and the number of electrodes. For threshold values ranging between 3.5 and 6, the algorithm typically yielded between four and seven clusters.

### 3.4 Testing Methods

The feature clustering algorithm was tested on two data sets to determine its value for the detection of both naturally occurring changes in mental state and intentional task-driven changes in mental state. The first data set came from a passive BCI study [161] where 11 participants completed a series of challenging mental tasks while self-reporting their perceived levels of fatigue, frustration, and attention. The subjective ratings for each state were reduced to a binary problem and a single-trial analysis was performed to differentiate
between low and high values for each of the three mental states. There were 320 trials for each participant, with lengths varying between 25 and 40 seconds. For each mental state, the class labels were split as evenly as possible between low and high.

The second data set came from an active BCI study [159] where 11 participants were trained to use a BCI based on the performance of either mental arithmetic, motor imagery (i.e. finger to thumb opposition), or word generation tasks. Each participant completed 60 trials of these three mental tasks, along with another 60 trials of rest. Each trial was five seconds long. Each non-rest task was independently differentiated from the rest task. One participant from this study was excluded from this analysis due to signal quality issues that interfered with feature clustering.

For both studies, EEG signals were recorded from 15 cortical locations (Fz, F1, F2, F3, F4, Cz, C1, C2, C3, C4, CPz, Pz, P1, P2, POz) using a B-Alert X24 wireless EEG headset (Advanced Brain Monitoring, Carlsbad, CA). Signals were filtered using a band-pass filter between 2 and 30 Hz and independent component analysis (ICA) was used to remove artefacts from eye blinks and eye movements [152]. A fast Fourier transform (FFT) was used to compute the frequency-domain representation of the EEG signal from each electrode for each trial.

Three algorithms were used to classify the data from each study:

1. Wide-band frequency algorithm (WB). In this algorithm, spectral power was computed at each electrode within typical ranges for each of the four major frequency bands (< 4 Hz for the delta band, 4-7 Hz for the theta band, 8-15 Hz for the alpha band, and 16-30 Hz for the beta band). This resulted in 60 features (four bands for each of 15 electrodes). A fast correlation-based filter (FCBF) was used to reduce this feature set to a dimensionality suitable for classification. Note that a suitable dimensionality for classification is considered as one that can be justified by the number of available samples per class [250].

2. Narrow-band frequency algorithm (NB). In this algorithm, spectral power was computed within 1-Hz ranges from 0-1 Hz to 29-30 Hz for the signal from each electrode, resulting in 450 features. Again, a FCBF was used to reduce this feature set to a dimensionality suitable for classification.

3. Feature clustering (FC) algorithm. The feature set computed for the NB algorithm was processed using the previously detailed feature clustering algorithm to construct an intermediate feature set.
With signals from 15 electrodes, the dimensionality of this intermediate feature set was generally between 60 and 100 features. A FCBF was used to reduce this intermediate set to a dimensionality suitable for classification.

The WB and NB algorithms were used for comparison because the FC algorithm represents a compromise between the hard-coded frequency bands of the WB algorithm and the high-dimensional feature set of the NB algorithm. The flexible nature of the FC algorithm should allow it to match the strengths of both the WB and NB algorithms.

For the data from the passive BCI study, between 2 and 20 features were tested for classification. For the data from the active BCI study, between 2 and 12 features were tested for classification. A linear discriminant analysis (LDA) classifier was used for each algorithm and the three algorithms were compared based on balanced classification accuracy. For the FC algorithm, a 10 x 10 (runs x folds) repeated cross-validation was used to estimate model performance for each classification problem. For the WB and NB algorithms, a permutation test was conducted for each classification problem by performing 100 iterations of the 10 x 10 repeated cross-validation with random initialization (i.e. random assignment of samples to folds) [173]. The resultant set of 100 classification accuracies was used to conduct non-parametric statistical testing. For a given classification problem and algorithm, the FC algorithm was considered statistically superior if its classification accuracy exceeded the maximum value observed during the permutation test. Conversely, the FC algorithm was considered statistically inferior if its classification accuracy was lower than the minimum value observed during the permutation test. Otherwise, the FC algorithm was considered statistically equivalent to the model being tested. More details on the data sets and protocols used for data collection can be found elsewhere [159, 161].

3.5 Results

3.5.1 Identification of Feature Clusters

Figure 3.1 depicts a set of results from the feature clustering algorithm. For each graph, every square on the grid represents a feature, with the frequency on the horizontal axis and the electrode on the vertical axis. Each clustering is from a different participant. The hypothetical clustering implied by the WB algorithm is
also included to allow comparison to the traditional EEG frequency bands. Threshold values of 5 and 4 were used for the mental state and mental task clusterings, respectively. These hyperparameters were determined by tuning the algorithm for each data set.

Figure 3.1: Sample clusterings. Each square within the grid represents a feature, with the x and y-coordinates representing the frequency range and electrode from which the feature was computed. Each different color represents a separate cluster. The color selections are arbitrary. Clusterings 1 through 3 were derived from the passive BCI study and Clusterings 4 through 6 from the active BCI study. The final clustering depicts the hypothetical division into four frequency bands per electrode that is implied by the WB algorithm. These sample clusterings show a range between 3 and 7 clusters, each of which is generally contiguous both in cortical location and frequency, although the frequency ranges identified are not always uniform across the cortex (e.g. Clustering 6). Clusterings 1 through 3 were obtained using a threshold of 5 while Clusterings 4 through 6 were obtained using a threshold of 4, demonstrating the effect of the threshold on the number of clusters formed.

Qualitatively, the obtained clusterings exhibit some similarities to the traditional EEG frequency bands.
There are cases in which the FC algorithm merges together multiple traditional bands (e.g. the delta and theta bands in Clustering 4), divides one band into multiple sub-bands (e.g. the alpha band in Clustering 5), forms multiple clusters within one frequency range by dividing the electrodes spatially (e.g. the alpha band in Clustering 2), and merges multiple frequency bands over only a subset of electrodes (e.g. the delta, theta, and alpha bands in Clustering 3). Overall, the clusters formed seem to represent meaningful topological groupings of cortical activity. This suggests that the FC algorithm can identify correlated EEG frequency bands that share predictive information in a data-driven, participant-specific way that avoids the necessity for hard-coded spectral boundaries. There are also clear differences between participants, as expected based on previous research [31, 61, 85].

Quantitatively, the Rand index [190] was used to calculate the consistency of clustering for each classification problem and also the extent to which the derived clusterings corresponded to the hypothetical clustering implied by the WB algorithm. Across all participants, clusterings were highly consistent, with average between-cluster Rand indices exceeding 0.88 for all classification problems. The derived clusterings also exhibited similarities to the hypothetical clustering, with Rand indices ranging between 0.76 and 0.83. Note that it was unnecessary to perform this analysis for each mental state from the first data set, since all samples were included for each state and class labels were ignored by the clustering algorithm.

Table 3.1: Average Rand indices across all participants when comparing all derived clusterings to each other (reflexive) or to the hypothetical WB clustering.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Classification</th>
<th>Reflexive</th>
<th>Hypothetical</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All</td>
<td>0.92</td>
<td>0.83</td>
</tr>
<tr>
<td>2</td>
<td>Mental Arithmetic</td>
<td>0.89</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Motor Imagery</td>
<td>0.88</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Word Generation</td>
<td>0.88</td>
<td>0.77</td>
</tr>
</tbody>
</table>

### 3.5.2 Mental State Classification

Figure 3.2 compares the classification accuracy of the FC algorithm to those of the WB and NB algorithms for the detection of fatigue, frustration, and attention. On balance, the FC algorithm was statistically superior to both the WB and NB algorithms. For fatigue detection, the FC algorithm was more accurate than the NB algorithm while matching the WB algorithm. For frustration detection, the FC algorithm was more accurate than the NB algorithm but less accurate than the WB algorithm for moderate dimensionalities. For attention detection, the FC algorithm was more accurate than the WB algorithm at all dimensionalities.
Figure 3.2: Comparison of each algorithm’s performance for each mental state classification problem. Fatigue detection is on the left, frustration detection in the center, and attention detection on the right. The performance of the FC algorithm is represented by the black line. For the WB and NB algorithms, a permutation test was used to repeat classification 100 times and the full range of observed classification accuracies was tabulated for each mental state. These ranges are represented by the blue and red shaded regions, respectively.

and more accurate than the NB algorithm for low dimensionalities. It is evident from Figure 3.2 that the FC algorithm was particularly effective at low feature set dimensionalities. This suggests that the FC algorithm provides an effective means of compressing EEG activity. Table 3.2 summarizes the statistical comparisons between the FC/WB and FC/NB algorithms across all dimensionalities for mental state classification.

3.5.3 Mental Task Classification

Figure 3.3 compares the classification accuracy of the FC algorithm to those of the WB and NB algorithms for the classification of mental arithmetic, motor imagery, and word generation. For mental arithmetic, the FC algorithm was more accurate than the NB algorithm at all dimensionalities and more accurate than the WB algorithm at high dimensionalities. For motor imagery, the FC algorithm was more accurate than both the WB and NB algorithms for nearly all dimensionalities. For word generation, the FC algorithm was more accurate than the NB algorithm while matching the WB algorithm. On balance, the FC algorithm was again statistically superior to both the WB and NB algorithms. Table 3.2 summarizes the statistical comparisons between the FC/WB and FC/NB algorithms across all dimensionalities for mental task classification.
Figure 3.3: Comparison of each algorithm’s performance for each mental task classification problem. Mental arithmetic is on the left, motor imagery in the center, and word generation on the right. The performance of the FC algorithm is represented by the black line. For the WB and NB algorithms, a permutation test was used to repeat classification 100 times and the full range of observed classification accuracies was tabulated for each mental task. These ranges are represented by the blue and red shaded regions, respectively.

### 3.5.4 Summary of Statistical Testing

Table 3.2 summarizes the results from Figures 3.2 and 3.3 by tallying the number of cases for which the FC algorithm was statistically superior to, equivalent to, and inferior to the NB and WB algorithms, as established by a permutation test. These comparisons were performed at all feature set dimensionalities, but, for practical purposes, the results at only the highest dimensionalities tested (20 features for mental state classification and 12 for mental task classification) are also summarized separately. Across all analyses, the FC algorithm was superior to the NB algorithm in 60 of 66 cases, equivalent in five cases, and inferior only once. Compared to the WB algorithm, the FC algorithm was superior in 29 of 66 cases, equivalent in 25, and inferior in 12. At the maximum dimensionalities, the FC algorithm was more accurate than the NB algorithm in all six cases. Similarly, the FC algorithm yielded higher accuracies than the WB algorithm in three cases and equivalent accuracies in the remaining three cases. Collectively, these results suggest that the FC algorithm can outperform both the WB and NB algorithms for the identification of discriminatory EEG spectral features.

### 3.6 Discussion

The FC algorithm was superior to the WB algorithm for nearly one half of the performed analyses and to the NB algorithm for nearly all analyses. These findings suggest that the FC algorithm yields more
Table 3.2: Summary of statistical testing for both analyses. Numbers in the FC+, FC=, and FC- columns represent the frequency with which the FC algorithm was statistically superior, equivalent, or inferior to the competing algorithm in terms of classification accuracy for the specified classification problem.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Algorithm</th>
<th>All Dimensionalities FC+</th>
<th>FC=</th>
<th>FC-</th>
<th>Max Dimensionalities FC+</th>
<th>FC=</th>
<th>FC-</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Mental State)</td>
<td>WB</td>
<td>14</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>NB</td>
<td>24</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 (Mental Task)</td>
<td>WB</td>
<td>15</td>
<td>17</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>NB</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Overall</td>
<td>WB</td>
<td>29</td>
<td>25</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>NB</td>
<td>60</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

discriminatory spectral representations than both the WB and NB algorithms, making it a valuable tool for frequency-domain EEG analysis. Moreover, the largest differences in classification accuracy (e.g. low-dimensionality attention and motor imagery classification) were in favor of the FC algorithm. In contrast, when the FC algorithm was inferior to the WB or NB algorithms, the difference in classification accuracy was small (e.g. frustration detection). This suggests that the statistical comparisons within Table 3.2 may actually underestimate the efficacy of the FC algorithm.

It is particularly interesting that the FC algorithm led to a marked enhancement in accuracy over both the WB and NB algorithms for low feature set dimensionalities. This may indicate that the FC algorithm is particularly effective at compressing EEG signals so that a great deal of information can be extracted from minimal numbers of features. The poor performance of the NB algorithm in these cases demonstrates the detrimental impact of wholly discarding the majority of the feature set. It is also intriguing that the difference between the FC algorithm and the WB and NB algorithms was more apparent for the second data set. Each trial of EEG activity was only five seconds long within this set, compared to the 25 to 40 second durations of each trial in the first data set. It is intuitive to assume that this decrease in trial duration led to more erratic features and less separable classification problems. Our results suggest that the FC algorithm is better suited to deal with these challenges than the WB or NB algorithms.

There is reason to believe that the analyses presented here understate the value of the FC algorithm in comparison to the WB algorithm. Both data sets originated from studies with small, relatively homogeneous populations. In the first study, 8 of 11 participants were female and ages ranged only between 20 and 30 years. In the second study, 9 of 11 participants were female and ages ranged only between 21 and 32 years. It is well-known that age has a significant effect on the EEG frequency bands, particularly in the case of the alpha band and the alpha peak frequency [61, 112], and gender may affect EEG activity as well [14, 41].
Consequently, the homogeneous populations in these studies may have minimized the importance of the flexibility of the FC algorithm. In studies with more heterogeneous populations, we expect that the value of the FC algorithm over the WB and NB algorithms would become more apparent.

For fatigue and frustration detection and for mental arithmetic and word generation classification, the WB algorithm was clearly superior to the NB algorithm. However, for attention detection and motor imagery classification, the NB algorithm was able to match or exceed the WB algorithm for many feature set dimensionalities. This implies that certain types of cognitive events (e.g. fluctuations in attention levels and the performance of motor imagery) are easier to detect using narrow-band spectral powers, while others (e.g. fluctuations in fatigue levels and the performance of mental arithmetic) are easier to detect using wide-band spectral powers. It is likely not coincidental that the FC algorithm clearly outperformed the WB algorithm for the former pair of classification problems. However, unlike the NB algorithm, the FC algorithm was also typically capable of matching the WB algorithm for the other four classification problems. This seems intuitive, as the FC algorithm represents a middle ground between the WB and NB algorithms. It starts with the narrow-band spectral powers used by the NB algorithm and automatically agglomerates them to construct features that represent the spectral power over wider frequency ranges, similar to the ones used by the WB algorithm. However, unlike the WB algorithm, the FC algorithm defines these wider frequency ranges automatically and flexibly based on data from each participant rather than rigidly based on manual specifications. Consequently, the FC algorithm possesses the benefits of both the NB and WB algorithms without their disadvantages.

The Rand index confirmed that the FC algorithm derives clusters of features that bear similarities to the traditional EEG frequency bands (see Table 3.1). However, greater similarity was observed during mental state classification than during mental task classification. This may be related to the threshold used for clustering. The hypothetical WB clustering contained only four clusters. During mental state classification, a threshold value of 5 was used and an average of 3.99 clusters were formed across all participants, closely matching the WB algorithm. During mental task classification, a threshold of 4 was used and an average of 6.05, 6.34, and 5.94 clusters were formed for fatigue, frustration, and attention detection, respectively. This increase in the number of clusters may have resulted in more variation from the WB clustering. The additional clusters formed when using a lower threshold value can reflect either non-uniformity in frequency bands across the cortex (see Figure 3.1f) or a dissection of the major frequency bands into smaller sub-bands (see Figure 3.1e). The latter is well-supported by the common practice of dividing some frequency bands,
particularly the alpha band, into upper and lower sub-bands for analysis [176].

The FC algorithm has several limitations. First, following feature clustering, dimensionality reduction was performed by extracting the average power within each cluster at each electrode. Although this method proved effective, it does not take cluster topology into account by averaging the power across both frequency ranges and electrodes. It may be possible to develop a more sophisticated method of extracting features from the clustering results if topological information is also taken into account. Second, the FC algorithm requires significant computational time. It is more complex than either the WB or NB algorithm, bearing a resemblance to agglomerative clustering using an average linkage criteria. Although these algorithms are often too slow for large data sets, the computational requirements remained tractable given the reasonable size of our feature set. Moreover, our algorithm was designed to merge only nearby unclustered features into existing clusters, taking advantage of a priori knowledge regarding the likelihood of correlations between features from similar spatial locations and frequencies. This limited the number of linkage criteria that had to be computed at each step, reducing computational time. The FC algorithm is sufficiently swift to be used for practical, real-time applications of EEG analysis. Future work will investigate such applications.

