Evaluating Human Performance in Virtual Reality Based on Psychophysiological Signal Analysis

by

Jillian M. Clements

Department of Electrical and Computer Engineering
Duke University

Date:_____

Approved:

_____
Leslie Collins, Chair

_____
Lawrence G. Appelbaum, Co-Chair

_____
Regis Kopper

_____
Marc Sommer

_____
Stacy Tantum

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering in the Graduate School of Duke University

2018

ABSTRACT

Evaluating Human Performance in Virtual Reality Based on Physiological Signal Analysis

by

Jillian M. Clements

Department of Electrical and Computer Engineering
Duke University

Date:_____

Approved:

_____
Leslie Collins, Chair

_____
Lawrence G. Appelbaum, Co-Chair

_____
Regis Kopper

_____
Marc Sommer

_____
Stacy Tantum

An abstract submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering in the Graduate School of Duke University

2018

Copyright by
Jillian M. Clements
2018

Abstract

Physiological signals measured from the body, such as brain activity and motor behavior, can be used to infer different physiological states or processes in humans. Signal processing and machine learning often play a fundamental role in this assessment, providing unique approaches to analyzing and interpreting physiological data for a variety of applications, such as medical diagnosis and human-computer interaction. In this work, these approaches were utilized and adapted for two separate applications: brain-computer interfaces (BCIs) and the assessment of visual-motor skill in virtual reality (VR).

The goal of BCI technology is to allow people with severe motor impairments to control a device without the need for voluntary muscle control. Conventional BCIs operate by converting electrophysiological signals measured from the brain into meaningful control commands, eliminating the need for physical interaction with the system. However, despite encouraging improvements over the last decade, BCI use remains primarily in research laboratories. One of the biggest obstacles limiting their daily in-home use is the significant amount of time and expertise that is often required to set up the biosensors (electrodes) for recording brain activity. The most common modality for brain recording is electroencephalography (EEG), which typically employs gel-based "wet" electrodes for recording signals with high signal-to-noise ratios (SNRs). However,

while wet electrodes record higher quality signals than dry electrodes, they often hinder frequent use because of the complex and time-consuming process of applying the electrodes to the scalp. Therefore, in this research, a signal processing solution was implemented to help mitigate noise in a dry electrode system to facilitate a more practical BCI device for everyday use in people with severe motor impairments. This solution utilized a Bayesian algorithm that automatically determined the amount of EEG data to collect online based on the quality of incoming data. The hypothesis for this research was that the algorithm would detect the need for additional data collection in low SNR scenarios, such as those in the dry electrode systems, and collect sufficient data to improve BCI performance. In addition to this solution, two anomaly detection techniques were implemented to characterize the differences between the wet and dry electrode recordings to determine if any additional types of signal processing would further improve BCI performance with dry electrodes. Taken as a whole, this research demonstrated the impact of noise in dry electrode recordings on BCI performance and showed the potential of a signal processing approach for noise mitigation. However, further signal processing efforts are likely necessary for full mitigation and adoption of dry electrodes for use in the home.

The second study presented in this work focused on signal processing and machine learning techniques for assessing visual-motor skill during a simulated marksmanship task in immersive VR. Immersive VR systems offer flexible control of an

interactive environment, along with precise position and orientation tracking of realistic movements. These systems can also be used in conjunction with brain monitoring techniques, such as EEG, to record neural signals as individuals perform complex motor tasks. In this study, these elements were fused to investigate the psychophysiological mechanisms underlying visual-motor skill during a multi-day simulated marksmanship training regimen. On each of 3 days, twenty participants performed a task where they were instructed to shoot simulated clay pigeons that were launched from behind a trap house using a mock firearm controller. Through the practice of this protocol, participants significantly improved their shot accuracy and precision. Furthermore, systematic changes in the variables extracted from the EEG and kinematic signals were observed that accompanied these improvements in performance. Using a machine learning approach, two predictive classification models were developed to automatically determine the combinations of EEG and kinematic variables that best differentiated successful (target hit) from unsuccessful (target miss) trials and high-performing participants (top fourth) from low-performing participants (bottom fourth). Finally, in order to capture the more complex patterns of human motion in the spatiotemporal domain, time series methods for motion trajectory prediction were developed that utilized the raw tracking data to estimate the future motion of the firearm controller. The objective of this approach was to predict whether the controller's virtually projected ray would intersect with the target

before the trigger was pulled to shoot, with the eventual goal of alerting participants in real-time when shooting may be suboptimal.

Overall, the findings from this research project point towards a comprehensive psychophysiological signal processing approach that can be used to characterize and predict human performance in VR, which has the potential to revolutionize the design of current simulation-based training programs for realistic visual-motor tasks.

Contents

Abstract iv

Contents viii

List of Tables xii

List of Figures xiii

Acknowledgements xvii

1 Introduction 1

2 Background 8

2.1 Electroencephalography 8

2.1.1 Signal Acquisition 9

2.1.2 Electrode Type 10

2.1.2.1 Wet Electrodes 11

2.1.2.2 Dry Electrodes 12

2.1.3 Signal Characteristics 14

2.1.4 Physiological Artifacts 17

2.2 P300 Speller Brain-Computer Interface 20

2.2.1 System Operation 20

2.2.2 Impact of Electrode Type on Performance 22

2.2.3 Bayesian Dynamic Data Collection 26

2.3 Visual-Motor Skill Learning 29

2.3.1 Mobile Brain/Body Imaging 31

2.3.2 Biomarkers of Skilled Performance 33
2.3.3 Biomarker Identification 36
2.3.4 Motion Trajectory Prediction 38
2.3.4.1 Controller Motion 39
2.3.4.2 Recursive Bayes Filters 41
2.3.4.3 Gaussian Process Regression 44
2.3.4.4 Recurrent Neural Networks 49
3 Improving the Performance of Dry Electrodes for the P300 Speller 54
3.1 Dynamic Data Collection with Wet and Dry Electrodes 55
3.1.1 Methods 55
3.1.1.1 Participants 55
3.1.1.2 Equipment 57
3.1.1.3 Software and Data Collection 58
3.1.2 Results 60
3.2 Identifying Artifacts with Anomaly Detection 67
3.2.1 Methods 68
3.2.2 Results 70
3.3 Discussion 73
4 Visual-Motor Skill Learning during Marksmanship Training in Virtual Reality 76
4.1 Brain Dynamics of Visual-Motor Skill Learning 78
4.2 Methods 79
4.2.1 Subjects and Equipment 79

4.2.2 Experimental Task 81
4.2.3 Performance Measures 84
4.2.4 Statistical Analysis and Outlier Removal 87
4.3 Results 88
4.3.1 Shot Accuracy and Positional Error 88
4.3.2 Reaction and Trigger Response Times 91
4.3.3 Movement Kinematics 93
4.3.4 Neural Dynamics 96
4.3.4.1 Preparatory Phase Spectral Power 96
4.3.4.2 Visual Evoked Potential 97
4.3.5 Neural Correlates of Behavior 100
4.4 Discussion 103
5 Biomarker Classification 105
5.1 Methods 106
5.1.1 Preprocessing and Feature Extraction 107
5.1.2 Classification and Feature Selection 109
5.2 Results 111
5.2.1 Classifier Performance 111
5.2.2 Feature Selection 114
5.3 Discussion 117
6 Controller Motion Trajectory Prediction for Real-time Performance Estimation 119
6.1 Short-Term Prediction using Kalman Filters 121

6.2 Long-term Prediction using Gaussian Processes and Bayes Filters 127
6.2.1 Gaussian Process + Particle Filter 130
6.2.2 Gaussian Process + Extended Kalman Filter 134
6.3 End-to-End Long-Term Prediction using Recurrent Neural Networks 139
6.4 Discussion 146
6.5 Future Work 147
7 Conclusion 150
References 155
Biography 166

List of Tables

Table 1. Demographic information of study participants with ALS/PLS. 56

Table 2: Average performance in able-bodied participants. 61

Table 3: Average performance in ALS/PLS participants. 63

Table 4: Survey results for study participants with ALS/PLS. 64

Table 5: Summary of datasets used for evaluating classifier performance. ***n*** indicates the number of trials used in each classification task. 108

Table 6: The five kinematic features selected by the SFFS-NB algorithm for the skill level and shot success classification tasks. 115

List of Figures

Figure 2.1: Images of a reusable wet electrode cap. 11

Figure 2.2: Images of the g.Sahara active dry electrode cap. 13

Figure 2.3: An averaged P300 waveform, where a postive voltage component appears around 300ms after the presentation of a rare but relevant stimulus. 16

Figure 2.4: Maximum amplitudes for five common physiological signals plotted across frequency. 18

Figure 2.5: A P300 speller user is presented with a character grid on a computer screen from which characters may be selected to spell. 21

Figure 2.6: Comparison of able-bodied participants' performance between wet and dry electrode systems in the static data collection environment 25

Figure 2.7: A flowchart representing the dynamic stopping algorithm described in Throckmorton et al. [48]. 29

Figure 2.8: Example of an ROC curve. 37

Figure 2.9: Cartoon example of the spherical target (orange) and controller (blue) positions for a single frame captured during the simulated marksmanship. 40

Figure 2.10: Examples of the spherical target (orange) and controller (blue) trajectories (all frames included) displayed in spherical coordinates 41

Figure 2.11: A 1D example of a Kalman filter prediction. 43

Figure 2.12: A synthetic 2D example of a trained GP motion model for trajectory prediction. 48

Figure 2.13: Example of the RNN structure. 50

Figure 2.14: Example of the LSTM cell architecture. 51

Figure 3.1: Screen captures of the character grids used in this study. 59

Figure 3.2: Comparison of able-bodied participants' performance with wet and dry electrode systems in the dynamic data collection environment 60

Figure 3.3: Comparison of ALS/PLS participants' performance with wet and dry electrode systems in the dynamic data collection environment..... 62

Figure 3.4: Comparison of P300 signal characteristics between wet and dry electrode systems recoded over electrode location Cz in able-bodied participants..... 65

Figure 3.5: Power spectral density (PSD) estimates for continuous P300 speller EEG data. 66

Figure 3.6: Number of possible artifacts detected during the 3.5 s inter-character EEG segments with an autoregressive model for different thresholds..... 71

Figure 3.7: Number of possible artifacts detected during the 3.5 s inter-character EEG segments with wavelet analysis for different thresholds..... 72

Figure 4.1: Participant performing the simulated marksmanship task in the Duke immersive Virtual Environment (DiVE), a six-sided CAVE-like VR system..... 80

Figure 4.2: Electrode montage consisting of the 13 EEG electrodes and 2 HEOG electrodes used for the study, placed according to the 10/20 International 81

Figure 4.3: (Left) The six target trajectories, shown slightly off-center to improve visibility, showing the orange spherical target in flight (all frames included) 82

Figure 4.4: Data collection for a single trial. The shaded regions indicate which data stream was analyzed at different times during the trial. 84

Figure 4.5: Movement of the firearm controller displayed for a single trial, time-locked to the target launch at 0 s..... 86

Figure 4.6: Accuracy differed across trajectories and with practice. 89

Figure 4.7: A significant interaction between day and horizontal direction was observed, where greater improvements occurred in the left and right horizontal trajectories 90

Figure 4.8: Shot error, expressed as the Euclidean distance from the center of the target, differed with practice and across trajectories. 91

Figure 4.9: Average reaction time of the controller displayed across blocks and split by day..... 92

Figure 4.10: Trigger response times varied across trajectories and with practice. 93

Figure 4.11: Peak speeds of the controller differed across blocks and trajectories. 94

Figure 4.12: A tradeoff occurred between the durations of the ballistic phase and refinement phase. 95

Figure 4.13: Beta band power during the preparatory phase significantly increased in channels Cz, P3, and Pz across blocks on Day 1 and 2. 97

Figure 4.14: Grand average VEPs for the left and right horizontal directions of the target trajectories, time-locked to the target launch at time 0 ms. 98

Figure 4.15: Topographic maps of the VEPs at 20 ms time steps displayed for the left target trajectories (top) and right target trajectories (bottom). 99

Figure 4.16: Significant decreases in VEP amplitude were observed across days in channels P3 and P4. 100

Figure 4.17: Scatter plots comparing the preparatory phase beta band power in channels Cz, P3, and Pz (x-axis) to behavior (y-axis) in terms of reaction time in the left column101

Figure 4.18: Scatter plots comparing the contralateral VEP amplitude (x-axis) to behavior (y-axis) in terms of reaction time in left column and trigger response time 102

Figure 5.1: Comparison of classifier performance in terms of AUC. 112

Figure 5.2: Comparison of average classifier performance in terms of AUC using randomly reallocated labels 113

Figure 5.3: Classifier performance during SFFS-NB. 114

Figure 6.1: Example of long-term controller motion prediction using a Kalman filter .. 124

Figure 6.2: The performance measure $\boldsymbol{q}$ was computed as the cumulative distribution of the prediction (large purple sphere) bounded by the faces of a cube whose center 125

Figure 6.3: Long-term controller prediction performance with the Kalman filter. 126

Figure 6.4: An example of a trained Gaussian process motion model for the controller trajectories in the upper left target trajectory direction. 129

Figure 6.5: Example of long-term prediction using a GP-PF with 20 particles, where the predicted subtrajectory started 60 samples (1 s) before the trigger pull. 132

Figure 6.6: Comparison of long-term controller trajectory prediction performance using the GP-PF (dotted line) and Kalman filter (solid line). 133

Figure 6.7: Example of long-term controller prediction using the GP-EKF 136

Figure 6.8: Comparison of long-term controller trajectory prediction performance of the GP-EKF (dashed line), GP-PF (dotted line), and Kalman filter (solid line). 137

Figure 6.9: Example of a single long-term controller prediction using the LSTM 141

Figure 6.10: Prediction error was calculated as the Euclidean distance between the controller's true position (blue) and the LSTM prediction (purple). 143

Figure 6.11: Within-subject comparison of long-term controller motion prediction for the GP-EKF posterior mean (blue line) and the LSTM (red line) in the upper left 144

Acknowledgements

The completion of this doctoral dissertation would not have been possible without the support of several people, to whom I am forever grateful for their scientific guidance and encouragement.

I would first like to thank the faculty members at North Carolina State University who inspired me to pursue an advanced degree in engineering. I am especially grateful to Dr. Michael Escuti for his guidance on my undergraduate research project and for teaching me how to present my work in an eloquent way.

I would like to gratefully acknowledge the various organizations that funded this research: the National Science Foundation's IGERT Program, the National Institutes of Health, and the Department of Electrical and Computer Engineering at Duke University. I would also like to acknowledge those who participated in my research experiments, as this work would not have been possible without their cooperation and patience.

I must express my sincere gratitude to my brilliant advisors, Drs. Leslie Collins and Gregory Appelbaum, for their mentorship throughout my time as a PhD student. Leslie, your scientific insight and guidance has given me the confidence to pursue my dream of using my skills in engineering to tackle problems in the field of health science. Thank you for admitting me into the program, believing that I could succeed, and supporting me along the way. Greg, your enthusiasm is contagious and kept me going

when I felt discouraged. Thank you for sharing your immense knowledge of neuroscience with me and helping me to become a better writer and researcher.

To my incredible mentors, Drs. Chandra Throckmorton and Boyla Mainsah, thank you for continuously keeping me on track to graduate. Sandy, thank you for maintaining a cheerful attitude while answering an infinite amount of questions as I struggled through my first year in graduate school. Boyla, our discussions and your creative suggestions were instrumental in my success as a PhD student. Thank you for always pushing me to do my best, meticulously editing my documents, and maintaining a calm demeanor that kept me on an even keel.

In addition to my mentors, I was fortunate to work with and learn from a group of students and faculty members from a variety of disciplines: Kevin Caves, Elayna Kirsch, Regis Kopper, Loren Nolte, Marc Sommer, Hrishikesh Rao, Stacy Tantum, and David Zielinski. Thank you all for sharing your knowledge and experience with me. I'm grateful for the constant exchange of ideas that helped build this document.

I was also lucky to be a member of a research lab at Duke that consisted of incredibly talented engineers. Ken Colwell and Nick Czarnek, thank you for helping me get started when I first arrived. Dima Kalika, thank you for keeping me up to date on the latest Reddit news related to obscure sports and machine learning. Lidea Shahidi, your optimism and boundless curiosity continue to inspire me. I will miss our morning coffee breaks and having someone clap after presentations in meetings. Kedar Prabhudesai,

thank you for always being willing to chat about the gamut of machine learning methods you've dabbled in over the years. To the other current and past members of the lab, Kevin Chu, Bohao Huang, Jordan Malof, Daniel Reichman, Joe Camillo, Jill Desmond, Sara Duran, Achut Manandhar, Rayn Sakaguchi, and Patrick Wang, thank you for providing feedback on my work and creating a comfortable environment that was conducive to learning.

I am forever grateful to my family and friends who were helpful in numerous ways. To my wonderful parents, Nancy and Trevor Clements, thank you for inspiring me to pursue a meaningful career. Mom, your success as professional engineer solving problems in the pharmaceutical industry was the impetus for my entrance into the field of engineering. After your ALS diagnosis, our bonding experience over your participation in a brain-computer interface study was the reason I applied to join the AML Lab at Duke. It is my hope that I will one day inspire young women in engineering as I have been inspired by you, Leslie, and Boyla. Dad, thank you for being my lifelong role model and providing me with an endless amount of encouragement and advice over the years. I could not have made it to this point without your generosity with your time and resources. I don't know how to thank you enough for helping me get to where I am today. To my older brother, Ben, and sister-in-law, Cady Mack, thank you for all the laughs during times of high stress. There's nothing quite like a good cat meme to express one's true feelings about any situation. To my grandparents, Almeda, Bob, Cindy, and Jack,

thank you for sharing the wisdom you've gained from your own endeavors to help me on this journey. To my step family, Cindy, Amie, Anna, and Matt, thank you for the inspirational pep talks that always seemed to make me feel better. To my parents-in-law, Beth and Chris Casper, thank you for your support and kind words of encouragement that were motivation to keep going.

Finally, I'd like to thank my dear husband, James Purcell. Jimmy, you are my pillar of strength during times when I feel weak. You always know how to make me laugh, whether it's dancing in the kitchen or inventing lyrics to songs. Marrying you was one of the best decisions I've ever made. Thank you for giving me the courage to go back to school and for not letting me quit when things got tough. I admire your strong drive and desire to accomplish great things. I'm excited to see what we can accomplish together in the years to come.

1 Introduction

Physiological processes generated by humans, such as brain activity and motor behavior, produce electrical and mechanical signals that can be captured by placing sensors on the body and analyzed to assess human function. These signals carry important information about the state of a specific biological system and are used in a variety of tasks, including medical diagnosis and human-computer interaction. The work presented in this document encompasses two distinct research projects on different applications of physiological data from human subjects. The overarching goal of both projects was to utilize novel signal processing and machine learning techniques for physiological-based analyses to produce and define meaningful results under the specified hypotheses. The objective for the first project was to improve the overall design of a brain-computer interface (BCI) system by creating a novel hardware/software architecture that integrated an innovative data collection algorithm with existing, easy-to-mount hardware for recording brain activity on the scalp using electroencephalography (EEG). In the second project, EEG was synchronized with motion tracking in virtual reality (VR) and analyzed to assess the psychophysiological mechanisms underlying visual-motor skill during a full-body marksmanship task.

BCI systems allow a user to control a device without the need for voluntary muscle control by translating encoded information in EEG signals into meaningful commands. While head and eye tracking systems provide alternative methods for controlling a device,

these systems still require some muscle function and can fail when this function is lost or cannot be sustained for effective device control (e.g., [1]). Thus, BCIs may provide communication assistance for people with severe neuromuscular disabilities, such as those with amyotrophic lateral sclerosis (ALS), that limit the utility of other communication aids due to their reliance on physical activity. However, BCIs are rarely used outside of a laboratory setting. One of the main challenges for BCI implementation is the weak signal-to-noise ratio (SNR) associated with the EEG signals that are used for BCI control. EEG data recorded from the scalp contain significant amounts of noise that arise from the surrounding environment and physiological activities that are unrelated to the neural signal of interest. Therefore, to address this low SNR issue, most BCIs rely on gel-based "wet" electrodes to lower the impedance between the electrodes and the skin. However, wet electrode systems generally take a substantial amount of time and expertise to set up. For people with severe disabilities who need to use the system daily, minimizing this set-up time is highly desirable.

As an alternative to wet electrode systems, dry electrode systems are more user-friendly and much faster to set up. Typical dry electrodes contain several metal pins with sufficient length to reach through hair to the skin on the scalp. However, without a layer of conductive gel to keep the electrode surface in good electrical contact with the skin, dry electrodes are more susceptible to external noise such as that induced by movement [2]. While hardware improvements have been explored in previous work, the performance of

dry electrode BCI systems remains inferior to the performance of wet electrode BCI systems (e.g., [3]–[5]). Therefore, in this work, an algorithmic solution (rather than a hardware solution) was investigated as a method to compensate for the lower SNR associated with dry electrode systems in an effort to raise their performance to the same level as that seen for wet electrode systems. Noise detection algorithms were also explored to substantiate the hypothesis that poor dry electrode system performance was the result of external noise and to help determine if the system would benefit from additional types of signal processing to further improve performance. This BCI research is essentially complete, and is in press (see [6]), although future directions for the research are suggested.

A broader application of EEG is in the field of cognitive and behavioral neuroscience. While neuroscience research can inform the development of BCI, it can also be used to gain a more complete picture of the psychophysiological processes underlying motor control during physical activity. Therefore, in the second part of this research, a novel experimental framework was created that fused EEG and motion tracking to probe the neural and kinematic underpinnings of visual-motor skill learning during a simulated marksmanship task in virtual reality (VR). Visual-motor skill refers to the ability to coordinate visual perception with fine motor control. These skills play an integral role in all human action, from everyday functional tasks such as writing or tying shoelaces to more complicated tasks with smaller margins for error such as professional sports or

surgery. Thus, there is a strong interest in understanding the psychophysiological mechanisms that enable visual-motor skill learning, which has the potential to inform the design of training programs to increase their effectiveness. However, most experimental protocols to study visual-motor skill for tasks that would naturally involve a dynamic target, such as catching a ball or shooting a moving target, have traditionally been modified to a static exercise task. While these modifications are needed to provide more experimental control in a laboratory setting and reduce noise in EEG recordings, static exercises provide an incomplete picture of the full cascade of the psychophysiological processes involved with performing a realistic motor task.

To overcome the limitations imposed by static EEG experiments, a new method known as mobile brain/body imaging (MoBI) was proposed by Makeig et al. [8] that uses a combination of EEG and motion sensors to investigate the links between brain and behavior as humans move naturally through an environment. In this work, the MoBI framework was extended to include fully immersive VR, which allows individuals to experience realistic simulations in a controlled 3D setting while full-body motions are tracked with sub-millimeter precision and at relatively high temporal rates. This fusion offers a powerful testbed for naturalistic neuroscience research that can be used to gain a better understanding of the psychophysiological processes associated with skilled motor control that occur during dynamic tasks.

The dynamic task selected for this study was marksmanship. Marksmanship is particularly useful for studying visually-guided movement because it can be tightly constrained in both space and time, produces discrete measures of performance (i.e., the bullet either hits or misses a target), and yet still requires the visual-motor skills that demand high mental and physical coordination. Participants in this study performed a simulation of a dynamic marksmanship task, modeled after Olympic Trap Shooting standards and conducted in a fully immersive virtual environment. Over three days of training, participants practiced shooting a simulated clay pigeon that was launched from behind a trap house using a firearm game controller while synchronized recordings of EEG and 3D motion were collected.

In order to investigate the psychophysiological mechanisms that enabled visual-motor skill learning during this task, three approaches to EEG and motion signal analysis were explored. In the first approach, biological markers (*biomarkers*) describing the psychophysiological processes that transpired during the task were extracted from the EEG and motion tracking data to analyze the within-subject changes that occurred through practice and correlated with improvements on the task (parts of this work were published in [9]). The second approach utilized these biomarkers in a machine learning classification algorithm to assess the combinations of biomarkers that were most predictive of marksmanship performance on a trial-by-trial basis. Finally, in the third approach, time-series techniques were explored for estimating single-trial success using

the raw motion tracking data for controller motion trajectory prediction. Given the repetitive and sequential structure of the motion signals collected during the experiment, the goal of this approach was to model movement patterns in the data that could be used to predict whether the firearm controller's virtually projected ray would intersect with the simulated target pigeon before the shot was taken. Taken as a whole, these three approaches demonstrate a first step towards the development of data-driven tools for psychophysiological signal analysis that can be used to characterize visual-motor skill and predict performance during dynamic tasks for simulation-based training in VR.

The structure of this document is as follows. Chapter 2 presents an overview of EEG and introduces the two application areas discussed in this document: a specific BCI system, known as the P300 speller, and the study of visual-motor skill in VR. Chapter 3 evaluates the performance impact of dry electrodes for use with the P300 speller and presents an empirical study to characterize the differences between EEG signals recorded with wet and dry electrodes. Chapter 4 introduces a novel experimental framework to study visual-motor skill learning during a dynamic marksmanship task in VR using a combination of EEG and motion tracking data. Chapter 5 investigates classification and feature selection as a method for the discovering biomarker combinations that are predictive of marksmanship performance. Chapter 6 explores approaches to modeling controller motion trajectories for real-time prediction of marksmanship success, along

with avenues for future work. Chapter 7 provides a summary of the primary contributions of this research.

2 Background

Due to its flexible design and sensitivity to millisecond changes in cognitive processing, electroencephalography (EEG) is perhaps the most commonly used method for recording brain activity. This chapter provides a brief overview of EEG and introduces two of its applications, the P300 speller BCI and the study of visual-motor skill in VR. These two applications are the subject of my research and are explored more deeply in the subsequent chapters of this document. In addition to EEG, motion signals were also investigated in the study of visual-motor skill in VR and a short summary of time series methods for motion signal processing is presented at the end of this chapter.

2.1 Electroencephalography

Communication between neurons in the brain can be measured as electrical activity on the scalp using EEG. The first human EEG was recorded in 1929 by Hans Berger [10]. This historical breakthrough produced a low-cost, non-invasive neurological tool for measuring cognitive information using small sensors (electrodes) placed in electrical contact with the skin on the scalp.

Although EEG continues to be the most widely used method among the different brain recording techniques, it suffers from poor spatial resolution. Due to the resistive layers of fluid, skull, skin, and hair between the brain and the electrodes, it is difficult to determine the exact location of electrical activity within the brain (often referred to as the inverse problem [11]). However, the high cost and physically large size of other non-

invasive recording devices that do not suffer from poor spatial resolution, such as magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) systems, make them infeasible for many uses. MEG, which measures magnetic fields produced by electrical currents in the brain, is also less sensitive to radial and deep sources in the brain than EEG due to the intense folds on the brain's surface [12]. Furthermore, indirect non-invasive recording modalities such as fMRI, which measures changes in blood flow that are associated with brain activity, are temporally limited by the speed of the underlying hemodynamic or metabolic processes, which evolve over seconds and are much slower than brain processes that unfold on a millisecond timescale [13]. Invasive modalities, such as electrocorticography (ECoG), require implanting micro-electrodes inside the skull and involve a significant amount of health risks (e.g., intracranial hemorrhage or infection [14]). Therefore, despite its limited spatial resolution, EEG remains one of the most valuable tools for a myriad of brain-related research experiments because it is non-invasive, readily accessible, inexpensive, and provides temporal resolution on a millisecond timescale.

2.1.1 Signal Acquisition

An EEG signal represents the potential difference between a scalp electrode and a designated reference electrode, usually placed over a non-cortical site (e.g., a mastoid). A ground electrode is then used to measure the voltage difference between the scalp and reference electrodes. Therefore, the minimal configuration of an EEG measurement must

consist of at least one scalp electrode, one reference electrode, and one ground electrode. Multiple electrode configurations can comprise up to 256 scalp electrodes when placed according to the International 10-20 electrode placement standard [15].

To perform a conventional EEG recording, a cap or headband containing multiple electrodes is typically placed on a user's scalp and conductive gel is applied to each electrode to ensure low skin-electrode impedance. The skin-electrode interface serves as a transducer, where an ionic current flow within the firing neurons of the brain becomes a small electronic flow within the electrode; thus, keeping the impedance level between the skin and electrode low (< 10kΩ) is recommended for recording high-quality signals [16]. Many EEG systems also benefit from skin abrasion below each individual electrode to further reduce impedance levels. The electrodes are connected to an amplifier that magnifies the microvolt-scale amplitudes of the analog EEG signals. An A/D converter then digitizes the amplified signals so that they can be stored and displayed on a personal computer (or other relevant device).

2.1.2 Electrode Type

Proper coupling of the skin and electrodes is essential for obtaining high-quality EEG data with minimal noise interference. In this section, the advantages and disadvantages of the two most common electrode types, gel-based (wet) and dry, and their skin coupling medium variations are discussed.

2.1.2.1 Wet Electrodes

Gel-anchored or "wet" electrodes are the most commonly used type of electrodes and are often considered to be the gold standard for EEG recording. Wet electrodes can be either reusable or disposable.

Reusable wet electrodes rely on an electrolytic gel that permeates through the hair to form a stable, low-impedance path between the electrode and the skin. These electrodes are usually made of silver-silver chloride, but other conductive metals such as gold or stainless steel can also be used [17]. An example of an EEG cap containing multiple reusable wet electrodes is shown in Figure 2.1. In addition to the application of gel, skin abrasion is often performed directly below each electrode to remove highly resistive parts of the epidermis.

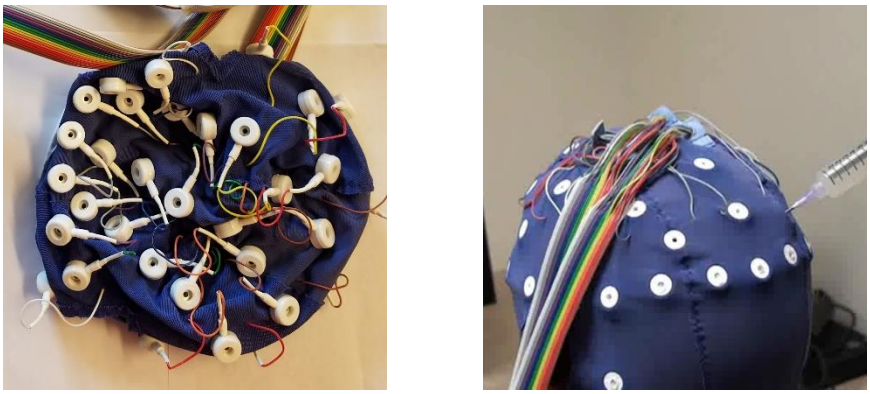

Figure 2.1: Images of a reusable wet electrode cap. (Left) The inside of the cap. (Right) Gel being applied to an electrode using a syringe.

Pre-gelled disposable electrodes are typically shaped like large flexible disks with adhesive glue on one side and a metal snap on the other, which connects to the lead wires

of an amplifier. While disposable electrodes are quick and easy to apply, they are limited to recording areas with little or no hair and, therefore, are typically used in conjunction with reusable electrodes that can penetrate areas on the scalp containing hair.

Although wet electrodes often produce high-quality EEG recordings, their preparation is time-consuming and wearing the cap for an extended period of time can be uncomfortable, which may frustrate users who would like to use the system daily. Another disadvantage to using wet electrodes is that the conductive gel can dry out in a matter of hours, disrupting long-term recordings until the gel is replenished. As a result, the use of dry electrodes that do not require skin preparation or gel is appealing to many users.

2.1.2.2 Dry Electrodes

The basic design of a dry electrode is based on a set of pins with sufficient length to make electrical contact with the scalp through hair (see Figure 2.2). Without a conductive gel present, the skin-electrode contact impedance typically ranges from 100 - 200kΩ [18]. These high impedance levels create large voltage drops across the skin-electrode interface, resulting in lower SNRs for the recorded voltage signals. In order to reduce current flow and thereby reduce this voltage drop, an ultra-high input impedance amplifier is typically directly connected to each electrode, creating what is known as an "active" electrode.

While active dry electrodes may reduce certain types of noise (i.e., non-physiological signals such as power line noise and cable movement), recordings from dry skin are still subject to increased physiological noise sensitivity due to the lack of a stabilizing gel layer that buffers the electrodes against movement [2]. Relative motions of the electrodes with respect to the body give rise to artifacts in the received signal. Thus, dry electrodes in their current state are unlikely to replace wet electrodes without further development, either in signal processing or hardware design, due to their reduced signal quality.

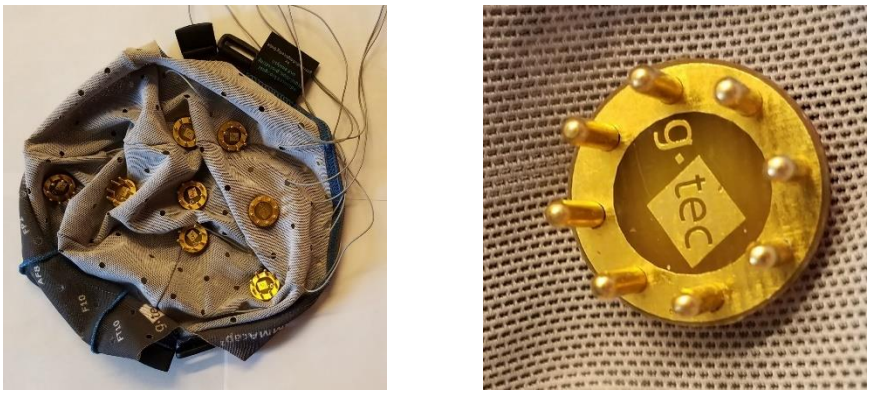

Figure 2.2: Images of the g.Sahara active dry electrode cap. (Left) The inside of the electrode cap. (Right) A close up of a single electrode.

Other variations of dry electrodes, such as needle and non-contact dry electrodes, also exist. Micro-fabricated needle electrodes penetrate the high-resistance layer of the skin to achieve electrical contact comparable to wet electrodes [19]. However, irritation and pain have been observed when using these electrodes. Non-contact dry electrodes avoid direct skin-electrode coupling altogether and operate via capacitive coupling.

Requiring zero preparation, the signal on the scalp is capable of coupling with the metal portion of the electrode through hair and clothing. While non-contact EEG electrodes have been shown to successfully record certain visual evoked potentials, the signal quality is significantly reduced when compared to dry contact electrodes [20]. Currently, motion artifacts appear to be an insurmountable problem for non-contact capacitive sensors and, therefore, "dry electrodes" in this document only refers to the dry contact electrodes.

2.1.3 Signal Characteristics

EEG signals are typically defined by their amplitude, frequency, shape, and location on the scalp [21]. Amplitudes vary from a few microvolts to 100μV, but can be much higher when corrupted by non-cerebral artifact activity. The frequency content of EEG ranges from .01 to 100 Hz, which is commonly divided into five bands: delta (.01 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 13 Hz), beta (13 – 30 Hz) and gamma (30 – 100 Hz). These bands are often associated with a certain stage of consciousness. For example, alpha waves normally appear with closed eyes or a lapse of attention, and are reduced with increased alertness [10].

In addition to frequency content, event-related potentials (ERPs) are another well-studied measure of brain activity that can be recorded by EEG. ERPs are cerebral waveforms that contain stable temporal relationships with a definable event (e.g., the presentation of a stimulus), and consist of a series of negative and positive voltage deflections that are often referred to as peaks or components (see [11] for a review). The

nomenclature for these voltage peaks uses the letters “P” and “N” to denote positive and negative polarity, followed by a number that indicates a peak's temporal position within the waveform. The amplitude of a peak is usually only a few microvolts and therefore, these waveforms are easily obscured by noise such as non-relevant brain activity, physiological signals, or external environmental interference. Thus, it is common practice to collect several instances of the ERP signal and average them together to increase SNR [22]. It is also common to apply baseline correction, which subtracts the voltage measured during a period before the event from the entire waveform so that the signal only reflects the voltage relative to the pre-event voltage.

An important mid-latency ERP component that has received a great deal of attention in EEG research is the P300. The P300 is a positive voltage peak that occurs around 300ms after a rare but relevant stimulus is presented and is speculated to reflect stimulus evaluation or categorization. Figure 2.3 illustrates an example of an averaged P300 component signal. A broad range of uses for the P300 exists, which includes lie detection [23], BCI control [24], and dementia diagnosis [25]. The latency (delay between stimulus onset and peak voltage) and amplitude characteristics of ERP components such as the P300 are thought to represent the processing of a stimulus in combination with higher processes such as attention or memory [26].

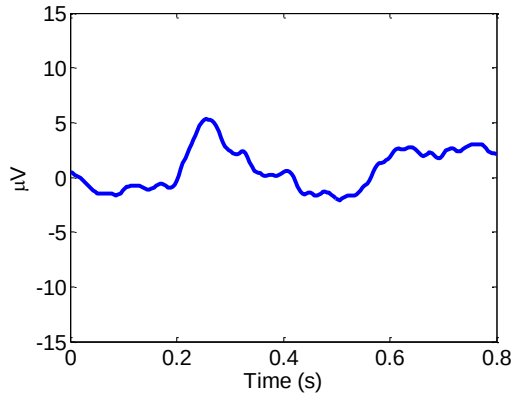

Figure 2.3: An averaged P300 waveform, where a postive voltage component appears around 300ms after the presentation of a rare but relevant stimulus.

Another important ERP in EEG research is the visual evoked potential (VEP), which is measured over the posterior EEG channels and typically contains the N100 ERP component (a negative voltage deflections that occurs around 150 ms post-simulus). Previous research has shown that VEPs reflect early visual processing of a target stimulus, which can be modulated by attention [27], [28]. In EEG studies that involve human learning, such as those that include training on visual search tasks, changes in the VEP over the course of the experiment suggest that learning is accompanied by reorganizations in the neural processes that contribute to improvements on the task [29]–[31].

In this work, P300 signals were used for control in a BCI system and VEPs were used to assess visual processing during marksmanship training in VR. However, due to the low amplitudes of these ERPs, unwanted components of EEG (artifacts) can easily

corrupt how this information is inferred. Thus, in order to successfully interpret and analyze the EEG data, careful consideration of artifacts in the recordings is required.

2.1.4 Physiological Artifacts

Contamination by artifacts is a well-known problem for both clinical and experimental EEG. Any recorded signal that is not related to the brain signal of interest can be termed as an artifact. EEG artifacts are divided into two categories: *physiological* and *non-physiological*. *Non-physiological* artifacts arise from external sources such as power lines and electrical equipment. *Physiological* artifacts are the result of body activity; ocular, muscular, and cardiac signals are among the most common types of physiological artifacts. While EEG systems are susceptible to both physiological and non-physiological artifacts, non-physiological interference can be easily reduced by recording in controlled areas, shielding cables, and properly attaching electrodes.

However, physiological artifacts can rarely be avoided and are difficult to filter as they often lie within the same frequency range as desired brain activity [32]. Figure 2.4 shows examples of the frequency ranges and maximum amplitudes for physiological artifacts whose frequency ranges are known to overlap with EEG signals. Ocular activity, such as eye movement and blinking, often appears in the frontal channels with amplitudes above 100μV, although they may extend into other locations on the scalp as well. Perspiration artifacts also appear at low frequencies and consist of longer waveforms that slowly shift the electrical baseline over durations of time. Large head and cable

movements can also cause artifacts to appear in the recording at lower frequencies, even if all of the electrodes are in good electrical contact with the skin. Muscle contractions, such as chewing, give rise to faster recurring potentials and repetitive muscle activities that can generate fast muscle and slow movement artifacts. Cardiac-related potentials, which usually appear over the posterior head region, can also be picked up by EEG below 100 Hz.

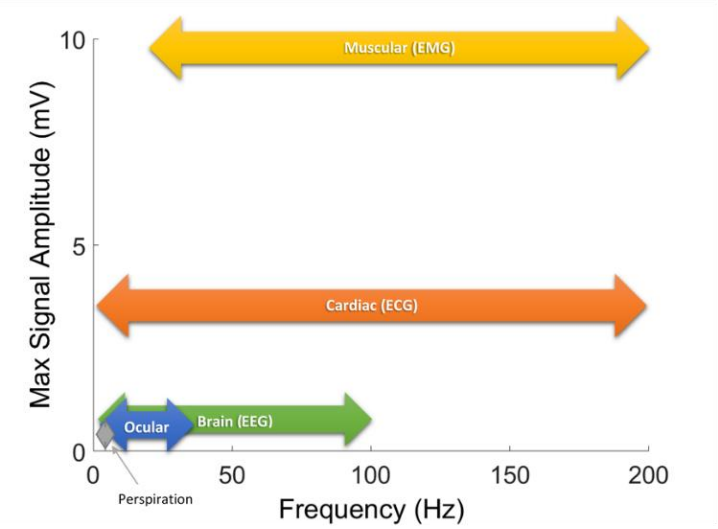

Figure 2.4: Maximum amplitudes for five common physiological signals plotted across frequency. The frequency bands of artifacts and brain overlap and, therefore, simple bandpass filters are not completely effective.

Mitigating artifacts includes methods for both *rejection* and *removal* (see [33] and [34] for a review). Artifact *rejection* methods, which use predefined criteria to identify the artifact and remove the entire segment from the EEG recording, are useful for occasional interference; however, it is often desirable to retain as much data as possible. Artifact

removal techniques seek to eliminate only the contribution from the interfering source. While many artifact removal techniques exist, the literature is very broad and researchers do not seem to agree on a standard method. Uriguen et al. [33] suggested that the lack of standardized approaches for artifact removal are due to at least three factors: (1) the wide variability in EEG and artifact signal characteristics, (2) public datasets are scarce and, hence, comparisons between methods differ among publications, and (3) performance measures are not consistent from one paper to another, making fair comparisons difficult to carry out. Novel methods for artifact removal are often extensions or combinations of the standard existing approaches including regression, filtering, and blind source separation. However, even the most popular techniques do not always completely separate cerebral and artifact sources (e.g., independent component analysis in [35]) and performance varies according to the EEG signal type, recording electrodes, and artifacts present.

In the first part of this work, a novel method to mitigate the effects of artifact contamination in noisy dry electrode recordings was investigated that uses data-driven Bayesian inference for online EEG data collection. Rather than rejecting or removing artifacts, this method improved SNR by automatically determining the amount of data to collect online using a probabilistic approach. These improvements were illustrated with a P300 speller BCI.

In the second part of this work, EEG was fused with motion tracking using body-mounted sensors to record movement-related activity during a task that was designed to include stationary and non-stationary periods of EEG recording. Given information about movement-related activity, EEG analyses were performed in a calculated way that avoided incorporating large artifacts.

2.2 P300 Speller Brain-Computer Interface

The P300 speller is an EEG-based BCI communication system whose operation relies on eliciting ERPs to enable a user to spell text on a computer letter by letter. The name P300 speller refers to the reliance on the P300 component to determine the BCI user's intent, although other ERP components may also be used for classification [36]. A standard method for eliciting the P300 component is the oddball paradigm, where two stimuli are presented in a random series with one of the two occurring infrequently [37]. In 1988, Farwell and Donchin used the oddball paradigm to create a novel BCI spelling device, and the P300 speller was established [24].

2.2.1 System Operation

To operate a P300 speller, users are first outfitted with an electrode cap for recording EEG signals. In a typical visual P300 speller, a computer screen containing a two-dimensional array of characters is presented to the user (see Figure 2.5). To spell a character, the user focuses on the desired target character as groups of characters are randomly flashed together. Characters are flashed in groups to increase the character

presentation rate. When the target character is flashed, since it is rare, a P300 component is elicited in the EEG response signal. A flash sequence occurs when all of the defined flash groups have been exhausted once. Due to the low SNR of a single ERP, several flash sequences are completed to collect multiple repetitions of the P300 component that can be averaged to improve the SNR. The speller assumes that the character the user intends to spell is the character corresponding to the response signals containing the largest P300 component.

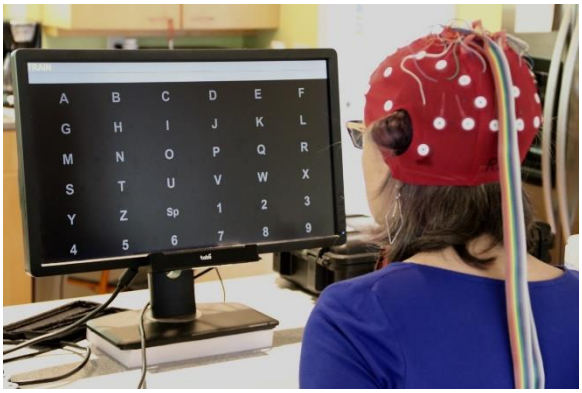

Figure 2.5: A P300 speller user is presented with a character grid on a computer screen from which characters may be selected to spell.

BCI systems such as the P300 speller need to be calibrated for a specific user due to variations in brain structure and neural responses. To calibrate the P300 speller, the user is asked to copy each letter exactly as it appears at the top of the character matrix in order to collect labeled training data. Users do not receive feedback during calibration. The training data are used to develop a learning algorithm that classifies future online EEG recordings to determine the likelihood of whether a flash response signal contains

the P300 component. Following calibration, BCIs are typically tested online using single-word copy-spelling tasks. During copy-spelling, the user is once again asked to copy each letter exactly as it appears at the top of the character matrix and is now provided with real-time feedback of the selected character (i.e. the character with the highest score from the trained classifier).

The performance of the P300 speller during online testing is typically evaluated using measures such as spelling accuracy, chance level performance, and communication rate. Accuracy is defined as the percentage of characters spelled correctly during a copy-spelling session; anecdotally, an accuracy of > 70% is typically considered the standard for a user being able to spell accurately. Chance level accuracy is the measure of whether or not a certain level of spelling accuracy was achieved through random character selection [38]. Bit rate is a measure of communication systems that incorporates accuracy, speed, and the number of selectable characters presented on the screen [39].

2.2.2 Impact of Electrode Type on Performance

Despite significant improvements in BCI performance over the last decade, the P300 speller remains primarily a research device rather than a home-based communication aid [40]–[42]. One limitation that inhibits widespread use is the complex and time-consuming process of applying conventional wet electrodes for EEG recording. In order to reduce skin impedance and ensure electrical contact with the scalp, wet electrodes require the application of conductive gel to each electrode and abrasive skin

preparation. Depending on the number of electrodes used, this process can take ten minutes or more and varies with regard to the ability and experience of the person setting up this system. For people with severe communication disabilities and their caregivers who would like to use the system daily, minimizing this set-up time is highly desirable [40].

In an attempt to reduce the required set-up time, the use of dry electrodes for P300-based BCIs has been explored ([3]–[5], [43]). Dry electrodes are much faster and easier to use, as they do not require the application of gel to each electrode. However, due to poor skin-electrode contact, dry electrode systems are much more sensitive to physiological artifact noise than wet electrodes. Signal quality can have a significant impact on BCI performance. The lower SNR can decrease the detectability of the ERPs used for BCI control, which may negatively affect both the speed and accuracy of the system.

Several studies have evaluated the performance of dry electrodes in comparison to wet electrodes for use with P300-based BCIs in able-bodied participants with mixed results. In a comparison of a hybrid dry electrode sensor array (HESA) system to a conventional wet electrode system, Sellers et al. [3] observed higher P300 speller accuracy with the wet electrode system for four out of eight participants, although the differences in accuracy were not statistically significant. Zander et al. [4] showed that the amplitude and temporal structure of the grand average P300 signal did not differ significantly between their novel dry electrode system and a conventional wet electrode system, but

offline analysis revealed significant differences existed between the classification accuracies. Grozea et al. [5] developed a novel dry electrode system using bristle sensors, but observed significant differences between the P300 signal characteristics recorded with their system and a conventional wet electrode system. A study performed by Guger et al. [43] compared P300 speller accuracies with their commercially available g.Sahara dry electrode system to accuracies obtained in a previous study [44] with a conventional wet electrode system. The wet electrode study reported an average accuracy of 91.0% ± 18.5 across 81 participants while the average accuracy in the dry electrode study was 90.4% ± 17.2 across 23 participants. However, although no significant difference was observed between the accuracies across participant pools for the two systems, intra-participant variation across the two participant pools may have been a confounding factor.

To determine whether similar wet and dry electrode system accuracies could be achieved using the same participant pool, a comparative analysis study was performed by the Brain-Computer Interface Laboratory at East Tennessee State University. This study compared the P300 speller performance of a conventional wet electrode system and the g.Sahara dry electrode system in sixteen able-bodied participants by collecting the same amount of data using both electrode types (termed *static* data collection). The performance results, measured by spelling accuracy and bit rate, are shown in Figure 2.6. Assuming a binomial distribution on the probability of correctly selecting a set of characters [38], chance level accuracy was 11.1%.

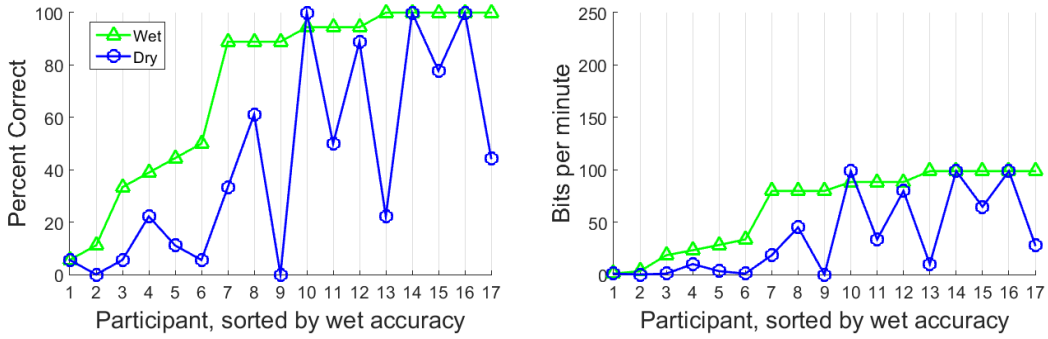

Figure 2.6: Comparison of able-bodied participants' performance between wet and dry electrode systems in the static data collection environment for (a) spelling accuracy, in percent correct, (b) bit rate, in bits per minute, and (c) the amount of data collected. Pairwise significance tests revealed the difference in performance between the electrode systems were statistically significant for both performance measures ($p < .001$).

Twelve out of the sixteen participants in this study performed worse with the dry electrode system when compared to the wet electrode system. Two Wilcoxon signed-rank tests revealed that the differences in performance between the dry and wet electrode systems were statistically significant for both accuracy and bit rate ($p < .001$). This suggests that improvements, either in signal processing or in hardware design, may be required to bring dry electrode systems to the same level of performance as wet electrode systems.

Dry electrodes would greatly simplify the set-up process for P300 speller systems. However, as demonstrated by the results in Figure 2.6, the low signal quality of dry electrode system recordings results in poor system performance, which can lead to user and caregiver frustration. One reason for poor performance may be due to the fact that dry electrodes are more sensitive to physiological artifact noise than wet electrodes.

In the standard static data collection environment, the P300 speller collects a predefined number of repetitions of the EEG response signal and averages them to improve the SNR of the ERP. This number is held constant across participants and characters spelled. All of the "wet versus dry" electrode comparison studies mentioned previously relied on static data collection for which the same amount of EEG data was collected from both electrode systems. However, given the poorer quality of the dry electrode recordings, these systems may have benefited from additional data collection to improve the SNR. In the next section, a novel algorithm for controlling the amount of EEG data collected online is discussed that provides the flexibility to adjust the number of repetitions collected based on the quality of the incoming signals (termed *dynamic* data collection).

2.2.3 Bayesian Dynamic Data Collection

Several methods for adaptively controlling data collection in a P300 speller have been proposed in literature (e.g., [45]–[48]). However, many of these methods depend on some form of past performance of the participants. For example, Serby et al. [45] and Lenhardt et al. [46] set a threshold for stopping data collection using training data averaged across a participant pool, which may not translate if the participant pool is changed. Jin et al. [47] bypassed this problem by using participant-specific stopping criteria, but also made the assumption that participant performance remains relatively constant throughout a spelling session. In contrast to these studies, Throckmorton et al.

[48] avoided depending on past performance by developing a probabilistic dynamic data collection algorithm that automatically determined the amount of data to collect on a subject-by-subject and character-by-character basis based on the quality of incoming data. If data were of poor quality, additional data was collected to improve accuracy, and if data were of high quality, the amount of data collection was reduced to increase spelling speed. This algorithm resulted in significant improvements in accuracy and spelling speed for able-bodied participants with a wet electrode system. These results were validated in a separate study with participants with amyotrophic lateral sclerosis (ALS), one of the target populations for P300 spellers [49].

The amount of data collected by the dynamic stopping algorithm is controlled by a threshold on the probability that each character in the grid is the target character (c^*). The initial probability of each character being the target is set to 1/N (where N is the total number of characters in the grid), i.e., no a priori knowledge is assumed. For each new EEG response signal ($\boldsymbol{x_j}$) obtained after a character group flash ($\boldsymbol{S_j}$ contains the characters in a group), a classifier score is calculated ($y_j = \boldsymbol{w} * \boldsymbol{x_j}$, where $\boldsymbol{w}$ represents the trained classifier weights) and stored in $\boldsymbol{Y_j} = [y_1, \dots, y_{j-1}, y_j]$. The probability for each character (c_n) in group $\boldsymbol{S_j}$ is then updated using the likelihood estimates from the training data (obtained from the probability density functions for the target, $p(y_i|H_1)$, and non-target, $p(y_i|H_0)$, response signals). The update equation is based on Bayes rule and implemented as

$$P(c_n = c^* | \boldsymbol{Y_j}, \boldsymbol{S_j}) = \frac{P(y_j | c_n = c^*, \boldsymbol{S_j}) P(c_n = c^* | \boldsymbol{Y_{j-1}}, \boldsymbol{S_{j-1}})}{\sum_{i=1}^{N} P(y_j | c_i = c^*, \boldsymbol{S_j}) P(c_i = c^* | \boldsymbol{Y_{j-1}}, \boldsymbol{S_{j-1}})} \quad (\textbf{2-1})$$

where $P(c_n = c^* | \boldsymbol{Y_j}, \boldsymbol{S_j})$ is the posterior probability of a character c_n being the target character, $P(y_j | c_i = c^*, \boldsymbol{S_j})$ is the likelihood of the classifier score given that c_i was/was not in group $\boldsymbol{S_j}$,and $P(c_i = c^* | \boldsymbol{Y_{j-1}}, \boldsymbol{S_{j-1}})$ is the prior probability of a character c_i being the target character.

The complete algorithm is summarized by the flowchart provided in Figure 2.7 (adapted from [48]). The top row demonstrates how the likelihood estimates are obtained using the classifier scores from the labeled training data. The scores are first grouped by whether a target was present/absent and the likelihood probability density functions are then created using kernel density estimation. The bottom row demonstrates the online implementation for dynamically controlling the collection of test data. Initially, a uniform distribution is assumed for each character in the grid. With each new flash, a new classifier score is obtained and the probabilities for the characters in this flash group are updated with the associated target likelihood estimate. The probabilities for the characters outside of the flash group are simultaneously updated with the associated non-target (target absent) likelihood estimate. After the character probabilities are updated, they are then compared to a threshold. If one of the character probabilities exceeds the threshold, that character is selected as the target character to be spelled. If not, a new flash is presented and the updating process is repeated.

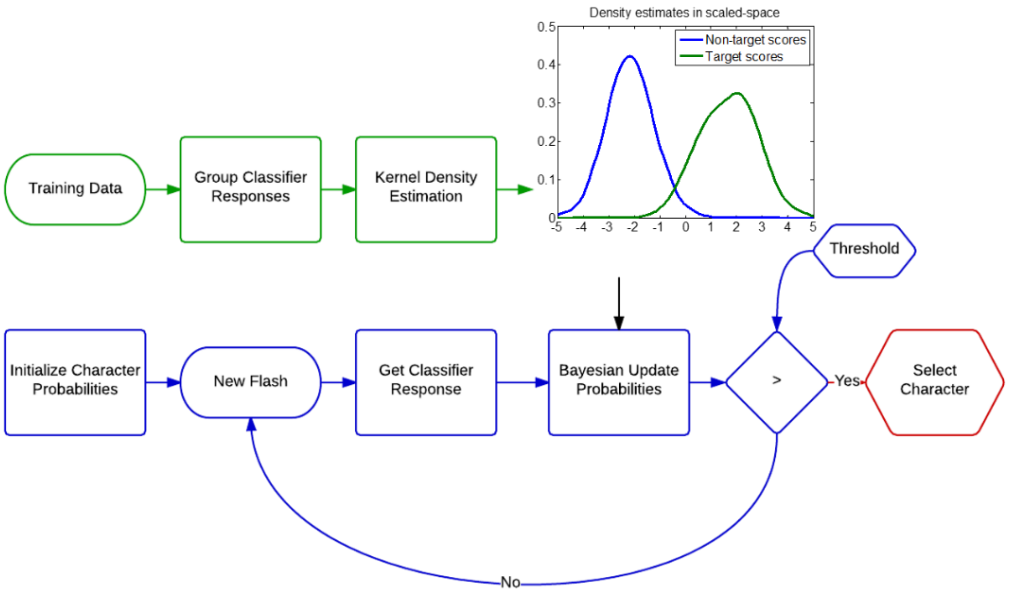

Figure 2.7: A flowchart representing the dynamic stopping algorithm described in Throckmorton et al. [48]**. The top row represents the offline calibration needed to obtain likelihood density estimates from labeled training data, while the bottom row shows how these estimates are used during online processing to dynamically contol data collection.**

In the next chapter, the use of this algorithm was investigated as a possible solution to help make dry electrodes in a P300 speller system feasible as a home-based communication aid. Additionally, an analysis of the differences in signal characteristics between the wet and dry electrode recordings was also conducted to help elucidate any observed performance results.

2.3 Visual-Motor Skill Learning

EEG has many uses for the scientific investigation of brain function. In the second part of this research, a fusion of EEG and motion tracking was used as a tool to study the ability of humans to integrate visual perception with motor control (i.e., visual-motor

skills) during a full-body marksmanship task in VR. Visual-motor skills play a key role in all human action, from everyday functional tasks such as handwriting to more complicated tasks such as competitive sports. Given the importance of these skills, there is a profound interest in understanding the psychophysiological changes accompanying skill learning during complex motor tasks. Knowledge such as this has the potential to provide useful information to inform the design of training programs, such as those for athletics, to accelerate skill learning.

Previous research has demonstrated that visual-based tasks require cognition that interacts with perception and motor action (e.g., [29]). Recording the cognitive processes involved in performing a visual-motor task requires a portable modality with sufficient temporal resolution to capture brain activity on the millisecond timescale of motor behavior, making EEG the most suitable choice. However, due to the large amount of motion artifacts recorded during active behavior, traditional EEG experiments to study visual-motor ability typically restrict participants to performing minimally-active motor responses (e.g., a button press or joystick control) to suddenly presented stimuli on a computer screen [7]. Given that much of human cognition is tightly coupled to our environment and physical actions, these experiments may not be capturing important brain activity involved with performing a more complex real-world motor task. Thus, the study of the brain processes supporting motor function requires a new approach to experimental protocols and data analyses. To overcome the restrictions of established

approaches, researchers have introduced a new technique known as Mobile Brain/Body Imaging (MoBI), which uses EEG in combination with motion sensors to study brain activity while users physically interact with their environment [8].

2.3.1 Mobile Brain/Body Imaging

The goal of MoBI research is to investigate the links between distributed brain dynamics and motivated natural behavior using synchronized recordings of EEG and movement tracking. Making use of these multiple data streams, this approach no longer considers movement-related activity as artifacts but as aspects of cognitive activity associated with a motor task.

Previous MoBI studies have focused on recording brain activity during stereotyped movements (e.g., walking [50]–[53], rhythmic finger movements [54], cycling [55]), where recurrent patterns of movement allow for motion artifacts to be extracted and then subtracted from the raw EEG data. While these studies clearly show that brain activity can be analyzed during repetitive movement, it is also important to investigate brain activity during nonstereotyped movements, which can be used to determine how much traditional EEG experiments that restrict participants' natural movements deviate from experimental results in actively moving participants.

A recent study by Jungnickel and Gramman investigated brain dynamics during a volatile nonstereotyped pointing movement [56]. Using a visual oddball paradigm, participants in the study were required to react to color changes in a moving target with

either a button press or by physically pointing towards the target. The results revealed faster reaction times and significantly higher P300 amplitudes in the EEG data for target stimuli in the pointing condition when compared to the button press condition, suggesting that differences in the brain dynamics may exist for various behavioral states. Jungnickel and Gramman demonstrated for the first time that MoBI is feasible for recording cognitive processes during nonstereotyped movement, which could generalize to a wide range of motor tasks involving natural interaction with a dynamic system, such as those involved in sporting events.

Relationships between brain activity and motor proficiency have been reported for sports-related tasks such as marksmanship [57]–[61], golf putting [62], [63], and archery [64]–[66] using EEG. However, these tasks are typically performed as self-paced tasks that require minimal movement to acquire a static target. Tasks that require larger natural movements to engage with a dynamic target (e.g., catching a ball, trap shooting) introduce additional brain processes that allow for the integration of proprioception (awareness of the position of one's body) and the vestibular system (sensory system that provides input about one's body movement and orientation in space) [56]. Thus, EEG experiments that limit mobility may not be capturing the full cascade of cognitive processing involved with performing a real-world, nonstereotyped motor action.

In this research, the MoBI framework was extended to include immersive VR to study visual-motor skill for nonstereotyped movement during a multi-day dynamic

marksmanship training regimen. This task was modeled after the Olympic Trap Shooting event, where shooters must prepare their gun for the launch of a clay pigeon and, upon release, track the moving target before pulling the trigger to intercept it as it moves away from them. The trap shooting task is particularly useful for studying visual-motor learning because it produces discrete measures of performance (participants either hit or miss the target) while still requiring proficient hand-eye coordination. In order to capture complex movements such as aiming during dynamic shooting, simulation technologies such as immersive VR provide precise position and orientation tracking that can easily be fused with EEG that creates a powerful testbed for naturalistic neuroscience research. Properly equipped VR systems also offer flexible control of an indoor 3D environment that allows for the easy modification of task parameters. Additionally, training in VR has been shown to translate improvements of motor skill to improvements in the real-world (e.g., [67]).

2.3.2 Biomarkers of Skilled Performance

One of the main objectives of this research was to identify the predictive EEG and kinematic biological markers (*biomarkers*) of performance for the simulated marksmanship task. The term biomarker commonly refers to a measurable indicator of a specific biological state, which can be used to describe the psychophysiological processes underlying visual-motor skill learning in this study.