3.7 Conclusions

This study investigated a new clustering algorithm for feature extraction and selection in frequency-domain EEG analysis. The newly proposed FC algorithm combined the flexibility of narrow-band EEG analysis with the robustness of wide-band EEG analysis. On balance, this new FC algorithm led to higher accuracies than the WB and NB algorithms in the detection of changes in mental state and the detection of mental task performance. The FC algorithm is thus a promising alternative to conventional EEG spectral analysis and may be considered in future passive and active BCI research.
Chapter 4

A Passive EEG-BCI for Single-Trial Detection of Changes in Mental State

The majority of this chapter has been reproduced from the following article, which is currently under revision at IEEE Transactions on Neural Systems and Rehabilitation Engineering: Myrden, A., and Chau, T. A passive EEG-BCI for single-trial detection of changes in mental state. Manuscript under revision, 2016.

4.1 Abstract

Traditional brain-computer interfaces often exhibit unstable performance over time. It has recently been proposed that passive brain-computer interfaces may provide a way to complement and stabilize these traditional systems. In this study, we investigated the feasibility of a passive brain-computer interface that uses electroencephalography to monitor changes in mental state on a single-trial basis. We recorded cortical activity from 15 locations while 11 able-bodied adults completed a series of challenging mental tasks. Using
a feature clustering algorithm to account for redundancy in EEG signal features, we classified self-reported changes in fatigue, frustration, and attention levels with $76.8 \pm 8.5\%$, $71.9 \pm 5.9\%$, and $86.1 \pm 7.1\%$ accuracy (mean ± std), respectively. Based on the most frequently-selected features across all participants, we note the importance of the frontal and central electrodes for fatigue detection, posterior alpha band and central delta band activity for frustration detection, and posterior alpha band activity for attention detection. Future work will focus on integrating these results with a traditional active brain-computer interface.

4.2 Background

Brain-computer interfaces (BCIs) have traditionally been defined as systems that allow control of an external device using thoughts alone [140]. Potential applications for these devices have been identified within a diverse array of fields, including access technologies for physically impaired individuals, motor recovery, and entertainment [148]. Conceptually, BCIs have three major components - a measurement modality that monitors some physiological variable related to cognitive activity, a decoding module that translates this signal into a prediction of user intent, and an activation module that produces the desired output (e.g. a simulated key press or mouse click) [227].

Existing BCIs use a variety of measurement modalities, including both invasive and non-invasive techniques. Invasive techniques, such as electrocorticography (ECoG) and single-cell recordings, require the implantation of electrodes within the brain [128]. Non-invasive techniques, such as electroencephalography (EEG), near-infrared spectroscopy (NIRS), and transcranial Doppler (TCD) ultrasound, can be safely recorded from the surface of the skull [163, 186, 188]. Most of these BCIs recognize and react to a specific voluntary cognitive activity. Typical cognitive activities for BCI applications include the modulation of visuospatial attention (e.g. the P300 response), the performance of covert mental tasks (e.g. motor imagery), and the control of slow cortical potentials [143].

Recently, an alternative to this standard BCI paradigm has been proposed. By focusing on voluntary cognitive activity, existing BCIs discard a wealth of information regarding the underlying mental state of the BCI user [171]. This information may include fatigue, attention, mental workload, and affective state [50, 155]. It has also been proposed that more complex aspects of user state, such as perceived loss of control over a system, may be detectable [249]. BCIs that focus on detecting and quantifying these
underlying variables have been defined as passive BCIs [247]. A diverse set of potential applications has been proposed for these devices, including cognitive state monitoring [157] and physiological computing [65]. Passive BCIs may also be useful as supplements to traditional active BCIs, providing information about the underlying user state to facilitate greater accuracy in the classification of voluntary cognitive activity [148]. This could be viewed as a type of context-awareness [246].

Previous active BCI studies have noted the apparent effects of some of these mental states on BCI performance. For example, users experiencing cognitive fatigue often struggle to complete the mental tasks required for BCI control [49]. This may lead to significant deterioration in classification accuracy. Similarly, users experiencing frustration or distraction may have more difficulty controlling a BCI - the former because the physiological changes related to frustration inhibit concentration and the latter because it diverts attention from BCI operation. It may be possible to design an active BCI that adapts to these changes to avoid compromised performance, provided that we can detect these changes with high accuracy. This adaptation could include modification of the learning algorithm itself or simply identification of an optimal range of mental states for BCI operation.

Several recent studies have investigated the detection of these mental states. For example, EEG-based fatigue detection has been the subject of both long-term sleep deprivation studies and shorter-duration studies that focus on mental task performance [206, 226]. Of these, the latter are more relevant to BCI applications. These studies have indicated that changes in fatigue levels can manifest as changes in spectral activity within each of the four major EEG frequency bands - delta, theta, alpha, and beta [105, 119]. Single-trial detection of these changes has often resulted in classification accuracies approaching 90%, although such studies have often relied on sleep deprivation protocols that may not generalize to BCI usage [206, 226, 251]. On the other hand, far fewer studies have focused on EEG-based detection of attention, and those that exist have achieved classification accuracies no higher than 75% [87, 135]. Automatic detection of frustration, meanwhile, has received little attention in previous research.

We investigated the changes in fatigue, frustration, and attention that occurred during the performance of three active mental tasks - mental arithmetic, anagram solution, and a short-term spatial memory task - and a rest task. This protocol was designed to replicate traditional active BCI studies as closely as possible. Each of the three active tasks was chosen based on its usage in previous work monitoring changes in mental state [17, 87, 191]. The rest task was chosen due to its importance to BCI usage. We hypothesized that
changes in each state could be detected reliably on a single-trial basis using EEG recordings. Furthermore, we hypothesized that using a variety of mental tasks would allow our algorithm to identify the changes in EEG that reflect fluctuations in underlying mental state, not simply those related to the performance of particular tasks.

4.3 Methods

4.3.1 Protocol

Eleven participants (eight female, mean age 25 years) were recruited from the population of students and employees at the Holland Bloorview Kids Rehabilitation Hospital. Participants had no history of brain injury, were fluent in English, and refrained from consuming caffeine for four hours prior to the experiment. Each participant completed four one-hour sessions. Within each session, participants performed a series of 80 trials, divided into four blocks of 20 trials each. During each trial, participants performed one of four mental tasks - mental arithmetic, anagram solution, a grid-recall (short-term memory) task, and rest.

During the mental arithmetic task, participants were prompted to complete an on-screen mathematical problem. Each problem involved a series of additions and/or subtractions with four operands. During the anagram task, participants were presented with a set of jumbled letters with only one English solution and prompted to enter the unscrambled word. During the grid-recall task, participants were presented with a five-by-five grid of squares. A pattern was highlighted on the grid by marking a subset of the squares red while leaving the remainder white. Participants were given five seconds to memorize the grid pattern, and, after a ten second break during which the grid was removed from the screen, were asked to recreate the initial pattern by manually highlighting squares on a blank grid. During the rest task, participants were simply instructed to relax and let their minds wander, mimicking the typical no-control state for BCI usage. Time limits were 30 seconds for mental arithmetic, 20 seconds for anagram solution, and 15 seconds for the grid-recall task.

There were five difficulty levels for each non-rest task, ranging from very easy to very hard/impossible. For the mental arithmetic task, difficulty was controlled by increasing the magnitude of and the number of significant digits in the operands. For the anagram task, difficulty was controlled by increasing the number of
letters (ranging between 5 and 8) in the scrambled word, and, for the highest level, providing a set of letters that did not have a solution. For the grid-recall task, difficulty was controlled by increasing the number of highlighted squares, thereby increasing the complexity of the pattern to be memorized. Fig. 4.1 depicts two examples of each task - one classified as very easy and one classified as very hard. A short pilot study was conducted to ensure that the assigned difficulty levels matched the experienced difficulty levels.

![Math - Easy](image1)

![Math - Hard](image2)

![Anagram - Easy](image3)

![Anagram - Hard](image4)

![Grid-Recall - Easy](image5)

![Grid-Recall - Hard](image6)

Figure 4.1: Examples of easy and difficult cases for each active task

There were an equal number of trials at each difficulty level for each task. Trials were presented in a pseudo-random order, with some intentional clustering of both easy and difficult trials. These sustained intervals during which all trials were of trivial or extreme difficulty were intended to cause peaks and troughs in each mental state over time. During each non-rest trial, participants attempted to enter the correct answer before the time limit expired. To minimize the effects of movement on the recorded EEG signals, participants were instructed to remain completely still until they were ready to enter a final answer. It was not possible
to completely avoid movement, as manual entry of answers was necessary to allow on-screen performance tracking. This direct feedback was used to motivate participants and also to induce changes in frustration when performance was poor.

After each trial, participants were prompted to self-report their perceived levels of fatigue, frustration, and attention. A five-point Likert scale was used for each measure. The perceived difficulty of each trial was also self-reported on a five point scale. Participants completed the expanded Positive and Negative Effect Schedule (PANAS-X) prior to and at the conclusion of each session [234].

4.3.2 Signal Acquisition

EEG data were collected using a wireless B-Alert X24 headset (Advanced Brain Monitoring, Carlsbad, CA, USA). Signals were acquired from electrodes placed at the F3, F1, Fz, F2, F4, C3, C1, Cz, C2, C4, CPz, P1, Pz, P2, and POz cortical locations by the international 10-20 system [102]. The sampling rate was 256 Hz, and signal quality was monitored throughout the experiment. EEG signals were acquired through a custom LabVIEW interface that also presented all required stimuli to participants.

4.3.3 Pre-Processing

Recorded EEG data were filtered using a FIR band-pass filter (1 to 30 Hz) to isolate the frequencies of interest. The ADJUST algorithm for independent component analysis (ICA) was used to remove eye movement and blink artefacts from the recorded signals [152]. Each individual trial was segmented from the recordings for further analysis using markers that were automatically inserted, during the experiment, into the recorded EEG data at the beginning (i.e. the presentation of the task stimuli to participants) and the end (i.e. the expiration of the time limit) of the trial.

4.3.4 Feature Extraction

A set of signal features was constructed for each trial. These features were based on frequency-domain analysis of the recorded EEG data. A fast Fourier Transform (FFT) was computed for the EEG signal from every electrode location for each trial. The resultant frequency spectra were used to compute the total
spectral power within each non-overlapping 1 Hz frequency range from 0-1 Hz to 29-30 Hz. Each of these spectral power measurements was used as a feature for classification. For each trial, there were a total of 450 features (30 different frequencies from 15 different electrodes).

Following feature extraction, a 10 x 10 (runs x folds) repeated cross-validation was used to estimate the accuracy of mental state detection. The same initial feature set was used for each mental state, but feature selection and classification were performed independently for each state, resulting in the construction of a unique classifier for each mental state.

### 4.3.5 Feature Selection

Two stages of dimensionality reduction were performed. First, a participant-specific feature clustering algorithm was used to group the features from each electrode into data-sensitive frequency bands [160]. This algorithm uses inter-feature correlations to identify highly similar clusters of features and derives individualized frequency bands rather than using the traditional definitions of the delta, theta, alpha, and beta bands. The 30 original frequency-domain features from each electrode were compressed into four to seven features per electrode, each of which represented the mean of a set of original features. Subsequently, a fast correlation-based filter (FCBF) [245] with an inter-feature threshold of 0.65 was used to select between 2 and 20 features for classification. Feature selection was performed using only the training data during each fold of the cross-validation.

### 4.3.6 Data-Sensitive Class Balancing

Due to the subjective and self-reported nature of the class labels, most participants had significant levels of class imbalance (i.e. certain classes had many more samples than others). Both class reduction and oversampling methods were used to address this issue.

Initially, data-sensitive class reduction was used to reduce classification to a binary problem (e.g. low vs. high fatigue). During each fold of cross-validation, the number of classes was reduced based on the training data. First, training samples from scarce (defined as class frequency less than 10%) classes were combined with the nearest non-scarce class, as determined by the Euclidean distance between the mean feature vector for
each class. Next, the smallest between-class Euclidean distance for the remaining classes was identified, and these two classes were combined. This procedure continued iteratively until only two classes remained. This process is depicted for one participant’s attention ratings in Fig. 4.2. The class transformations resulting from this algorithm were then applied to the test data. Following class relabeling, the training data were cleaned by removing outliers. Any points of either class for which the nearest five neighbours all belonged to the opposite class were removed from the training set.

![Figure 4.2: Iterative data relabeling for Participant 1 attention levels. The top row displays only the class distributions while the bottom row shows all data. Nearby classes are iteratively merged until only two classes remain. The final graph is prior to the removal of outlying data points and resampling.](image)

Next, an oversampling approach based on the Borderline-SMOTE algorithm was employed to balance the class distribution [89]. A k-nearest neighbours algorithm (k = 5) was used to classify each training point. A training point belonging to the minority class was defined as a ‘dangerous point’ if it was misclassified (i.e. if it had more neighbours belonging to the majority class than the minority class). Synthetic training points were then generated for the minority class based on these ‘dangerous points’. A dangerous point was randomly selected, and a synthetic point was then generated at the mean of the feature vectors of five of the ten nearest training points from the minority class, selected at random. This procedure was repeated until there were an equal number of training points in each class.
4.3.7 Classification

Three different classifiers were investigated - linear discriminant analysis (LDA), support vector machines (SVM), and Gaussian naive Bayes (NB). For the SVM classifier, a radial basis function kernel was used. Values were set for $\sigma$ and the regularization constant $C$ using a grid search over relevant values of each parameter ($2^2 \leq \sigma \leq 2^4$, $2^{-3} \leq C \leq 2^{-1}$) during an inner cross-validation on the training set. Target dimensionalities between two and 20 features were used for all classifiers. Classifier performance was measured using the balanced classification accuracy (i.e. the mean of sensitivity and specificity). The cortical areas and frequencies most often used for classification of each state were identified.

Two alternative cases were considered for practical reasons. In the first, smaller subsets of electrodes were used for classification to investigate the feasibility of fatigue, frustration, and attention detection using less complex EEG headsets. In the second, the inter-participant generalizability of mental state detection was investigated by performing feature selection for each participant based only on the remaining 10 participants, leaving classifier training as the only participant-dependent step in the classification algorithm.

4.4 Results

4.4.1 Mental State Detection

Fig. 4.3 shows the relationship between number of features and balanced classification accuracy for each classifier. Based on the adjusted Wald interval, the upper bound of the 95% confidence interval for the balanced classification accuracy of each mental state on the participant level (320 trials) and the study level ($11 \times 320 = 3520$ trials) were 55.45% and 51.65%, respectively [19]. The upper bounds of the 95% confidence interval for balanced classification accuracy was also computed for the 10-feature LDA classifier using a 100-run permutation test with randomly shuffled class labels, resulting in values of 51.45% for fatigue, 52.82% for frustration, and 51.31% for attention [173]. This closely corresponded with the adjusted Wald interval, confirming that theoretical chance levels were approximately 50%. Due to the computational complexity of performing a permutation test for each combination of mental state, classification algorithm, and feature set dimensionality, the adjusted Wald interval was used to gauge significance for all analyses. All three classifiers exceeded chance performance levels for all feature set dimensionalities.
Table 4.1 presents the balanced classification accuracies attained by each algorithm at the maximum feature set dimensionality of 20 features. Using the Wilcoxon rank-sum test, the LDA algorithm was statistically superior ($p < 0.05$) to the NB algorithm for the classification of all three mental states and statistically superior ($p < 0.05$) to the SVM algorithm for the classification of fatigue. The NB classifier was thus omitted from further analysis.

Table 4.1: Mean and standard deviation of balanced classification accuracy across all participants for each mental state using 20-feature LDA, SVM, and NB classifiers.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Fatigue</th>
<th>Frustration</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>76.8 ± 8.5</td>
<td>71.9 ± 5.9</td>
<td>86.1 ± 7.1</td>
</tr>
<tr>
<td>SVM</td>
<td>72.9 ± 8.8</td>
<td>70.7 ± 5.7</td>
<td>84.9 ± 7.2</td>
</tr>
<tr>
<td>NB</td>
<td>68.6 ± 9.3</td>
<td>68.2 ± 5.5</td>
<td>81.1 ± 8.6</td>
</tr>
</tbody>
</table>

Fig. 4.4 depicts the balanced classification accuracy for each participant and mental state using both the LDA and SVM classifiers. Balanced classification accuracies using both LDA and SVM exceeded chance levels for all participants and mental states. Since the LDA classifier exhibited the highest balanced classification accuracies for all three mental states, it was selected for further analysis.
4.4.2 Feature Analysis

We sought a physiological understanding of changes in fatigue, frustration, and attention through an analysis of the most frequently selected features for detection of all three states. The original EEG feature set comprised 450 features, each of which represented the spectral power at one electrode within a specific frequency range. The frequency of feature selection was computed for each of these original features as follows. Within the 10 x 10 repeated cross-validation, 100 classifiers were trained for each mental state. Each of these classifiers was trained using a reduced set of features selected by an FCBF. Due to the clustering algorithm that we employed, each feature used for classification was the sum of some set of original features. The feature selection frequency for each original feature was defined as the proportion of classifiers for which that original feature was used to compute one of the features used for classification. These feature selection frequencies were then averaged across all participants to determine which features, electrodes, and frequencies were most predictive of changes in each mental state.