Previous studies on marksmanship have identified several potential psychophysiological biomarkers that differentiate between novice and expert marksmen; however, the focus of this literature was primarily on static marksmanship, where participants fired at fixed targets at predetermined distances [57]–[61]. In these studies, the pre-shot routines of expert marksmen were characterized by an increase in EEG alpha power (8 – 12Hz) in the seconds leading up to the trigger pull, indicating that expert marksmen have reduced cortical activation when compared to novice marksmen. For marksmanship studies involving a dynamic target, only kinematic behavioral differences between experts and novices were reported [68], [69]. Expert marksmen exhibited significantly faster eye reaction times, lower peak velocities for gun barrel movements, shorter durations between peak velocity and the shot, shorter total movement times, and smaller displacement of the gun.

While all of these previous studies demonstrated that certain elements of preparatory attention and motor behavior are enhanced in expert shooters, much less is known about development of visual-motor skill and expertise as a function of training. Although there has been a growth in interest and research to gain a better understanding of the basic psychophysiological mechanisms that drive skill learning during dynamic tasks (e.g., [70]–[74]), these studies typically involve minimally-active laboratory tasks where the same principles may not apply to motor learning in more complex situations. Thus, one of the goals of this research was to gain a more complete picture of the

systematic changes in the psychophysiological mechanisms that accompany performance improvements through the practice of the full-body dynamic marksmanship task.

This work builds upon a past study by Rao et al. [75], which demonstrated that performance improvements on a similar marksmanship task in VR were correlated with significant changes in the movement kinematics over a single day of practice. Participants in the study exhibited longer, slower, and more precise ballistic movements (i.e., the large initial controller movement to advance towards the general vicinity of the target) that replaced effort spent on movement corrections and refinement before the trigger was pulled to shoot. While the discovery of these kinematic changes during early learning were interesting in themselves, the findings also indicated that it may be possible to shed light on their neural basis through the addition of neurophysiological monitoring with EEG.

Therefore, in this study, a fusion of EEG and motion tracking in VR was used to study the relationships between the brain and body while participants performed the simulated marksmanship task over three days of practice. The addition of EEG recording in the VR training environment added a new dimension into the investigation of biomarkers that reflect motor skill learning. With this novel framework, potential biomarkers were extracted from EEG and motion tracking data that reflected the temporal evolution of pre-launch preparatory brain oscillations, neural response in the visual

hemisphere induced by the launch of the target stimulus (VEPs), and kinematic response that transpired over the course of a single trial.

2.3.3 Biomarker Identification

In recent years, there has been an increasing interest in the identification of biomarkers to assess health and performance in sports-related research [76], [77]. Validated biomarkers can aid coaches, athletes, and researchers in the design of training programs that can increase their effectiveness. As such, one of the key objectives in this research was to develop and identify biomarkers that were predictive of success in the simulated marksmanship task.

In general, there are two main methods for identifying biomarkers. The first method utilizes univariate statistical hypothesis tests, such as the repeated measures analysis of variance (ANOVA), to identify biomarkers that provide a significant amount of information in isolation from other biomarkers. However, it is expected that most biological processes involve complex multivariate interactions that univariate techniques might miss. Therefore, another method for biomarker identification is to utilize a machine learning classification approach that highlights multivariate combinations of predictive biomarkers.

A binary classifier discriminates between two classes of features (i.e., variables or biomarkers that describe the observed data). The classifier is trained with a set of features thought to represent a set of unseen test features with the same distribution. In order to

evaluate how well the classifier differentiated between the two classes, the receiver operator characteristic (ROC) curve is typically used. ROC curves are a method for comparing the sensitivity (i.e., the number of true positive predictions divided by total the number of true positive and false negative predictions) to the specificity (i.e., the number of false positive predictions divided by the total number of false and true negative predictions) across all possible decision thresholds. An example of an ROC curve is shown in Figure 2.8. A summary of the ROC curve can be obtained by computing its area under the curve (AUC) and used as a single measure of classifier performance. Higher AUCs (i.e., closer to 1) indicate better prediction performance.

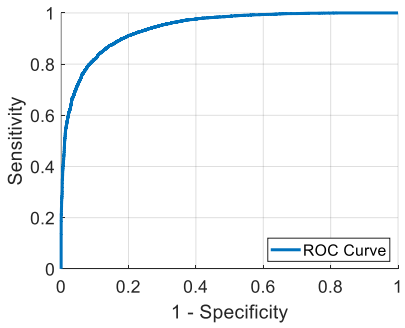

Figure 2.8: Example of an ROC curve. The curve represents a tradeoff between sensitivity, or probability of true positive, and specificity, or probability of false positive.

Biomarker classification is often preceded by techniques for feature selection to identify the combinations of biomarkers with the most predictive power. In this work, a greedy feature search method known as sequential forward floating selection (SFFS) was utilized to obtain a subset of features that is likely to contain the most predictive biomarkers [78]. Greedy search methods such as SFFS iteratively form a best feature

subset by maximizing the performance (e.g., AUC) of the chosen classifier. The SFFS algorithm starts from the empty set and adds a new best feature on each forward step. After each forward step, however, the algorithm performs a backwards step by subtracting the worst feature that least reduces performance. If subtracting the worst feature improves performance, this step is repeated. The addition of the backwards step reduces the chance of obtaining a locally optimum solution.

In Chapters 4 and 5, these methods are used to identify EEG and kinematic biomarkers that are predictive of performance in the simulated marksmanship task. However, while hand-crafted biomarkers such as these are easy to interpret, they often ignore the more complex spatiotemporal dependencies within the data. Therefore, in the next section, an alternative approach for predicting marksmanship success is introduced that uses time series methods to model the continuous motion patterns of the controller, which eliminates the need to manually design biomarkers that may fail to expose the more predictive underlying properties of the data.

2.3.4 Motion Trajectory Prediction

Time series prediction is the process of using a model to generate sequential predictions of future events based on known past events. Given the repeated temporal structure of the psychophysiological processes that transpired over the course of a trial during the marksmanship task, it may be possible to use time series prediction methods to estimate how a participant will react in the future. By adding order dependence to the

predictions using the time dimension, these methods have the potential to estimate whether the controller's virtually projected ray will intersect with the target well before they pull the trigger to shoot. A prediction model such as this could be used in a closed-loop feedback approach that alerts individuals in real-time when shooting might be suboptimal.

Several approaches to modeling the motion trajectories of moving objects have been reported for vehicles [79]–[82], robots [83]–[85], pedestrian paths [86]–[89], and hand movements in the form of computer mouse motion [90]–[92]. In Chapter 6, these methods were adapted to build 3D motion models of the firearm controller's motion trajectory during the simulated marksmanship task.

2.3.4.1 Controller Motion

The tracking system used in this study measured the 3D position vector, $\boldsymbol{p}_{sensor} = [x, y, z]$, and unit direction vector, $\boldsymbol{d}_{sensor} = [a, b, c]$, for the controller-mounted sensor during the simulated marksmanship task. These values were combined to create a virtually projected pointer that was used to aim and intersect targets throughout the task. The virtually projected ray was determined by

$$\vec{\boldsymbol{r}} = \boldsymbol{p}_{sensor} + \beta \boldsymbol{d}_{sensor} \qquad (\textbf{2-2})$$

where $\vec{\boldsymbol{r}}$ is ray and β is a scalar value that varies to represent points along the ray. In order to determine whether the target was hit or missed at the time of the trigger pull, the closest point on the ray to the center of the target was calculated as

$$\beta_{controller} = \left(\boldsymbol{p}_{target} - \boldsymbol{p}_{sensor}\right) \cdot \boldsymbol{d}_{sensor} \qquad (2\text{-}3)$$

$$\boldsymbol{p}_{controller} = \boldsymbol{p}_{sensor} + \beta_{controller}\boldsymbol{d}_{sensor} \qquad (2\text{-}4)$$

where $\boldsymbol{p}_{target}$ is the 3D position of the center of the target and $\boldsymbol{p}_{controller}$ is the closest 3D position on the ray to the center of the target. If the Euclidean distance between $\boldsymbol{p}_{controller}$ and $\boldsymbol{p}_{target}$ was less than or equal to 1 ft at the time of the trigger pull, the shot was considered a hit. An example of these positions for a single frame of the task is shown in Figure 2.9.

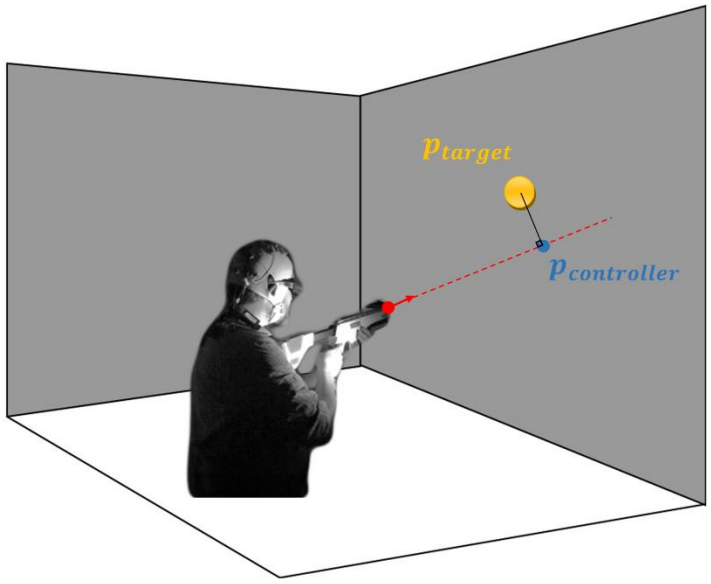

Figure 2.9: Cartoon example of the spherical target (orange) and controller (blue) positions for a single frame captured during the simulated marksmanship. The dashed red line represents the virtually projected ray emitted by the handheld firearm controller.

Throughout the remainder of this document, controller motion trajectories are illustrated using the closest position on the ray to the center of the target. Additionally, 3D Cartesian positions are mapped to 2D spherical positions (azimuth and elevation

angles) for visualization purposes only. Examples of the 2D target and controller trajectories are shown in Figure 2.10.

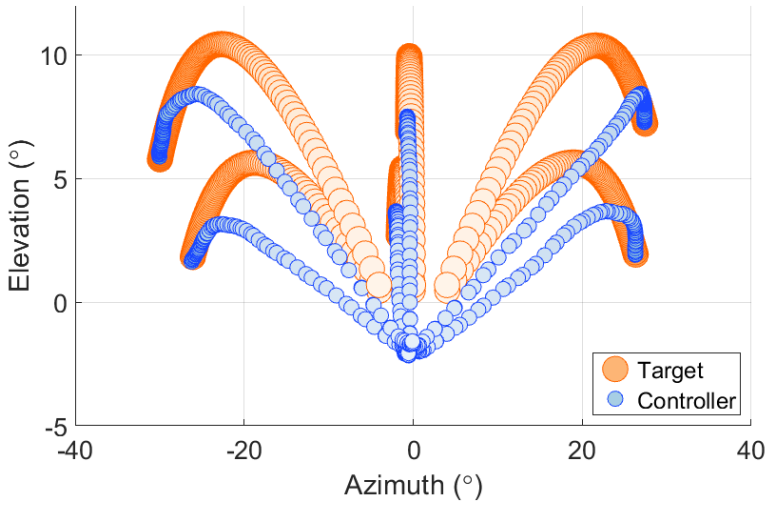

Figure 2.10: Examples of the spherical target (orange) and controller (blue) trajectories (all frames included) displayed in spherical coordinates. Each point represents data sampled at a time frame (~16.67ms). The color gradient of a trajectory changes from light to dark to represent the start and end of the trajectory, respectively.

In the following subsections, brief descriptions of the basis for each of the methods used to predict $\boldsymbol{p}_{controller}$ in Chapter 6 are provided.

2.3.4.2 Recursive Bayes Filters

A state is a group of numbers that describes the underlying configuration of a system (e.g., position and orientation). A general framework for recursively estimating the state of a dynamic system at time t using information about its previous state at time $t-1$ is known as the Bayes filter algorithm. Techniques for implementing the Bayes filter algorithm are commonly used to estimate motion trajectories in a variety of different

tracking problems, such as mobile robotics and target tracking in surveillance systems. For the simulated marksmanship task, controller motion at a discrete time step t can be estimated as a multivariate state, $\boldsymbol{s}_t$, defined by the controller-mounted sensor's 3D position (x_t, y_t, z_t) and orientation expressed as azimuth and elevation angles (θ_t, ϕ_t) along with its linear and angular instantaneous velocities.

One of the most widely used parametric implementations of the Bayes filter is the Kalman filter [93]. The Kalman filter is a relatively simple algorithm that applies a linear operator to the estimate of a current state to predict estimates of future states. The Kalman filter assumes these variables are normally distributed with a mean $\boldsymbol{\mu}$ (the expected value of the state) and a covariance matrix Σ (uncertainty). This assumption is essential because the application of a linear operator to a normally distributed state produces another normally distributed state, which allows the filter to operate recursively.

The Kalman filter algorithm is typically implemented in two distinct stages for prediction and measurement correction. In the prediction stage, the current state at time $t-1$, defined by $\boldsymbol{\mu}_{t-1}$ and Σ_{t-1}, is used to predict an estimate of the next state at time t. An example of the prediction stage is shown in Figure 2.11. In this figure, the current state is modified according to a linear motion model, which is represented in matrix form using state transition matrix F. The state transition matrix is determined using *a priori* knowledge of the linear equations that describe the system, such as the basic kinematic equation to describe objects in motion (e.g., $x_t = x_{t-1} + \Delta x_{t-1}$ where x is position and Δx

is instantaneous velocity). Unknown uncertainties from the environment (e.g., jerk movements between sensor measurements) are modeled by adding a process noise covariance matrix Q to the current state covariance Σ_{t-1}. In practice, the covariance values in Q are tuned to small values because selecting overly large values increases uncertainty in the prediction, which implies that the model produces poor predictions of the trajectory. The output of the prediction stage is the estimated state at the next time step t, where the new best estimate $\hat{\boldsymbol{\mu}}_t$ is a prediction made from the previous best estimate $\boldsymbol{\mu}_{t-1}$ and the new uncertainty $\hat{\Sigma}_t$ is a combination of the previous uncertainty Σ_{t-1} and uncertainty from the environment Q. Repeated application of the prediction stage allows for the predicted state to be calculated many time steps into the future and, for long-term prediction applications such as controller trajectory prediction, the measurement correction stage is ignored.

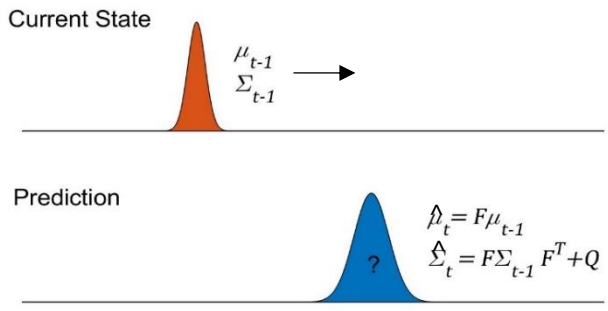

Figure 2.11: A 1D example of a Kalman filter prediction. Linear combinations of the transition matrix F with the mean and covariance of the current state (a 1D Gaussian defined by $\boldsymbol{\mu_{t-1}}$ and $\boldsymbol{\Sigma_{t-1}}$), plus some process noise Q, are used to predict the mean and covariance of the next state (another 1D Gaussian defined by $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\Sigma}}_t$).

Given that motion is rarely linear in practice, the linear transition assumption of Kalman filter often does not hold. Variations of the Kalman filter that relax the linear assumption include the extended Kalman filter and the unscented Kalman filter, which allow for nonlinear transition functions via linearization methods [83]. However, without an accurate transition model, inaccuracies and errors propagate in the predictions and produce poor estimates of the state as time progresses. Additionally, due to the complex nature of human motion, parametric models might be too restrictive when varying degrees of uncertainty are present in the data.

2.3.4.3 Gaussian Process Regression

A more flexible nonparametric method for modeling trajectories that relies on previously seen trajectories is Gaussian process (GP) regression [94]. A GP $f(\boldsymbol{s})$ is a collection of random variables where the marginal distribution over any finite subset has a joint Gaussian distribution. One can think of a GP as defining a distribution over functions that describe a set of data, where the properties of possible functions are fully defined by a mean $m(\boldsymbol{s})$ function and a covariance function $k(\boldsymbol{s}, \boldsymbol{s}')$ based on the following equations:

$$m(\boldsymbol{s}) = E[f(\boldsymbol{s})] \quad \textbf{(2-5)}$$

$$k(\boldsymbol{s}, \boldsymbol{s}') = E\big[\big(f(\boldsymbol{s}) - m(\boldsymbol{s})\big)\big(f(\boldsymbol{s}') - m(\boldsymbol{s}')\big)\big] \quad \textbf{(2-6)}$$

Assuming a normal distribution, the GP can be written as:

$$f(\boldsymbol{s}) \sim \mathcal{N}\big(m(\boldsymbol{s}), k(\boldsymbol{s}, \boldsymbol{s}')\big) \quad \textbf{(2-7)}$$

The output of the GP is a normal distribution that is described by $f(\boldsymbol{s})$ for some input $\boldsymbol{s}$. Moreover, the corresponding vector of outputs for any vector of inputs $[\boldsymbol{s}_1, \ldots \boldsymbol{s}_n]$ is a Gaussian $[f(\boldsymbol{s}_1), f(\boldsymbol{s}_2), \ldots, f(\boldsymbol{s}_n)] \sim \mathcal{N}(\boldsymbol{\mu}, K_{S,S})$, with mean $\boldsymbol{\mu}$ such that $\boldsymbol{\mu}_i = m(\boldsymbol{s}_i)$ and covariance matrix $K_{S,S}$ that satisfies $(K_{S,S})_{ij} = k(\boldsymbol{s}_i, \boldsymbol{s}_j)$.

In GP regression, the distribution over functions is used as a prior for Bayesian inference that can be converted into a posterior distribution over functions after some data are observed. The form of possible functions is determined through the choice of a kernel function for covariance function k and the hyperparameters within the kernel function. A kernel function is used because the posterior distribution is defined solely in terms of inner products in the input space, which means the input can be projected into a higher dimension feature space by replacing occurrences of those inner products with the kernel function $k(\boldsymbol{s}, \boldsymbol{s}')$. This technique can also lead to computational savings in situations where it is cheaper to compute then kernel than the explicit computation of the input vectors themselves.

Given a new test input, s_*, the posterior distribution f_* over the predicted values is obtained by conditioning on the observed training data where the inputs $S = [\boldsymbol{s}_1, \ldots \boldsymbol{s}_n]$ have scalar targets $\boldsymbol{y} = [y_1, \ldots, y_n]$:

$$f_* \mid s_*, S, \boldsymbol{y} \sim N(k(s_*, S)k(S, S)^{-1}\boldsymbol{y}, \\ k(s_*, s_*) - k(s_*, S)k(S, S)^{-1}k(S, s_*)^T) \qquad \textbf{(2-8)}$$

In this work, GP regression was used to model the controller trajectories using instantaneous velocity flow fields. The controller state at a discrete time point, $\boldsymbol{s}_t$, was defined by its position and orientation. This is different from the controller state defined in the previous section for the Kalman filter, which included velocity.

For this method, instantaneous velocity ($\Delta\boldsymbol{s}_t = \boldsymbol{s}_{t+1} - \boldsymbol{s}_t = (\Delta x_t,\ \Delta y_t,\ \Delta z_t,\ \Delta\theta_t,\ \Delta\phi_t)$) is treated as the target variable where each dimension is assumed to be independent. A Gaussian process prior was placed over $f(\boldsymbol{s})$ which lead to a prediction model with the following form:

$$p(\Delta x_t|\boldsymbol{s}_t, D_x) \approx N(GP_{\mu}(\boldsymbol{s}_t, D_x), GP_{\Sigma}(\boldsymbol{s}_t, D_x)) \qquad \textbf{(2-9)}$$

where the training dataset $D_x = \{\boldsymbol{s}_i = (x_i, y_i, z_i, \theta_i, \phi_i), \Delta x_i\}_{i=1\ldots n}$ represented a set of trajectories containing n position and instantaneous velocity observations, and similar but independent models were used to predict $\Delta\boldsymbol{y}, \Delta\boldsymbol{z}, \Delta\boldsymbol{\theta}$, and $\Delta\boldsymbol{\phi}$.

The squared exponential kernel function was used for the covariance function k (2-10), which enforces smoothness between neighboring inputs and outputs. The choice of this kernel for this application was intuitive because it is expected that neighboring points along the controller's trajectory leads to movement in similar directions. The squared exponential kernel function is defined by:

$$k(\boldsymbol{s}_i, \boldsymbol{s}_j) = \sigma_f^2\, exp\left(-\frac{1}{2}\left|\left|\boldsymbol{s}_i - \boldsymbol{s}_j\right|\right|_{\Lambda}^2\right) + \delta_{ij}\sigma_n^2 \qquad \textbf{(2-10)}$$

where the hyperparameters in the covariance function (σ_f, Λ, and σ_n) express prior information about the expected form of the motion model. Inferences about the

hyperparameters can be made from the observed training data. The log likelihood of the data given the hyperparameters for the observed training targets (e.g., $\Delta \boldsymbol{x}$) at inputs S can be expressed as

$$log\ P\left(\Delta \boldsymbol{x} \middle| S, \sigma_f, \Lambda, \sigma_n\right) = -\frac{1}{2}\Delta \boldsymbol{x}^T k(S,S)^{-1}\Delta \boldsymbol{x} - \frac{1}{2} log\ |k(S,S)| - \frac{n}{2}\ log\ 2\pi \quad \textbf{(2-11)}$$

The hyperparameters of the maximum log likelihood are found by optimizing this equation, which is typically done using standard optimization routines such as iterative conjugate gradients [95].

An example of a GP motion model trained using 2D trajectory data containing only (x, y) positions is shown in Figure 2.12, where the training trajectories start in the bottom left corner and moves upwards towards the right corner. The mean instantaneous velocity predictions obtained from the GP model follow the same direction as the training trajectories and uncertainty in the predictions (GP variance) increases further away from the training data.

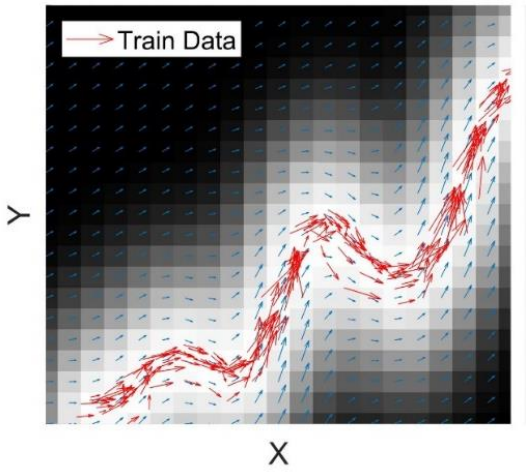

Figure 2.12: A synthetic 2D example of a trained GP motion model for trajectory prediction. The training trajectory data points are shown in red. The blue arrows correspond to the mean predicted instantaneous velocity and the shading represents the variance of the prediction, where darker shades indicate higher uncertainty.

However, GPs are limited to learning only pairwise correlations between the input observations, making it difficult to account for long-term dependencies and often ignoring more complex temporal structures in the data. Fusing GPs with Bayes filters has the potential to address this issue because Bayes filters can recursively propagate the predicted state estimation through the velocity flow field (e.g., [84], [86]). Therefore, in this work, combinations of GP regression and Bayes filters for long-term controller trajectory predictions were explored for the simulated marksmanship task, with the traditional Kalman filter model serving as the baseline to evaluate performance.

2.3.4.4 Recurrent Neural Networks

Another approach to learn nonlinear motion models from previously seen data is through recurrent neural networks (RNNs), which have seen a recent explosive growth in interest due to their success in applications with sequential data [96]. Like GPs, RNNs also use information about previously observed trajectories to make more informed predictions about future motion. One of the main advantages of RNNs over GPs is that they are able capture temporal dependencies between data points along the trajectory, which creates an end-to-end model that no longer requires the incorporation of recursive methods such as Bayes filters into their framework. RNNs operate in a similar sequential manner to Bayes filters, where predictions can be made at each time step using information from previous inputs and the current input. However, unlike Bayes filters, RNNs do not make assumptions about the type of motion, meaning they are able to capture both linear and nonlinear motion. Another difference between RNNs and GPs is that RNNs are discriminative models, meaning that the output is single value rather than a distribution.

The basic structure of an RNN is a unit that takes an input $\boldsymbol{s}_t$ at discrete time t and outputs $\boldsymbol{h}_t$ (see Figure 2.13). The information derived in the unit at time t is looped back into the unit at time $t + 1$, giving it memory of previous inputs.

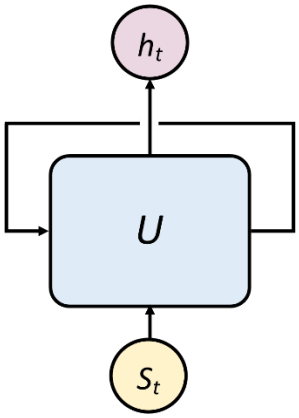

Figure 2.13: Example of the RNN structure. Computations derived by unit U from input s_t at discrete time t are output to h_t and fed back into the network at the next time step. This allows the network to store memory over time, which is ideal for modeling sequential tasks with long term dependencies.

The most common RNN is the long short-term memory network (LSTM). First introduced in 1997 by Hochreiter and Schmidhuber [97], LSTMs were explicitly designed to avoid a common long-term dependency problem with traditional RNNs known as the vanishing gradient problem. At the core of the LSTM is the cell state (C), which gives it the ability to learn and reproduce sequences over long periods of time. An example of an LSTM unit is shown in Figure 2.14. The cell state is a vector that allows past information to flow through the LSTM units with only minor linear interactions. The LSTM can add or remove information from the cell state using "gates" that consist of sigmoid functions (σ). Sigmoid functions map their input to a number between zero and one, which equates to how much information should be let through each gate.

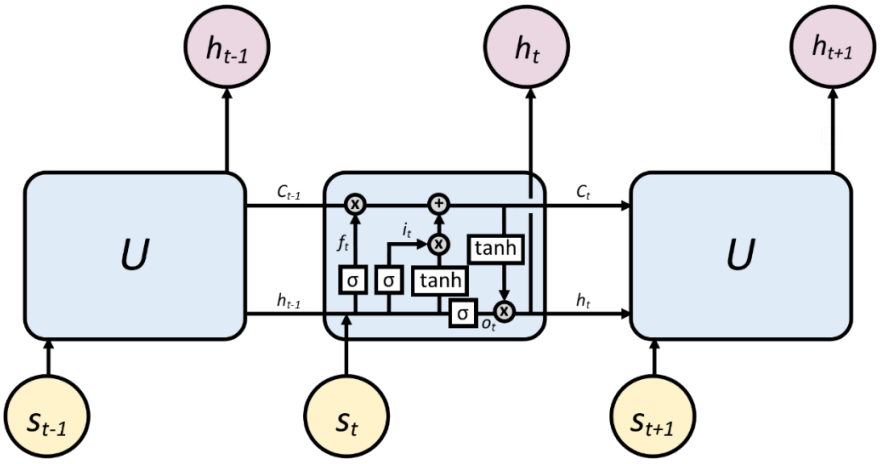

Figure 2.14: Example of the LSTM cell architecture. The cell state (*C*) and hidden state (*h*) are fed back into the unit at each time step.

A single LSTM unit consists of six layers: (1) the forget gate layer f, (2) the candidate state layer $\tilde{C}$, (3) the input gate layer i, (4) the output gate layer o, (5) the hidden state layer h, and (6) the cell state layer C. The equations for each layer are as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, s_t] + b_f\right) \quad \textbf{(2-12)}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, s_t] + b_C) \quad \textbf{(2-13)}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, s_t] + b_i) \quad \textbf{(2-14)}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad \textbf{(2-15)}$$

$$o_t = \sigma(W_o[h_{t-1}, s_t] + b_o) \quad \textbf{(2-16)}$$

$$h_t = o_t * \tanh(C_t) \quad \textbf{(2-17)}$$

The forget gate f_t takes the input into the LSTM (s_t) at time t, along with the hidden state from the previous h_{t-1} time step and decides whether to keep or forget the cell state C_{t-1} ($f_t = 0$: forget entirely, $f_t = 1$: keep entirely). The LSTM then decides what new information to store in the cell state by creating the candidate state $\widetilde{C_t}$. The sigmoid

function in input layer i_t determines how the cell state C_t is updated with the candidate state $\widetilde{C_t}$ ($i_t = 0$: don't update, $i_t = 1$: update). Finally, the hidden state h_t is updated according to o_t ($o_t = 0$: don't update, $o_t = 1$: update) with a filtered version of the cell state using the $tanh$ function to push the values to between -1 and 1. The cell state C_t and the hidden state h_t are then passed to the next time.

The weight vectors W, along with the bias vectors b, in the LSTM layer equations are determined during training using backpropagation through time (BPTT) [98]. BPTT is a method that modifies these vectors in order to minimize the error between the network outputs and the expected outputs from the training data. At a high level, this method propagates the training inputs through the network, calculates the accumulated errors across each time step, and adjusts the weights to minimize error. The BPTT algorithm repeats this method until a given stopping criterion is reached.

Within the last few years, LSTMs have been shown to successfully learn the properties of vehicle trajectories [81] and pedestrian [88] with minimal assumptions using what is known as the *encoder-decoder* LSTM [99]. The encoder-decoder LSTM is a model that involves mapping an input sequence to an output sequence using two separate LSTMs, one for encoding the input one step at a time and another for decoding the output. For example, Alahi et al. [88] showed that they could predict ("decode") a pedestrian's trajectory 4.8 s into the future using encoded information about how they had moved in the previous 3.2 s.

The first LSTM in the encoder-decoder LSTM framework encodes the input sequence into the final fixed-length hidden state vector h and cell state vector C. The length of both vectors is equal to the number of hidden units the LSTM's architecture, which are predefined and part of what determines the model's complexity. These vectors are then used to initialize the state and cell vectors of the second LSTM. This means that the length of the hidden state and cell state layers in both LSTMs must also be equal. The output of the second LSTM then passes through a fully connected dense layer to produce an outcome at each future time step. For a more detailed review of the encoder-decoder LSTM, the reader should refer to Cho et al. [99].

In Chapter 6, the performance of the encoder-decoder LSTM is compared to the performance of the GP-BayesFilter model for predicting future positions of the controller during the simulated marksmanship task, with the Kalman filter serving as the baseline. The goal of this work was to develop a model that estimated whether the controller's virtually projected ray would intersect with the simulated clay pigeon well before the shot was taken.

3 Improving the Performance of Dry Electrodes for the P300 Speller

While recent research on BCI systems such as the P300 speller highlights their potential to aid people with communication disabilities, their use remains almost entirely in the research laboratories and a majority of the work is comprised of data gathered from able-bodied participants [42]. Thus, design considerations that can arise when testing in people with communication disabilities may not be evident when testing in able-bodied individuals due to differences in user experience based on the level of disability. In a survey of 61 target BCIs users with communication disabilities, Huggins et al. [40] revealed that "simplicity of BCI setup" is the second most desired design feature by target end-users behind accuracy.

The time-consuming and complex process of traditional gel-based "wet" electrode application for EEG recording makes it difficult for BCIs to be utilized in daily life. Dry electrodes have an advantage over wet electrodes by providing quicker set-up time; however, the potentially poorer skin-electrode contact with dry electrodes can result in noisier recordings. Previous studies provide evidence that additional work, either in hardware design or signal processing, is needed before dry electrodes become a feasible alternative to wet electrodes for P300-based BCI systems (see Chapter 2 for a review).

In this chapter, the impact that additional noise due to the use of dry electrodes has on the P300 speller was examined. A signal processing approach for mitigation of these deleterious effects was investigated in both able-bodied participants and people

with ALS, who are among the targeted end-users of BCIs. The goal of this approach was to automatically detect and respond to the need for additional data collection in low signal-to-noise ratio (SNR) scenarios, such as those with dry electrode recordings. Finally, the differences in the noise characteristics between EEG data recorded with wet and dry electrodes were also analyzed to determine the types of signal processing most likely to be beneficial for further improvements in the development of dry electrodes for use with the P300 speller system.

3.1 Dynamic Data Collection with Wet and Dry Electrodes

The dynamic stopping algorithm proposed by Throckmorton et al. [48] utilized a probabilistic approach to online data collection in the P300 speller, which relied on the quality of the EEG data to control data collection. In this study, the dynamic stopping algorithm was assessed as a potential method to compensate for the lower SNR in EEG data recorded with a dry electrode system by automatically determining the amount of data to collect online in both able-bodied participants and participants with communication disabilities.

3.1.1 Methods

3.1.1.1 Participants

Ten able-bodied individuals from Duke University and eleven individuals with communication disabilities from the Duke ALS Clinic were recruited to participate in this study. Three participants with ALS were dropped from the study due to vision loss or

caregiver unavailability. Details of the study participants with communication disabilities used for data analysis can be found in Table 1. Participants in each experiment completed two P300 speller sessions; one using a wet electrode system and one using a dry electrode system. The two sessions were performed between one and 54 days apart for the able-bodied participants and between one and 95 days apart for the participants with communication disabilities. The electrode type used in the first session was counter-balanced across participants. Participants were asked to remove any jewelry or electronic devices that may interfere with the system and to avoid any type of movement during both sessions. This is imperative for recording with dry electrodes, which are more sensitive to these types of interference. P300 speller sessions with able-bodied participants were performed in a laboratory setting, while sessions with participants with disabilities were performed in either the Duke ALS clinic or the participant's home. All experiments were approved by Duke University IRB board.

Table 1. Demographic information of study participants with ALS/PLS.

Participant Number	ALSFRS-R	Experiment Location	Additional Information
1	4	Clinic	Movement via wheelchair, respirator-dependent, feeding tube
2	25	Home	Movement via wheelchair
3	20	Home	Movement via wheelchair
4	20	Home	Ambulatory
5	46	Home	Ambulatory
6	8	Home	Movement via wheelchair, respirator-dependent, feeding tube
7	39	Home	Ambulatory
8	41	Clinic	Ambulatory

3.1.1.2 Equipment

EEG signals were measured using electrodes positioned according to the International 10-20 system and connected to a computer via a 16-channel g.USBAMP amplifier. The signals were sampled at 256Hz and bandpass filtered (0.1–30 Hz). The dry electrode cap utilized the g.Sahara active dry system, which was comprised of 8-pin golden alloy coated electrodes with 7mm pin lengths (Figure 2.2). The wet electrode cap and electrolyte solution were purchased from Electro-Cap International. The dry electrode cap included a chin strap and the wet electrode cap did not. Three cap sizes (small, medium, and large) were available for both electrode types. Head circumference and the distance from the nasion to inion were measured to determine an appropriate cap size that applied gentle pressure to the electrodes on the scalp for each participant. Dry electrodes were turned back and forth a few times to remove any hair between the electrode and the skin, ensuring proper electrical contact. Grounding bracelets were worn by both the experimenter and the participant during dry electrode recordings to avoid electrostatic charges. Eight electrodes (Fz, Cz, P3, Pz, P4, P07, P08, and Oz), with ground and reference electrodes attached to each mastoid, were used for data collection and classification. These electrodes have been demonstrated to provide adequate information for P300 speller communication [100].

3.1.1.3 Software and Data Collection

The BCI2000 software package [101] was used for stimulus presentation and data collection, with additional functionality added for the application of the dynamic stopping algorithm. Able-bodied participants were presented with a 9 x 8 grid of characters on a computer screen while participants with communication disabilities were presented with a 6 x 6 grid (see Figure 3.1). These grids match the grids used in Throckmorton et al. [48] and Mainsah et al. [49] for participants without disability and with ALS, respectively. Both grids were flashed based on a checkerboard paradigm to eliminate errors resulting from the same character flashing twice in succession (double flash errors) and flashes of non-target characters near the target character that may attract users' attention (character adjacency-distraction errors) [102]. Each character in the grid was flashed twice in a sequence of 24 flashes. The flash duration (i.e., the character illumination duration) was 62.5ms followed by an inter-stimulus interval of 62.5ms, with an inter-character interval of 3.5s. EEG signals were segmented into epochs of 800ms following each flash and preprocessed according to according to Krusienski et al. [100]. These epochs were used as observations to train a classifier during the calibration phase and for character selection during online testing.

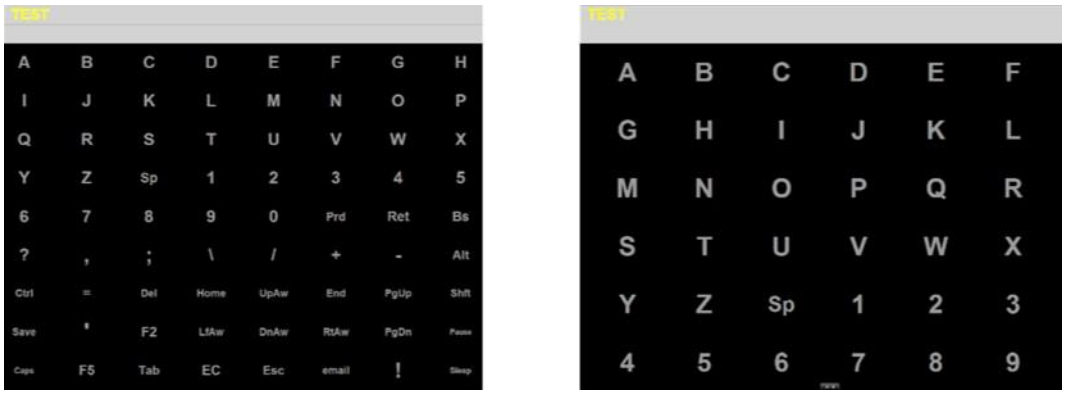

Figure 3.1: Screen captures of the character grids used in this study. (Left) The 9 x 8 grid used for able-bodied participants. (Right) The 6 x 6 grid used for participants with ALS.

Five calibration runs and five online test runs were recorded during both P300 speller sessions. For each calibration run, the participant was asked to copy-spell a six-character token randomly selected from the English language. EEG signals were collected for seven flash sequences, or 168 flashes (7 sequences x 24 flashes/sequence = 168 flashes), per character. After the fifth calibration run, stepwise linear discriminate analysis (SWLDA) was performed on the data to train a classifier for online testing. Although several learning algorithms are known to adequately classify P300 speller data, a statistical analysis of several classifiers performed by Krusienski et al. [103] indicated that the SWLDA classifier provides sufficient accuracy and the added complexity of nonlinear methods is not necessary.

For the five online test runs, the participant was once again asked to copy-spell a six-character token from the English language. However, instead of using a pre-set number of flash sequences for data collection, the dynamic stopping algorithm

automatically determined the necessary amount of data to collect for each target character. The amount of data collected was controlled by a threshold of 90% on the probability that each character in the grid was the target character. Data collection stopped once one of the character probabilities in the grid exceeded this threshold.

3.1.2 Results

Figure 3-2 shows the spelling accuracy, bit rate (related to spelling speed), and amount of data collected online for the able-bodied participants using the dry and wet electrode systems. The average performances are listed in Table 2. Assuming a binomial distribution on the probability of correctly selecting a set of characters, chance level accuracy was 10.0%.

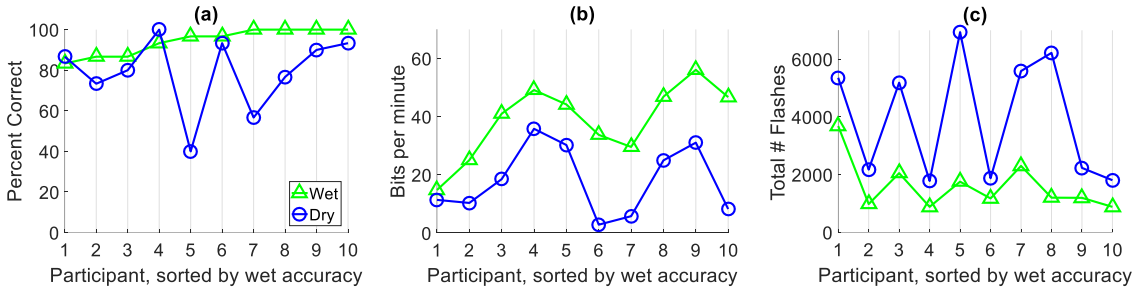

Figure 3.2: Comparison of able-bodied participants' performance with wet and dry electrode systems in the dynamic data collection environment for (a) spelling accuracy, in percent correct, (b) bit rate, in bits per minute, and (c) the amount of data collected. The differences in performance between the electrode systems were statistically significant for each measure ($p < .05$).

Table 2: Average performance in able-bodied participants.

	Accuracy (%)	Bit Rate (Bits/minute)	# Flashes
Dry System	79.0	17.9	3,915
Wet System	94.3	38.7	1,615

While the differences in accuracy between the wet and dry electrode systems appear small for the majority of subjects in Figure 3-2a (< 13%), they were statistically significant with lower dry system accuracy for eight out of the ten participants ($p < .05$). As hypothesized, the dynamic stopping algorithm increased the amount of data collected for the dry system across all participants (Figure 3-2c), indicating that the algorithm can indeed detect and respond to the need for additional data collection in low SNR scenarios. However, collecting more data did not increase the performance of the dry electrode system to the same level of performance as the wet electrode system. A significant difference between the systems was also observed for bit rate ($p < .01$), with lower bit rates occurring in the dry electrode system for all ten participants (Figure 3-2b).

Although performance for the dry electrode system remained lower than performance for the wet electrode system despite using dynamic data collection, the performance difference appears smaller than those observed using static data collection (see Figure 2.6 for the static data collection results). In order to compare the static data collection results to the dynamic data collection results across the two participant pools, a subset of participants was selected from the static data collection experiment based on

their performance with the wet electrode system. Eleven participants in the static data collection study had spelling accuracies above 80% with the wet electrode system, similar to all ten of the participants in the current dynamic data collection study. The average accuracy for this subset of participants with the wet electrode system was 95.5%, similar to the 94.3% average accuracy observed for the participants in the dynamic data collection study. However, the average accuracy with the dry electrode system for this subset participants was 61.6%; much lower than the 79% average accuracy observed in the current dynamic data collection study, suggesting that dynamic data collection may mitigate some of the lost performance associated with dry electrode recordings.

These results in Figure 3.2 were confirmed for participants with communication disabilities and are displayed in Figure 3.3 and Table 3.

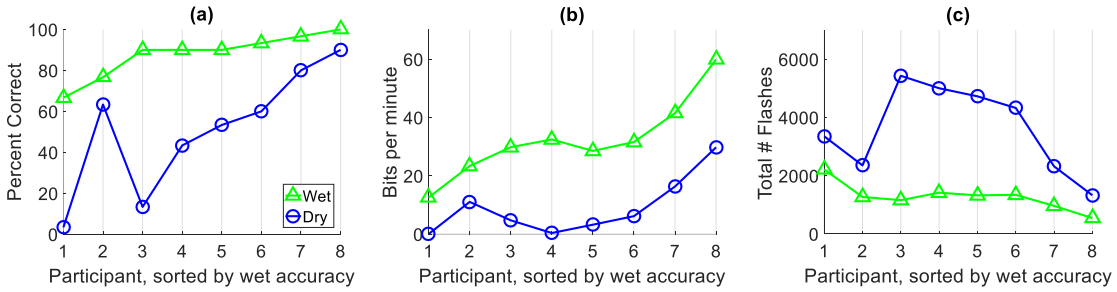

Figure 3.3: Comparison of ALS/PLS participants' performance with wet and dry electrode systems in the dynamic data collection environment for (a) spelling accuracy, in percent correct, (b) bit rate, in bits per minute, and (c) the amount of data collected. The differences in performance between the electrode systems were statistically significant for each measure ($p < .05$).

Table 3: Average performance in ALS/PLS participants.

	Accuracy (%)	Bit Rate (Bits/minute)	# Flashes
Dry System	50.9	8.9	3,604
Wet System	87.9	32.5	1,278

Once again, the differences in accuracy and bit rate were statistically significant ($p < .01$) and the amount of data collected increased across all participants in the dry electrode system when compared to the wet electrode system. This further implies that the dynamic stopping algorithm detected the need to collect more data to compensate for the low SNR in the dry electrode system recordings, but performance was still degraded when compared to the wet electrode system. Thus, improvements in signal processing or hardware design may still be required to bring dry electrode systems to the same level of performance as that observed with wet electrode systems.

Participants with communication disabilities were also given a short survey after both sessions. The survey consisted of six statements and participants were asked to rate each statement with a number between one (strongly disagree) and seven (strongly agree). The average responses, shown in Table 4, suggest that while the participants agreed that the setup of the dry electrode system was quick and not agitating, they also found the dry system to be slow and uncomfortable.

Table 4: Survey results for study participants with ALS/PLS.

Statement (1 = Strongly Disagree, 7 = Strongly Agree)	Wet (Average)	Dry (Average)
1) The time required to setup the cap was reasonable	5.70 (±2.20)	6.50 (±0.79)
2) The setup of the cap was frustrating/agitating	2.28 (±1.89)	1.00 (±0.00)
3) I found the cap comfortable	5.42 (±1.81)	4.00 (±1.95)
4) I could see myself wearing this cap for more than two hours at a time	5.00 (±2.08)	2.50 (±2.07)
5) The speed and accuracy of the system was acceptable	5.71 (±1.11)	3.25 (±2.70)
6) If this was my primary method of controlling a computer I would use it daily	5.14 (±2.27)	2.75 (±1.98)

The question of whether the raw data recorded from the two electrode types were similar was also addressed. In Figure 3.4, P300 latency and amplitude values were compared for the signals measured over the electrode location Cz in able-bodied participants. These values were selected using the maximum peak amplitude from a window of each participant's grand average P300 signal between 200 – 400 ms. The grand averages included signals from both the training and testing stages; therefore, more data were averaged for the dry electrode grand averages than for the wet electrode grand averages.

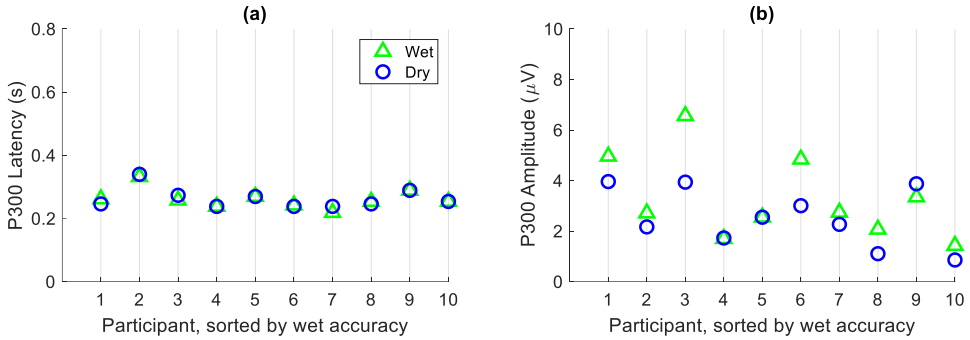

Figure 3.4: Comparison of P300 signal characteristics between wet and dry electrode systems recoded over electrode location Cz in able-bodied participants. The x-axis represents the ten participants in this study and the y-axis represents (a) P300 latency, in seconds, and (b) P300 amplitude, in microvolts. A significant difference in amplitude between the electrode systems was observed ($p < .01$).

The average latencies for the P300 peak amplitudes recorded in the wet and dry electrode systems were 262 ± 3 ms and 263 ± 3 ms, respectively, and the difference in latency between the two systems was not statistically significant ($p > .05$). However, the difference in P300 peak amplitude was statistically significant ($p < .05$) with lower amplitudes recorded in dry electrode system (2.55μV ±1.14) when compared the amplitudes recorded in the wet electrode system (3.29μV ±1.65). This suggests that techniques other than additional data collection may be needed to produce more robust P300 components in the dry electrode recordings. However, in order to determine the signal processing techniques that are most likely to be beneficial for improving the performance of dry electrode systems, the difference(s) between wet and dry electrode recordings need to be characterized.

Without a conductive gel to serve as buffer to bridge the gaps between the skin and the electrode as a participant moves during a recording session, dry electrode systems are more sensitive to artifact interference arising from movement. Given that motion artifacts are known to impact the frequency content of EEG recordings, it was hypothesized that increases in spectral amplitude occurred more often in the dry electrode recordings than in the wet electrode recordings. Therefore, the power spectral densities (PSDs) were estimated and compared for each participant's continuous dry and wet electrode recordings. PSDs were computed using Welch's method [104] with a 256 sample hamming window and 50% window overlap. Figure 3-5 displays the PSD estimates for a single electrode channel, Cz, in both able-bodied participants and participants with communication disabilities for the dry and wet electrode systems.

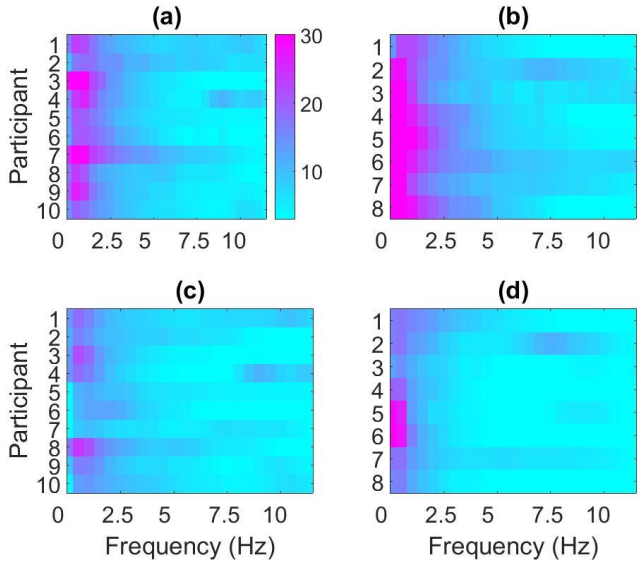

Figure 3.5: Power spectral density (PSD) estimates for continuous P300 speller EEG data. PSDs for the dry electrode recordings are displayed for (a) able-bodied participants and (b) participants with communication disabilities. PSDs for the wet

electrode recordings are displayed for (c) able-bodied participants and (d) participants with communication disabilities. The color axis displays power in dB, where the values are mapped to a color according to the color bar on the right in (a).

Differences in the frequency content were examined for the four main frequency bands of human EEG: (1) Delta (0.1 - 4 Hz), (2) Theta (4 - 8 Hz), (3) Alpha (8 - 13 Hz), and (4) Beta (13 -30 Hz). Wilcoxon signed-rank tests were performed on all 8 channels for each power band to determine the significant differences between the two systems. These tests revealed that the differences in delta and theta band power were significant for 8 out of the 8 electrode channels ($p < .05$). Differences in alpha band power were significant for 7 channels and differences in beta band power were significant for 3 channels. This suggests that as frequency decreased, the differences between the electrode system recordings increased across channels, which implies low-frequency artifacts may be negatively impacting dry system performance given that the P300 component also lies in this frequency band [105].

3.2 Identifying Artifacts with Anomaly Detection

In the previous section, it was demonstrated that the dynamic stopping algorithm detected the need to collect more data in a noisy dry electrode system, but poorer P300 speller performance was still observed when compared that obtained with a wet electrode system. Spectral differences were also observed between the two systems, with more significant differences occurring at lower frequencies. Given that the frequency content of

EEG data is easily impacted by motion artifacts, it is possible that slow-wave motion artifacts negatively impacted dry electrode system performance. Therefore, in this section, the differences in the number of possible motion artifacts recorded by the wet and dry electrode systems were analyzed using two anomaly detection techniques: autoregressive models and wavelet analysis.

3.2.1 Methods

To quantify possible motion artifacts, two techniques that have been well-studied in seizure detection literature (e.g., [106]–[108]) were used to identify voltage deflections in the EEG recordings during the 3.5s time periods between character spelling (i.e., the inter-character intervals). Characters were not flashing during this period, eliminating the possibility of inadvertently detecting ERPs.

The first technique used an autoregressive (AR) model, where a window of previous observations of the EEG signal were used to predict the signal value at the next time step [104]. The assumption of the AR model is that the signal is stationary and can be described by the following linear difference equation:

$$\hat{x}(t) = d_0\, w(t) + \sum_{k=1}^{P} a_k\, x(t-k) \qquad \textbf{(3-1)}$$

where $x(t-k)$ for $k = 1 \ldots P$ are the previous observations of the signal, $w(t)$ is white noise, and $\hat{x}(t)$ is the predicted value at time t. The constant regression coefficients d_0 and $\boldsymbol{a}$ are used to determine the magnitude of the effect of noise and previous observations. In this study, equation (3-1) was used to model the inter-character EEG

signals as stationary signals. Given that most artifacts are non-stationary phenomena, the model detected peaks in the autoregressive error (i.e., $x(t) - \hat{x}(t)$) that exceeded a predefined threshold when artifacts may have occurred. A model order of $P = 7$ was selected, which implies that each EEG sample depends on 7 prior samples (or ~27.3 ms) and has been shown to produce adequate estimates for EEG patterns [109]. The model's coefficients d_0 and $\boldsymbol{a}$ were estimated with the Yule Walker equations [104].

The second technique used for artifact detection was wavelet analysis, where wavelet transforms were used to decompose the inter-target EEG signals into different frequency ranges using a convolution of the signal with a set of functions generated by a mother wavelet (see [110] for a review on wavelet decomposition). The benefit of using a wavelet transform over a conventional Fourier transform is that it provides simultaneous information on both the frequency and time localization of the signal characteristics, which is relevant for analyzing non-stationary signals such as artifacts. The convolution of the EEG signal with the functions generated by the mother wavelet produces a series of detail coefficients (high-pass filtered) and approximation coefficients (low-pass filtered). For anomaly detection, peaks in the lower frequency wavelet approximation coefficients (0.1 - 32 Hz) above a threshold were measured as artifacts. A Daubechies-4 mother wavelet was selected as the transform function based on its use in prior work with EEG signals (e.g., [108], [111], [112]), where tests were performed with different types of

wavelets and the maximum efficiency for the particular application was achieved with a Daubechies wavelet.

After modeling the inter-target EEG segments using both the autoregressive and wavelet approaches, a hysteresis thresholding technique (i.e., thresholding by two thresholds) was used to determine whether an artifact was identified or not. This technique utilized two thresholds to determine the beginning and end of the artifact. For example, two artifacts were detected in a signal if the signal exceeded a given threshold at time t, then fell below the threshold at time $t + 10$, and then exceeded the threshold again at time $t + 15$. This means that it is possible for each EEG segment to contain a different number of artifacts.

3.2.2 Results

For the autoregressive model, a hysteresis threshold was applied to the prediction error signal to look for non-stationary artifacts in the stationary inter-character EEG signal. The mean number of possible artifacts identified by the errors in the participant-specific autoregressive models for a given set of thresholds between 0 and 200 μV is displayed in Figure 3.6. Although the difference between the number of possible artifacts detected with the wet and dry electrode systems for able-bodied participants appears small (Figure 3.6a and Figure 3.6b), two Wilcoxon sign-ranked tests revealed the difference was statistically significant ($p < .001$). Perhaps more noticeable is the difference for participants with ALS in Figure 3.6c, which was also significant ($p < .001$).

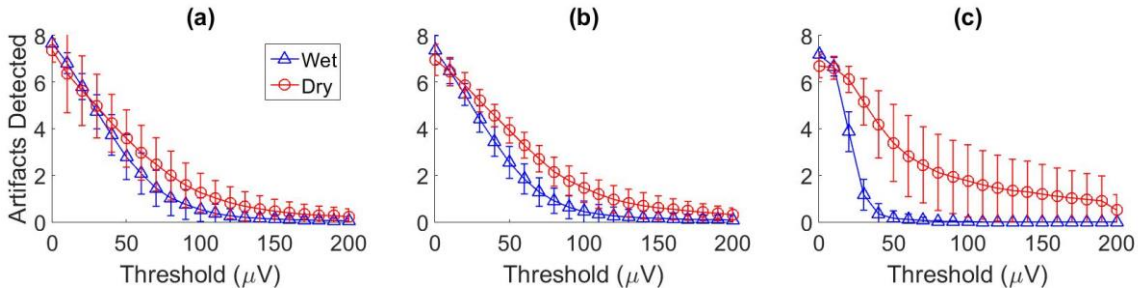

Figure 3.6: Number of possible artifacts detected during the 3.5 s inter-character EEG segments with an autoregressive model for different thresholds. Results were averaged across participants and EEG segments in (a) the able-bodied static data collection experiment, (b) the able-bodied dynamic data collection experiment, and (c) the ALS/PLS dynamic data collection experiment. The differences between the two systems were statistically significant, with a higher number of artifacts detected in the dry electrode recordings.

Similar results were observed using the wavelet approach and are displayed in Figure 3.7. For this approach, the hysteresis threshold was applied to the wavelet approximation coefficients to look for low-frequency artifacts. Once again, the differences in the number of possible artifacts detected between the two systems for able-bodied participants appear small (see Figure 3.7a and Figure 3.7b), but were indeed statistically significant ($p < .001$). It is also clear that this difference is significant for participants with ALS in Figure 3.7c ($p < .001$).

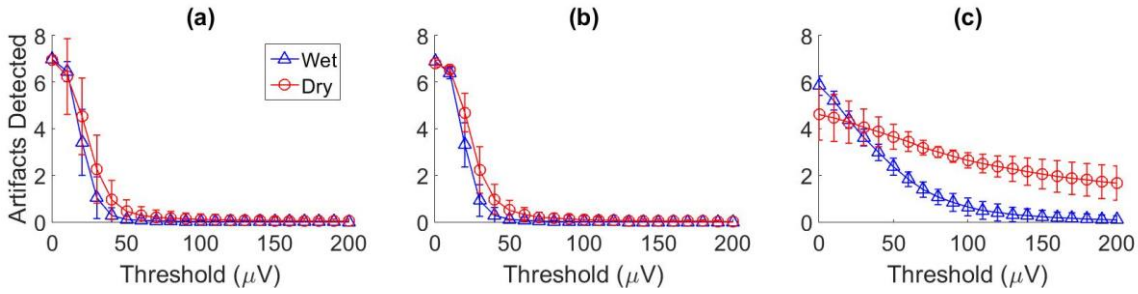

Figure 3.7: Number of possible artifacts detected during the 3.5 s inter-character EEG segments with wavelet analysis for different thresholds. Results were averaged across participants and EEG segments in (a) the able-bodied static data collection experiment, (b) the able-bodied dynamic data collection experiment, and (c) the ALS/PLS dynamic data collection experiment. The differences between the two systems were statistically significant, with a higher number of artifacts detected in the dry electrode recordings.

The results in Figure 3.6 and Figure 3.7 showed that for thresholds above ~30μV, both techniques detected a larger number of possible artifacts in the dry electrode recordings when compared to the wet electrode recordings. The results for participants with communication disabilities (Figure 3.6c and Figure 3.7c) revealed much larger differences between the number of possible artifacts detected in the wet and dry electrode systems. Given that these experiments were performed either in the participant's home or a clinical setting (as opposed to the able-bodied participants who performed experiments in a laboratory setting), the larger differences may be due to an increase in environmental noise that dry electrodes, with their poor skin-electrode contact, may be more sensitive to. Taken as a whole, these results suggest that there may be a benefit to applying signal

processing approaches that target artifacts without attenuating lower frequency ERPs such as the P300 component.

3.3 Discussion

Using a dry electrode system with the P300 speller would greatly reduce the complexity and time it takes to set up the system, creating a more practical home-based communication aid for people with communication disabilities. While some previous studies have suggested that dry electrode recordings are comparable to wet electrode recordings for P300-based BCI applications, the performance results presented in this chapter suggest that dry electrode recordings may suffer from lower SNRs. These lower SNRs negatively impact classification accuracy and reduce the potential utility of dry electrodes for BCI systems as home-based communication aids.

Using a data-driven dynamic stopping algorithm partially compensated for the additional noise in the dry electrode recordings by collecting more data, but performance was still reduced when compared to the wet electrode system. Although a variety of commercially available dry electrode systems exist (e.g., Cognionics, Brain Products, Emotiv and Wearable Sensing), the results presented in this study would likely translate to these systems because the issue of poor skin-electrode contact remains. However, certain design criteria might have reduced some of the artifact noise. A wireless dry system that eliminates the possibility of recording cable movement noise might have been beneficial for ambulatory participants in this study. Flexible dry electrodes that conform

to the scalp could have helped provide the right amount of contact pressure for each individual sensor, minimizing participant discomfort and reducing the impact of motion artifacts in cases where the cap was ill-fitted. Future work should include applying the dynamic stopping algorithm to dry electrode systems other than the g.Sahara system to validate the results presented in this chapter.

A possible future approach to make dry electrodes in P300 speller systems feasible is to develop methods to mitigate the effects of excessive low-frequency noise. While several artifact removal techniques exist, their application in BCI is limited because they may attenuate the desired signal (e.g., linear filtering [69]), require a priori knowledge of the desired/artifact signal (e.g., adaptive filtering [69]), not detect small amplitude artifacts (e.g., thresholding [70]) or require that the artifact and desired signal are minimally correlated or maximally independent (e.g., blind source separation [70]).

It may be possible to use machine learning to separate undesirable artifacts from ERP responses and thereby improve BCI performance with dry electrodes. The current standard learning algorithm operates under the assumption that an increase in EEG signal energy only occurs in response to the presentation of relevant stimuli, which ignores voltage changes that can originate from the user's physical activity or external movements in the surrounding environment. Dry electrodes, with their higher contact impedance, may be particularly sensitive to these types of motion artifact interference. Therefore, an additional classification stage could be developed that separates detected artifacts into

two classes, ERP responses and artifacts, using features that describe the spatiotemporal characteristics of the artifact. The exploration of features beyond spatiotemporal characteristics, such as statistical moments or power spectral density information, may help increase artifact classification performance. Unsupervised learning techniques could also facilitate further discrimination between different types of artifacts, leading to customized detection methods. Future work could also include incorporating these artifact detection methods into the dynamic stopping algorithm to reduce the system's response to false energy responses and produce a more accurate learning algorithm in dry electrode systems.

4 Visual-Motor Skill Learning during Marksmanship Training in Virtual Reality

The ability to coordinate visual information with motor output is essential to a great number of human endeavors. Activities such as those in sports, surgery, and law enforcement often rely on efficient reciprocal interactions between visual perception and motor control, allowing individuals to execute precise movements under time-limited, stressful situations. As such, developing a better understanding of the psychophysiological mechanisms that underlie visual-motor control, and characterizing how these change with practice, offers fundamental new insight into skilled performance, which may be useful for the development of better training programs that improve their effectiveness.