Figure 4.5 depicts feature selection frequency for each feature and mental state. Each feature is represented by an oval at the appropriate point on the grid, with the spectral frequency on the x-axis and the electrode on the y-axis. The size of the oval represents the frequency of feature selection, with larger ovals representing features that were selected more frequently. The vertical and horizontal lines represent the average feature selection frequency for all features from the same electrode and frequency, respectively. This visualization was obtained using the results from 2-feature LDA classification of mental state to ensure that only the most
predictive features were identified.

Figure 4.5: Most important frequencies and electrode locations for predicting changes in (a) fatigue, (b) frustration, and (c) attention. For each feature, the value on the x-axis represents spectral frequency and the value on the y-axis represents cortical location. The size of each oval represents how frequently that feature was used for classification across all participants on average. Fatigue was most frequently classified using features from the frontal and central electrodes within the delta, theta, alpha, and beta frequency bands. Frustration was most frequently classified using alpha band features from the posterior electrodes and delta band features from the central electrodes. Attention was most frequently classified using alpha band features from the posterior electrodes.

Clear differences can be observed between mental states. The features used for fatigue detection most commonly originated from the frontal and central electrodes, particularly those offset from the midline of the cortex. There was a broad range of frequencies amongst these features, including alpha band activity from the central electrodes and activity from all four major frequency bands from the frontal electrodes. Both frustration and attention were frequently classified using features that represented alpha band activity from the posterior electrodes, although frustration detection was also dependent upon features from some
other electrodes in the central and frontal regions.

4.4.3 Effects of Electrode Removal

The extreme electrode dependence of mental state classification depicted by Figure 4.5 encouraged an investigation of the effects of electrode removal on balanced classification accuracy. Classification was performed using 20-feature LDA for six different subsets of electrodes, each of which was compared to the full set of 15 electrodes. These subsets are summarized by Table 4.2. The feature set derived from each electrode subset was reduced using feature clustering and a FCBF before classification.

Table 4.2: Electrode subsets. Classification was performed independently for each subset to gauge the effects of electrode removal.

<table>
<thead>
<tr>
<th>Label</th>
<th>Electrodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Fz, F1, F2, F3, F4</td>
</tr>
<tr>
<td>C</td>
<td>Cz, C1, C2, C3, C4</td>
</tr>
<tr>
<td>P</td>
<td>CPz, Pz, POz, P1, P2</td>
</tr>
<tr>
<td>FC</td>
<td>F ∪ C</td>
</tr>
<tr>
<td>FP</td>
<td>F ∪ P</td>
</tr>
<tr>
<td>CP</td>
<td>C ∪ P</td>
</tr>
<tr>
<td>FCP</td>
<td>F ∪ C ∪ P</td>
</tr>
</tbody>
</table>

The balanced classification accuracies attained using each electrode subset for each mental state are presented in Fig. 4.6. Comparisons were conducted between classification results using the Wilcoxon rank-sum test. As expected, using the full FCP set of electrodes provided the highest balanced classification accuracies for each state. However, some electrode subsets provided statistically equivalent performance to the FCP set despite their reduced dimensionality. For fatigue detection, the FC subset (p = 0.98) and the FP subset (p = 0.19) were statistically equivalent to the FCP set. For frustration detection, the FC (p = 0.21), FP (p = 0.078), and CP subsets (p = 0.31) were statistically equivalent to the FCP set. For attention detection, the FCP set was statistically superior (p < 0.05) to all other electrode subsets.
4.4.4 Generalization Between Participants

In the previous analyses, every stage of the classification algorithm was participant-dependent. It is of practical interest to investigate methods of rendering this algorithm participant-independent. We examined the generalizability of the algorithm using a participant-level block cross-validation. Data from each participant were set aside in turn while feature clustering and feature selection were performed using data from all other participants. A classifier was then trained for each participant within a 10 x 10 repeated cross-validation. This allowed us to investigate the consistency of the neural correlates of each mental state across the population of the study. However, individualized LDA classifiers were still trained for each participant within a 10-fold cross-validation to accommodate individual differences in the importance of each selected feature to classification.

Fig. 4.7 shows the relationship between number of features and balanced classification accuracy for detection of each mental state. Results from the initial analysis with participant-dependent feature clustering and feature selection are also shown for comparison. Balanced classification accuracies for both classifiers at 20 features are presented in Table 4.3. Using the Wilcoxon rank-sum test, there were no significant differences between the participant-dependent and participant-independent algorithms for this dimensionality.
Figure 4.7: Comparison between LDA balanced classification accuracy after participant-dependent (black line) and participant-independent (blue line) feature selection. The participant-dependent algorithm was much more accurate for low feature set dimensionalities but this difference was less pronounced as the number of features increased. For both algorithms, participant-dependent classifiers were trained after feature selection.

Table 4.3: Mean and standard deviation of balanced classification accuracy across all participants for the 20-feature LDA classifier for participant-dependent (PD) and participant-independent (PI) feature selection.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Fatigue</th>
<th>Frustration</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD</td>
<td>76.8 ± 8.5</td>
<td>71.9 ± 5.9</td>
<td>86.1 ± 7.1</td>
</tr>
<tr>
<td>PI</td>
<td>75.2 ± 8.7</td>
<td>70.6 ± 6.5</td>
<td>85.2 ± 7.8</td>
</tr>
</tbody>
</table>

Fig. 4.8 depicts the results of a feature analysis for the participant-independent analysis. Unsurprisingly, there appeared to be less diversity for this analysis than for the participant-dependent analysis in Fig. 4.5. Fatigue was primarily detected using features that originated from F4 at frequencies between 10 and 20 Hz, frustration by features that originated from frontal electrodes between 4-7 and 10-20 Hz, and attention by features that originated from Fz below 5 Hz and P1 between 7 and 12 Hz.
Figure 4.8: Most important frequencies and electrode locations for predicting changes in (a) fatigue, (b) frustration, and (c) attention when participant-independent feature selection was used. There was much less variance than exhibited by the participant-dependent case in Fig. 4.5. Fatigue was most frequently classified by using alpha band power from F4, frustration by using theta and alpha band power from the frontal electrodes, and attention by using delta band power at Fz and alpha band power from P1.

4.5 Discussion

4.5.1 Feasibility of Detecting Changes in Fatigue, Frustration, and Attention

This study has demonstrated the ability to predict changes in three mental states - fatigue, frustration, and attention - with classification accuracies exceeding 70%. These changes were detected during the performance of four entirely different mental tasks, suggesting that the neuroelectric manifestations of changes
in these states are consistent across a range of cognitive processes. The demonstrated level of mental state
detection sets the stage for practical applications of mental state monitoring. In particular, since it has been
hypothesized that traditional BCIs are susceptible to changes in mental state [148, 207], pairing this passive
BCI with a traditional BCI may allow the design of an adaptive BCI that is more robust to fluctuations in
mental state.

We found no advantage to using SVM or NB rather than LDA for classification. The inferiority of SVM
implies that the features used for classification were linearly separable, which is unsurprising given that the
metric used for feature selection was a measure of linear correlation. We hypothesize that the NB algorithm
may have performed poorly because the features used for classification were not statistically independent,
violating the conditional independence assumption that is central to this classifier.

One limitation of this study was that task order was not randomized for each participant, with the conse-
quence that this could have had systematic effects on the results. However, we eschewed randomization due
to the necessity of including sequences of trivial and extremely difficult tasks to induce changes in mental
state. Moreover, although participants all completed the same series of tasks, they each had unique reac-
tions to each task due to variations in their performance, their aptitude for each task, and their subjective
enjoyment of each task. Consequently, we believe that any systematic effects on our results from this aspect
of the experimental protocol are minimal.

4.5.2 The Continuous Nature of Mental State

In this study, we chose to model the subjective experience of the user, as captured by their self-reported
ratings of fatigue, frustration, and attention. Two disadvantages of this approach were the inability to ensure
balanced target classes for each mental state and the potential inaccuracy of self-reported ratings [187]. The
former issue was resolved by using an oversampling algorithm to balance the classes prior to training. The
latter issue was addressed by using a data-sensitive relabeling approach that combined similar classes and
removed outlying training data to attempt to correct inaccuracies in the self-reported ratings. This had the
effect of reducing our classification problem from five classes to two, comparing low and high values for each
state.

Although this approach is useful, it is also an artificial interpretation of naturally continuous mental states.
Some EEG-based fatigue studies have employed three-class models [225], and moving in this direction for all three states may be valuable during future integration with an active BCI. Much of the raw data that we collected during this experiment indicate that such an approach may be viable. Fig. 4.2 displayed a simple two-dimensional interpretation of the different self-reported attention levels for Participant 1 as data reduction was performed. Even in such a low-dimensional space, it is possible to identify a continuous trend from low to high attention levels. This trend was seen across all participants and for all mental states.

This underlying structure suggests that with more training data it may be possible to increase the sensitivity of the algorithms detailed within this paper. This could be achieved either through a multi-class problem where each mental state is approximated with three or more states or a regression-based approach where the concept of classes is discarded and each mental state is estimated on a continuous scale. It may be helpful to design an alternative data reduction algorithm that automatically clusters redundant classes without iterating until the classification problem becomes binary.

4.5.3 Cortical Correlates of Mental State

EEG is an undeniably coarse modality for an analysis of the cortical regions in which changes in mental state manifest, particularly after feature clustering has been performed. However, it is worthwhile to summarize the results detailed here and relate them to previous findings. Based on our analysis of the most frequently selected features in Fig. 4.5, we conclude that changes in fatigue manifested within delta, theta, alpha, and beta band activity in the frontal electrodes, as well as alpha band activity within the central electrodes. On the other hand, frustration detection was most dependent upon delta band activity from electrodes in the central region and posterior alpha band activity, while attention detection relied heavily upon posterior alpha band activity. Participant-independent detection of changes in mental state generally relied upon similar features, particularly for fatigue (frontal alpha) and attention (posterior alpha) detection.

The relationships identified for fatigue detection are reasonably consistent with previous literature. The onset of fatigue has been characterized by an increase in the ratio of slow wave to fast wave EEG, potentially explaining why features from most frequency bands were used for classification in the participant-dependent case [105, 118]. Strijkstra et al. noted that alpha activity across the cortex and theta activity within the centro-frontal region of the cortex correlated with subjective sleepiness, while Trejo et al. found that differences in spectral power within these regions and frequency ranges were predictive of large modulations in
fatigue [215, 226]. Our results confirm these relationships while also indicating that these changes in cortical activity can be used for single-trial prediction of small fluctuations in subjective fatigue. It is possible that other cortical areas may also be implicated in the modulation of fatigue levels, but our analysis focuses only on the most predictive combinations of cortical area and frequency, potentially ignoring regions or frequencies in which less significant changes occurred.

While no existing studies have investigated EEG-based frustration detection, some studies have elucidated the roles of the prefrontal cortex and the parietal lobe in frustration [1, 55, 194], lending credence to our observations. Most existing studies regarding the neural basis of attention have focused on sustained vigilance tasks, which may not generalize perfectly to the protocol used here. However, one recent study found that failure in a vigilance task may be predicted by an increase in alpha activity over parieto-occipital regions [141]. This connection was also observed during this study, as both the participant-dependent and participant-independent analyses found that parieto-occipital alpha activity was a strong predictor of fluctuations in attention level. Moreover, the importance of fronto-parietal activity for attention classification is supported by the presence of the dorsal attention network within these regions of the cortex [232].

4.5.4 Electrode Removal

Classification results following the removal of some electrodes from the feature set (see Fig. 4.6) are encouraging for the design of practical passive BCIs. For both fatigue and frustration detection, reduced subsets of electrodes (FC and FP for fatigue; FC, FP, and CP for frustration) resulted in classification accuracies statistically equivalent to those attained using the full FCP set. Even for attention detection, classification accuracies exceeding 80% were attained when using the FP and CP subsets. This is highly encouraging. The ability to reliably detect fluctuations in mental state with just a few EEG electrodes may allow mental state monitoring to be more easily integrated with practical systems. It may also be worthwhile to investigate data-driven methods of constructing electrode subsets, rather than simply using electrodes from one cortical region. This may allow interactions between different cortical areas to be used for detection of changes in mental state.

It is interesting to compare the electrode subset classification accuracies from Fig. 4.6 to the feature selection results in Figs. 4.5 and 4.8. For instance, fatigue detection was typically most accurate for subsets that included the frontal electrodes, as both the participant-dependent and participant-independent feature
selection analyses would suggest. A similar relationship is apparent between attention detection and the posterior electrodes. Only for frustration detection is such a relationship not present, as the FC, FP, and CP subsets provided extremely similar classification accuracies. This may explain the inconsistency between the participant-dependent feature selection analysis (which implicated the posterior electrodes) and the participant-independent analysis (which implicated the frontal electrodes) for frustration detection. It is possible that fluctuations in frustration manifest over a wider cortical region than fluctuations in fatigue and attention.

Finally, it is worthwhile to note that the central electrodes do not appear to unduly influence classification for any mental state. This is an important observation, as the motor movements required to answer the mental arithmetic, anagram solution, and grid-recall tasks would have produced event-related synchronization and desynchronization in these electrodes [179]. Since accurate performance was maintained even when the central electrodes were removed from the analysis, it is clear that actual fluctuations in mental state were detected, rather than solely the presence of motor activation.

4.5.5 Generalizability of Feature Selection

The usage of participant-independent feature selection had a significant effect on classification accuracy. This was particularly obvious for low feature set dimensionalities, where the difference in classification accuracy between participant-dependent and participant-independent feature selection approached 10% for all three mental states. This suggests that there is significant inter-participant variability in the cortical regions and frequency ranges that are most predictive of changes in mental state. However, the superiority of participant-dependent feature selection became less apparent as feature set dimensionality was increased. When 20 features were used, there were no significant differences between the participant-dependent and participant-independent algorithms. It is possible that there are a few cortical regions and frequencies that are typically useful for classification of each mental state, and increasing the dimensionality of the feature set makes it easier for participant-dependent classifier training to identify the most predictive features within this set for each individual.

Based on this hypothesis, the varying speeds with which the participant-independent algorithm converged with the participant-dependent algorithm for each mental state may represent the amount of variance between participants. Frustration and attention converge very quickly, implying that there are a few specific features
that encapsulate information regarding these mental states for most individuals. Fatigue, on the other hand, converges more slowly, suggesting that there was significant variability between participants in how changes in fatigue levels manifested. This variability may be partially due to the various ages of study participants, as the neural correlates of fatigue and drowsiness are affected by age [118].

### 4.5.6 Comparison to Previous Literature

The classification accuracies attained for fatigue detection are lower than those reported in previous literature. However, this may have occurred because the experimental protocol used within this study inherently made detection more difficult. For example, Liu et al. classified fatigue with 84% accuracy but ignored fatigue levels other than unfatigued and exhausted [134], Shen et al. classified five fatigue levels with 91% accuracy during a 25-hour sleep deprivation study [206], and Trejo et al. attained 97% accuracy but again considered only alert and exhausted conditions [225]. In contrast, we considered all fatigue levels rather than solely extreme cases, and the short duration of our study limited the extent to which fatigue could be induced. These characteristics of the experimental protocol made the classification of fatigue much more challenging, leading to reduced classification accuracies. Moreover, using a diverse set of cognitive tasks and including rest periods may also have limited the effects of fatigue. However, these shortcomings were deliberate, as this study was designed with the intent of replicating BCI studies as closely as possible. The results from this study are more practically applicable to future passive BCI development than those from studies with prolonged exhaustion protocols.