Cognitive processes that precede and accompany motor action unfold on a millisecond timescale. Recording brain activity to study these cognitive processes during a motor task requires a portable signal acquisition modality with high temporal resolution, making EEG the most suitable choice. However, traditional EEG experiments to study the brain dynamics involved with performing a motor action typically limit physical responses to a button press or joystick control to avoid artifact contamination. However, when movement is restricted, important information reflecting active cognition from motor behavior may be absent in the recording (e.g., [10]). Therefore, an approach as known mobile brain/body imaging (MoBI) [8] was developed to overcome the limitations of traditional EEG experiments by using EEG in combination with motion

sensors to study brain activity while users physically interact with a real or simulated environment.

In this chapter, a MoBI-style approach was implemented to study the psychophysiological processes underlying visual-motor skill learning during a marksmanship task in immersive VR. Modeled after Olympic Trap Shooting standards [113], the simulation required participants to track and shoot a digital clay pigeon launched from behind a trap house while synchronized recordings of their brain activity and full-body movements were collected with EEG and motion tracking technology. The goal of this work was to establish how visual-motor skill learning was manifested in the EEG and motion tracking data by identifying biomarkers that changed as participants improved their performance on the task through the course of training.

In previous work, a similar marksmanship task in VR was used by Zielinski et al. [114] to study image persistence and by Rao et al. [75] to investigate changes in the movement kinematics during a single-day experimental training session. Despite the short training duration, Rao et al. provided evidence that novice participants modified their movements as they improved on the task over the course of the session. The work presented in this chapter expands on the study by Rao et al. through the addition of two extra training days to examine skill retention and by including EEG to investigate the neural basis of motor control during the task.

4.1 Brain Dynamics of Visual-Motor Skill Learning

Previous EEG research on skill learning in sporting environments has focused on self-paced tasks that are typically characterized by spectral power during a motionless preparatory period ("pre-shot routine") before skilled movement occurs. While significant differences in preparatory phase EEG between experts and novices have been reported for tasks such as static marksmanship and archery (see [115] for a review), much less is known about the brain activity related to an individual's skill development. Furthermore, additional cognitive processes that might also contribute to skill acquisition, such as visual information processing or proprioception, have yet to be explored in realistic dynamic sporting environments. Therefore, one of the aims of this study was to improve upon previous research by considering the brain activity immediately following the launch of the clay pigeon, in addition to the traditional spectral analysis of the preparatory phase preceding the target launch. EEG data were analyzed to calculate visual evoked potentials (VEPs) that occurred after the digital target was launched and prior to skilled movement. This approach is an extension of past skill learning research with EEG during static visual search tasks that were designed to exploit the neurophysiological contralateralization of the sensory and motor systems (i.e., information in the left visual field is processed in the right side of the brain and vice versa) to derive lateralized VEPs that changed through practice [29], [30].

4.2 Methods

4.2.1 Subjects and Equipment

Twenty-four healthy participants (13 female) from Duke University were recruited for this study. Four participants were excluded from analyses: two left-handed participants (due to differences in brain activity when compared to right-handed participants [19]), one participant with extensive marksmanship experience, and one participant who did not complete the experiment. The ages of the remaining twenty participants ranged from 18 to 35 years and all were right-handed. Participation was voluntary, and participants were compensated $10/hour for their involvement in the study. The experimental protocol was approved by the Duke University Institutional Review Board.

The study was conducted in the Duke immersive Virtual Environment (DiVE), a six-sided cave automatic virtual environment (CAVE) VR system [9], where participants stood in the center of the room-size cube with projectors directed at each of the cube's six walls. The back-facing wall of the DiVE display was not active during the study, effectively making it a 5-wall system. The projectors were run at 120 Hz with a total resolution of 1920 x 1920 pixels per wall.

Participants wore 3D shutter glasses operating at 60 Hz to display stereoscopic graphics. A head tracking device, which was mounted on the shutter glasses, controlled the system viewpoint according to the participant's head movements. An Xbox Top Shot

Elite firearm game controller was used for target shooting. The controller, furnished with a 6-DOF tracking sensor, was held in the participant's right hand and stabilized with the opposite hand placed along the barrel of the controller (see Figure 4.1). An Intersense IS-900 tracking system was used to record the position and orientation of the controller and head throughout the experiment. Data from both tracking sensors were sampled at 60 Hz.

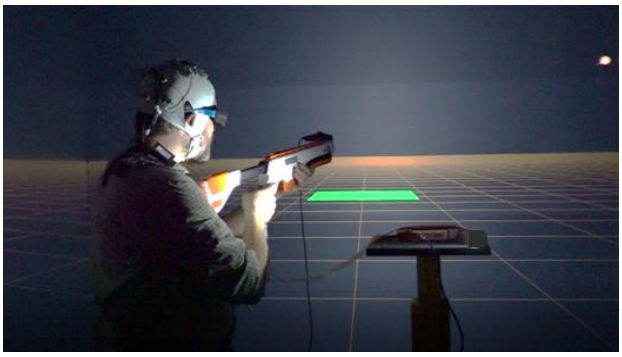

Figure 4.1: Participant performing the simulated marksmanship task in the Duke immersive Virtual Environment (DiVE), a six-sided CAVE-like VR system. For clarity, the picture was taken with monoscopic graphics.

Participants' EEG signals were recorded using actiCAP active wet electrodes connected to a computer via a 16-channel BrainVision V-Amp system. Thirteen active electrode channels (F3, Fz, F4, C3, Cz, C4, T5, T6, P3, Pz, P4, O1 and O2) were placed along the scalp according to the 10-20 Electrode Placement system with a linked mastoid reference [16]. Two additional electrodes were placed on the right and left outer canthi of the eyes to record horizontal eye movements using horizontal electrooculography (HEOG). All electrode impedances were kept below 10 kΩ. Data were sampled at 1000 Hz. The dominant electrical artifact at 60Hz (power line frequency) was attenuated using

a 0.1-30 Hz bandpass filter. Furthermore, the use of active electrodes containing built-in amplifiers reduced environmental noise at the recording site by converting high impedance input signals into low impedance output signals.

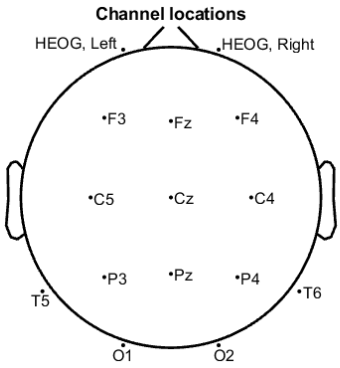

Figure 4.2: Electrode montage consisting of the 13 EEG electrodes and 2 HEOG electrodes used for the study, placed according to the 10/20 International Electrode Placement system.

4.2.2 Experimental Task

The marksmanship task used in this study was modeled after International Shooting Sport Federation standards [113]. To mimic realistic target flight times, trajectories, and distances observed in real trap shooting events, the design of the simulation incorporated the physics of projectile motion such as gravitational pull, air resistance, and lift force.

Participants entered the DiVE wearing the EEG cap and the 3D shutter glasses. To begin a trial, the participant aimed the controller towards a trap house, which was displayed as a rectangle on the ground 54 ft in front of them in simulated space. After an

initial 500 ms waiting period, the trap house changed color from red to green and a second waiting period began (variable between 1 and 1.5 s). During the initial waiting period, if the participant aimed the controller away from the trap house before the color changed from red to green, the timer was reset and did not begin again until the participant aimed the controller back towards the trap house. At the end of the second waiting period, a target was launched in one of six possible trajectories.

The six possible target trajectories, illustrated in Figure 4.3, consisted of three horizontal directions relative to the center of the trap house (left = -45°, center = 0°, right = 45°) and two elevations relative to the ground plane (upper = 25.17°, lower = 12.95°). To increase ecological validity (e.g., fluctuations in outdoor environmental conditions such as wind currents), a random horizontal jitter ranging from -3°to 3° was added to each trajectory. The target was displayed as an orange sphere of radius 1 ft that traveled at a speed of 95.34 ft/s. The maximum flight times for the target were 1.772 s and 3.085 s for the upper and lower trajectory elevations, respectively.

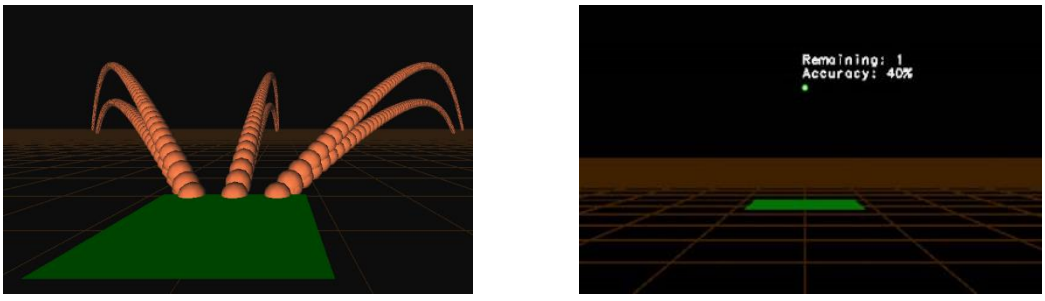

Figure 4.3: (Left) The six target trajectories, shown slightly off-center to improve visibility, showing the orange spherical target in flight (all frames included) and the green trap house from which the targets were launched. (Right) Feedback given immediately following the trigger pull, which included the number of trials remaining

in a given block and cumulative accuracy in terms of the percent hits out of shots taken. The target changed color from orange to green for a hit and orange to red for a miss.

Participants completed the simulated marksmanship task on three separate days within one week. On each day, the experiment was split into 8 blocks of 60 trials each. Before each block, participants stood with their eyes open for 30 seconds to record resting state EEG data prior to beginning the first trial. Within a block, all six target trajectories were presented 10 times in a random order. The target acquisition task was done using a raycasting technique [23], but rather than a visible ray, only a white dot was shown at the target depth, to mimic a laser sight. For a given trial, participants were allowed only one attempt to hit the target. If the controller's ray was in contact with the target at the time of the shot, the screen would freeze and the target would change color from orange to green. If the ray was not in contact with the target at the time of the shot, the screen would freeze and the target would change color from orange to red. After each shot, the participant's shot location and the target were displayed on the screen until they indicated that they were ready for another target by aiming the controller back over the trap house. This gave participants the opportunity to infer their relative precision as a potential feedback signal to guide learning. Participants were also given feedback on their cumulative accuracy (% hits) for the block and how many trials remained in that block via text on the screen (see Figure 4.3).

4.2.3 Performance Measures

Data were analyzed using block, day, trajectory elevation, and trajectory horizontal direction as independent variables to study the movement-related dependent variables extracted from the EEG and motion tracking data in order to determine the psychophysiological modifications that occurred through practice. Figure 4.4 illustrates the time frames during a single trial from which the EEG and kinematic variables were extracted. EEG data were only analyzed before movement due to the large artifacts present in the data.

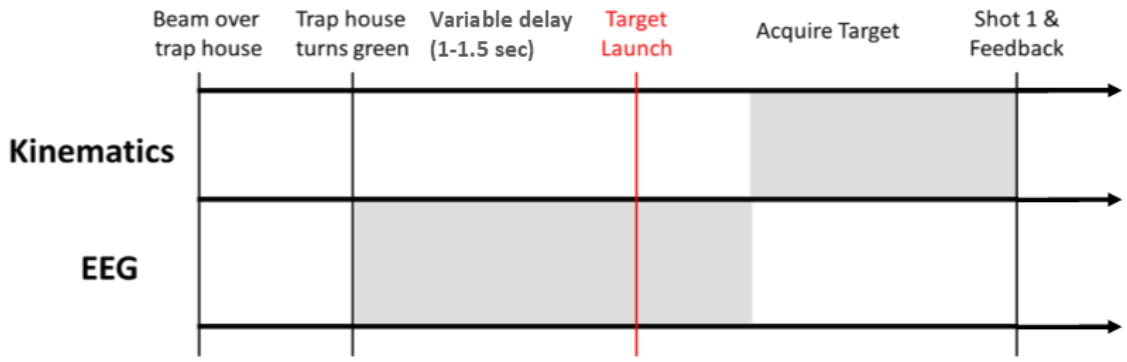

Figure 4.4: Data collection for a single trial. The shaded regions indicate which data stream was analyzed at different times during the trial.

Data collected from the Intersense trackers and HEOG sensors were used to extract the timing of key behavioral events after a target launch, such as reaction time of the movement and trigger response time. Reaction time - the elapsed time from the target launch to the start of movement - was calculated for the movement of the eyes, controller, and head. Eye reaction time was calculated using a rectified sum of the two HEOG

channels, where the beginning of a voltage deflection resulting from changes in eye position was detected when a threshold of 3 standard deviations above the baseline mean ([0 100] ms post-launch) was reached. Due to the lack of a vertical eye measurement, eye reaction times were only computed for the left and right horizontal trajectories. Head and controller reaction times were calculated using 10% of the peak acceleration, as measured with the Intersense trackers. Acceleration was computed as the derivative of the speed trace, after the speed trace was smoothed with a 7th order FIR filter.

After reacting to the target launch, movements were categorized into two phases: (1) an initial ballistic phase, characterized by the large accelatory and decelatory motions as the participant proceeded toward the general vicinity of the target, and (2) a refinement phase that reflected finer error-correction movements prior to pulling the trigger. The beginning and end of the ballistic phase were calculated as the controller reaction time and the time when the acceleration trace returned to 10% of the peak acceleration, respectively. The refinement phase began immediately following the end of the ballistic phase and ended when the trigger was pulled to shoot the target. Examples of the kinematic measures obtained during a single trial are illustrated in Figure 4.5.

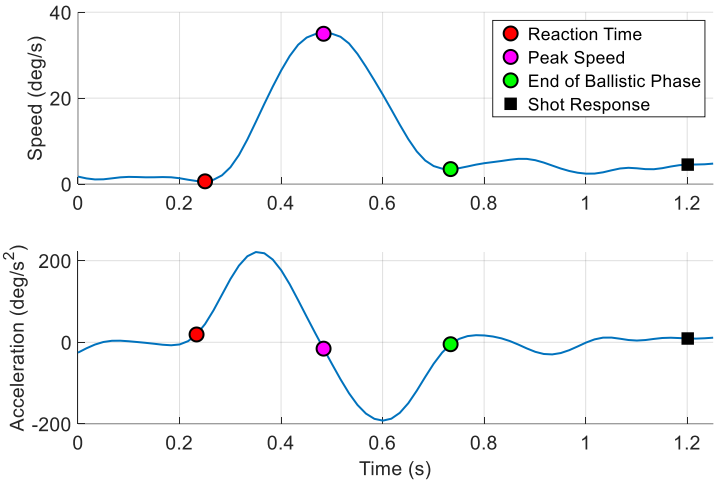

Figure 4.5: Movement of the firearm controller displayed for a single trial, time-locked to the target launch at 0 s. (Top) The speed trace, smoothed using a zero-phase FIR filter. (Bottom) The acceleration trace, calculated by taking the derivative of the speed trace. Dependent variables are indicated by the circular markers along the traces.

EEG data were analyzed in two separate groups of epochs to calculate preparatory phase spectral power during the 1s period before the target launch and VEPs in the 200 ms window immediately following target launch. The preparatory phase epochs were extracted time-locked to the beginning of second waiting period, while the VEP epochs were extracted time-locked to the beginning of the target launch. Both sets of epochs were baseline corrected using the mean value from a 50 ms baseline. Preparatory phase power spectral density estimates were calculated using Welch's method with a window size of 500 ms and 50% overlap and divided into bands by frequency bands: delta (0.1 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 13 Hz), and beta (13 – 30 Hz). VEP epochs were calculated by averaging over trials on each block/day for a given participant in order to attenuate noise

so that the brain signal could be seen more easily. Channels P3 and P4 were selected for VEP analysis due to the posterior locations of the electrodes over the left and right hemispheres of the visual cortex, respectively.

4.2.4 Statistical Analysis and Outlier Removal

Statistical differences were computed using 3-way repeated measures analysis of variance (ANOVA), where the main effects were computed across the independent variables (blocks, days, trajectory elevations, and trajectory horizontal directions). Data were tested for sphericity using Mauchly's Test for Sphericity and, if the assumption of sphericity was violated, a Greenhouse-Geisser correction was used. ANOVA results are reported in the format $[F(DOF_{conditions}, DOF_{error}) = F-statistic,\ p > or < threshold]$, where the F-statistic was calculated by dividing the mean sum of squares for the independent variable by the mean sum of squares for error. Correlations between dependent variables were computed and tested for statistical significance using Pearson's correlation coefficient.

Trials were excluded from both kinematic and EEG analyses if the participant did not shoot (1.67%, 482 trials). Additionally, if movement was initiated too quickly for a given trial, defined as less than 16.667 ms, the trial was removed from the kinematic analyses (0.39%, 23 trials).

Trials containing artifacts that exceeded a threshold of 200 μV in the preparatory phase EEG data were excluded from the analyses (0.97%, 279 out of 28800 trials). Trial

rejection for the VEP was based on two calculations done within a very specific spatial (contralateral visual responses in P3/P4) and temporal (launch-locked before the behavioral response) window. Epochs were removed from the VEP analyses if they exceeded a threshold of 40 μV (3.02%, 580 out of 19200 trials) or contained data outside of 5 standard deviations of the joint probability distribution observed at each time point (0.35%, 67 out of 19200 trials). The use of artifact removal techniques (as opposed to rejection) would have been severely challenged due to the lack of clear biological templates (e.g., ocular artifact correction) to base removal on. Moreover, based on the low prevalence of rejected VEP trials (3.37%), it was evident that the signal under consideration offered an unimpeded view of the neural activity that was meant to be scrutinized in the planned hypothesis tests.

4.3 Results

EEG and motion tracking data were analyzed to determine within-subject changes through learning. Therefore, in the figures presented, variables were averaged across participants for visualization purposes only. Error bars were excluded for ease of interpretation and to avoid conflating within-subject and between-subject distributions, which are a result of individual variability.

4.3.1 Shot Accuracy and Positional Error

The average shot accuracy – the percentage of hits out of total shots taken – for the 20 participants in this study was 59.14% on Day 1, 69.61% on Day 2, and 72.18% on Day

3. Participants significantly improved their accuracy through practice, with main effects of day ($F(2, 38) = 71.355$, $p < .001$) and block on Day 1 ($F(7, 133) = 22.789$, $p < .001$) and Day 2 ($F(7, 133) = 10.041$, $p < .001$). However, there was not a main effect of block on Day 3 ($F(7, 133) = 1.540$, $p > .05$), suggesting that performance may have begun to plateau. Main effects of elevation ($F(1, 19) = 13.032$, $p < .01$) and horizontal direction ($F(2, 38) = 196.03$, $p < .01$) were also observed, where the best performance occurred in the upper and central trajectories and symmetrically decreased for the side directions. Figure 4.6 displays the average accuracy across time and trajectories, along with individual accuracy across days.

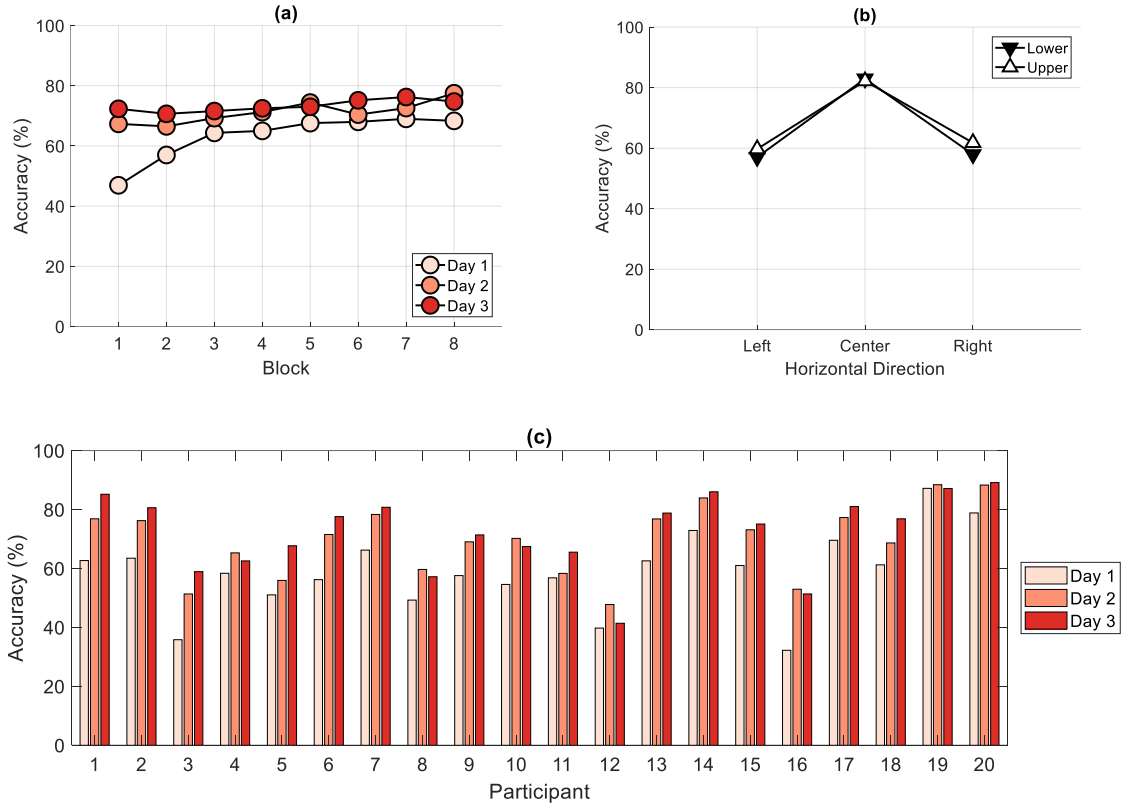

Figure 4.6: Accuracy differed across trajectories and with practice. (a) Average accuracy improvements displayed across blocks and split by day. Significant improvements were observed across days and blocks on Day 1 and 2. (b) Average

accuracy displayed across horizontal directions for the target trajectories and split by elevation. Significantly better performance was observed in the upper and central trajectories. (c) Individual accuracies displayed for each particpant across days.

There was a significant interaction effect of horizontal direction with day for accuracy ($F(4, 76) = 3.476$, $p < .05$), indicating that improvements over days were not uniform across trajectories. This interaction is illustrated in Figure 4.7, where larger increases in accuracy across days were observed for the left and right trajectories when compared to the center trajectories. The mean accuracy for the side trajectories increased by 14.6% from Day 1 to Day 3, while the mean accuracy for the center trajectories only increased by 10.35%.

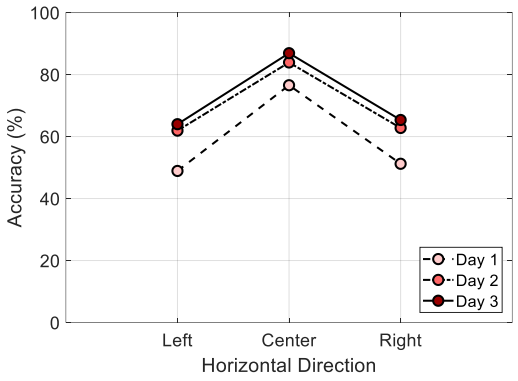

Figure 4.7: A significant interaction between day and horizontal direction was observed, where greater improvements occurred in the left and right horizontal trajectories when compared to the center trajectories.

Shot error – the Euclidean distance between the shot and the center of the target – significantly decreased across days ($F(2, 38) = 23.854$, $p < .01$) and blocks on Day 1 ($F(7,$

133) = 26.915, p < .001), Day 2 (F(7, 133) = 11.272, p < .001), and Day 3 (F(7, 133) = 3.269, p < .05), indicating that participants became more precise with practice. Significant main effects were also observed for elevation (F(1, 18) = 56.762, p < .01) and horizontal direction (F(2, 38) = 125.024, p < .01), where better performance (i.e., lower shot error) occurred in the upper and central trajectories. Figure 4.8 displays the mean error values across time and trajectories.

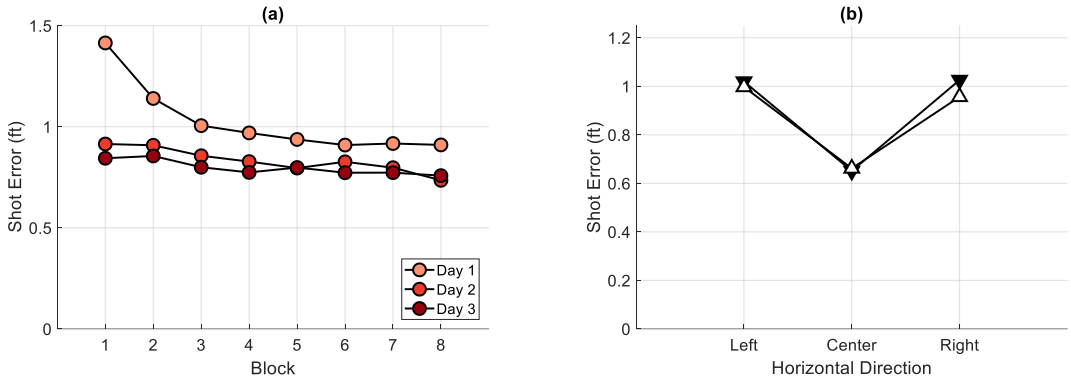

Figure 4.8: Shot error, expressed as the Euclidean distance from the center of the target, differed with practice and across trajectories. The maximum distance for a hit was 1 foot (i.e., the radius of the target). Lower shot error indicates better performance. (a) Average shot error displayed across blocks and split by day. Shot error significantly decreased across days and blocks on each day. (b) Average shot error displayed across horizontal directions and split by elevation. Lower errors were observed in the upper and central directions.

4.3.2 Reaction and Trigger Response Times

Reaction time – the elapsed time between the target launch and the start of movement – was computed for horizontal eye movement (via HEOG), controller rotation, and head rotation. The average reaction times were 0.194 ± 0.04s for eye movement, 0.206

± 0.04s for controller rotation, and 0.287 ± 0.12s for head rotation. A significant decrease in the reaction time of the controller was observed across days ($F(2, 38) = 28.247$, $p < .01$) and blocks on Day 1 ($F(7, 133) = 13.270$, $p < .001$), Day 2 ($F(7, 133) = 7.132$, $p < .001$), and Day 3 ($F(7, 133) = 5.748$, $p < .001$), indicating that hand movements occurred earlier through practice. These trends are illustrated in Figure 4.9. There was also a significant decrease in the reaction time of head across days ($F(2, 38) = 3.917$, $p < .01$), but not across blocks. There was not a main effect of day or block for eye reaction time.

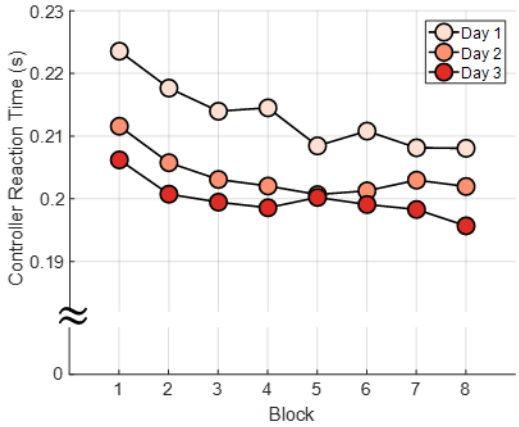

Figure 4.9: Average reaction time of the controller displayed across blocks and split by day. The y-axis has been scaled to show small, but significant decreases in reaction time across days and blocks on each day.

Trigger response time – the elapsed time between the target launch and the time the trigger was pulled – was significantly different by day ($F(2, 38) = 4.218$, $p < .05$), elevation ($F(1, 19) = 46.86$, $p < .01$), and direction ($F(2, 38) = 116.886$, $p < .01$). Earlier trigger responses occurred across days and for the lower and central trajectories, as illustrated in

Figure 4.10. However, a significant increase was observed across blocks on Day 1 ($F(7, 133) = 3.030$, $p < .05$), suggesting that there was an early learning phase where participants tuned their technique before it became a learned behavior. Additionally, there was a significant main interaction between day and elevation, where larger decreases in trigger response time occurred across days for the upper trajectories when compared to the lower trajectories ($F(2, 38) = 5.745$, $p < .01$).

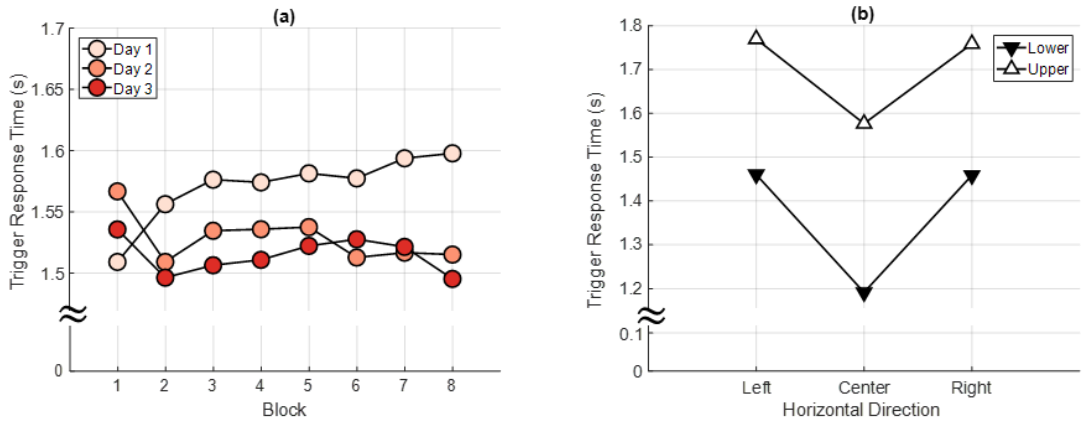

Figure 4.10: Trigger response times varied across trajectories and with practice. (a) Average trigger response time displayed across blocks and split by day. Response times significantly decreased across days but increased across blocks on Day 1. (b) Average trigger response time displayed across horizontal directions and split by elevation. Response times were later for the upper and side trajectories. Participants rarely shot before 1 s (3.1% of trials); therefore, the y-axis has been scaled to show small, but significant changes.

4.3.3 Movement Kinematics

Dependent variables describing movement patterns, such as speed and phase duration, were also modified through training and across trajectories. The peak speed of

the controller significantly decreased across blocks on Day 1 ($F(7, 133) = 3.708$, $p < .05$) and Day 3 ($F(7, 133) = 2.621$, $p < .05$), but did not decrease across days ($F(2, 38) = 2.392$, $p > .05$). The peak speed of the controller was also significantly different by elevation ($F(1, 19) = 20.114$, $p < .001$) and direction ($F(2, 38) = 612.389$, $p < .001$), with higher speeds occurring in the upper elevation and side directions. Moreover, the latency of the peak speed was earlier in the upper elevation ($F(1, 19) = 202.951$, $p < .001$) and side directions ($F(2, 38) = 69.129$, $p < .001$). There was not a main effect of block or day for the latency of the peak speed.