For attention detection, our reported classification accuracies exceed most of those previously reported (e.g. 57% in [131], 76% in [135]). Our results are nearly on par with the 89% reported in [87] even though that study differentiated attentive and non-attentive tasks in general rather than fluctuations in attention level across the same set of tasks. Our results provide a foundation for further investigation of multi-level attention detection. Likewise, our single-trial EEG-based frustration detection system has no antecedent of which we are aware, and provides a basis for future passive BCI research.
4.6 Conclusions

This study investigated the ability to detect, on a single-trial basis, fluctuations between low and high values of fatigue, frustration, and attention during the performance of challenging mental tasks. The maximum classification accuracies for these three states were $76.8 \pm 8.5\%$, $71.9 \pm 5.9\%$, and $86.1 \pm 7.1\%$ accuracy, respectively. Our findings suggest that real-time monitoring of these mental states may be possible. This may allow the development of hybrid brain-computer interfaces that are capable of detecting functional brain activity as well as neuroelectrical manifestations of the user’s psychological disposition [177]. Future work should investigate the potential benefits of this hybrid approach for practical BCI applications.
Chapter 5

Effects of User Mental State on EEG-BCI Performance

The majority of this chapter has been reproduced from the following journal article: Myrden, A., and Chau, T. Effects of user mental state on EEG-BCI performance. *Frontiers in Human Neuroscience*, 9 (308), 2015. Since Frontiers in Human Neuroscience is an open-access journal, no permission was required from the published to reproduce the article here. This article can be found on the publisher’s website at [http://journal.frontiersin.org/article/10.3389/fnhum.2015.00308/abstract](http://journal.frontiersin.org/article/10.3389/fnhum.2015.00308/abstract).

5.1 Abstract

Changes in psychological state have been proposed as a cause of variation in brain-computer interface performance, but little formal analysis has been conducted to support this hypothesis. In this study, we investigated the effects of three mental states - fatigue, frustration, and attention - on BCI performance. Twelve able-bodied participants were trained to use a two-class EEG-BCI based on the performance of user-specific mental tasks. Following training, participants completed three testing sessions, during which they used the BCI to play a simple maze navigation game while periodically reporting their perceived levels of fatigue, frustration, and attention. Statistical analysis indicated that there is a significant relationship
between frustration and BCI performance while the relationship between fatigue and BCI performance approached significance. BCI performance was 7% lower than average when self-reported fatigue was low and 7% higher than average when self-reported frustration was moderate. A multivariate analysis of mental state revealed the presence of contiguous regions in mental state space where BCI performance was more accurate than average, suggesting the importance of moderate fatigue for achieving effortless focus on BCI control, frustration as a potential motivating factor, and attention as a compensatory mechanism to increasing frustration. Finally, a visual analysis showed the sensitivity of underlying class distributions to changes in mental state. Collectively, these results indicate that mental state is closely related to BCI performance, encouraging future development of psychologically adaptive BCIs.

5.2 Introduction

Brain-computer interfaces allow information to be conveyed to an external device, such as a computer, using cognitive activity alone [140]. Originally envisioned simply as a means of communication and environmental control for individuals with disabilities [240], more and more prospective applications of BCIs have been proposed in recent years for both healthy and disabled individuals. BCIs have been harnessed for recreational purposes in gaming and virtual reality applications, where they provide an alternative input modality by which a simulation can be controlled [32, 122]. BCIs have also been used to enable creative expression by translating cognitive activity into music and visual art [151], to track changes in cognitive states such as alertness [247], as a neurofeedback tool to achieve altered states of consciousness via meditation [47], and in neurorehabilitation for individuals who have lost motor control due to stroke [9].

Most current BCIs use electroencephalography (EEG) as a tool to infer mental state and communicative intent [139]. EEG provides a low-resolution spatial map of electrical activity on the cortex [166]. Despite this low resolution, it has generally been favored for BCI applications due to its relatively simple setup and low cost. Some recent research has also investigated the usage of hemodynamic imaging technologies such as near-infrared spectroscopy [209] and transcranial Doppler ultrasound [163], but these technologies cannot currently match the information transfer rate of EEG-BCIs. However, these hemodynamic imaging technologies may be useful in combination with EEG. Recent work on hybrid EEG-NIRS BCIs has shown that simultaneous measurement of electrical and hemodynamic activity on the cerebral cortex may allow for more accurate BCI operation by combining features from both modalities [68, 121]. More complex
arrangements are also possible - Liu et al. have shown that attention measured based on NIRS may improve the reliability of an EEG-BCI [136], while Koo et al. showed that NIRS can be used to detect whether motor imagery has been performed while EEG is used to differentiate different types of motor imagery, allowing the development of a self-paced BCI [115].

EEG-BCIs that are used for communication and control typically employ one of two paradigms. The first depends upon involuntary neuronal reactions to presented stimuli, and has been described as a “reactive BCI” [247]. This includes BCIs that detect the P300 response to anticipated visual, auditory, or tactile stimuli [99] and steady-state visually evoked potential (SSVEP) BCIs that detect the flicker frequency of the stimulus on which the user is fixated [44]. Both of these types of BCIs allow the user to choose one option from a grid of stimuli and are most commonly used as the basis for a spelling system. The second paradigm depends upon the detection of a voluntary cognitive activation, typically produced by performing a specific mental task. This has been described as an “active BCI” [247]. Active BCIs differentiate two or more mental tasks from each other, allowing the user to use each task to communicate a different message. Differentiating more than two tasks from each other typically incurs a decrease in classification accuracy, and it is rare for more than four mental tasks to be used [62, 204]. Mental tasks used for active BCIs in previous research have included a rest state, motor imagery, mental arithmetic, and a verbal fluency task, among others [49, 163, 181].

One pervasive challenge in BCI research is the tendency for BCI accuracy to decrease over time due to the non-stationarity of the signals used [207]. It is well-known that class distributions tend to change over time, and maintaining high BCI performance during long sessions and across weeks and months of usage is typically difficult [207]. This inconsistent performance is a significant impediment to the adoption of BCIs as access modalities for individuals with disabilities and may also be a significant risk factor for the abandonment of BCIs by these individuals [183]. It has been proposed that one cause of this inconsistent performance may be fluctuations in psychological variables such as alertness and distraction [49, 148]. Systems that track this type of involuntary ongoing cognitive user state can be categorized as passive BCIs [247]. Examples include estimation of task engagement and attention [11, 17, 91, 93, 155], mental workload [17, 97, 114], fatigue [206], and emotional state [32, 208]. These passive BCIs each use either EEG, NIRS, or fMRI, allowing them to be integrated with an active or reactive BCI that uses the same modality (or a complementary modality in the case of a hybrid BCI). Such a combination may allow adaptation to fluctuations in mental state, mitigating the observed variation in BCI performance over time. However, it is first imperative to verify that these
fluctuations in mental state are related to variation in BCI performance, as to the best of our knowledge, this hypothesis has not been formally tested.

This paper investigates the effects of user mental state on BCI performance. Three mental states of particular interest were identified based on previous work - cognitive fatigue, frustration, and attention [49]. Subjective self-reported estimations of these three mental states were gathered from BCI users while playing a simple maze navigation game. These ratings were compared to BCI performance to identify relationships between mental state and classification accuracy. A multivariate analysis was also performed to identify a region in mental state space for optimal BCI performance. Finally, the class distributions of the rest and active tasks were analyzed in the feature space to determine the effects of changes in mental state on the individual signal features used for classification.

5.3 Material & Methods

5.3.1 Population

Twelve able-bodied participants (two male, average age 27.7 years) were drawn from graduate students and staff at Holland Bloorview Kids Rehabilitation Hospital. Participants had normal or corrected-to-normal vision and refrained from consuming caffeine for four hours prior to each session. Participants provided written informed consent, and the experimental protocol was approved by the Holland Bloorview Research Ethics Board.

5.3.2 Instrumentation

During each session, electrical signals from the cortex were recorded using a B-Alert X24 wireless EEG headset (Advanced Brain Monitoring, Carlsbad, CA, USA). Signals were recorded from the Fz, F1, F2, F3, F4, Cz, C1, C2, C3, C4, CPz, Pz, P1, P2, and POz locations according to the international 10-20 system [102]. Signals were band-pass filtered between 2 and 30 Hz, and artefacts resulting from eye movements were removed using independent component analysis [152].
5.3.3 Training Sessions

Participants completed two training sessions on separate days. The goal of these sessions was to identify a mental task that could reliably be differentiated from rest for each participant. Four candidate mental tasks were considered - mental arithmetic, motor imagery, music imagery, and word generation. Multiple tasks were considered because the most effective BCI task typically varies across participants [83]. The mental arithmetic task required participants to perform a repeated addition or subtraction task in their head. The motor imagery task required participants to imagine performing finger-to-thumb opposition with the hand of their choice. The music imagery task required participants to sing a song of their choice in their heads. The word generation task required participants to silently think of as many words as possible that began with a given letter.

During each training session, participants completed 150 trials. These were evenly divided between the four candidate mental tasks and a rest task, during which participants were instructed to relax and let their minds wander. Each trial was 15 seconds long, consisting of a five-second preparation period during which a visual cue was displayed to indicate the required task for the trial; a five-second task period during which the participant performed the required task; and a five-second cool-down period before the next trial began. Participants were instructed to remain silent and still during the preparation and task periods to minimize motor artefacts. The visual cue for each trial was displayed on a computer monitor using a custom LabVIEW interface. To avoid mislabeling training data, participants were prompted to report whether they had successfully completed each trial at the end of the cool-down period. This was done through an on-screen dialog box that had to be completed before the next trial could begin. At the end of each training session, participants ranked the four candidate mental tasks in order of preference for future BCI usage.

5.3.4 BCI Development

Following the completion of the training sessions, a BCI was trained to differentiate each candidate mental task from the rest task. There were a total of 60 trials for each task. For each signal from each electrode, the spectral power within the signal in 1-Hz increments (from 0-1 Hz to 29-30 Hz) was estimated by summing the squares of the corresponding Fourier coefficients. These local spectral power estimates yielded 450 individual features (30 frequencies from 15 electrodes) for each trial.
Two feature selection methods were used to reduce the dimensionality of the feature set to between 1 and 12 features for classification. In the first method, a fast correlation-based filter (FCBF) directly reduced the dimensionality from 450 to the target number of features. This resulted in most of the feature set being discarded. In the second, this 450-dimensional feature set was reduced by clustering highly correlated features and performing principal component analysis (PCA) on each cluster to compute 75 intermediate features before using a FCBF to arrive at the target number of features. The latter approach was included to accommodate tasks that elicited widely distributed cortical activation at varying frequencies.

For each feature space dimensionality, a linear discriminant analysis (LDA) classifier [25] was trained for each candidate task and feature selection method. Ten runs of ten-fold cross-validation were performed and the average classification accuracy across the folds was computed. This resulted in a set of 24 different classifiers for each task. The classifier that yielded the highest classification accuracy was identified for each task and the tasks were then ranked by their maximum accuracy. Participants were presented with these accuracy-based task rankings along with their own rankings of task preference. Based on this information, they were allowed to choose which active task they wanted to use for the remainder of the study.

5.3.5 Testing Sessions

Participants completed three testing sessions on separate days. During these sessions, they used a BCI based on the task that they selected at the end of the training sessions to play a simple maze navigation game that was programmed in LabVIEW. Participants attempted to complete a series of 10 mazes. Each session started with the first (and simplest) maze. Mazes grew more difficult as the session progressed, but this was primarily due to the number of intersections between the origin and destination rather than the cognitive difficulty of plotting a path through the maze.

Participants navigated through the maze by moving from intersection to intersection. Their current position was indicated by an image of a person, while the destination was indicated by an image of a door. At each intersection, there were between two and four potential directions (labeled as north, south, east, and west) in which movement to another intersection was possible. Participants were prompted to select the direction in which they wanted to travel from an on-screen window. Subsequently, the potential navigational directions were highlighted one at a time for five seconds each, constituting four task periods. Each task period was punctuated with a five-second break. When a direction was highlighted, appropriate task cues were shown,
namely the cue for the active task for the selected direction of travel and the cue for the rest task for all other directions. The selected direction of travel was recorded only to label data for future analysis and to ensure that appropriate task cues were presented during each task period. An example of the game, depicting an initial intersection, a BCI decision, and the subsequent intersection, is shown in Figure 5.1.

Figure 5.1: Maze navigation. The position of the participant is represented by the image of a person and the destination by the image of a door. The participant chose to move east at Intersection 1, so they are shown a cue for the word generation task when the arrow pointing east is highlighted. A cue for the rest task was shown when the north and west arrows were highlighted. The BCI analyzed the task period from each direction, predicting that the north and west task periods represented rest tasks (green bars) while the east task represented the word generation task (red bar). Consequently, the BCI moved the participant to Intersection 2, ignoring the intermediate intersection where the only options would have been to continue moving in the same direction or to go back to the last intersection. At the new intersection, the participant chose to move west to continue approaching the exit, so a new word generation cue was shown when that direction was highlighted.

The EEG recording from each five-second task period was classified in real-time by the BCI. When the task period had been completed for each potential direction, the BCI decided which task period was most likely to represent the active task rather than the rest task, and the image on screen was moved in the corresponding direction. However, before this movement was displayed, the participant was prompted to self-report their perceived levels of fatigue, frustration, and attention. Each of these mental states was rated by moving a slider on a continuous scale from 0 to 1 with textual anchors at either end (i.e. “Least fatigued” and “Most...
fatigued”). A new maze was automatically loaded when the current maze was completed (i.e. when the participant navigated to the door) and the session was terminated either after completion of the tenth maze or once 50 minutes had elapsed.

5.4 Results

5.4.1 Choice of BCI Task

Offline performance for each participant during the training sessions is summarized in Table 5.1. Word generation was selected as the optimal BCI task by nine participants, motor imagery by two participants, and music imagery by the final participant. The average classification accuracy of the selected task was 72.4%. This exceeded the minimal BCI performance criteria of 70% despite the short task duration and relatively small selection of electrodes. Since the analysis focused on the effects of mental state on BCI performance, it was not necessary to obtain extremely high accuracy. In fact, high accuracy may have inhibited the analysis, as ideal performance may have limited the observed fluctuations in each mental state (e.g. high levels of frustration would be unlikely).

Table 5.1: Classification accuracies for each participant during the training sessions for the mental arithmetic (MA), motor imagery (MI), music imagery (MuI), and word generation (WG) tasks. Word generation was selected by nine participants, motor imagery by two participants, and music imagery by the remaining participant. Entries labeled as ‘N/A’ indicate occasions when a mental task was removed from the second training session for a participant due to both poor classification accuracy during the first training session and placement as the least preferred task for that participant following the first training session.

<table>
<thead>
<tr>
<th>Participant</th>
<th>MA%</th>
<th>MI%</th>
<th>MuI%</th>
<th>WG%</th>
<th>Selected Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>87</td>
<td>95</td>
<td>67</td>
<td>74</td>
<td>MI</td>
</tr>
<tr>
<td>2</td>
<td>63</td>
<td>60</td>
<td>63</td>
<td>70</td>
<td>WG</td>
</tr>
<tr>
<td>3</td>
<td>58</td>
<td>75</td>
<td>68</td>
<td>50</td>
<td>MI</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>64</td>
<td>66</td>
<td>70</td>
<td>WG</td>
</tr>
<tr>
<td>5</td>
<td>80</td>
<td>77</td>
<td>62</td>
<td>82</td>
<td>WG</td>
</tr>
<tr>
<td>6</td>
<td>62</td>
<td>55</td>
<td>57</td>
<td>60</td>
<td>WG</td>
</tr>
<tr>
<td>7</td>
<td>N/A</td>
<td>69</td>
<td>61</td>
<td>69</td>
<td>WG</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>71</td>
<td>65</td>
<td>71</td>
<td>WG</td>
</tr>
<tr>
<td>9</td>
<td>56</td>
<td>61</td>
<td>59</td>
<td>66</td>
<td>WG</td>
</tr>
<tr>
<td>10</td>
<td>82</td>
<td>81</td>
<td>73</td>
<td>83</td>
<td>WG</td>
</tr>
<tr>
<td>11</td>
<td>69</td>
<td>N/A</td>
<td>71</td>
<td>64</td>
<td>MuI</td>
</tr>
<tr>
<td>12</td>
<td>66</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>WG</td>
</tr>
</tbody>
</table>
5.4.2 Online BCI Performance

Participant 7 was excluded from this analysis as he or she was not able to control the BCI during the testing sessions, and Participant 8 was unable to attend the testing sessions. One testing session for each of Participants 2 and 3 could not be analyzed due to signal quality issues. Three testing sessions were analyzed for all other participants. Although retraining the BCI after each testing session for all participants would have resulted in higher accuracies, it was avoided to minimize the number of factors affecting classification accuracy. Consequently, BCIs were retrained after testing sessions only when the experimenters felt it absolutely necessary in order to maintain motivation for participants. This occurred only for the final testing sessions for Participants 11 and 12. All other participants used the same BCI for all testing sessions.

Two metrics were considered: the balanced individual classification accuracy, referring simply to the average of sensitivity and specificity for the individual tasks; and the collective classification accuracy, referring to the proportion of maze intersections at which the BCI correctly identified the intended direction of transit. There were typically three or four potential directions of transit, so the collective accuracy was expected to be lower than the individual accuracy.