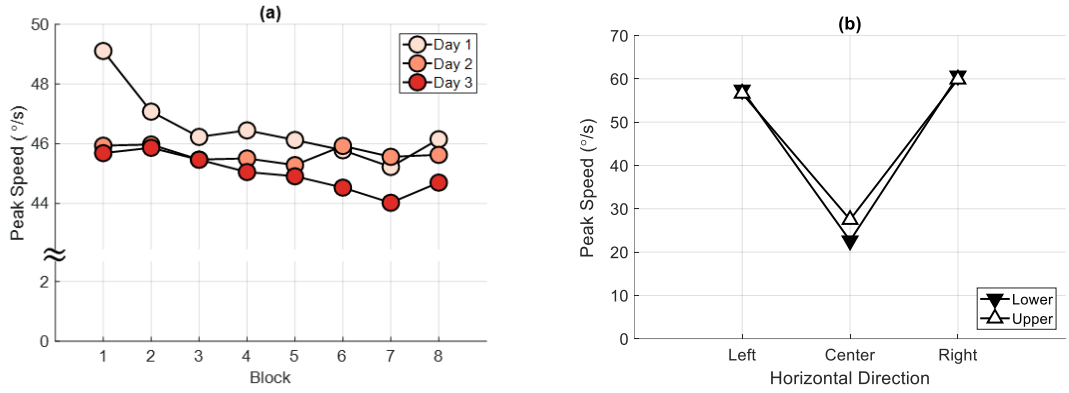

Figure 4.11: Peak speeds of the controller differed across blocks and trajectories. (a) Average peak speed of the controller displayed across blocks and split by day. Peak speeds significantly decreased across blocks on Day 1 and 3. (b) Average peak speed displayed across horizontal directions of the target trajectories and split by elevation. Higher speeds were observed in the upper and side trajectories.

As training progressed, differences in the duration of the ballistic phase and the the refinement phase were also observed. The duration of the feedback-based refinement phase significantly decreased across days ($F(2, 38) = 4.219$, $p < .05$), suggesting that

participants needed less time for error-correction as they practiced. On Day 1, the duration of the ballistic phase increased across blocks ($F(7, 133) = 3.504$, $p < .05$). On Day 2, the duration of the ballistic phase increased across blocks ($F(7, 133) = 2.21$, $p < .05$) and the duration of the refinement phase decreased across blocks ($F(7, 133) = 2.597$, $p < .05$), indicating a tradeoff existed between elongated ballistic movements and shorter feedback mediated refinements. There was no main effect of block on Day 3 for either phase duration. These trends are illustrated in Figure 4.12.

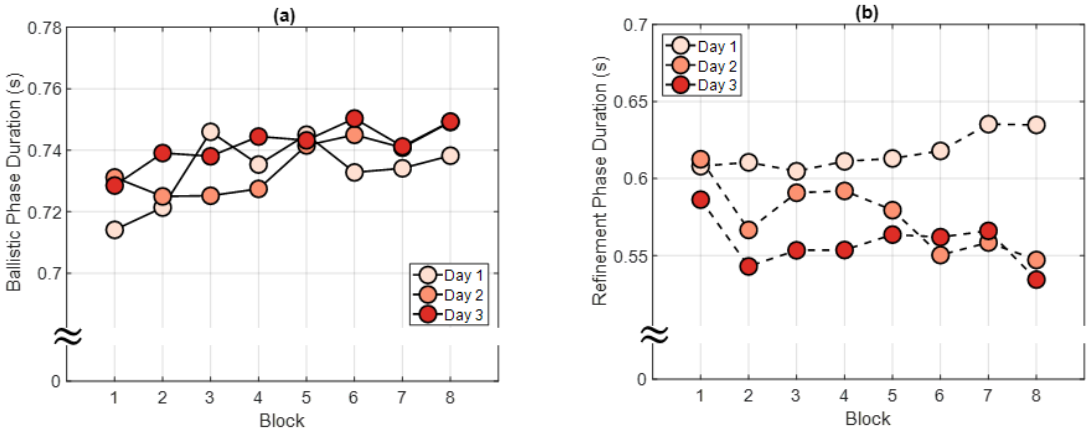

Figure 4.12: A tradeoff occurred between the durations of the ballistic phase and refinement phase. (a) Average duration of the ballistic phase displayed across blocks and split by day. Significant increases in the duration of the ballistic phase occurred across blocks on Day 1 and 2. (b) Average duration of the refinement phase displayed across blocks and split by day. Significant decreases in the duration of the refinement phase occurred across days and blocks on Day 2. The y-axis in both plots has been scaled to show small but significant changes.

Across trajectories, a main effect of elevation was observed with shorter phase durations occurring in the lower trajectories (ballistic: [$F(1, 19) = 31.933$, $p < .001$];

refinement: [$F(1, 19) = 31.702$, $p < .001$]). Significantly shorter durations of the ballistic phase were also observed for the central trajectories ($F(2, 38) = 588.895$, $p < .001$).

4.3.4 Neural Dynamics

4.3.4.1 Preparatory Phase Spectral Power

Power spectral density estimates were grouped into the standard EEG bandwidths for the analysis of the preparatory phase epochs. Epochs were measured for 1 s after the cue onset for the second waiting period (i.e., when the trap house changed color from red to green). ANOVA tests were performed to determine within-subject changes in band power across days and blocks.

A main effect of block was observed for preparatory phase beta band power (13 – 30 Hz) in channels Cz ($F(7, 133) = 4.722$, $p < .01$), P3 ($F(7, 133) = 3.661$, $p < .05$), and Pz ($F(7, 133) = 5.424$, $p < .01$) on Day 1 and 2. Beta band power, which is thought to reflect active cognition (e.g., [116]), increased across blocks near the sensory-motor area of the brain, suggesting that the neural hierarchy of preparatory motor processes might be modulated through training. Figure 4.13 illustrates this trend in all three channels. There was not a main effect of day or block for delta, theta, or alpha band power.

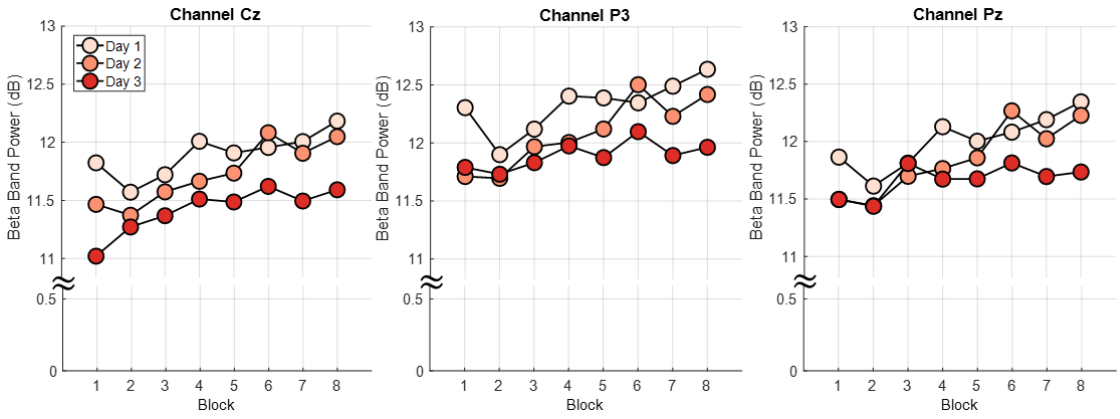

Figure 4.13: Beta band power during the preparatory phase significantly increased in channels Cz, P3, and Pz across blocks on Day 1 and 2. The y-axis has been scaled to show the small but significant changes.

4.3.4.2 Visual Evoked Potential

To study brain responses elicited by the launch of the target, EEG data were analyzed using the time window between the target launch and the onset of movement for each trial (0 ms to ~200 ms). During this timeframe, an early positive ipsilateral VEP followed by a late negative contralateral VEP was observed over the visual cortex for the left and right horizontal directions of the target trajectories. This simply means that if the target was launched leftward, a positive potential could be seen over the left hemisphere of the visual cortex followed by a stronger negative potential in the right hemisphere. Conversely, if the target was launched rightward, a positive potential could be seen over the right hemisphere of the visual cortex followed by a stronger negative potential in the left hemisphere.

Figure 4.14 displays the grand average VEPs (averaged across subjects and days) on separate lines for the left and right horizontal directions of the target trajectories, time-locked to the target launch. The electrodes selected for analysis, P3 and P4, are located over the left and right posterior hemispheres of the brain, respectively. The early ipsilateral positive VEP appears to begin at ~100 ms, followed by the onset of the larger contralateral negative VEP at ~110-120 ms that ends before the onset of eye movement at ~190 ms. Parametric statistical tests, corrected for multiple comparisons using Bonferroni correction, show that significant differences ($p < .05$) between the VEPs for the left and right trajectories existed in the range of ~100 ms and ~175 ms. These differences are displayed as black bars below the VEPs in Figure 4.14. Figure 4.15 displays the grand average scalp topographies for the left and right target trajectories.

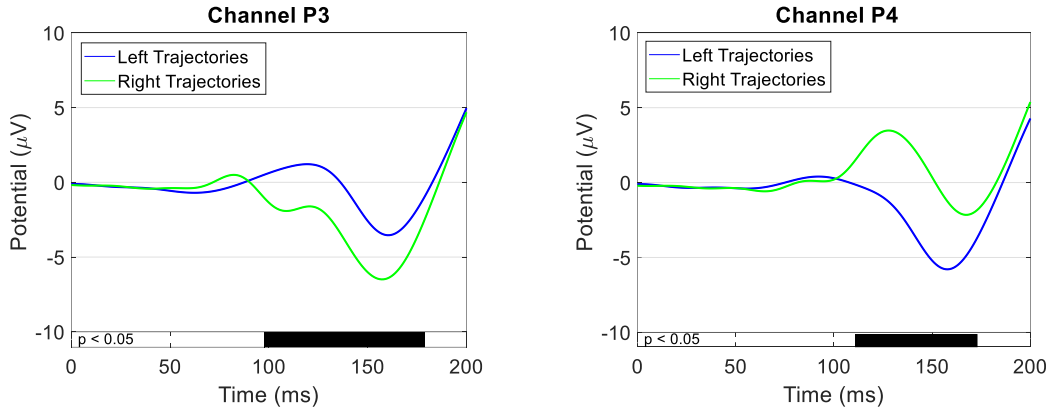

Figure 4.14: Grand average VEPs for the left and right horizontal directions of the target trajectories, time-locked to the target launch at time 0 ms. (Left) VEPs displayed for channel P3 over the left cortex. (Right) VEPs displayed for channel P4 over right cortex. Significant differences in amplitude between the left and right target trajectories are indicated by the black bars below the VEPs.

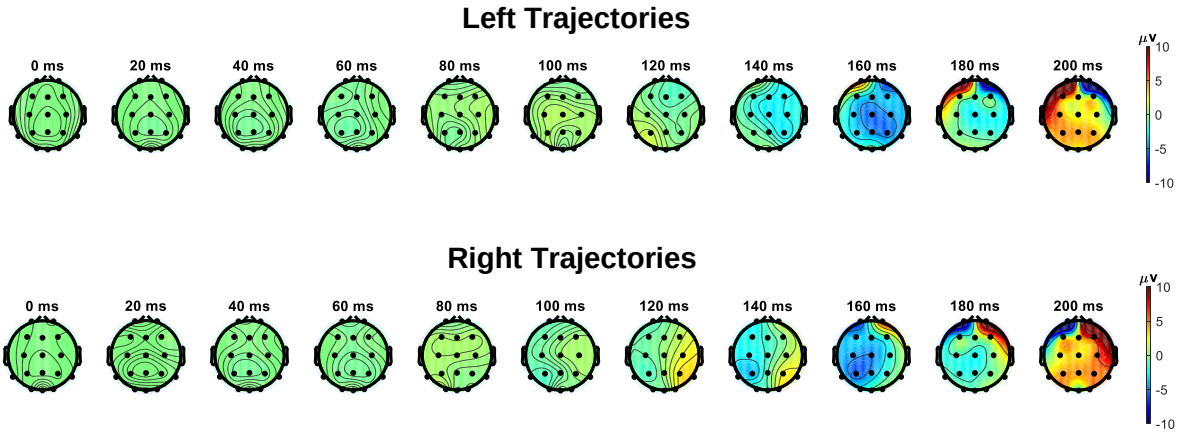

Figure 4.15: Topographic maps of the VEPs at 20 ms time steps displayed for the left target trajectories (top) and right target trajectories (bottom). Electrodes are displayed as black dots on the scalp. The color axis displays the voltage (in μV), where the values are mapped to a color according to the color bar on the far right.

An ANOVA test performed on the participant-specific VEP amplitudes revealed a significant decrease across days for both channels (P3: $F(2, 38) = 4.555$, $p < .05$; P4: $F(2, 38) = 8.481$, $p < .01$), indicating that more negative amplitudes occurred with practice. VEP amplitudes were determined by selecting the most negative peak in the participant's grand average on a given day and calculating the mean of the [-10 10] ms time window surrounding the peak.

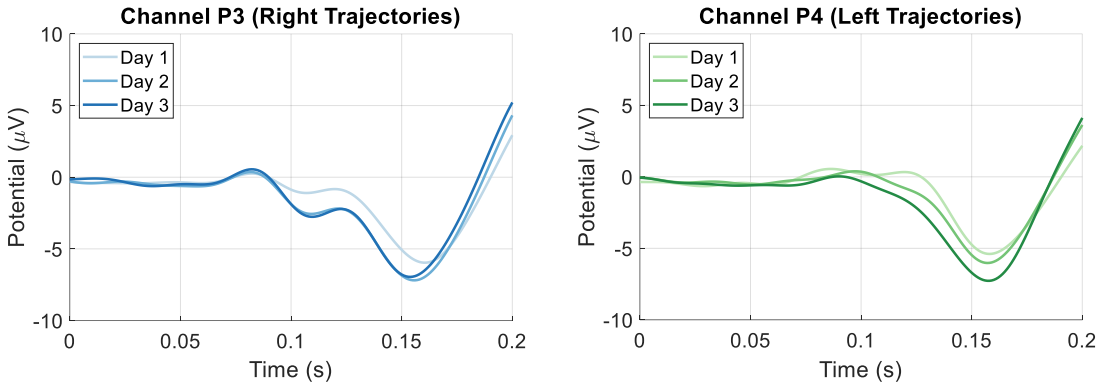

Figure 4.16: Significant decreases in VEP amplitude were observed across days in channels P3 and P4. (Left) Grand average contralateral VEPs plotted for channel P3 and split by day. (Right) Grand average contralateral VEPs plotted for channel P4 and split by day.

4.3.5 Neural Correlates of Behavior

An important goal of this study was to link changes in the EEG data to improvements in behavioral performance on the marksmanship task. To assess this, preparatory phase beta band power was evaluated in channels Cz, P3, and Pz while the mean VEP amplitudes were evaluated in channels P3 and P4 for the contralateral target trajectories. These values were then correlated with reaction time and trigger response time by averaging across trials on a given block, producing 20 (subjects) by 3 (days) by 8 (blocks) data points for each variable.

Significant negative correlations were observed for the preparatory phase beta band power, with lower reaction and response times seen for higher beta band power. Correlations with controller reaction time were significant in all three channels on each

day. Correlations with trigger response time were observed in channels Cz (Day 1 and 3), P3 (Day 1 and 3), Pz (Day 1 and 3). These correlations are illustrated in Figure 4.17 for each day.

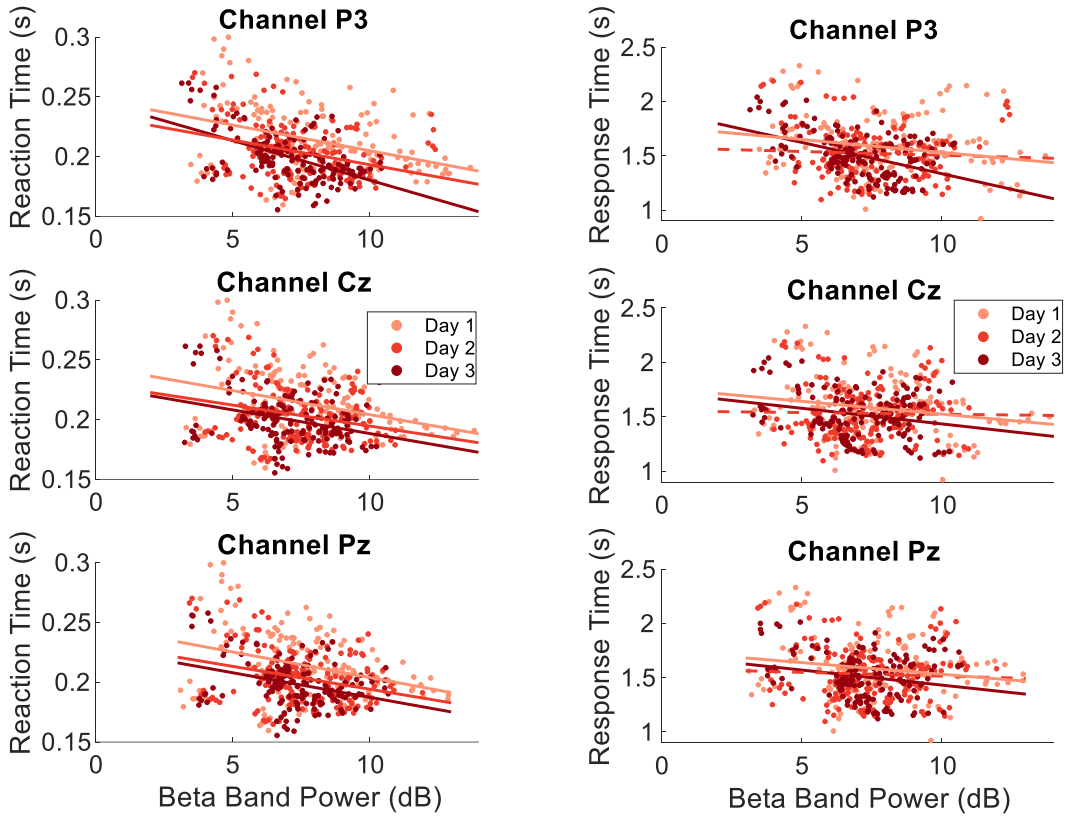

Figure 4.17: Scatter plots comparing the preparatory phase beta band power in channels Cz, P3, and Pz (x-axis) to behavior (y-axis) in terms of reaction time in the left column and trigger response time in the right column. Data points represent 20 (subjects) x 3 (days) x 8 (blocks). A least squares line of best fit indicates the correlation between the variables on each day, where a dashed line indicates that the correlation was not significant ($p > .05$).

The positive correlations between contralateral VEP amplitudes and behavior on each day are illustrated in Figure 4.18. Correlations for channel P3 are shown for the right target trajectory directions, while correlations for channel P4 are shown for the left target trajectory directions. Significant correlations were observed with lower reaction and response times seen for more negative amplitudes in all but one case (trigger response time on Day 3).

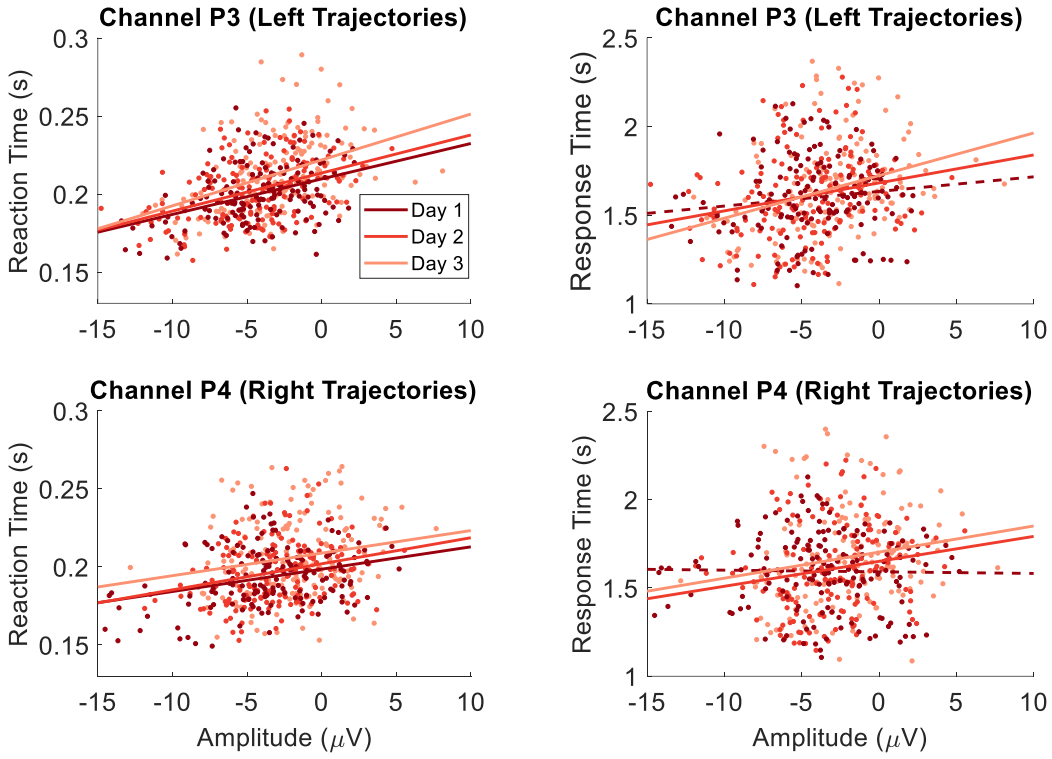

Figure 4.18: Scatter plots comparing the contralateral VEP amplitude (x-axis) to behavior (y-axis) in terms of reaction time in left column and trigger response time in the right column. Data points represent 20 (subjects) x 3 (days) x 8 (blocks). A least squares line of best fit indicates the correlation between the variables on each day, where a dashed line indicates that the correlation was not significant ($p > .05$).

4.4 Discussion

In this study, a simulated trap shooting task was used to investigate the psychophysiological processes underlying motor skill learning in twenty novice participants. Repetitive natural movement patterns were measured using motion tracking as participants practiced shooting at moving targets while their brain activity was recorded with EEG. Over three days of training, participants improved their accuracy by an average of 13.03%.

Analysis of behavior and movement kinematics revealed that performance improvements were accompanied by faster controller reaction times, shorter refinement phase durations, and earlier trigger response times, indicating that less time was needed for motor planning, execution, and error-correction as training progressed. A similar marksmanship task without EEG was performed by Rao et al. [75] in a single-day training experiment with twenty participants, where improvements in accuracy over blocks were accompanied by a significant decrease in the duration of the refinement phase. Our results show that these trends continue over days, with the addition of faster reaction and trigger response times.

This study also revealed important new insights into the cognitive processes accompanying full-body movement during marksmanship training in VR. First, it was shown that a MoBI-style approach is feasible for recording and analyzing EEG during a trap shooting event when conducted in fully immersive virtual reality. Second, an

increase in preparatory phase beta band power observed over blocks during the timeframe before the target launch suggests that cortical activation was modulated through practice. Third, post-launch contralateral VEPs observed in the EEG data over the posterior channels revealed that visual processing of the target occurred before the onset of eye, hand, and head movements. The amplitudes of the VEPs decreased over days, indicating that more robust visual processing occurred through practice. Moreover, preparatory phase beta band power and VEP amplitudes were strongly correlated with behavior, suggesting that better performance was preceded by stronger beta oscillations and more negative VEP amplitudes.

In the next chapter, these EEG and kinematic variables are investigated as features for a machine learning approach to identify biomarkers that are predictive of marksmanship performance on a trial-by-trial basis. Since the psychophysiological processes involved with performing a motor task are expected to involve complex relationships between one another, this approach also highlights predictive combinations of biomarkers that univariate techniques, such as the ANOVA, fail to capture.

5 Biomarker Classification

The overall goal of the simulated marksmanship task presented in the previous chapter was to combine EEG, motion tracking, and VR technology to gain a better understanding of the psychophysiological mechanisms underlying visual-motor skill learning. Visual-motor skill is crucial for the successful execution of tasks that demand a high level of precision, especially in stressful environments such as those in sports, combat, and surgery. As such, a key area of interest in this study was to identify the EEG and kinematic biological markers (*biomarkers*) that describe the psychophysiological processes that are predictive of task success. The findings in the previous chapter showed that significant within-subject differences in the variables extracted from the EEG and motion tracking data occurred through practice and correlated with improvements on the task. However, this approach was only able to identify biomarkers in isolation from the other variables and did not provide predictions on a trial-by-trial basis. Therefore, in this chapter, a machine learning approach was developed to highlight biomarker interactions and to automatically identify the most predictive biomarker combinations of trial-by-trial marksmanship performance.

In machine learning, binary classification models are used to predict which of two classes a set of observed data points belong to. Classification models use features associated with the relevant data to characterize the phenomenon being observed. Using similar variables to those extracted from the EEG and motion tracking data in the previous

chapter as features, two biomarker classification models were developed to categorize single-trial data according to a participant's skill level (top 25% or bottom 25% of participants in terms of accuracy) and shot success (target hit or miss). The goal of these models was to automatically identify the biomarkers associated with better or worse performance on a trial-by-trial basis.

Moreover, biomarker classification models can also be used in conjunction with data-driven feature selection methods to find the combinations of biomarkers that best separate the two classes. By selecting a group of biomarkers that yields the highest classification accuracy, it may be possible to gain important knowledge about the biomarker interactions underlying marksmanship performance. Therefore, in this chapter, the performance of three different classifiers were first evaluated, which was followed by the application of a feature selection method to obtain a final subset of biomarkers that was most predictive of marksmanship performance.

5.1 Methods

Two binary classification tasks were explored using single-trial observations to assess the feasibility of features extracted from the EEG and motion tracking data as potential biomarkers of marksmanship performance. The goals for these classification tasks were to distinguish *skilled* from *unskilled* performance and successful (*hit*) from unsuccessful (*miss*) trials. To do this, data from the five best participants (total accuracy > 75%) and the five worst participants (total accuracy < 60%) were used for skill level

classification, while data from all twenty participants were used for shot success classification.

5.1.1 Preprocessing and Feature Extraction

Data collected during the simulated marksmanship task were divided into three consecutive datasets for separate analyses, where each dataset represented a different psychophysiological process that occurred during a single trial: (1) pre-launch *preparatory phase* EEG data that represents the participant's mental preparation before the target was launched and is characterized by spectral power, (2) post-launch *VEP response* EEG data that reflects visual processing of the target before movement occurred, and (3) post-launch *kinematic response* data recorded as the participant moved to acquire the target. Figure 4.4 illustrates the time frames for which EEG and motion tracking data were analyzed during a single trial. EEG data were not analyzed during movement due to the large artifacts present in the data.

Each of the three datasets were used in separate classification models to highlight the predictive power of the unique biomarkers at each stage of the trial. Additionally, due to significant differences in the post-launch responses between the horizontal trajectory directions (see Figures 4-9 through 4-12), the VEP and kinematic response features associated with side and center trajectories were analyzed separately to avoid bimodal feature distributions. A summary of all five datasets is displayed in Table 5. Any observation with missing data was excluded from the analyses.

Features from the EEG and motion tracking data were extracted in the same manner as the dependent variables in the previous chapter (for details, refer to Section 4.2).

Table 5: Summary of datasets used for evaluating classifier performance. *n* indicates the number of trials used in each classification task.

	Features	*n* (skill level)	*n* (shot success)
Preparatory Band Power	52	14,395	28,521
VEP (side)	100	9,473	18,553
VEP (center)	100	4,675	9,331
Kinematics (side)	15	9,280	18,520
Kinematics (center)	15	4,741	9,476

Preparatory phase EEG data were extracted from the 1s time window following the first cue (i.e., the trap house turning green) before the target was launched. These data were processed to calculate power spectral densities (PSDs) estimates that were then separated into the four standard EEG frequency bands: delta (0.1 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 13 Hz), and beta (13 – 30 Hz). The band powers obtained for each of the 13 EEG channels were concatenated across the channels to generate a vector of 52 spectral features. To capture the VEP response in the post-launch EEG data, features were extracted from the EEG signal from a time window of 200 ms in each channel following the target launch. These signals were first concatenated across the 13 channels to create a features vector of 2600 samples. Due to the high dimensionality of this vector, principal component analysis (PCA) was used to reduce the number of the features. PCA is a technique for dimension reduction that linearly transforms the feature matrix into a lower-dimensional space so

that the variance of the data in the low-dimensional representation is maximized [117]. This created a smaller set of new, linearly uncorrelated VEP features that accounted for > 95% of the variance in the data using 100 principal components.

The kinematic response features were as follows: reaction time (controller and head), peak speed (controller), latency of the peak speed (controller), duration of the ballistic phase, latency of the end of the ballistic phase, angular error at the end of the ballistic phase, speed at the end of the ballistic phase (controller and head), duration of the refinement phase, trigger response time, speed at the trigger response time (controller and head), and total rotation (head).

Finally, an imbalance existed between the classes for shot success (participants hit the target more than they missed) and so a bootstrap resampling technique was applied to oversample the minority class (misses) to avoid biased classification results. Feature standardization was also used to make the values of each feature fit a standard normal distribution with a mean of zero and unit variance.

5.1.2 Classification and Feature Selection

The performance of three different classifier algorithms for use with the five feature sets were compared: logistic regression (LR), naïve Bayes (NB), and random forest (RF) (see [117] for a review). Using Occam's razor principal (use the least complicated algorithm that addresses the problem), LR was selected to provide a baseline for comparison. LR is simple linear classifier that seeks to find a linear decision boundary that

best separates the two classes. NB is also a linear classifier but, unlike LR, it creates a generative model rather than a discriminative model. A generative model learns the joint probability of both the data and labels to explicitly model the distribution of each class (i.e., it models how the data are generated), whereas a discriminative model learns the probability of the label given the data. Generative models such as NB typically perform better than discriminative models when only a limited amount of training data is available by reducing the model's variance, or sensitivity to small fluctuations in the training data, that leads to overfitting. However, as the training dataset grows, discriminative models such as LR usually perform better by providing lower bias, or error from the erroneous assumptions in the algorithm, that leads to underfitting [118].

RF was selected to represent a modern nonlinear classifier. Nonlinear classifier algorithms seek to find more complex nonlinear decision boundaries between the two classes than linear classifiers. RF is a discriminative model, which means it also provides low bias but can create more a complicated decision boundary than LR. RF is an ensemble algorithm that uses a "bagging" method to train several weak classifiers that are then aggregated to create a single strong classifier. In general, the bagging technique used by RF makes it more robust to outliers and noise than LR or other nonlinear classifiers.

To train and test each classifier, a leave-one-subject-out ("subject-wise") cross-validation technique was used to avoid creating dependencies between the training and test sets. Classifier performance was measured using area under the ROC curve (AUC),

where an AUC of 0.5 indicates chance level performance and 1.0 indicates perfect prediction. Classification models were compared using McNemar's test for significance on the aggregate of left-out subject predictions [119].

Finally, the sequential forward floating search (SFFS) algorithm was used for feature selection and the feature subset with the highest AUC was reported. The sequential forward search (SFS) is a simple greedy search algorithm that starts with an empty feature set and iteratively adds the feature that maximizes AUC when combined with the features that have already been selected [120]. The floating variant, SFFS, includes an additional step that allows features to be removed after they have been included so that a larger number of subsets can be tested [78]. While this technique does not guarantee the global optimality of the feature set, it is an efficient method for feature selection with low computational cost. The best feature subsets selected for both classification tasks were compared and linked to the results from hypothesis tests in the previous chapter.

5.2 Results

5.2.1 Classifier Performance

Figure 5.1 compares the performance of the three classifiers in terms of AUC for the skill level and shot success classification tasks. Each classifier was trained and tested with the five separate feature sets described in Table 5 using leave-one-subject-out cross-validation.