Figures 5.2 and 5.3 depict the individual and collective accuracies, respectively, for each participant during each online session. Despite the non-adaptive classifier, four of ten participants exceeded the 70% threshold for individual classification accuracy during one or more testing sessions. Furthermore, five of ten participants achieved a collective classification accuracy that exceeded this threshold during one or more sessions.

![Figure 5.2: Balanced individual classification accuracies for all participants during each session.](image)

Figure 5.2: Balanced individual classification accuracies for all participants during each session.
Both individual and collective classification accuracy increased in the second and third sessions, suggesting that participants became more proficient with the BCI over time. For the third session, both individual and collective classification accuracy neared the 70% threshold when averaged across all participants.

### 5.4.3 Effects of Fatigue, Frustration, and Attention on BCI Performance

For each participant, self-reported ratings for each mental state during each session were quantized to two levels, defined as, for example, ‘low fatigue’ and ‘high fatigue’. The cut point for quantization was varied for each participant and mental state to ensure that each level contained as close to the same amount of trials as possible. Each individual trial was categorized as either low or high for each mental state and the classification accuracy at each level across all participants was computed. By ensuring that each session was equally represented in the low and high categories for each mental state, this approach controlled for learning effects. These results are presented in Table 5.2 for the individual classification accuracy.

<table>
<thead>
<tr>
<th>Level</th>
<th>Fatigue</th>
<th>Frustration</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>62.6</td>
<td>61.7</td>
<td>63.8</td>
</tr>
<tr>
<td>High</td>
<td>65.0</td>
<td>65.7</td>
<td>64.2</td>
</tr>
</tbody>
</table>
Classification accuracy at low fatigue was about 2.5% inferior when compared to high fatigue, a result that approached significance using the chi-squared test (p = 0.088). The 4% difference in classification accuracy between low and high frustration was statistically significant (p = 0.0038). There was no significant difference between classification accuracy at low attention and high attention levels. However, these findings collectively indicate that there is a significant relationship between BCI performance and mental state.

To investigate whether the choice of mental task influenced the relationship between mental state and BCI performance, participants were split into two categories - those who chose word generation as the active task and those who did not - and the preceding analysis was repeated. The results for each group are depicted in Table 5.3. While frustration appeared to affect the two groups similarly, fatigue had more impact on the WG group and attention on the not-WG group. However, due to the small sample sizes incurred by splitting the group in two, further research with control groups of equal size for each task would be necessary to draw significant conclusions.

Table 5.3: Classification accuracies at low and high levels for each mental state. Classification accuracies were based on 4814 trials across ten participants, and each level contained roughly the same number of trials.

<table>
<thead>
<tr>
<th>Task</th>
<th>Level</th>
<th>Fatigue</th>
<th>Frustration</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>WG</td>
<td>Low</td>
<td>61.5</td>
<td>60.7</td>
<td>62.3</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>64.0</td>
<td>64.4</td>
<td>63.3</td>
</tr>
<tr>
<td>Not WG</td>
<td>Low</td>
<td>66.7</td>
<td>64.3</td>
<td>68.8</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>67.2</td>
<td>68.8</td>
<td>66.0</td>
</tr>
</tbody>
</table>

Since the two-level quantization of each rating was a simplistic means of investigating these effects, further analysis was conducted using normalized values for each mental state. For each session, the self-reported ratings for each mental state were normalized to zero mean and unit variance. For each mental state, all trials across all participants were then sorted by their normalized rating. Since each trial was also associated with a classification result (i.e. either a correct or incorrect decision by the BCI), this allowed the construction of a binary sequence representing BCI performance over the full range of normalized ratings. This sequence was smoothed to minimize the noise produced by the usage of individual classification results, resulting in a classification accuracy curve $C_{\text{actual}}$ for each state.

Since ratings from each participant were not uniformly distributed within the range of ratings for each state, $C_{\text{actual}}$ was biased due to individual variations in classification accuracy. To mitigate this, an expected classification accuracy curve $C_{\text{expected}}$ was constructed by replacing the actual classification result from each trial with the average classification accuracy from the session within which each trial originated. The same
smoothing process was performed, and the effects of mental state on BCI performance were assessed based on the difference between $C_{\text{actual}}$ and $C_{\text{expected}}$, as depicted in Figure 5.4. Through random sampling, the difference between actual and expected classification accuracy was observed to follow a normal distribution with a 90% confidence interval of $0 \pm 0.032$. The bounds of this confidence interval are depicted in Figure 5.4.

These figures support the results from Table 5.2 while also providing a higher-resolution view of the trends in classification accuracy. Performance was poor for low levels of self-reported fatigue, with the difference between actual and expected accuracy surpassing -7%, well outside of the 90% confidence interval. For frustration, higher classification accuracies than expected were exhibited for moderate values, peaking at approximately +7%. In contrast, the difference between actual and expected classification accuracy over the full range of attention ratings was small, remaining almost entirely within the 90% confidence interval.
Figure 5.4: Difference between actual and expected classification accuracy across all participants for normalized values of self-reported fatigue, frustration, and attention ratings. The distributions of fatigue and frustration ratings were positively skewed while the distribution of attention ratings was negatively skewed, causing the apparent variation in range between states. Dotted lines represent the 90% confidence interval for the mean of the deviation between actual and expected classification accuracy.
5.4.4 Multivariate Analysis

The previous analysis considered only the effects of each individual mental state. A multivariate analysis of the effects of mental state on BCI performance was also conducted by analyzing the effects of each combination of two states. The raw self-reported ratings were extracted for each participant without quantization. A grid was constructed across the full range from 0 to 1 for each mental state with a resolution of 0.01. At each point within the grid, the nearest 500 trials, regardless of participant, were identified. For this set of 500 trials, the actual classification accuracy $C_{\text{actual}}$ was computed. In addition, the expected classification accuracy was computed as:

$$C_{\text{expected}} = \sum p w_p C_{p,s}$$

(5.1)

where $w_p$ represents the proportion of the nearest 500 trials which originated from Participant $p$ and $C_{p,s}$ represents the overall classification accuracy of all trials originating from session $s$ for Participant $p$. The difference between actual and expected classification accuracies was used to characterize this point. The results of this process are depicted in Figures 5.5 through 5.7 for fatigue and frustration, fatigue and attention, and fatigue and attention, respectively. Again, the difference between actual and expected classification accuracy was compared to the 90% confidence interval, established previously as $0 \pm 0.032$.

Figure 5.5: Two-dimensional view of the difference between actual and expected classification accuracies as a function of fatigue and frustration. The middle graph depicts the variation in classification accuracy as shown on the legend on the left, and the right graph shows only regions for which the difference exceeded 3.2% (in green) or was less than -3.2% (in red).
The fatigue-frustration and fatigue-attention graphs reveal relatively contiguous optimal regions for BCI control. For fatigue-frustration, two optimal regions are apparent - one from moderate to high fatigue and low to moderate frustration and one from low to moderate fatigue and moderate to high frustration. Of these, the former is larger and more consistent across a wide region in mental state space. For fatigue-attention, there is an optimal region for moderate to high fatigue and attention. Again, these findings corroborate the univariate analyses despite the usage of raw ratings rather than quantized or normalized ratings.

Figure 5.6: Two-dimensional view of the difference between actual and expected classification accuracies as a function of fatigue and attention. The middle graph depicts the variation in classification accuracy as shown on the legend on the left, and the right graph shows only regions for which the difference exceeded 3.2% (in green) or was less than -3.2% (in red).

The graph for frustration-attention is more equivocal. The area within mental state space that exceeded the bounds of the 90% confidence interval was small and located in the high frustration and high attention region. This may imply that high attention is necessary to compensate for high frustration. However, given the small size of this region, this could also potentially be a result of random variation.
5.4.5 Effects of Mental State on Class Distributions

The effects of mental state on BCI performance were also analyzed from a signal feature perspective. As in the univariate analyses, all trials for each participant were split into low and high categories for each mental state. In this case, the LDA classifier trained for each participant after the training sessions was used to project each trial to one dimension. The separability of the rest and active tasks were then estimated by computing the Fisher score for each projection. The Fisher score is defined as [73]:

$$ J = \frac{|\mu_1 - \mu_2|^2}{s_1^2 + s_2^2} $$

(5.2)

where $\mu_1$ and $s_1$ represent the mean and the variance of the projected values for the rest class and $\mu_2$ and $s_2$ represent the mean and the variance of the projected values for the active class. The results of this analysis are depicted in Figure 5.8.
These results suggest that the classifiers trained based on the training sessions were much less effective during the testing sessions, accentuating the importance of frequent retraining. However, the average differences between Fisher scores for low and high ratings for each mental state also suggest that, for some participants, mental state may affect class distributions in feature space.

To verify this, the class distributions for the rest and active tasks under different mental state conditions were inspected based on two of the features used for classification. Figure 5.9 shows the class distributions of each task for Participant 4 under low and high attention conditions. The center of each ellipse represents the class mean under that condition while the size of the ellipse represents the 67% confidence interval for the class, oriented along the eigenvectors of the covariance matrix. Even in this low-dimensional space (the classifier for this participant used 10 features), it is evident that modulations in mental state affect class distributions, as the two classes are nearly separable when attention is high but inseparable when attention is low.
Figure 5.9: Class distributions for the rest and active task for Participant 4 during the testing sessions. Each ellipse represents the distribution of one class (the rest task for solid lines and the active task for dashed lines) under one categorization of attention levels (low in blue lines, high in red lines). While classes were nearly separable when attention was high, they were unseparable when attention was low.

5.5 Discussion

5.5.1 Optimal Mental State for BCI Control

It is clear from our univariate analyses that mental state and BCI performance are closely intertwined. However, some of our observations were surprising. Online BCI performance was significantly less accurate during the trials for which participants reported the lowest fatigue levels and significantly more accurate during the trials for which participants reported high frustration levels. This was observed when each trial was categorized by quantized ratings for each state and also when ratings were normalized within each session. Based on the same analyses, attention seemed to have little impact on BCI performance. Analysis based on the choice of active task suggests that there may be task-related effects, but further investigation would be necessary for statistical verification.

Self-reported mental state ratings were quantized and normalized for these initial analyses in an attempt to account for the fact that different participants may have anchored their ratings differently on the continuous
scales used for each state. However, one shortcoming of this approach was that it ignored differences in average mental state. Since one participant reporting a higher average fatigue level than another could be either a result of variation in anchoring or a legitimate difference in fatigue levels, the raw ratings were used for the multivariate analysis in order to compare the results.

This multivariate analysis suggested the presence of optimal mental state regions for BCI control. The most interesting observation came from the fatigue-attention analysis, which showed that the highest accuracies occurred when moderate values were maintained for fatigue and moderate to high values for attention. BCI performance decreased markedly when these states varied, particularly for low fatigue and high attention. The multivariate analysis also presented a more nuanced portrait of the effects of each mental state on BCI performance. The fatigue-frustration and frustration-attention cases both showed interactions between pairs of states. In general, the former analysis showed that optimal performance occurred for either moderate fatigue and high frustration or high frustration and low fatigue. In contrast, the low fatigue and low frustration case exhibited notably poor performance.

Although there is no pre-existing BCI literature for comparison, some support for these results can be found in other disciplines. The state of psychological flow has been identified as a requirement for excellent performance in many fields [53, 54, 104]. Flow is characterized by what Romero describes as effortless attention, a state of deep concentration where perceived effort is generally lower than would be expected [196]. This is contrasted with effortful attention, in which the perceived effort to achieve focus is quite high and individuals must fight to maintain deep concentration. We hypothesize that, due to the high perceived effort, effortful attention is likely to be characterized by higher self-reported fatigue and potentially higher self-reported attention than effortless concentration. Since optimal performance can be expected during effortless attention, this could produce the pattern seen in Figure 5.6.

The role of frustration has also attracted attention in previous research. In learning studies, it has been observed that frustration, in moderation, is not necessarily a negative factor [12]. In fact, the presence of frustration during a difficult task may simply represent motivation, which is a factor likely to improve performance. However, studies have also observed that high frustration induces boredom, reducing attention and leading to poor performance [58]. This implies that optimal performance may be associated with moderate frustration, corroborating our findings, particularly those depicted in Figure 5.4.

There is one important caveat regarding this study. It has been shown that fluctuations in mental state are
related to fluctuations in BCI performance. However, it stands to reason that since these fluctuations affect the underlying class distributions of the active and rest tasks, as seen in Figure 5.9, the classifiers used for each BCI were dependent upon the mental state experienced by each participant during the training sessions. Consequently, it may be that the optimal mental state for BCI control is simply that which most closely approximates the mental state from the training sessions. However, given the length of each training session, the commensurate unlikelihood that mental state was consistent throughout, and the unusual topography of the optimal mental state regions in Figures 5.5 through 5.7, it is more likely that the results were affected by both the mental state during the training sessions and the inherent superiority of certain psychological conditions for BCI control. Regardless of which factor is most responsible for the relationship between mental state and BCI performance, these results strongly suggest that such a relationship does exist. This motivates future investigation of psychologically adaptive BCIs.

5.5.2 Towards Psychologically Adaptive BCIs

It has been proposed that there are two ways in which a computer system can adapt to information regarding the cognitive state of a user. These are overt adaptation, in which the adaptation is apparent to the user, and covert adaptation, in which it is not apparent to the user [65]. These definitions can also be applied to the design of psychologically adaptive BCIs.

Overt adaptation, although potentially more effective than covert adaptation for modifying user state, also has a potentially higher cost [65]. For a BCI, overt adaptation would require an attempt to modify user mental state to bring it closer to the optimal region, likely taking the form of an adaptive user interface [220]. Such an interface could use targeted stimuli or helpful feedback to mitigate undesirable changes in mental state [65, 220]. The interface could also take more drastic steps, potentially going so far as to automatically deactivate the BCI when extremely low attention is detected, reactivating only when the user’s attention has returned. The interface could even modify the timing variables of the BCI, extending task durations when it is likely that classification will be inaccurate, a psychologically-driven approach with some similarities to the evidence accumulation algorithms that are often used for online classification. The danger of overt adaptation lies in the potential for false alarms [65]. Explicit interventions that are not required may actually further inhibit BCI control by inducing additional frustration or distraction. It may be wise to use overt adaptation sparingly [65].
Covert adaptation, on the other hand, could involve modifications to the classifier itself. There are several potential methods by which this could be implemented. First, we observed in Figures 5.2 and 5.3 that there was little difference between individual and collective classification accuracy even though the individual accuracy was based on a binary decision and the collective accuracy on a decision that typically involved three or four options. This implies, for the LDA classifiers that were used, that there was more difficulty locating an appropriate value for the bias parameter than for the weight vector. Thus, an adaptive bias parameter based on mental state may allow for covert adaptation without repeated classifier retraining.

Second, given the effects of mental state on class distribution observed in Figure 5.9, it is possible that selective online resampling of the training set and retraining of a simple classifier (such as LDA) could be implemented. This would require online estimation of user mental state, the selection of the training points most closely matching this current mental state, and the training of a classifier based on this subset of the training data. Since these adaptations would go unnoticed by the BCI user, they could be employed as frequently as necessary [65].

There are two significant limitations for any psychologically adaptive BCI. First, it is obviously necessary to achieve accurate detection of these changes in mental state. Our group is currently working on achieving reliable differentiation between low and high values of the three mental states in question. Second, any psychological adaptation that is implemented must be adaptive in itself. Significant differences were observed across participants in terms of the reactivity of BCI performance to changes in mental state, and it is unlikely that a “one size fits all” approach will be sufficient.

5.6 Conclusions

In this study, we investigated the effects of mental state on BCI performance. We observed that the relationships between these variables were complex, rather than monotonic. There appear to be optimal operating conditions where fatigue, frustration, and attention levels are most appropriate for effective control of an EEG-BCI. Moreover, signal features are affected by changes in mental state, potentially necessitating classifier adaptation. Future work should consider the development of BCIs that display both overt adaptation to keep user mental state within the optimal region and covert adaptation that automatically modifies the BCI classification algorithm to adapt to changes in mental state. This will allow the development of BCIs that are more robust to changes in mental state.
Disclosure/Conflict-of-Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Chapter 6

Practical Considerations for EEG Classification of a Word Generation Task

Chapter 5 detailed the online performance of an EEG-BCI that included user-selected mental tasks. We noted that the word generation task was selected much more frequently than the mental arithmetic or motor imagery tasks despite being much less common in BCI literature. This short addendum details some relevant observations for the usage of the word generation task within a BCI paradigm.

6.1 Background

Discussion of mental task-based EEG-BCIs typically begins and ends with the sensorimotor rhythms induced by motor imagery tasks such as finger-to-thumb opposition and grasping. However, these tasks are not without their drawbacks. Numerous studies have noted that motor imagery classification is not a fait accompli [83, 231]. In studies with able-bodied participants, a small but significant proportion of participants have been unable to generate reliable motor imagery activations [6, 29]. This shortcoming may be exacerbated
when EEG headsets with lower signal-to-noise ratios and sparse arrays of electrodes are used, particularly due to the localized nature of these rhythms within the cerebral cortex. Regardless of individual ability to produce motor imagery, previous BCI studies have noted the benefits of personalized mental tasks for BCI usage [236].

Application within the target population of individuals with disabilities is beset by further difficulties. It is an open question whether participants with long-established motor dysfunctions can be expected to imagine motor movements when they have never been able to physically perform these motor movements. Kubler and Birbaumer found that individuals presenting as locked-in were unable to reliably activate a motor imagery BCI, although it could not be ascertained whether this reflected an inability to produce motor imagery or a reduction in the capacity for goal-directed action [117]. It seems intuitive that, in such scenarios, it may be helpful to have potential alternatives to the motor imagery task.

One such alternative is the word generation task detailed in the preceding chapter, which requires the BCI user to imagine words that begin with a given letter. This task has been used within numerous metabolic BCI studies that employed TCD or NIRS [163, 236], but has only occasionally been used as the foundation for an EEG-BCI [133]. In our study, the majority of participants expressed a clear preference for word generation over motor imagery as a BCI control task. This motivates an investigation of some practical considerations for the classification of word generation - namely the identification of the cortical regions and frequency bands that provide the best discrimination between word generation and rest.

6.2 Methods

The data used for this analysis were collected during the study detailed by Chapter 5. Eight participants from this study chose to use word generation during the online testing sessions. For these participants, all rest and word generation trials completed during these online sessions were segmented from the recorded EEG. Each participant completed between 120 and 200 five-second trials of the word generation task, along with nearly twice as many five-second trials of the rest task.

The feature clustering algorithm detailed in Chapter 3 was used to extract features. A 10 x 10 (runs x folds) repeated cross-validation was performed. Prior to each run of the cross-validation, a random subset of rest tasks was drawn from each participant’s data to ensure that there were equal numbers of rest and
A FCBF was used to select five features for classification. Plots of feature selection frequency were generated using the methods detailed in Chapter 4 to identify the most frequently used electrodes and frequency ranges for the classification of the word generation task.

### 6.3 Results

The mean classification accuracy of word generation across all participants was $71.9 \pm 6.7\%$. Individual classification accuracies are depicted in comparison to chance levels for each participant in Figure 6.1. Note that chance levels varied across participants due to the differing number of trials. For each participant, the upper bounds of the 95% confidence interval for chance classification accuracy was computed using the adjusted Wald interval [19]. All participants significantly exceeded chance levels.

![Figure 6.1: Comparison of classification accuracy to chance levels for all participants.](image)

For each EEG signal feature (i.e. each combination of electrode and spectral frequency), the normalized frequency of feature selection was computed. Figure 6.2 depicts these results.

For each participant, the upper bounds of the 95% confidence interval for chance classification accuracy was computed using the adjusted Wald interval [19]. All participants significantly exceeded chance levels.
Figure 6.2: Most frequently selected features for the classification of the word generation task. Each oval represents an EEG signal feature, with the cortical location on the y-axis and the spectral frequency on the x-axis. The size of the oval represents the normalized frequency with which each feature was selected. The vertical and horizontal lines represent the frequency with which features from each electrode and spectral frequency were chosen for classification, respectively.

Some dominance of the left hemisphere was noted from these results. EEG activity within the delta band from F3 and P1 was very frequently used for classification. Moreover, features originating from C3 and C1 were very frequently selected for classification across a wide range of spectral frequencies, including the delta, theta, alpha, and beta bands. In contrast, C4 was the sole electrode from the right hemisphere to be frequently used for classification. Across the entire cortex, activity within the delta and alpha bands was most frequently used for classification.

The importance of each electrode for the classification of word generation was further investigated using the following approach. Every possible subset of five electrodes was used independently for classification, again using a 10 x 10 repeated cross-validation. Feature clustering was omitted due to the computational requirements of this analysis. This produced a population of 3003 (15 choose 5) classification accuracies, which formed a Gaussian distribution with a mean of 69.45% and a standard deviation of 0.43%. Classifiers for which the z-score exceeded 2 were isolated, and the normalized frequency with which each electrode was part of the subsets used by these classifiers was tabulated. The results are depicted in Figure 6.3. To estimate chance levels for these normalized frequencies, the same process was repeated after randomly shuffling class labels. The minimum and maximum normalized frequencies with which any electrode was part of the subsets used for the most accurate classifiers (again defined as those with classification accuracies more than two
standard deviations above the mean) were identified.

Figure 6.3: Electrodes that were most frequently part of the electrode subsets that attained the most accurate classification accuracies for word generation detection. Classification accuracy was calculated for every possible combination of five electrodes (left). The 53 electrode subsets that produced classification accuracies more than two standard deviations above the mean were isolated and the normalized frequency with which each electrode was a member of these subsets was computed (right). The horizontal lines on the right graph indicate the minimum and maximum frequencies with which each electrode was part of these optimal subsets when the class labels were randomly permuted prior to classification.

Three electrodes were selected more frequently than chance levels - F3, C2, and POz. F3 and POz were also among the most frequently selected electrodes in Figure 6.2, while C2 was actually one of the least frequently selected electrodes in that analysis.

6.4 Significance

Accurate classification of word generation was attained using both small subsets of five electrodes and the full set of 15 electrodes. The average five-electrode subset classification accuracy of 69.45% was only marginally lower than the 71.9% performance observed when all electrodes were used. This is particularly notable since the feature clustering algorithm, which was very effective for the classification of motor imagery in Chapter 3, was not used with the smaller electrode subsets.

Based on our results, we conclude that F3 and POz are integral cortical locations for the classification of word generation, as these electrodes were implicated by both analyses of feature selection frequency. Some other cortical regions that may play key roles in the classification of word generation include Fz, C3, C2, and P1, each of which was implicated by one of the two analyses. The results depicted by Figure 6.3 must
be taken in context, as variables other than sheer predictive value affect the formulation of an ideal feature subset, a topic discussed in Chapter 3. For instance, the extremely high frequency with which POz was part of the most accurate electrode subsets may have suppressed the selection of nearby electrodes such as P1 and Pz that likely exhibited significant redundancy. This may explain why the three most frequently used electrodes are well-distributed across the cortex, including one electrode from the frontal lobe, one from the central lobe, and one from the parieto-occipital region.

These results are supported by neuroscience research. Broca’s area, a part of the cerebral cortex linked to speech production, can be found in the left frontal lobe, close to the F3 electrode site in this study [13]. Wernicke’s area, a region of the cerebral cortex linked to language comprehension, can be found in the left temporo-parietal region [172], closer to C3 and P1, each of which was frequently selected for classification. The word generation task has been described as lateralized to the left hemisphere by both BCI and functional neuroscience research, supporting the dominance of features from the left hemisphere apparent in Figure 6.2 [163].

6.5 Conclusions

We differentiated the word generation task and a rest task with approximately 72% accuracy across eight participants, exceeding chance levels for all individuals and for the population as a whole. Classification appeared to rely heavily on the left frontal lobe and the left centro-parietal region of the cortex, matching the anatomic locations of Broca’s and Wernicke’s areas. Delta and alpha band activity appeared to be the strongest predictors of the word generation task. Our results indicate that the word generation task is a promising alternative for participants who struggle to produce reliable motor imagery activations.
Chapter 7

Towards Psychologically Adaptive
Brain-Computer Interfaces

The majority of this chapter has been reproduced from the following article, which is currently under review at the Journal of Neural Engineering: Myrden, A., and Chau, T. Towards psychologically adaptive brain-computer interfaces. Manuscript under review, 2016.

7.1 Abstract

Brain-computer interface (BCI) performance is sensitive to short-term changes in psychological states such as fatigue, frustration, and attention. This paper explores the design of a BCI that can adapt to these short-term changes. Eleven able-bodied individuals participated in a study during which they used a mental task-based EEG-BCI to play a simple maze navigation game while self-reporting their perceived levels of fatigue, frustration, and attention. In an offline analysis, a regression algorithm was trained to predict changes in these states, yielding Pearson correlation coefficients in excess of 0.45 between the self-reported and predicted states. Two means of fusing the resultant mental state predictions with mental task classification were investigated. First, single-trial mental state predictions were used to predict correct classification by the BCI during each trial. These predictions significantly exceeded chance levels. Second, an adaptive BCI was
designed that retrained a new classifier for each testing sample using only those training samples for which predicted mental state was similar to that predicted for the current testing sample. Although this resulted in a smaller training set, 5 of 11 participants exhibited significantly higher classification accuracy using this adaptive BCI while the remaining six participants exhibited no significant differences. Collectively, these findings indicate that adaptation to psychological state may allow for the design of more accurate BCIs.

7.2 Introduction

Brain-computer interfaces, or BCIs, provide a potential means of communication and environmental control for individuals with disabilities [240]. Most current BCIs employ electroencephalography (EEG) to record electrical activity on the cerebral cortex, providing a means by which changes in cognitive activity can be monitored [7]. EEG-BCIs can detect fluctuations in electrical activity that are characteristic of certain cognitive events. For example, P300 and SSVEP BCIs detect the modulation of attention [174, 205] while task-based BCIs detect the performance of mental tasks such as motor imagery [178]. When these cognitive events are reliably detected, BCI users can employ them to control spellers [243], mobility devices [123], and other systems [157].

Practical usage of BCIs has been hindered by their instability. EEG signals are non-stationary both between and within sessions [207]. The former type of non-stationarity can occur due to small variations in electrode impedance and positioning from day to day or changes in the experimental protocol between training and testing sessions, while the latter may be caused by changes in underlying mental state during BCI operation (e.g. the fatigue and attention levels of the BCI user) [137, 207]. This combination of between- and within-session factors can cause inconsistent BCI performance. However, adaptive BCI algorithms may be capable of mitigating the effects of such non-stationarities.

Previous adaptive BCI research has focused primarily on between session non-stationarity. For example, Shenoy et al. explored several approaches for adapting a linear discriminant analysis classifier based on data recorded during each new testing session [207]. Li et al. investigated the usage of covariate shift adaptation to pursue a similar goal without the need for labeled data from the new testing session [130]. More recently, Nicolas-Alonso et al. have used kernel discriminant analysis to gradually enhance classification
during testing sessions [165]. Comparatively little attention has been devoted to methods of adapting to within-session non-stationarity.

It has been hypothesized that psychological conditions such as fatigue and attention influence BCI performance [38, 49]. These latent mental states can change rapidly during BCI usage and can affect BCI performance both directly, by causing EEG signal fluctuations that increase the difficulty of detecting the required control task (e.g. motor imagery), and indirectly, by inhibiting the BCI user’s ability to effectively complete the required control task. Recent work suggests that at least three mental states - fatigue, frustration, and attention - have statistically measurable effects on BCI performance [159]. However, adaptation to these states is complicated by their subjective nature and the difficulty of monitoring them with high accuracy.

Unlike the intentional cognitive tasks typically detected by BCIs, changes in mental states such as fatigue are involuntary. Systems that are used to monitor such involuntary changes in cognitive activity have been defined as passive BCIs to differentiate them from the more traditional active and reactive BCIs [248]. Some recent examples of passive BCI research include monitoring mental workload [153, 197] and affective state [96]; detecting the perceived loss of control over a system [246]; and detecting interaction errors by active BCIs [125]. In our own work, we have demonstrated the ability to detect coarse modulations in fatigue, frustration, and attention during the performance of mental tasks similar to those used by many BCIs [161]. Combining this passive detection of mental state with a traditional task-based BCI may allow the design of a hybrid BCI that adapts to changes in psychological state.

Approaches to between-session adaptation have generally assumed that the separability between BCI classes (e.g. left and right hand motor imagery) is not lost between sessions, but rather, the distributions of each class migrate within the feature space, necessitating an updated classifier [207]. It is not clear that this assumption holds for within-session adaptation, as certain psychological conditions (e.g. complete loss of attention) may preclude accurate BCI classification. This motivates two approaches to psychological BCI adaptation. In the first, predicted mental state can be used to assess the risk that a given BCI trial will be classified incorrectly by a static, or non-adaptive, BCI. This approach assumes that separability is irreversibly lost under certain conditions. In the second, predicted mental state can be used to directly adapt the BCI, implicitly assuming that separability is retained but the classifier must be retrained to recognize it.

This paper proposes and evaluates a psychologically adaptive two-class EEG-BCI based on personalized
mental tasks. This BCI encompasses both a passive BCI that predicts the current levels of fatigue, frustration, and attention and an active BCI that differentiates between two mental tasks. The effects of using predicted mental state to assess the likelihood of task misclassification and to directly increase the accuracy of task classification are independently investigated. These approaches were compared to a typical non-adaptive BCI that ignores psychological state.

7.3 Methods

7.3.1 Protocol

11 able-bodied participants (two male, average age 27.4 years) completed five sessions, each of which was approximately one hour in duration. The first two sessions were offline training sessions while the remaining three sessions were online testing sessions. Participants had normal or corrected-to-normal vision and refrained from consuming caffeine for four hours prior to each session. Participants provided written informed consent, and the experiment was approved by the Holland Bloorview Research Ethics Board.

7.3.2 Training Sessions

During each training session, participants completed 30 five-second trials of each of five mental tasks - four active tasks (motor imagery, mental arithmetic, music imagery, and word generation) and an unconstrained rest task. For each participant, two-class BCIs were trained to differentiate each of the four active tasks from the unconstrained rest task. One active task was selected for each participant on the basis of participant preference and the accuracy of each BCI. Eight participants selected the word generation task, two selected the motor imagery task, and one selected the music imagery task. The BCI that differentiated the selected task from the unconstrained rest task was used during the testing sessions. Full details regarding the protocol of the training sessions can be found in [159].
7.3.3 Testing Sessions

During each testing session, participants used the selected BCI to play a simple maze navigation game. BCI performance was intentionally hindered by omitting any between session adaptation to ensure that significant changes in psychological state were induced. Participants periodically self-reported their perceived levels of fatigue, attention, and frustration on a continuous scale between 0 and 1. Full details regarding the protocol of the testing sessions, the online performance of the BCI, and the observed relationships between BCI performance and mental state can be found in [159].

7.3.4 Data Acquisition, Signal Processing, and Feature Extraction

During each session, electrical activity on the cerebral cortex was recorded from 15 locations (Fz, F1, F2, F3, F4, Cz, C1, C2, C3, C4, CPz, Pz, P1, P2, and POz by the international 10-20 system [102]) using a B-Alert X24 wireless EEG headset (Advanced Brain Monitoring, Carlsbad, CA). Each recorded signal was band-pass filtered between 2 and 30 Hz. Independent component analysis was used to remove signal artefacts caused by eye movements and blinking [152].

Each five-second trial during which either the rest or the active task was performed was extracted from the recorded EEG signals. Only data from the testing sessions were analyzed. A frequency-domain feature set was used to characterize each trial due to the short trial duration and the importance of cortical activity within the four major EEG frequency bands (i.e. delta, theta, alpha, and beta) for monitoring mental state [2, 118]. For each trial, a fast Fourier transform was used to convert the recorded EEG signals into the frequency domain. The frequency spectra for each signal were compressed by computing the total power within each non-overlapping 1 Hz frequency range from 0-1 Hz to 29-30 Hz. Each of these powers was used as a feature, yielding a feature set with 450 features (the power within 30 frequency ranges from each of the 15 electrodes). This feature set was further compressed during cross-validation using a feature clustering algorithm that derived participant-specific frequency bands for each electrode [160].
7.3.5 Mental State Prediction

Data from the online sessions were analyzed offline to investigate the ability to predict self-reported levels of fatigue, frustration, and attention based on the recorded EEG data. Previous work on detection of these states focused on binary differentiation between low and high levels of each state [161]. In contrast, during this study each state was self-reported on a continuous scale between 0 and 1. This provided a more nuanced measure of mental state but also necessitated a supervised regression approach.

A 10 x 10 (runs x folds) repeated cross-validation was performed for each participant. Least-squares regression was performed using lasso regularization [222]. The quality of mental state prediction was then quantified by computing the Pearson correlation coefficients between the self-reported and predicted values of each mental state.

7.3.6 Adaptive Brain-Computer Interface

Two methods of adapting to changes in mental state were investigated: mental state-based reliability prediction, and mental state-based classifier adaptation. For each approach, five features were selected for classification using a fast correlation-based filter (FCBF) while a linear discriminant analysis (LDA) classifier was used to differentiate the rest and active tasks [25, 245].