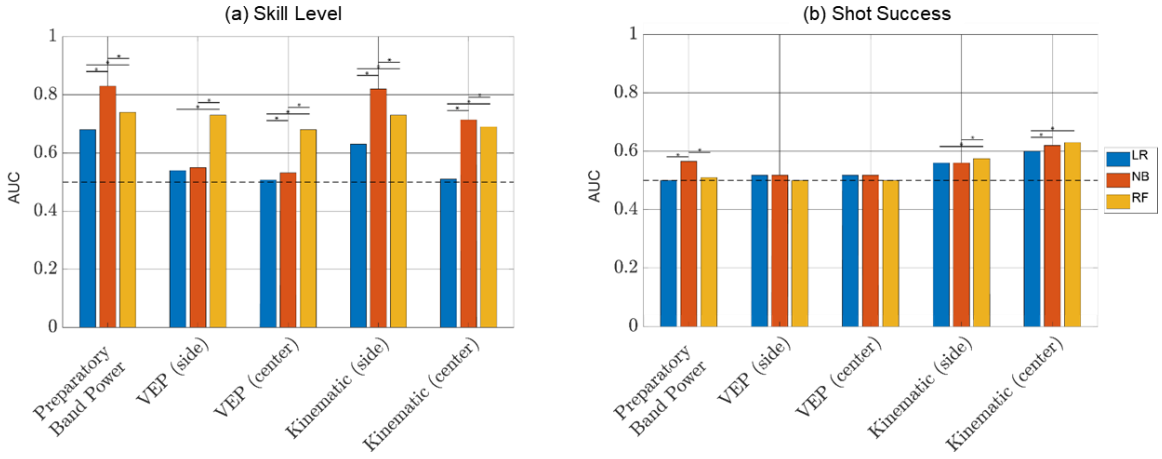

Figure 5.1: Comparison of classifier performance in terms of AUC. The height and color of each bar represent the AUC and classifier type, respectively. Each set of bars on the x-axis represents one of the five feature sets for two classification tasks: (a) participant skill level and (b) shot success. Significant differences in performance are indicated by the lines above each set of bars ($p < .05$). The dashed line indicates chance level performance.

The best performance for the skill level classification task in Figure 5.1a was obtained with NB for the preparatory band power and kinematic features and with RF for the VEP features. For the shot success classification task results in Figure 5.1b, near chance-level performance (AUC ≈ 0.5) was observed for all three classifiers with the VEP features. However, above-chance performance was observed for the shot success classification models with the preparatory band power and kinematic features, indicating that they may contain some predictive power for determining single-trial shot success.

Given the limited number of participants used to train each of the models and the potentially high inter-subject variability of the data, it was important to verify that the

classification models performed better than random guessing. To confirm that the results in Figure 5.1 were not due to chance, participants were randomly relabeled as skilled or unskilled for the skill level classification task and trials were randomly relabeled as hit or miss for the shot success classification task. The same leave-one-subject-out cross-validation method was used for training and testing. This process was repeated 50 times and the average results are displayed in Figure 5.2. A higher AUC for randomly reallocated labels suggests that the AUC with ground-truth labels was due to random guessing.

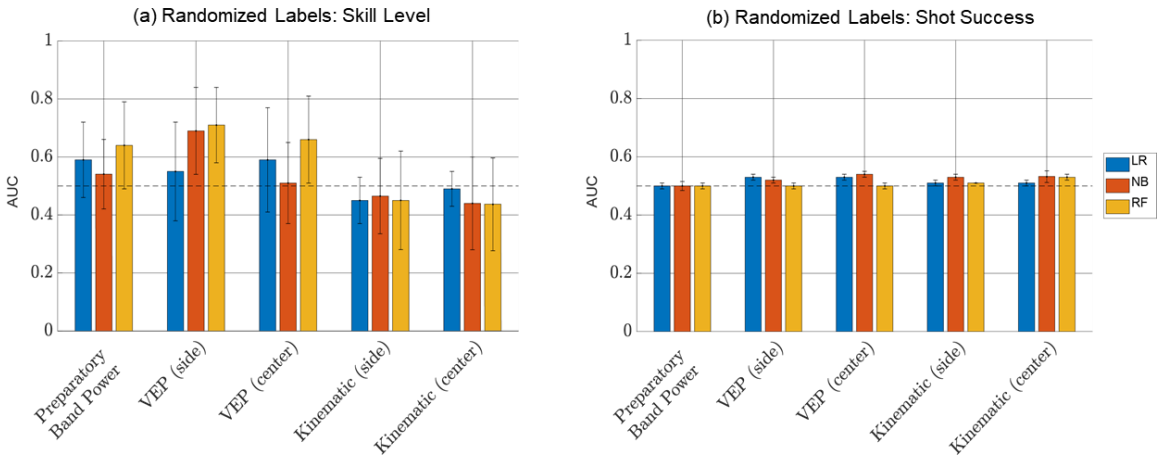

Figure 5.2: Comparison of average classifier performance in terms of AUC using randomly reallocated labels. Error bars indicate ±1 standard deviation of the population of AUCs generated from 50 runs. Each set of bars on the x-axis represents one of the five feature sets for two classification tasks: (a) participant skill level and (b) shot success. Higher AUCs suggest that the classification models with ground-truth labels were randomly guessing.

Comparing Figure 5.2 to Figure 5.1 reveals that label randomization reduced the AUCs for the preparatory band power and kinematic features, indicating that these

models are indeed capturing real differences between the two classes. However, the AUCs for the skill level classification models with VEP features increased with label randomization, suggesting that the models learned differences between individual subjects rather than between the two skill level groups.

5.2.2 Feature Selection

The next step in this investigation was to use the SFFS algorithm for feature selection to identify the combinations of biomarkers that were the most predictive of marksmanship. The NB classifier was selected for use with the SFFS algorithm (SFFS-NB) based on the results from Figure 5.1 and its computational efficiency. Figure 5.3 illustrates how classifier performance changed as features were sequentially added to the feature subset during SFFS. The poor performance results of both classification models with VEP features did not improve with feature selection and, therefore, are not included in this section.

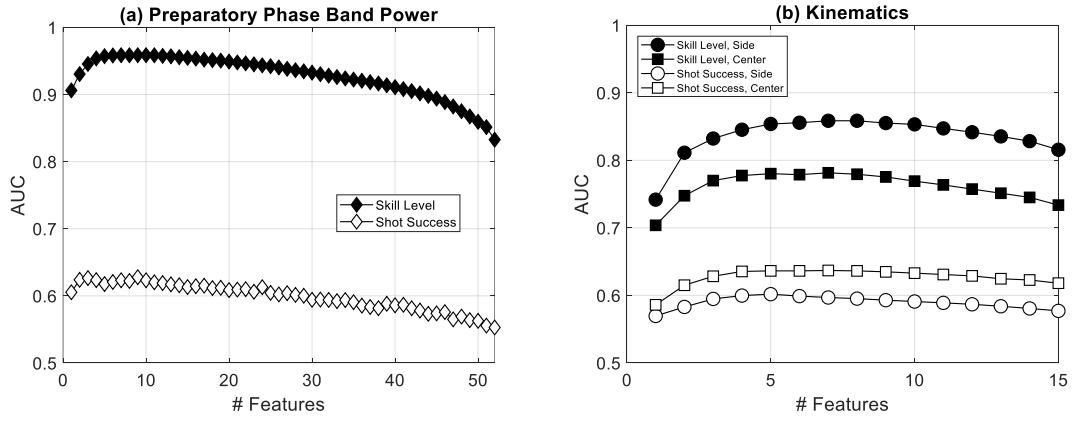

Figure 5.3: Classifier performance during SFFS-NB. The x-axis denotes the number of features selected during the greedy SFFS algorithm and the y-axis denotes

the AUC for the NB classifier. (a) SFFS-NB with preparatory EEG features for the skill level task (black) and shot success tasks (white). (b) SFFS-NB with kinematic features for the skill level (black) and shot success tasks (white), split by horizontal trajectory direction for the side (circle) and center (square) target trajectory directions.

From the performance curves in Figure 5.3, it is clear that the feature subsets obtained with SFFS produce higher AUCs than single features in isolation (1 on the x-axis) and the full feature set (far right on x-axis). The curves reached a maximum AUC at around five features for both classification tasks. The kinematic features in these subsets are listed in Table 6. The preparatory phase band features had many close-to-optimal feature subsets that achieved similar results and, therefore, are not listed.

Table 6: The five kinematic features selected by the SFFS-NB algorithm for the skill level and shot success classification tasks. The numbers represent the stage in the SFFS algorithm when the feature was added.

		1	2	3	4	5
Skill Level	**Kinematic (Side)**	Reaction Time (Controller)	Peak Speed (Controller)	Time of Peak Speed (Controller)	Duration of Ballistic Phase	Speed at Response Time (Head)
	Kinematic (Center)	Reaction Time (Controller)	Reaction Time (Head)	Time of Peak Speed (Controller)	Duration of Refinement Phase	Speed at Response Time (Controller)
Shot Success	**Kinematic (Side)**	Reaction Time (Controller)	Reaction Time (Head)	Duration of Refinement Phase	Speed at Response Time (Controller)	Speed at Response Time (Head)

	Kinematic (Center)	Reaction Time (Controller)	Peak Speed (Controller)	Duration of Ballistic Phase	Duration of Refinement Phase	Speed at Response Time (Controller)

The overlapping kinematic features between the subsets in Table 5 were as follows: Reaction time of the controller was the first feature selected in all four subsets. Using an unpaired two-sample t-test to compare the class distributions, skilled participants had significantly faster reaction times than unskilled participants for both the center and side trajectory directions ($p < .001$). Similarly, the reaction times were significantly faster for successful trials than unsuccessful trials ($p < .001$). Duration of the refinement phase and the controller's speed at the trigger response time appeared in three of the four kinematic feature subsets, not including the side trajectories for the skill level classification task. Significantly shorter durations and lower peak speeds at the response time were observed for skilled participants and successful trials when compared to unskilled participants and unsuccessful trials ($p < .001$). The speed of the head at the response time was selected for both tasks with center trajectory features, with lower speeds occurring for skilled participants and successful trials ($p < .01$). Finally, the time of the controller's peak speed appeared in both feature subsets for skill level classification, where skilled participants reached the peak speed sooner than unskilled participants ($p < .001$).

5.3 Discussion

In this chapter, it was demonstrated that efficient feature extraction and classification, when combined with feature selection, has the potential to identify combinations of EEG and kinematic biomarkers that are predictive of marksmanship performance on a trial-by-trial basis. These findings suggest that subsets of preparatory phase band power and kinematic features may possess predictive power for classifying marksmanship performance, particularly for differentiating between skilled and unskilled participants.

Perhaps the most feasible biomarker of marksmanship performance identified in this work was the reaction time of the controller, which was shown to significantly decrease with practice in Chapter 4 (Figure 4.9) and appeared as the first feature selected by the SFFS-NB in all the feature subsets (Table 6). In addition, the duration of the refinement phase also significantly decreased with practice in Chapter 4 (Figure 4.12) and appeared in three of the four selected feature subsets with SFFS-NB. An obvious trend did not appear to be present in the preparatory band power features selected by the SFFS-NB algorithm. However, given the high performance of the skill level classifier using these features, it is possible that further analysis could reveal more distinctive patterns of brain activity in proficient shooters. Additionally, it's possible that other feature selection methods (e.g., filtering or embedded methods) would select different feature subsets with

a different number of features and, therefore, should be explored to substantiate these features as biomarkers.

For the shot success classification task, poorer performance was observed with only slightly above chance-level AUCs for both the EEG and kinematic feature sets. This suggests that either the differences in preparatory brain states, VEPs, and kinematic behavior do not change on a trial-by-trial basis or that more data are needed to mitigate the effects of noise in the biosensor data. Due to the sensitivity of biosensors to interference and the complexity of human motion, most EEG and motion tracking data contain large amounts of noise. Learning from noisy input data is not trivial because these erroneous patterns can obscure patterns that are pertinent to the task at hand. Classifiers can sometimes end up learning these noise patterns, which reduces their ability to generalize to unseen data. Thus, future studies with more training data are needed to validate the stability of the results presented in this chapter. Furthermore, single-trial prediction may require a different approach in order to capture the minute intricacies of human motion that determine whether a participant will hit the target or not. Therefore, in the next chapter, methods for modeling the continuous movement signals are investigated that do not rely on hand-crafted features that may not be suitable for single-trial prediction of marksmanship success.

6 Controller Motion Trajectory Prediction for Real-time Performance Estimation

Biomarkers describing the psychophysiological processes observed during the simulated marksmanship task have the potential to be used to characterize visual-motor skill, which may be useful for the design of closed-loop feedback systems that can accelerate skill learning during training by delivering instructions to trainees either offline or in real-time (e.g., in static marksmanship [57]). However, in the previous chapter, the poor trial-by-trial classification performance of the psychophysiological biomarkers extracted from the EEG and motion tracking data suggested that these biomarkers may not have captured the more complex characteristics of human motion. Therefore, in this chapter, an alternative approach to estimating single-trial success was explored that uses trajectory prediction techniques to model the data in a spatiotemporal domain, which avoids some of the limitations posed by hand-crafted features.

The goal of trajectory prediction is to estimate a moving object's future position using knowledge of its past movement. In previous research, methods for trajectory prediction have been used for estimating vehicle patterns [79]–[82], robotic motion [83]–[85], and pedestrian paths in surveillance videos [86]–[89]. In the context of this research, trajectory prediction was used to build a model that estimated future positions of the firearm controller to determine whether the controller's virtually projected ray would intersect with the simulated clay pigeon before the shot was taken. Effectively, these position estimates produce predictions of single-trial success (i.e., target hit or miss) using

the raw tracking data directly, eliminating the need to manually design biomarkers that may not fully describe the complex data.

Perhaps the simplest method for trajectory prediction is the Kalman filter, which is widely used in the field of linear robotics for motion planning and control [83]. The Kalman filter is a parametric technique that uses *a priori* knowledge of the movement dynamics to develop equations that calculate predictions. However, parametric methods such as the Kalman filter are often too restrictive when varying degrees of uncertainty are present in the data: human motion (as opposed to robotic motion) can vary greatly based on environmental factors and sensory feedback. Therefore, it is likely that more complex nonparametric methods are needed to capture the dynamics of controller motion that are not known *a priori*.

Other approaches for trajectory prediction include nonparametric and semi-nonparametric methods such as Gaussian process (GP) regression [94] and recurrent neural networks (RNN) [96]. These methods provide more flexibility in modeling variable human motion data as they rely on previously observed motion trajectory patterns to make more informed predictions. In applications that involve modeling pedestrian paths [86], [88] and vehicle motion [80], [81], both methods have been shown to provide better long-term prediction performance than traditional methods such as the Kalman filter. Therefore, in this work, adaptations of the GP and RNN models for long-term controller

motion trajectory predictions were investigated, with the Kalman filter model serving as the baseline to evaluate performance.

The development of an efficient trajectory prediction model could have important implications for visual-motor skill learning during simulation-based training tasks with repetitive movements such as trap shooting. Practice by itself doesn't ensure task success because repetition with poor technique can lead to prolonged performance deficiencies; hence, an effective closed-loop feedback system that uses trajectory prediction to alert trainees in real-time about when shooting is suboptimal might have the ability to correct these deficiencies and thereby accelerate skill learning. The findings from this work illustrate the potential for a trajectory prediction model to provide early estimates of single-trial marksmanship success by predicting whether the trajectory of the controller's virtually project ray will intersect with the simulated clay pigeon before the shot is taken. Therefore, this work represents a first step towards the ultimate goal of using trajectory prediction in a sensor-based feedback system to highlight areas for improvement and relate the information to the trainee in real-time.

6.1 Short-Term Prediction using Kalman Filters

Estimating the future state of a dynamic object (e.g., position, orientation, and velocity) is a fundamental problem in many real-world applications. Bayes filters are one approach to this estimation problem because they recursively estimate a posterior probability distribution over the object's state to predict future states with some amount

of uncertainty. Perhaps the most well-studied parametric implementation of a Bayes filter is the Kalman filter, which represents the posterior distribution over the estimated state by assuming a Gaussian distribution.

At a high level, the Kalman filter produces estimates of a moving object's future state by applying a linear operator to the current state in order to estimate a new state (see Section 2.3.3.2 for a more detailed review). The application of a linear operator to a Gaussian state produces another Gaussian state; hence, each state in the process can be represented by the mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ of its distribution. Repeated application of this process allows for estimates of the state to be calculated many time steps into the future.

Given its simple and computationally efficient design, the Kalman filter was used as a baseline prediction method to estimate future states of controller during the simulated marksmanship task. The state of the controller ($\boldsymbol{s}_t$) at discrete time step *t* consisted of the sensor's position (x_t, y_t, z_t), orientation in spherical coordinates defined by azimuth and elevation angles (θ_t, ϕ_t), linear instantaneous velocity $(\Delta x_t, \Delta y_t, \Delta z_t)$, and angular instantaneous velocity $(\Delta\theta_t, \Delta\phi_t)$.

$$\boldsymbol{s}_t = [x \;\; y \;\; z \;\; \theta \;\; \phi \;\; \Delta x \;\; \Delta y \;\; \Delta z \;\; \Delta\theta \;\; \Delta\phi]_t^T \qquad \textbf{(6-1)}$$

The linear state transition function for the Kalman filter was expressed by

$$\boldsymbol{s}_{t+1} = F\boldsymbol{s}_t + w_{\boldsymbol{t}} \qquad \textbf{(6-2)}$$

where F represents the transition matrix and w represents additive Gaussian noise with mean zero and variance $Q = 1e - 7$. The value for Q was tuned manually to optimize prediction performance. The transition matrix F for the marksmanship task was determined using the standard kinematic equation (i.e., $\boldsymbol{s}_{t+1} = \boldsymbol{s}_t + \Delta\boldsymbol{s}_t$) and had the following form:

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \textbf{(6-3)}$$

The goal of the utilizing the Kalman filter in the marksmanship task was to estimate the state of the controller at the time of the trigger response (time step r) on a given trial. Prediction performance was evaluated using k different initial starting points leading up to the trigger response at time r. This simply means that the initial state for each starting point ($\boldsymbol{s}_{r-k}$) occurred k time steps before the trigger was pulled and, therefore, the goal was to predict the subtrajectory from time $r - k$ to time r. Performance was measured using the closest position on the controller's ray to the center of the target at time r in Cartesian coordinates ($\boldsymbol{p}_r = [h_r, k_r, l_r]$), which was calculated using the position and orientation of the controller's sensor in $\boldsymbol{s}_r$ (see Section 2.3.4.1 for the derivation of $\boldsymbol{p}$). An example of a controller motion subtrajectory prediction with the

Kalman filter is shown in Figure 6.1, where the initial starting point occurred 1 s ($k = 60$ time steps) before the trigger response at time r.

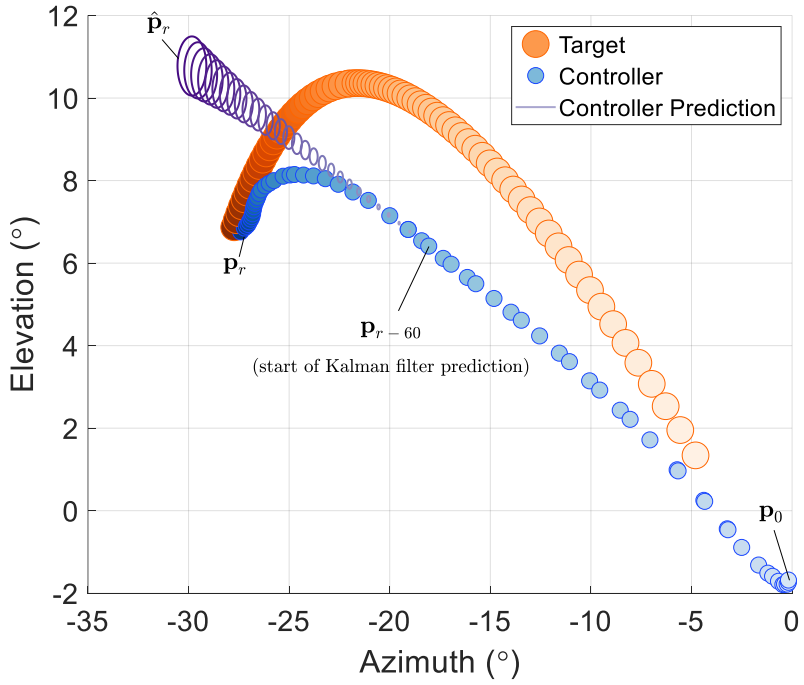

Figure 6.1: Example of long-term controller motion prediction using a Kalman filter to estimate the subtrajectory at starting 60 samples (1 s) before the trigger pull at time r. The purple ellipses display the 2σ confidence bound on the predicted controller position at each time step, where σ represents standard deviation. The large orange circles and smaller blue circles display the target positions and true controller positions, respectively. The color gradient of each trajectory changes from light at time 0 (bottom right corner) to dark at time r (upper left corner) to represent the start and end of the trajectory, respectively.

In order to compare the Kalman filter's prediction performance across all k starting points, the trivariate normal distribution representing the predicted controller position ($\hat{\boldsymbol{p}}_r = [\hat{\boldsymbol{h}}_r, \hat{\boldsymbol{k}}_r, \hat{\boldsymbol{l}}_r] \sim N_3(\boldsymbol{\mu}, \Sigma)$) was integrated over a cubic cell with an edge length of 4 ft, centered around the true position of the controller ($\boldsymbol{p}_r = [h_r, k_r, l_r]$) at time r by

$$q = \int_{l_r-2}^{l_r+2} \int_{k_r-2}^{k_r+2} \int_{h_r-2}^{h_r+2} f(\hat{h}_r, \hat{k}_r, \hat{l}_r)\, \partial\hat{h}_r \partial\hat{k}_r \partial\hat{l}_r \qquad \textbf{(6-4)}$$

where $f(\hat{h}_r, \hat{k}_r, \hat{l}_r)$ represents trivariate normal probability density function over the prediction and the bounds are a function of the true closest position $\boldsymbol{p}_r$. For efficiency, this triple integral was approximated using the method proposed in [121]. Figure 6.2 illustrates how q was computed at time r. A comparison of the average prediction performance across starting points is displayed in Figure 6.3.

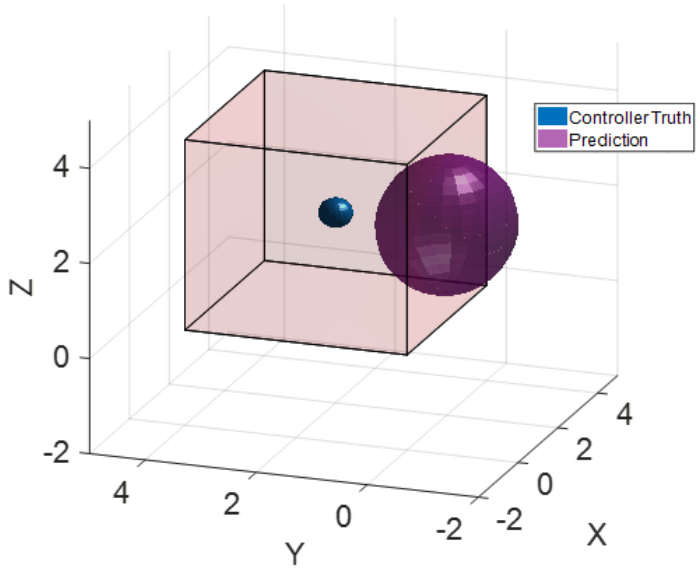

Figure 6.2: The performance measure q was computed as the cumulative distribution of the prediction (large purple sphere) bounded by the faces of a cube whose center is controller's true position (small blue sphere).

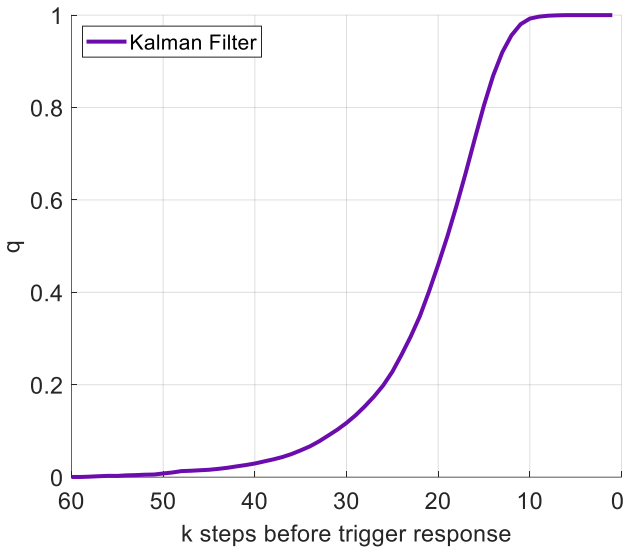

Figure 6.3: Long-term controller prediction performance with the Kalman filter. k on the x-axis represents the number of prediction steps before the trigger was pulled (i.e., the length of the predicted subtrajectory). q on the y-axis represents the cumulative distribution over a cell around the controller's true position at time r, where a higher value indicates better performance. Results were averaged across trials, participants, and trajectories.

From the example in Figure 6.1, it is clear that the Kalman filter did not capture the nonlinear patterns of controller motion that were needed to facilitate long-term prediction for 1s leading up to the trigger response. The errors caused by the linear motion and constant velocity assumptions propagated in the estimates, which led to more inaccurate predictions as time progressed.

The negative impact of these assumptions on long-term prediction is illustrated in in Figure 6.3, where the Kalman filter failed to accurately predict the state of the controller beyond ~.5 s ($k = 30$ time steps). However, in the short timeframe right before the trigger

response (i.e., closer to $k = 0$), the Kalman filter performed relatively well. This short-term success was due to the limited amount of activity occurring during the refinement phase of the controller movement, in which case the Kalman filter's linear and constant velocity assumptions were accurate. However, when the filter was initiated during the earlier and more volatile ballistic phase of the movement, these assumptions led to noticeably incorrect predictions, which was expected given the nonlinear shape of the full trajectories.

Thus, despite its efficient design, the Kalman filter was not able to accurately predict the complicated controller trajectory patterns due to its restrictive assumptions and vague parametric description, especially during the unstable early stages of the movement. In order to capture the more complex relationships needed for long-term prediction, more sophisticated models for trajectory prediction are needed.

6.2 Long-term Prediction using Gaussian Processes and Bayes Filters

To overcome the limitations associated with parametric methods, the use of nonparametric Gaussian processes (GPs) for prediction has been considered in a variety of applications with trajectory data (e.g., vehicles [80], mobile robots [84], pedestrian paths [86]). One of the main advantage of GPs is that they are able learn a distribution over all of the underlying functions that pass through a given data point using a set of training data. This simply means that instead of using a parametric function to describe the data, the output of the GP model is a Gaussian distribution representing the expected (most

likely) function with some confidence value that is consistent with the training data. This is particularly useful for modeling complex human motion, where uncertainty in movement can be explicitly modeled. For a more comprehensive description of the GP model structure and training procedure, see Section 2.3.4.3.

For the simulated marksmanship data, GP models were trained to predict the controller's instantaneous velocities ($\Delta \boldsymbol{s}_t = \boldsymbol{s}_{t+1} - \boldsymbol{s}_t = [\Delta x_t, \Delta y_t, \Delta z_t, \Delta \theta_t, \Delta \phi_t]$) given its position and orientation ($\boldsymbol{s}_t = [x_t, y_t, z_t, \theta_t, \phi_t]$) at time t and some training data $D = \{\boldsymbol{s}_i, \Delta \boldsymbol{s}_i\}_{i=1\ldots n}$ where n is the number of training trajectories (note: the definition of the state $\boldsymbol{s}_t$ for the GP is different from the definition of the state for the Kalman filter). The posterior distribution over the predicted velocity values was expressed as

$$P(\Delta \boldsymbol{s}_t | \boldsymbol{s}_t, D) = \mathcal{N}\left(GP_\mu(\boldsymbol{s}_t, D), GP_\Sigma(\boldsymbol{s}_t, D)\right) \qquad \textbf{(6-5)}$$

Here, GP_μ and GP_Σ are the posterior process mean and variance functions that represent the prediction of the controller's instantaneous velocity at time t and the uncertainty of this prediction. An example of the outputs from a trained GP controller motion model for the upper left target trajectory is illustrated in Figure 6.4, where the training trajectories began in the bottom right and moved towards the upper left of the figure. In this figure, the mean predictions of the instantaneous velocity follow the same direction as the training trajectories, while uncertainty in the predicted velocity increases farther away from the training data.

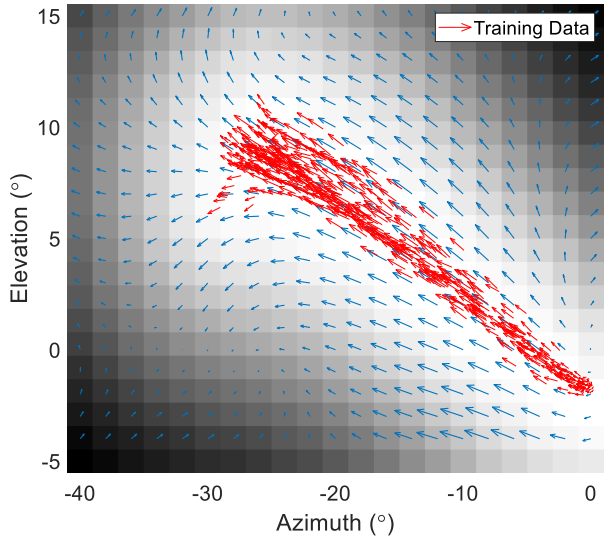

Figure 6.4: An example of a trained Gaussian process motion model for the controller trajectories in the upper left target trajectory direction. The arrows represent the predicted instantaneous velocity and the shading represents the variance of the prediction, where lighter areas indicate lower uncertainty. Training trajectory data are shown in red, which begin near the bottom right and end near the upper left of the figure.

Using the posterior distribution of the trained GP prediction model in (6-5), estimates of the controller's state at the next time step $t + 1$ could be computed by

$$P(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, D) = \mathcal{N}\left(\boldsymbol{s}_t + GP_\mu(\boldsymbol{s}_t, D), GP_\Sigma(\boldsymbol{s}_t, D)\right) \quad \textbf{(6-6)}$$

However, (6-6) has no numerical solution unless the state at time t is known exactly. When making a prediction that is more than two time steps into the future, the state at the previous time step is not known. To solve this problem, Ko and Fox [84] introduced the idea of GP-BayesFilters that integrates the GP model into different instances of Bayes filters for recursive estimation of the state. The Gaussian distribution over the predicted

state provided by the output of the GP model allows it to easily be incorporated into the recursive framework of a Bayes filter, which is then able to propagate the distribution through the GP model.

In this section, two instances of GP-BayesFilters were investigated for long-term controller motion trajectory prediction. Both instances used the same trained GP models for integration with a different Bayes filter. GP models were trained using participant-specific data for each target trajectory direction on a single day, creating 20 (subjects) x 3 (days) x 6 (trajectory directions) datasets of 80 controller motion trajectories each. Each GP model was trained on the first 40 trajectories in each dataset, which were downsampled by 2 to reduce computational complexity during training. The latter 40 trajectories were used for testing with the GP-BayesFilter structure to produce predictions of their subtrajectories.

6.2.1 Gaussian Process + Particle Filter

The GP motion model describes the complex paths that the controller can take during the simulated marksmanship task. In this section, a method based on the nonparametric particle filter (PF) [83], also known as Sequential Monte Carlo, was used to recursively propagate the predicted posterior distribution over the controller's state through the GP motion model.