Prediction of BCI Reliability

This approach assumes that the predicted mental state contains predictive information regarding the success of mental task classification. An inner cross-validation was performed on the training data to predict the class of each training sample using LDA. Let \( x_n \in \mathbb{R}^d, d = 5, n = 1, \cdots, N \) denote the \( n \)th training sample, where \( N \) represents the total number of samples. Projections \( p_n \in \mathbb{R} \) for each sample were computed as:

\[
P = [p_1 \ p_2 \ \cdots \ p_N] \tag{7.1}
\]

\[
p_n = w^T \cdot x_n + c, \quad 1 \leq n \leq N \tag{7.2}
\]
where $w$ and $c$, respectively, are the weight vector and constant specified by Fisher’s linear discriminant [25]. Each training sample $x_n$ was classified as reliable ($r_n = 0$) or unreliable ($r_n = 1$) according to the following:

\[
\begin{align*}
  r_n &= 0 \quad \forall n \mid p_n < -h \ast \text{med}(|P|), \quad C_n = 0 \\
  r_n &= 0 \quad \forall n \mid p_n > h \ast \text{med}(|P|), \quad C_n = 1 \\
  r_n &= 1 \quad \forall n \mid p_n > -h \ast \text{med}(|P|), \quad C_n = 0 \\
  r_n &= 1 \quad \forall n \mid p_n < h \ast \text{med}(|P|), \quad C_n = 1
\end{align*}
\]

where $C_i = 0$ indicates that the sample represented a rest task and $C_i = 1$ indicates that the sample represented an active task. The threshold $h \in \mathbb{R}$ controlled the aggressiveness of the algorithm. For $h = 0$, training samples were classified as unreliable only if they were misclassified during the inner cross-validation. For $h > 0$, training samples were also classified as unreliable if they were classified correctly but were projected close to the boundary between classes. For $h < 0$, training samples were only classified as unreliable if they were both misclassified and projected far from the boundary. The threshold $h$ was scaled by the median of the absolute magnitude of the projection vector $P$ to achieve invariance to the magnitude of projections between participants.

Previous work suggested that the relationship between mental state and BCI performance may be complex [159]. Consequently, a support vector machine (SVM) classifier was used to predict BCI reliability [25]. This classifier was trained using the set of three predicted mental states as a feature set and the training sample reliabilities $R = [r_1 \, r_2 \, \cdots \, r_N]$ as the set of target labels. The resultant classifier was used to predict BCI reliability for the testing set. The relationships between the three predictive algorithms involved in this analysis are depicted conceptually in Figure 7.1.

For the testing set, the task label targets are referred to as $T$, the task label predictions as $\hat{T}$, and the predictions of BCI reliability as $\hat{R}$. Each testing sample can be classified as either a true positive (TP), false positive (FP), true negative (TN), or false negative (FN) based on:

104
From these categorizations, six evaluation criteria were used to evaluate the efficacy of the reliability predictions - the sensitivity (Se), specificity (Sp), balanced classification accuracy (BCA), positive predictive value (PPV), negative predictive value (NPV), and overall predictive value (OPV), defined as:

\[
TP = \{n \mid (r_n = 1) \cap (t_n \neq \hat{t}_n)\} \tag{7.7}
\]

\[
FP = \{n \mid (r_n = 1) \cap (t_n = \hat{t}_n)\} \tag{7.8}
\]

\[
TN = \{n \mid (r_n = 0) \cap (t_n = \hat{t}_n)\} \tag{7.9}
\]

\[
FN = \{n \mid (r_n = 0) \cap (t_n \neq \hat{t}_n)\} \tag{7.10}
\]

\[
Se = \frac{|TP|}{|TP| + |FN|} \tag{7.11}
\]

\[
Sp = \frac{|TN|}{|TN| + |FP|} \tag{7.12}
\]

\[
PPV = \frac{|TP|}{|TP| + |FP|} \tag{7.13}
\]

\[
NPV = \frac{|TN|}{|TN| + |FN|} \tag{7.14}
\]

\[
BCA = \frac{1}{2}(Se + Sp) \tag{7.15}
\]

\[
OPV = NPV - (1 - PPV) \tag{7.16}
\]

We note that the OPV represents the difference between the classification accuracy of the BCI for ‘reliable’ samples and the classification accuracy of the BCI for ‘unreliable’ samples. A 10 x 10 repeated cross-validation was used to estimate BCA and OPV. Chance levels for each metric were computed using a permutation test by randomly permuting the training set reliability predictions R prior to training the SVM classifier that predicted BCI reliability [173]. This procedure was repeated 1000 times to establish a probability distribution for both BCA and OPV.
Figure 7.1: Conceptual overview of how each prediction algorithm is trained. For each learning algorithm, the input feature set is represented by the arrow arriving at the top of the associated rectangle (yellow lines), the labels by the arrow arriving on the side of the associated rectangle (green lines), and the outputs by the arrows departing the associated rectangle (orange lines). For each training sample, the mental task performed is predicted using LDA and mental state is estimated using regularized regression. The mental task predictions are used to estimate BCI reliability (see Equations 3 through 6) and the estimated mental state is used to predict these reliabilities. All three learning algorithms are then used for the testing data within each fold of cross-validation.

**Adaptive Classification**

This approach assumes that the separability of the active and rest tasks is retained throughout mental state space, but that the class distributions of each task move within the feature space as the BCI user moves within mental state space. To address this, a new classifier was trained for each sample in the testing set using only the training samples closest to that test sample within mental state space. Mental state space can be visualized as a three-dimensional space within which each training or test sample can be located using the associated predictions of fatigue, frustration, and attention. However, previous work indicates that the importance of each of these states varies between individuals [159]. Consequently, it was necessary to individualize the concept of mental state space by removing mental states that did not affect classification for each participant. This was performed within each fold of cross-validation based on the training data alone.

A random subset of data, balanced between the rest and active classes, was sampled from the training set. A fast correlation-based filter [245] was used to reduce the feature set to five features, and an LDA classifier was
trained to differentiate the two classes. The weight vector for the classifier was used to project the training
data to the single dimension used for classification. For a set of feature vectors $X$ with corresponding class
labels $C$, projections $P$, and mental state predictions for any arbitrary state $M$, scores for each class were
computed as:

$$ S_{rest} = |\text{cov}(P_i, M_i)| \quad \forall i \mid C_i = 0 \quad (7.17) $$

$$ S_{active} = |\text{cov}(P_j, M_j)| \quad \forall j \mid C_j = 1 \quad (7.18) $$

These scores represent the degree to which the projection of each class varies with the predicted mental state.
The relationship between each mental state and the separability of the two classes was also quantified by
randomly sampling 250 pairs of one rest task and one active task and, for each pair, computing the Euclidean
distance between their projections and the distance between their predicted mental states. Concatenating
these variables across all pairs to construct vectors of projection distance $D_P$ and mental state distance $D_M$,
another score was computed as:

$$ S_{diff} = |\text{cov}(D_P, D_M)| \quad (7.19) $$

Each score was computed independently for all three mental states. Bootstrapping was used to enhance
stability. The size of the randomly sampled subset was varied between 40% and 100% of the size of the
entire training set, and 50 iterations were completed at each size. The average value of each score was
computed for each mental state across all runs at all sizes. Scores were scaled by dividing by the maximum
values of $S_{rest}$, $S_{active}$, and $S_{diff}$ observed for any state. For each state, a composite score was computed by
summing the three scaled scores. An optimal set of states was identified by choosing the two states with the
highest and second highest composite scores. The third state was included only if its score exceeded 70% of
the maximum value. Subsequent analysis used only the states that were included in this optimal set.

The proportion of the training set used to train the BCI within each fold was determined through a repeated
inner cross-validation on the training data only. For each fold of this inner cross-validation, a classifier was
trained for each test sample using 40% to 100% of the data from the remaining folds, sampled based on their

107
proximity to that test sample. At each of these proportions, both the classification accuracy and the Fisher score between the projections of the rest task and those of the active task were computed. The resultant accuracies and Fisher scores were smoothed and the proportions that yielded the maximum values of both variables were located. The average of these two proportions was used to train classifiers for each testing point in the outer cross-validation. Figure 7.2 provides a conceptual overview of the stages of the adaptive classification algorithm while Figure 7.3 depicts the locations of one test sample and all training samples in mental state space, along with visualizations of three possible settings for the proportion of the training set used for classification.

Figure 7.2: Conceptual overview of the adaptive classification algorithm. The optimal mental states and proportion are chosen using an inner cross-validation on the training data and then the training data are resampled for each testing sample to train a new classifier.

The classification accuracy attained by this adaptive LDA algorithm was compared to those attained by two non-adaptive BCIs - one that randomly sampled the training set to match the size of the training set used by the adaptive BCI, and one that used the entire training set. A 10 x 10 repeated cross-validation was performed for the adaptive BCI. A permutation test was used to estimate the performance of the non-adaptive BCI by performing 100 iterations of a 10 x 10 repeated cross-validation [173]. Each run was randomly initialized to ensure a unique set of samples was selected for each fold. All participants had more examples of the rest class than the active class, but classifiers were provided with balanced data for training. For the adaptive BCI, this was achieved by randomly sampling the subset of training samples from the rest class that were as close to each testing point as the most distant training sample from the active class.
Figure 7.3: Visualization of 25%, 50%, and 75% cutoffs for training set inclusion. The filled circle represents the predicted mental state for the testing sample while the open circles represent the predicted mental state for each point in the training set.

7.4 Results

7.4.1 Mental State Prediction

The correlation coefficients between the self-reported and predicted values of each mental state are presented in Table 7.1. The average correlation ranged between .46 for attention and .56 for frustration, indicating that the mental state prediction algorithm was moderately accurate for the population as a whole.

Table 7.1: Correlation between predicted and self-reported values of each mental state for all participants.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Fatigue</th>
<th>Frustration</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.25 ± .13</td>
<td>.39 ± .12</td>
<td>.42 ± .13</td>
</tr>
<tr>
<td>2</td>
<td>.61 ± .13</td>
<td>.51 ± .13</td>
<td>-.01 ± .16</td>
</tr>
<tr>
<td>3</td>
<td>.52 ± .11</td>
<td>.69 ± .09</td>
<td>.43 ± .14</td>
</tr>
<tr>
<td>4</td>
<td>.87 ± .03</td>
<td>.78 ± .09</td>
<td>.72 ± .09</td>
</tr>
<tr>
<td>5</td>
<td>.67 ± .08</td>
<td>.32 ± .13</td>
<td>.33 ± .12</td>
</tr>
<tr>
<td>6</td>
<td>.71 ± .06</td>
<td>.63 ± .08</td>
<td>.49 ± .10</td>
</tr>
<tr>
<td>7</td>
<td>.46 ± .12</td>
<td>.45 ± .11</td>
<td>.59 ± .08</td>
</tr>
<tr>
<td>8</td>
<td>.31 ± .12</td>
<td>.47 ± .11</td>
<td>.29 ± .11</td>
</tr>
<tr>
<td>9</td>
<td>.51 ± .10</td>
<td>.47 ± .12</td>
<td>.65 ± .15</td>
</tr>
<tr>
<td>10</td>
<td>.06 ± .14</td>
<td>.80 ± .05</td>
<td>.67 ± .07</td>
</tr>
<tr>
<td>11</td>
<td>.53 ± .09</td>
<td>.62 ± .07</td>
<td>.46 ± .16</td>
</tr>
</tbody>
</table>

Avg         | .50 ± .23 | .56 ± .16  | .46 ± .21 |
7.4.2 Reliability Prediction

Figure 7.4 depicts the effects of the threshold value \( h \) on both balanced classification accuracy (BCA) and overall predictive value (OPV) across all participants. Performance for both metrics was relatively stable until the threshold exceeded 0.5, after which they began to decrease. A threshold value of 0.4 was selected for all subsequent analyses. This choice of threshold falls within the stable region for both statistics while also ensuring that a significant number of samples are flagged as unreliable.

![Figure 7.4: Effects of threshold value on balanced classification accuracy and overall predictive value.](image)

Figure 7.5 presents the balanced classification accuracies and overall predictive values for each participant for the aforementioned threshold. The average values of BCA and OPV across the entire population were 54.2\% and 7.9\%, respectively. The observed results for both statistics significantly exceeded chance levels, as shown by Figure 7.6.

Although balanced classification accuracy exceeded chance levels on average, it is still quite low at only 54\%. Consequently, it is unlikely that mental estimation can be used in isolation for error detection and correction. However, mental state clearly does contain information pertinent to BCI reliability, as evidenced by the average OPV of nearly 8\%. This suggests that predicted mental state can be used online to provide some indication of BCI reliability.
Figure 7.5: Balanced classification accuracy (left) and overall predictive value (right) for all participants. The average BCA across participants was 54.24% and the average OPV across participants was 7.9%.

Figure 7.6: Comparison between observed BCA and OPV and random results from the permutation test. The observed values for both statistics significantly exceeded chance values. Note that this would also hold true for any point on either curve in Figure 7.4.

7.4.3 Adaptive Classification

Figure 7.7 compares the classification accuracies obtained by the adaptive BCI for all combinations of mental states and proportions to those obtained by the non-adaptive BCI when random sampling was used to extract identical proportions of the training set. For most participants, the adaptive BCI outperformed the non-adaptive BCI for smaller training set sizes and approached the performance of the non-adaptive BCI when the entire training set was used, as would be expected. This suggests that there are benefits to sampling training data based on mental state rather than at random.
Figure 7.7: Comparison between the non-adaptive BCI and the adaptive BCI as the size of the training set is varied. The dashed red line represents performance of the non-adaptive BCI while the thin black lines each represent one of the seven possible combinations of mental states used by the adaptive BCI. When the entire set is used, there is little difference between the two algorithms. When the size of the training set is reduced through resampling, the adaptive BCI displays superior performance for two thirds of participants.

Practically, it is more interesting to compare the performance of the adaptive BCI at all training set sizes to that of the non-adaptive BCI when the entire training set is used. A 95% confidence interval for the non-adaptive BCI classification accuracy under the latter condition was computed for each participant based on the results of the permutation test. Figure 7.8 compares these confidence intervals to the performance of the adaptive BCI using only the combination of mental states that had the highest average classification accuracy across all proportions. It is clear that some participants (i.e. Participants 1, 2, 4, 6, and 10) exhibit performance that exceeds the upper bounds of the confidence interval for a wide range of training set sizes.
Again, this suggests that psychological BCI adaptation is useful for some participants.

Figure 7.8: Accuracy of the most effective adaptive BCI for each participant compared to that of the non-adaptive BCI using the entire training set. Each graph depicts the 95% confidence interval for the accuracy of the non-adaptive BCI (dotted lines) and the accuracy of the adaptive BCI for a range of different training set sizes (blue line). The adaptive BCI either exceeded or approached the upper limit of the non-adaptive confidence interval for the majority of participants. The legend for each plot identifies the combination of fatigue (Fa), frustration (Fr), and attention (At) used for the adaptive BCI.

Finally, the mental state and proportion selection algorithms were used to identify an ideal combination of mental states and training set size for each participant. The frequency with which each combination of mental states was selected for each participant is presented in Table 7.2. On average, each mental state was selected approximately 80% of the time (78.4%, 80.0%, and 82.2% for fatigue, frustration, and attention, respectively).
Table 7.2: Frequency with which each combination of mental states was chosen for each participant. Fa refers to fatigue, Fr to frustration, and At to attention. The overall frequencies with which each mental state was selected for each participant, either independently or in combination with any other state, are presented in the final three columns.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Fa-Fr</th>
<th>Fa-At</th>
<th>Fr-At</th>
<th>Fa-Fr-At</th>
<th>Fa</th>
<th>Fr</th>
<th>At</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.98</td>
<td>0.98</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>0</td>
<td>0.08</td>
<td>0.72</td>
<td>0.92</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.9</td>
<td>1</td>
<td>0.9</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0.02</td>
<td>0.9</td>
<td>0.06</td>
<td>0.02</td>
<td>0.94</td>
<td>0.1</td>
<td>0.98</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>6</td>
<td>0.6</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>7</td>
<td>0.74</td>
<td>0</td>
<td>0</td>
<td>0.26</td>
<td>1</td>
<td>1</td>
<td>0.26</td>
</tr>
<tr>
<td>8</td>
<td>0.06</td>
<td>0</td>
<td>0.8</td>
<td>0.14</td>
<td>0.2</td>
<td>1</td>
<td>0.94</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0.34</td>
<td>0.66</td>
<td>0.66</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0.14</td>
<td>0</td>
<td>0.58</td>
<td>0.28</td>
<td>0.42</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The adaptive classification accuracy and average training set size for each participant are summarized in Table 7.3. Five of 11 participants exhibited adaptive classification accuracies that exceeded the upper limit of the 95% confidence interval for the non-adaptive classification accuracy. Overall, the adaptive BCI exhibited an accuracy of 73.2% while the non-adaptive BCI exhibited an accuracy of 72.6%. All participants exhibited adaptive classification accuracies that exceeded the lower limit of the 95% confidence interval, although Participants 9 and 11 were close to this limit.

Table 7.3: Adaptive and non-adaptive classification accuracies for each participant. The performance of the adaptive BCI exceeded the limits of the 95% confidence interval for the performance of the non-adaptive BCI for participants denoted with an asterisk. The proportion of the training set used by the adaptive BCI, as determined by the algorithm detailed in 2.6.2, is listed in the final column.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Non-Adaptive</th>
<th>Adaptive</th>
<th>Training Set Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>90.2</td>
<td>91.2</td>
<td>0.80</td>
</tr>
<tr>
<td>2*</td>
<td>70.5</td>
<td>72.6</td>
<td>0.62</td>
</tr>
<tr>
<td>3</td>
<td>62.6</td>
<td>63.5</td>
<td>0.75</td>
</tr>
<tr>
<td>4*</td>
<td>73.2</td>
<td>74.9</td>
<td>0.64</td>
</tr>
<tr>
<td>5*</td>
<td>68.8</td>
<td>71.2</td>
<td>0.62</td>
</tr>
<tr>
<td>6</td>
<td>77.2</td>
<td>77.2</td>
<td>0.61</td>
</tr>
<tr>
<td>7</td>
<td>65.3</td>
<td>65.3</td>
<td>0.89</td>
</tr>
<tr>
<td>8</td>
<td>73.4</td>
<td>73.2</td>
<td>0.77</td>
</tr>
<tr>
<td>9</td>
<td>84.6</td>
<td>84.0</td>
<td>0.78</td>
</tr>
<tr>
<td>10*</td>
<td>67.1</td>
<td>67.8</td>
<td>0.59</td>
</tr>
<tr>
<td>11</td>
<td>65.4</td>
<td>64.5</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Average 72.6 73.2 0.72

Figure 7.9 compares the performance of the adaptive BCI with automatic mental state and proportion selection to the 95% confidence interval for the non-adaptive BCI and the performance of the most frequently
selected combination of states for all proportions of training set size.

Figure 7.9: Adaptive BCI performance for each participant. The solid horizontal and vertical lines identify respectively the mean classification accuracy and training set size for the adaptive BCI. The performance of the most frequently selected combination of mental states is represented by the blue dashed line and the 95% confidence interval for the non-adaptive classification accuracy by the two dashed horizontal lines. Participants 1, 2, 4, 5, and 10 exhibited significantly better performance with the adaptive BCI while no participants exhibited significantly worse performance. The settings of the adaptive BCI indicate that mental state and proportion selection effectively identified favorable operating conditions for most participants.
7.5 Discussion

7.5.1 Predictive Value of Mental State for BCIs

The results presented here suggest that predictions of mental state can be used both to estimate BCI reliability and to directly improve the accuracy of classification. When mental state was used to classify the reliability of BCI decisions for unseen trials, we observed an 8% difference in BCI classification accuracy between trials classified as reliable and trials classified as unreliable. When mental state was used to directly adapt the BCI classifier, five of eleven participants exhibited a significant increase in classification accuracy. These findings represent an initial step towards the design of BCIs that are capable of adapting to short-term changes in psychological state within a session. It may be possible to use this method in combination with methods that compensate for dramatic non-stationarity between training and testing sessions [216] or those that slowly update a classifier during a session [230].

One implication of this analysis is that under some circumstances a small set of relevant BCI training data is preferred over a larger set that also includes less relevant data. This emphasizes the importance of classification algorithms that match each new testing sample to the most relevant data in the training set, such as the dynamically weighted ensemble of classifiers detailed by Liyanage et al [137]. During BCI development, it is important to focus not just on collecting a substantial amount of training data, but also on identifying optimal ways to leverage these data for each new sample.

The importance of relevant training data also has obvious implications for between-session adaptation. When retraining a BCI during a new session, it is common to use a limited subset of the previously collected data [189]. This limits the size of the training set and prevents the new data from being overwhelmed by pre-existing data from previous sessions. Figure 7.7 suggests that the method by which this pre-existing data are sampled may affect BCI performance. For small training set sizes, mental state-based sampling clearly outperformed random sampling for most participants. Likewise, when discarding previous data to retrain a classifier, it is clear that care should be taken to ensure that this is done systematically and that the relevance of the remaining data is maximized.

Offline analyses such as those performed here are often over-optimistic regarding their conclusions. However, in this case, there is reason to believe that the merits of psychological adaptation in BCIs have been under-
stated. Short-term modulations in mental state are inevitable during BCI usage and may negatively affect
performance [159]. However, it is unclear to what extent these changes in mental state are also driven by
BCI performance. The relationship between these variables is likely to be cyclical, as fluctuations in mental
state may elicit a deterioration in BCI performance that, in turn, evokes further fluctuations in mental state.
As a result, the true efficacy of psychological adaptation may only be apparent in online analyses, where this
adaptation can potentially be used to suppress this positive feedback cycle between BCI performance and
mental state. This will be a focus of future research.

7.5.2 Contrast with Error Correction Using the Error-Related Potential

The error-related potential (ERP) has previously been used to detect and correct errors during BCI in-
teraction [213]. While the mental state-based prediction of BCI reliability detailed here cannot match the
effectiveness of ERP-based error detection, several characteristics of this approach distinguish it from ERPs.
Accurate detection of ERPs requires a participant to remain focused on the BCI so that they realize when it
has erred, while mental state-based reliability prediction should remain effective when participants are inat-
tentive. In fact, the approach detailed here is particularly effective when participant focus is inconsistent,
rendering it complementary to traditional ERPs. Furthermore, while ERP detection and the associated error
correction require a BCI trial to be terminated and a classification decision to be issued, mental state-based
reliability prediction can be performed online while the BCI trial continues. Consequently, intermediate
actions can be taken to reduce the risk of an error occurring at all. For example, in an online test, mental
state-based classifier reliability could be computed online and a BCI trial could be lengthened in real-time
if the estimated mental state indicates that classifier performance is uncertain. Alternatively, in the specific
case that disengagement from the BCI is detected, the BCI could either shut down to prevent uninten-
tional input or produce an attentional cue to recapture attention. This provides more versatility than the
straightforward error correction offered by ERPs.

7.5.3 Variation Between Reliability Prediction and Classifier Adaptation

The individual differences in the performance of the reliability prediction and classifier adaptation approaches
are intriguing. For example, reliability prediction was most accurate for Participant 4 and mostly uninforma-
tive for Participant 2. However, adaptive classification provided statistically significant improvements over non-adaptive classification for both participants. Meanwhile, some participants exhibited no significant differences between the adaptive and non-adaptive BCIs despite excellent reliability prediction (e.g. Participant 9). Overall, there does not appear to be a significant relationship between the effectiveness of reliability prediction and that of the adaptive BCI. Although effective reliability prediction does not necessarily translate to adaptive BCI success due to the possibility that class distributions may simply be inseparable under some psychological conditions, it is less clear why the converse is untrue. Regardless, these results suggest that different adaptation strategies are effective for different individuals, highlighting the importance of flexible BCI design.

7.5.4 Limitations and Future Directions

Several limitations are apparent within this study. The first is the offline nature of the analysis, which complicates the estimation of the value of mental state-guided adaptation, as previously discussed. It is crucial that future research consider online implementations of psychological adaptation. Second, the accuracy of mental state prediction clearly imposes some limits on the efficacy of adaptation. Although all three mental states were predicted with moderate accuracy, more accurate predictions may allow greater improvement in adaptive BCI performance.

Given a passive BCI that measures mental state and an active BCI that detects mental task performance, psychological adaptation requires the interactions between these systems to be defined. This is a non-trivial task. Two approaches have been investigated here, one of which limits the passive BCI to providing a supplementary measure of BCI reliability and one of which directly intertwines the two systems by using the output of the passive BCI to control the active BCI. However, these are not the only potential architectures for psychological adaptation, and they are not mutually exclusive. Combining the two approaches may allow all users to benefit from psychological adaptation, and there is no doubt that there remain many other ways to fuse passive and active BCIs.
7.6 Conclusions

This study investigated two methods by which an EEG-BCI can adapt to changes in mental state. Fatigue, frustration, and attention levels during BCI usage were predicted using least-squares regression with lasso regularization. The Pearson correlation coefficients between self-reported mental state and the predicted values approached or exceeded 0.5. These predictions of mental state were used to estimate BCI reliability with an accuracy exceeding chance levels. An 8% difference in classification accuracy was uncovered between trials classifiers as reliable and those deemed unreliable. Mental state estimations were also used to directly adapt an active BCI; the classifier was retrained using a mental state-guided resampline of the training set. Five of eleven participants exhibited significantly improved classification accuracies using the adaptive algorithm while no participants exhibited significant decreases. These results suggest that psychological adaptation may provide a means of improving online EEG-BCI performance.
Chapter 8

Conclusions

8.1 Summary of Contributions

This thesis makes several original contributions to the field of biomedical engineering, particularly in the domains of biomedical signal processing and electrical brain-computer interfacing. Specifically, in this thesis I have:

1. Developed a signal processing algorithm that uses a clustering of features to automatically identify information-bearing frequency bands within electroencephalographic signals [160]. This algorithm combines the adaptability of narrow-band frequency-domain EEG analysis with the versatility of wide-band frequency-domain EEG analysis using the traditional delta, theta, alpha, and beta bands (Chapter 3).

   (a) Showed that, when used as a dimensionality reduction method, this feature clustering algorithm provides statistically superior performance to both wide-band and narrow-band EEG analysis for the classification of three different mental tasks and three different mental states.

2. Demonstrated that binary fluctuations in mental fatigue, frustration, and attention can be reliably classified on a single-trial basis using an non-regularized LDA classifier [161]. Respective classification accuracies of 76.8%, 71.9%, and 86.1% were attained for these three mental states. Performance for
each mental state was in excess of chance for each individual participant and for the study population as a whole. The observed classification accuracies were in excess of or comparable to previous research, even though most previous studies have differentiated only dramatic changes in mental state (Chapter 4).

3. Identified cortical regions and frequency ranges within which neuroelectrical activity is most predictive of changes in fatigue, frustration, and attention [161]. Specifically, fatigue is most effectively predicted by fronto-central cortical activity within the delta, theta, alpha, and beta bands. Frustration is most effectively predicted by frontal and posterior activity within the theta and alpha bands. Attention is most effectively predicted by parieto-occipital activity within the alpha band (Chapter 4).

(a) Showed that accurate detection of binary fluctuations in these three mental states can be maintained when only small numbers of EEG electrodes are used and when the EEG signal features used for mental state classification are selected based on an agglomeration of data from other individuals. This suggests that a small number of cortical areas and frequencies are most predictive for each mental state but that the importance of each area and frequency band varies on an individual basis.

4. Demonstrated that the online performance of an active mental-task based EEG-BCI exhibited statistically significant correlations with the self-reported mental state of the BCI user [159]. Perceived frustration has a statistically significant relationship with BCI classification accuracy, with higher classification accuracies experienced when frustration increases, potentially highlighting the importance of moderate frustration as a motivating factor. The relationship between perceived fatigue and BCI performance approached significance, with higher fatigue tending to accompany more accurate mental task classification (Chapter 5).

(a) Identified statistically significant relationships between BCI performance and two-dimensional slices of user mental state. Of particular significance, BCI performance significantly exceeded expected BCI performance when low frustration and moderate to high fatigue were experienced, as well as when moderate to high attention and moderate fatigue were experienced. These findings suggest that BCI control is not optimal under the low fatigue-low frustration-high attention conditions as one might intuitively expect. Rather, some level of fatigue and frustration are helpful to BCI performance, potentially because these cognitive conditions more commonly resemble those
experienced during BCI training.

5. Demonstrated that a verbal generation task previously used predominantly for metabolic brain-computer interfaces is an effective alternative to the dominant motor imagery task for EEG-BCI control [159]. This has significant implications for future EEG-BCI research involving individuals who struggle to reliably generate detectable motor imagery activation (Chapter 6).

6. Demonstrated that perceived mental state can be predicted on a continuous scale using a regression algorithm based on frequency-domain EEG activity [162]. The average Pearson correlation coefficients between perceived and predicted fatigue, frustration, and attention were 0.50, 0.56, and 0.46, respectively. These results contrast with the binary classification scheme most commonly used for mental state detection within previous literature (Chapter 7).

7. Demonstrated that predicted fatigue, frustration, and attention can be used to improve BCI performance [162] (Chapter 7).

(a) Showed that predicted mental state can be used to predict BCI performance. Using a support vector machine classifier, predicted mental state was used to predict whether a BCI would successfully classify a given trial. BCI performance was 8% higher when reliable performance was predicted than when unreliable performance was predicted, significantly exceeding chance levels. This suggests that mental state, though not a perfect predictor of BCI performance, contains a statistically significant amount of information regarding BCI performance.

(b) Showed that predicted mental state can be used to directly adapt a BCI. A novel BCI algorithm was designed to resample the training set and retrain a classifier by using only previous samples closely mirroring the current levels of fatigue, frustration, and attention. Using this algorithm, 5 of 11 participants demonstrated significant individual increases in classification accuracy while no participants exhibited significant decreases in classification accuracy.

8.2 Future Work

The contributions of this thesis represent an initial step towards BCIs that truly adapt to the psychological state of their users. The following are general recommendations to the reader regarding further steps in this
8.2.1 Online Implementation

The immediate next step for this research must be online implementation and testing of the results detailed by this thesis. Although the relationships between perceived mental state and BCI performance were identified based on online BCI usage, the time requirements for this thesis precluded the online testing of the adaptive methods detailed in Chapter 7. This is a critical step towards the realization of psychologically adaptive BCIs and should be performed as soon as possible.

8.2.2 Combining Tonic and Phasic Adaptation

The adaptation considered within this thesis can be considered as a type of tonic adaptation. Fatigue, frustration, and attention fluctuate in real-time during BCI usage. However, they do so much more slowly than event-related potentials such as the error-related potential. Previous research cited elsewhere in this thesis has demonstrated the merits of what we might call phasic BCI adaptation - adaptation based on these nearly-instantaneous event-related potentials [70, 71]. This thesis, on the other hand, has demonstrated the potential merits of BCI adaptation based on the recognition of comparatively slow changes in cognitive conditions such as fatigue. There is reason to believe that the benefits of these two types of adaptation are not entirely overlapping. For instance, changes in mental state provide a preemptive opportunity to adapt a BCI, while event-related potentials can only be used retroactively to attempt to correct BCI operation. Since these methods function uniquely but are acquired using the same instrumentation, there is no limitation to the ability to combine them. Consequently, future research regarding psychologically adaptive BCIs should recognize the differences between these modes of adaptation and attempt to combine them.

8.2.3 Expanding Context-Awareness

We wrote in Chapter 1 that adaptation to mental state can be considered as a type of context awareness. However, we also wrote that contextual information need not only relate to the state of the BCI user. Ample contextual information can also be made available regarding the state of the BCI itself and the state of the
physical and emotional environment in which the BCI is being used. This information could also be used to increase BCI usability. For instance, how do environmental conditions or the presence of other individuals affect the ways in which a BCI user is likely to initiate or respond to communication? Investigating these questions may allow for a more complete treatment of context and contextual awareness with respect to BCI usage. Moreover, there are pressing reasons why this is desirable. Most BCI research focuses primarily on iterative advances in signal processing and the accuracy of mental task detection. Such improvements provide no antidote for the issues of fatigue and user-unfriendliness that trouble most BCIs. Intelligent context awareness presents a potential means by which BCI control can be made easier and more intuitive for BCI users themselves.

8.3 Resulting Publications

The following publications comprise the majority of this thesis (Chapters 3, 4, 5, and 7):

A. Myrden and T. Chau. Feature clustering for robust frequency-domain classification of EEG activity. *Journal of Neuroscience Methods*, manuscript under revision.


The following secondary publications were completed during the course of this Ph.D. They are not directly related to the central topic of this thesis, but represented a significant time investment and have been included for completeness:


H. Faulkner, A. Myrden, M. Li, K. Mamun, and T. Chau. Sequential hypothesis testing for automatic
Bibliography


[215] Strijkstra, A. M., Beersma, D. G., Drayer, B., Halbesma, N., and Daan, S. Subjective sleepiness correlates negatively with global alpha (8–12 Hz) and positively with central frontal theta