The approximate posterior distribution for the PF is represented by a set of samples ("particles") drawn from this posterior, which allows for a wide range of

distributions to be represented (e.g., non-Gaussian). Such representations are not restricted by a parametric form, which is ideal for estimating posterior distributions that can sometimes be intractable. This flexibility makes the PF the most versatile of all Bayes filters.

The integration of the PF algorithm into the GP motion model, termed GP-PF, was used to facilitate long-term controller trajectory prediction. To begin, an initial starting point on the trajectory (k time steps before the trigger response at time r) was represented by a set of 250 particles drawn from $P(\boldsymbol{s}_{r-k}|\boldsymbol{s}_{r-k-1}, D)$. At each consecutive time step until time r, each particle was updated by drawing a sample from its posterior distribution provided by the GP model in (6-5) and adding it to its previous value. An example of how the particles evolved during the marksmanship task is illustrated in Figure 6.5, where the initial starting point of the filter occurred 1 s ($k = 60$ time steps) before the trigger response.

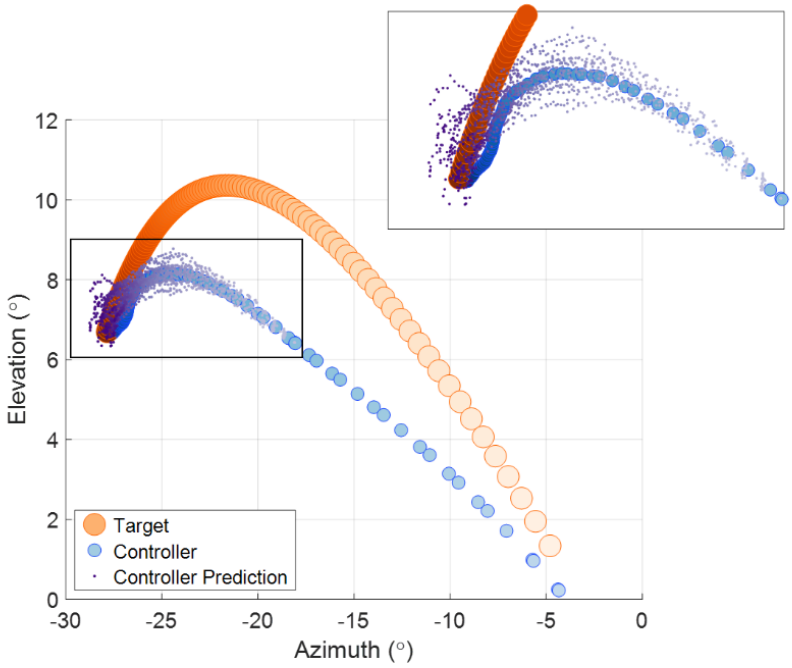

Figure 6.5: Example of long-term prediction using a GP-PF with 20 particles, where the predicted subtrajectory started 60 samples (1 s) before the trigger pull. The inset in the upper right corner displays a zoomed-in version of the predictions. The small purple dots represent the approximate distributions of the particles over the predicted controller position at each time step. The large orange circles and smaller blue circles display the target positions and true controller positions, respectively. The color gradient of each trajectory changes from light at time 0 (bottom right corner) to dark at time r (upper left corner) to represent the start and end of the trajectory, respectively.

Prediction performance for each trial was measured as the fraction of the particles residing in the 4 ft cube surrounding the true position of the controller at time r. The average performance results are displayed across k different starting points in Figure 6.6. The Kalman filter results are also included for comparison.

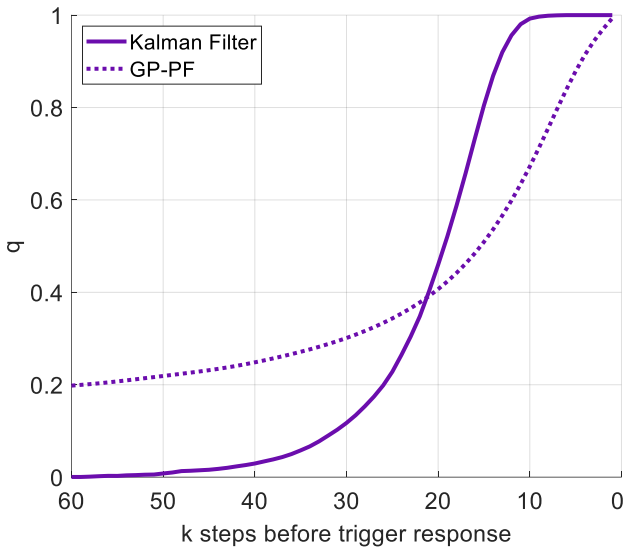

Figure 6.6: Comparison of long-term controller trajectory prediction performance using the GP-PF (dotted line) and Kalman filter (solid line). k on the x-axis represents the number of prediction steps before the trigger was pulled at time r (i.e., the length of the predicted subtrajectory). q on the y-axis represents the cumulative distribution over the cubic cell, or fraction of particles, around the controller's true position. Higher values of q indicate better prediction performance. Results were averaged across trials, participants, and trajectories. The GP-PF performed better than the Kalman filter for long-term predictions beyond $k \approx 21$ time steps.

In the example in Figure 6.5, the general direction of the particles remained fairly consistent with the controller's true nonlinear motion even though the particles dispersed as they moved through the motion model due to the growing uncertainty in the prediction. This is only a single example of controller motion trajectory prediction with the GP-PF algorithm, but it gives some insight as to how much of an improvement in prediction performance this algorithm achieved over the traditional Kalman filter.

From the results in Figure 6.6, it is clear that the GP-PF model was able to make more informed long-term predictions of controller motion than the Kalman filter beyond ~.35 s ($k = 21$ time steps) before the trigger pull, which was the goal of the model. However, these results were still fairly poor considering less than half of the particles remained in the cell beyond .25 s ($k = 15$). Additionally, the high computational cost of performing the GP mean and variance computation for each particle at every time step makes it a poor choice for integration with a real-time feedback system. This cost was defined by Ko and Fox [84] as

$$C_{pf} = M(C_{\mu} + C_{\Sigma}) \qquad \textbf{(6-7)}$$

where M is the number of particles, C_{μ} is the cost of computing the GP mean, and C_{Σ} is the cost of computing the GP covariance. As the covariance of the predicted state grows during long term prediction, the number of particles needed to fill the larger state space also grows and the computational cost increases rapidly. Therefore, a more cost-effective approach is needed since one of the objectives of this model is to eventually use it for real-time prediction during training.

6.2.2 Gaussian Process + Extended Kalman Filter

An alternative to the computationally expensive PF is the more efficient extended Kalman filter (EKF), which is a parametric method that represents the predicted state with a Gaussian distribution [83]. The EKF is an extension of the traditional Kalman filter that

allows for nonlinear transformations from one state to the next, which is achieved through a linearization of the nonlinear transformation function.

The method described by Ko and Fox [84] for integrating the EKF into the GP model was adapted for the controller motion trajectory prediction task as follows. The predicted mean for the controller state $\boldsymbol{s}_{t+1}$ using the GP-EKF algorithm was given directly by

$$\widehat{\boldsymbol{\mu}}_{t+1} = \boldsymbol{\mu}_t + GP_{\mu}(\boldsymbol{\mu}_t, D) \quad \textbf{(6-8)}$$

Here, GP_{μ} represents the mean function defined by the GP model, which was defined in Chapter 2 as

$$GP_{\mu}(\boldsymbol{\mu}_t, D) = k(\boldsymbol{\mu}_t, D_s)^T k(D_s, D_s)^{-1} D_{\Delta \boldsymbol{s}} \quad \textbf{(6-9)}$$

where k represents the kernel function, D_s represents the training data input (position and orientation), and $D_{\Delta s}$ represents the training data output (instantaneous velocity).

For the predicted covariance, a linearization of GP_{μ} (i.e., the nonlinear transformation function) was required to facilitate recursion. To do this, GP_{μ} was approximated by a linear function that was tangent to it at input $\boldsymbol{\mu}_t$. This approximation was computed by taking Jacobian of GP_{μ} with respect to the current state as

$$J_{t+1} = \frac{\partial GP_{\mu}(\boldsymbol{\mu}_t, D)}{\partial \boldsymbol{s}_t} = \frac{\partial k(\boldsymbol{\mu}_t, D_s)}{\partial \boldsymbol{s}_t} k(D_s, D_s)^{-1} D_{\Delta \boldsymbol{s}} \quad \textbf{(6-10)}$$

which was then used to compute the predicted covariance at time $t + 1$ by

$$\hat{\Sigma}_{t+1} = (I + J_{t+1}) \Sigma_t (I + J_{t+1}) + GP_{\Sigma}(\mu_t, D) \quad \textbf{(6-11)}$$

Here, the identity matrix I was added to J_{t+1} because GP_{μ} only represents the change in state from one time step to the next. Repeated application of (6-8) and (6-11) allow the state predictions to propagate through the GP motion model. Figure 6.7 shows an example of how the predicted controller state was propagated through the motion model using the GP-EKF algorithm with an initial starting point of 1 s ($k = 60$ time steps) before the trigger response.

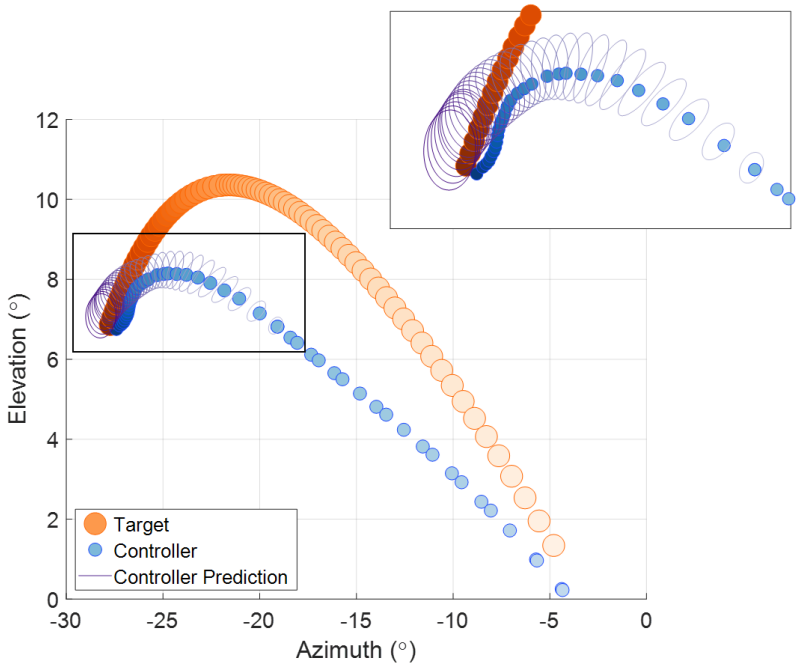

Figure 6.7: Example of long-term controller prediction using the GP-EKF, where the predicted subtrajectory started 60 samples (1 s) before the trigger pull. The inset in the upper right corner displays a zoomed-in version of the predictions. The purple ellipses display the 2σ confidence bound on the predicted controller position at each time step, where σ is standard deviation. The large orange circles and smaller blue circles display the target positions and true controller positions, respectively. The color gradient of each trajectory changes from light at time 0 (bottom right corner) to dark at time r (upper left corner) to represent the start and end of the trajectory, respectively.

Prediction performance on the whole dataset was measured using the same method as used for the traditional Kalman filter in (6-4), where the trivariate normal distribution of the estimated controller position $\hat{\boldsymbol{p}}_{\boldsymbol{r}}$ was approximated over a cubic cell around the true position of the controller $\boldsymbol{p}_{\boldsymbol{r}}$. The average performance results are displayed across k different starting points in Figure 6.8. The performance results for the GP-PF and Kalman are displayed for comparison.

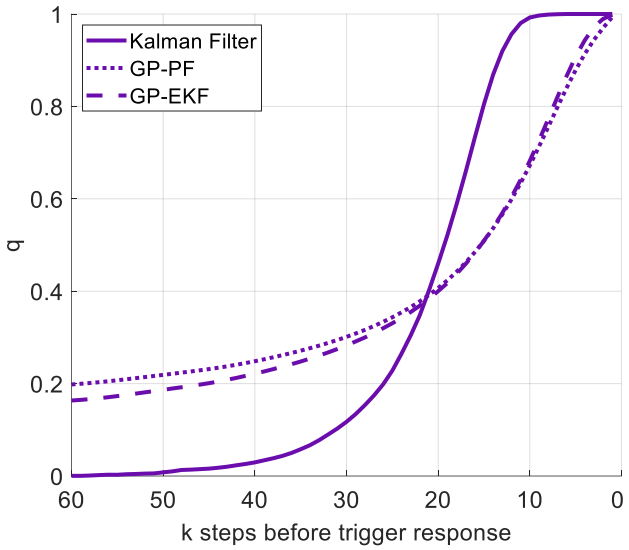

Figure 6.8: Comparison of long-term controller trajectory prediction performance of the GP-EKF (dashed line), GP-PF (dotted line), and Kalman filter (solid line). k on the x-axis represents the number of prediction steps before the trigger was pulled (i.e., the length of the predicted subtrajectory). q on the y-axis represents the cumulative distribution, or fraction of particles for the GP-PF, over a cubic cell around the controller's true position. Results were averaged across trials, participants, and trajectories. On average, the GP-EKF performed slightly better than the GP-PF for long-term predictions beyond $k \approx 20$, but the difference was not significant. The Kalman filter still performed better in the short term ($k \leq 20$).

In the example in Figure 6.7, uncertainty in the predicted mean increased as the state propagated through the GP motion model, as indicated by the growing covariance ellipses. However, a clear improvement over the traditional Kalman filter can still be seen by the nonlinear direction of the state propagation that was consistent with the controller's true trajectory.

The results displayed in Figure 6.7 show that prediction performance did not significantly change with the GP-EKF when compared to the GP-PF. However, the GP-EKF was substantially faster than the GP-PF at test time, which was expected. The computational cost of the GP-EKF was as follows:

$$C_{EKF} = C_{\mu} + C_{\Sigma} + C_{linerization} \qquad \textbf{(6-12)}$$

where C_{μ} is the cost of computing the GP mean, C_{Σ} is the cost of computing the GP covariance, and $C_{linerization}$ is the cost of the linearization operation in (6-10). The GP-EKF required only one GP mean and covariance computation compared to the M particle computations required by the GP-PF in (6-7); thus, the computational cost of the GP-EKF was relatively cheap compared to the cost of the GP-PF because M was much larger than the dimensionality the state space.

However, despite its flexible and computationally efficient design, the GP-EKF only slightly outperformed the traditional Kalman filter in the long term. Over half of the predicted state distributions did not lie within the cell around the controller's true position after ~.25 s ($k = 15$ time steps). Thus, there is still substantial room for performance

improvements to be made using more complex nonlinear methods for trajectory prediction.

6.3 End-to-End Long-Term Prediction using Recurrent Neural Networks

Another nonparametric technique for learning nonlinear motion models that has been gaining traction in recent years is the recurrent neural network (RNN) (e.g., [81], [82], [88], [122]). Like GPs, RNNs use information about previously observed trajectories to make more informed predictions about future trajectories. Additionally, RNNs do not make assumptions about the type of motion, meaning they are able to model both linear and nonlinear motion. One of the main advantages of RNNs over GPs is that they are able to capture the long-term temporal dependencies between points along the trajectory, eliminating the need to incorporate recursive Bayesian methods into their framework (i.e., the model is "end-to-end"). Another advantage of RNNs is that simple modifications to their basic structure can create more complex connections between the input and output, which is ideal for modeling complicated trajectories such as human motion. However, unlike GPs, RNNs are discriminative models, meaning their output is single value rather than a distribution and that uncertainty in controller motion is not explicitly modeled.

The most widely used RNN is the long short-term memory network (LSTM). LSTMs can learn and reproduce long complex sequences, making them particularly effective for trajectory prediction. For the simulation marksmanship task, it was expected that the LSTM would be able learn more complex relationships between the early

controller movements during the ballistic phase and the later movements closer to the time of the trigger pull, leading to enhanced long-term prediction accuracy over the GP-EKF model.

However, the prediction task for the LSTM was framed differently than in the previous section using GPs. In the previous section, the length of the predicted subtrajectory (k) was varied to compare the performance results for both short-term and long-term prediction of the controller's position at the trigger response time r. In this section, k was not varied because that would have required new LSTM models to be trained for each value of k, which was not in the scope of this work given the substantial amount of time it takes to train a single LSTM model. Therefore, instead of predicting the controller position k time steps before time r, the goal of the LSTM was only to predict a future 60 sample (1 s) subtrajectory given the first 30 samples (.5 s) of the full trajectory. This type of LSTM is known as an encoder-decoder LSTM [99]. The input to the network (the first 30 samples) was encoded one sample at a time to learn a single representation of the temporal relationship between the samples. This representation was then decoded one step at time by the network to produce a sequential output (i.e., the prediction of the subsequent 60 samples of the trajectory).

Another difference between the LSTM and GP-EKF models was how data were split into training and testing datasets. Since the LSTM model requires a much larger amount of training data to reliably determine the weights in the network, a leave-one-

subject-out method was employed for training and testing (i.e., train on all subjects except for one and then test on the left out subject). Finally, to avoid overfitting, the dimension of input was reduced to the Cartesian coordinates of the closest position on the controller's ray to the center of the target ($\boldsymbol{p}_t$) instead of the deriving it from the controller's sensor position and orientation ($\boldsymbol{s}_t$).

A manual tuning strategy was used to determine the structure of the LSTM. The best overall performance was obtained with 75 hidden units and the Adam optimizer [123]. It was observed that fewer hidden units did not fully capture the complexity of the movement and that additional hidden units led to overfitting the training data.

An example of the predicted subtrajectory produced by the LSTM is shown in Figure 6.9.

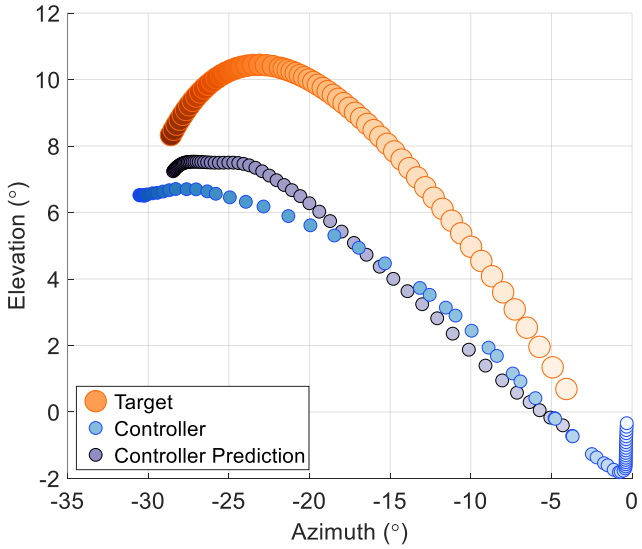

Figure 6.9: Example of a single long-term controller prediction using the LSTM, where the predicted subtrajectory started 30 samples (.5 s) after the target launch. The purple circles display the predicted controller position at each time step. The large

orange circles and smaller blue circles display the target positions and true controller positions, respectively. The color gradient of each trajectory changes from light at time 0 (bottom right corner) to dark at time $t = 90$ (upper left corner) to represent the start and end of the trajectory, respectively.

The LSTM produced single-value predictions of the controller's position rather than distributions over the predictions. Therefore, the performance of the model was measured using the average Euclidean distance between the predicted closest position and the true closest position at time steps 31 through 90 (see Figure 6.10 for an illustrative example). Once again, this was different from the results presented in previous section where the cumulative distribution of the prediction was computed over a cell around the truth at the trigger response time step r. In order to compare the performance LSTM to the GP-EKF, only posterior means of the GP-EKF predictions were used, where the initial starting point for the GP-EKF began 30 samples after the target launch.

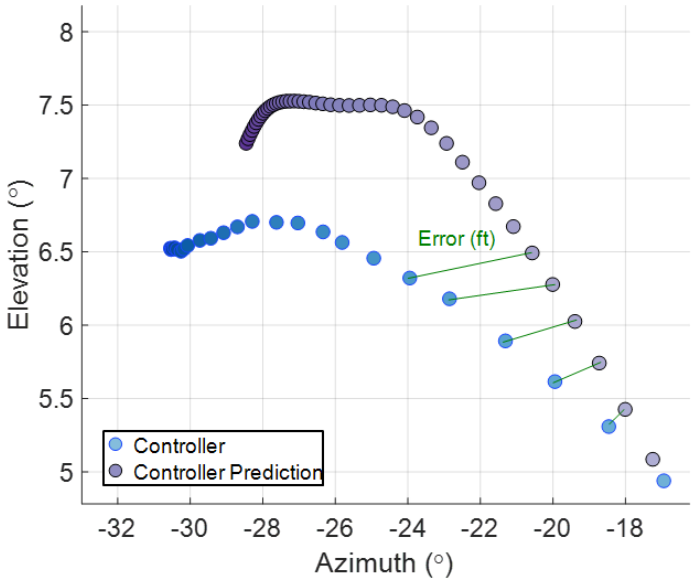

Figure 6.10: Prediction error was calculated as the Euclidean distance between the controller's true position (blue) and the LSTM prediction (purple).

The preliminary performance results for the LSTM and GP-EKF models are shown in Figure 6.11 for the upper left trajectory data. Trajectories shorter than 90 samples (1.5 s) were not included in these results. Two of the twenty total participants pulled the trigger before 1.5 s on every trial and, therefore, were excluded from this analysis.

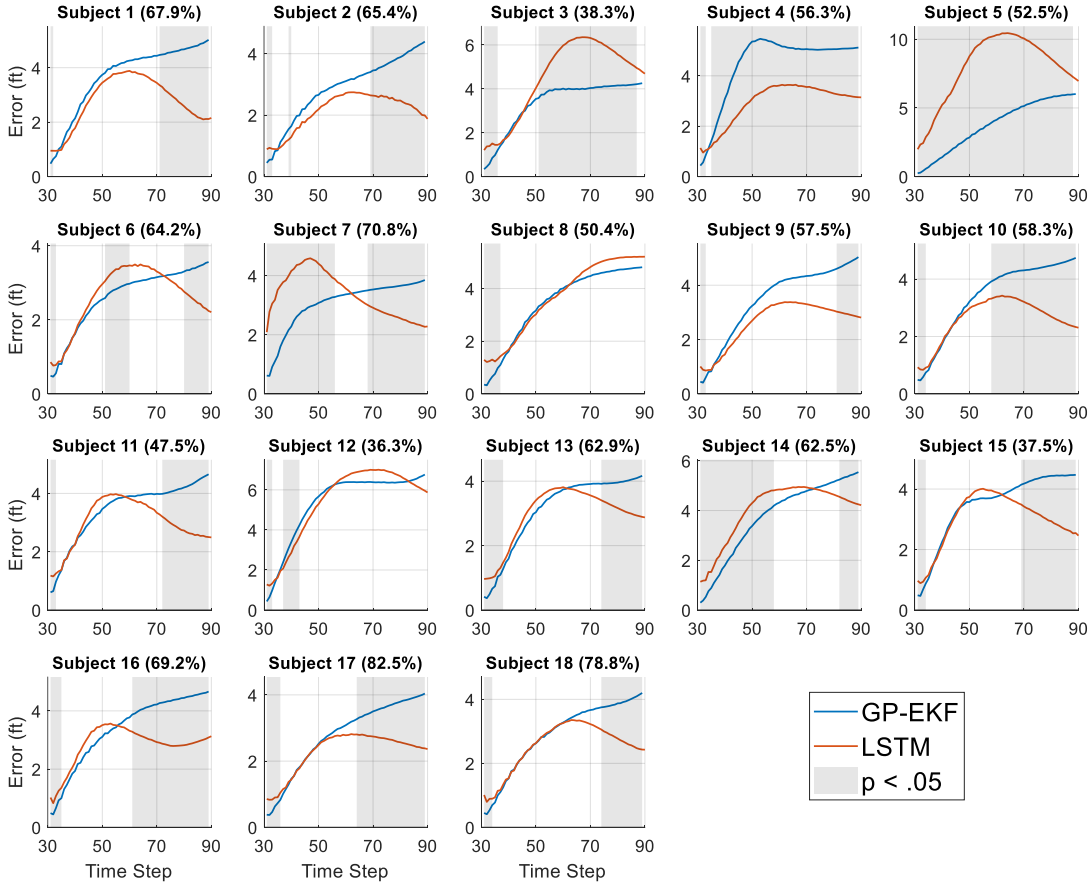

Figure 6.11: Within-subject comparison of long-term controller motion prediction for the GP-EKF posterior mean (blue line) and the LSTM (red line) in the upper left target trajectory direction. The x-axis represents the prediction step (1 step ≈ 16.67 ms) and the y-axis represents prediction error in terms of Euclidean distance between the predicted and true controller positions, where a lower error value indicates better performance. The shaded regions represent significant differences in prediction performance. The title of each subplot includes the participant's overall accuracy in parentheses. Results were averaged across trials.

The results in Figure 6.11 for the upper left trajectory data show that more accurate long-term predictions were obtained for time steps 70 through 90 with the LSTM when

compared to the GP-EKF for fourteen of the eighteen participants. Further analysis of the participants' data where the LSTM performed significantly worse (i.e., participants #3, #5, #7, and #14) revealed physiological differences between their data and the rest of the participants that may have led to poorer performance using leave-one-subject-out cross-validation. For instance, participant #3 utilized a shooting strategy where the controller was angled differently than the other participants during the ballistic phase of the movement, which meant that the movement pattern for this strategy did not exist in the training data. Additionally, participants #5 and #7 exhibited significantly different speeds when compared to the other participants. The average peak speed for participant #5 (35.9 deg/s) was significantly slower than the average peak speed of the group (53.7 deg/s), while the average peak speed for participant #7 (66.7 deg/s) was significantly faster. Participant #14 also exhibited a much slower average peak speed (45.2 deg/s) than the other participants, though the difference was not as drastic as participant #5.

Furthermore, the LSTM error curves in Figure 6.11 appear convex. Each curve reached a maximum error around time step 60 (1s after the target launch), which was around the same time that the refinement phase of the movement was initiated (0.95 s $\pm$ .19). This indicates that prediction error for the LSTM decreased as participants slowed down to begin the less volatile feedback-mediated movements before pulling the trigger to shoot.

6.4 Discussion

Trajectory prediction models can be incorporated into a closed-loop feedback approach that provides individuals with real-time feedback during simulation-based marksmanship training to alert them when shooting might be suboptimal. In this chapter, four different frameworks for predicting firearm controller motion trajectories during a simulated marksmanship task were evaluated. Model performance was compared in terms of prediction accuracy and efficiency, where the goal was to predict the trajectory as early in trial as possible.

The models developed in this work were inspired directly by methods used in previous literature for trajectory prediction, such as those for vehicle and pedestrian tracking (e.g., [79]–[88]). These methods were adapted for a novel application of trajectory prediction to estimate controller motion during the simulated marksmanship task presented in Chapter 4. A search of relevant literature did not yield similar research for this application; thus, this work represents the first instance in which trajectory prediction models were developed for predicting 3D controller motion in VR.

Furthermore, the results from this work demonstrated that nonparametric motion models (GPs and LSTM) outperformed the baseline parametric Kalman filter model by making use of previously observed trajectories to generate more informed nonlinear predictions of long-term controller motion. One of the key benefits of using GPs for long-term prediction was that they were able to explicitly model uncertainty in the predictions,

which is a necessary due the inherent variability of human motion. However, preliminary results for the encoder-decoder LSTM framework suggest that it might be a more accurate method for predicting controller motion than the GP-EKF. Even so, there are several limitations to consider.

It is likely that a more accurate LSTM model could be obtained by making simple modifications to add complexity to the model (e.g., increasing the number of hidden units). However, with the current dataset, adding complexity to the model quickly overfit the training data which led to poorer performance on the testing data. Therefore, a larger dataset is likely needed to ensure a more complex LSTM model's ability to generalize. Another drawback of the LSTM model is that it was not able to model and propagate uncertainty over time. In order to use this model in a closed-loop feedback system during training, knowing the uncertainty associated with the movement is essential for providing accurate real-time instruction to the trainee. Thus, while the preliminary results presented in this chapter showed the potential for LSTMs to learn more accurate motion models than GPs, further exploration is needed before it can be used as an effective training tool.

6.5 Future Work

Methods for improving the performance of the LSTM model for controller motion trajectory prediction should be investigated in future work. First and foremost, if additional data collection is not an option, data augmentation techniques could be explored to create realistic synthetic trajectory data to train more a complex LSTM model

that is capable of learning more complicated motion patterns. For example, DeVries and Taylor [124] showed that extrapolation of the feature space representation in an LSTM autoencoder could be used to generate new sequential data with the same distribution as the real data. LSTM models trained on their augmented dataset outperformed models trained on the original dataset.

Another technique for enhancing the LSTM model is to quantify and propagate predictive uncertainty, which provides a measure of how reliable a particular prediction is. Several approaches to address this issue have been proposed in recent years. Iter et al. [122] trained two LSTMs in a multi-target tracking task, one to predict the next position and another to predict the variance of that prediction. Gupta et al. [89] used a generative adversarial network to generate a variety of "socially-acceptable" trajectories for predicting pedestrian paths in a scene. Al-Shedivat et al. [125] proposed a technique that learned kernels with the LSTM structure for use with a Gaussian process, effectively creating a Bayesian LSTM. While each of these methods takes a different approach to modeling uncertainty, the fundamental idea for developing a predictive distribution remains and a comparison of their uncertainty predictions on the marksmanship task (or any trajectory prediction task, really) could be of interest to the research community.

Finally, methods for incorporating the preparatory phase EEG data into the model should also be explored. Given that the ultimate objective of this research is to develop a closed-loop feedback approach to alert individuals as early as possible when shooting

might be suboptimal, the preparatory phase EEG data might provide the first warning sign of an unsuccessful trial. In Chapters 4 and 5, differences in preparatory phase beta band power were observed through the course of training and between skilled and unskilled participants. This suggests that it may be possible to use this information, along with an enhanced LSTM model for controller trajectory prediction, to create an intelligent closed-loop feedback system that provides real-time instruction to individuals during simulation-based training in VR.

7 Conclusion

The goal of this research was to develop innovative tools using novel signal processing and machine learning techniques for analysis and inference in two separate applications that utilize physiological data collected from human subjects. In Chapter 3, a method for improving the practicality of a BCI communication system using dry electrodes (as opposed to wet) was evaluated in order to reduce the setup time and facilitate daily in-home BCI use for people with severe motor impairments. In Chapter 4, a novel experimental framework was developed and implemented to study the psychophysiological processes underlying human performance and skill learning during a simulated marksmanship task in VR. Biomarkers describing the brain activity and movement kinematics that occurred during the task were extracted from synchronized recordings of EEG and motion tracking data to analyze the within-subject psychophysiological changes that occurred through training. In Chapter 5, a machine learning approach was developed to investigate the combinations of biomarkers that were most predictive of trial-by-trial marksmanship performance in two binary classification tasks for differentiating between participant skill levels and shot success. Finally, in Chapter 6, four time-series methods were evaluated for predicting single-trial shot success using the raw motion tracking data collected during the marksmanship task, where the goal was to estimate whether the controller's virtually projected ray would intersect with the target in real-time.

Chapter 3 focused on improving the performance of a P300 speller system with dry electrodes for EEG recording. Dry electrodes are faster and easier to set up than conventional gel-based wet electrodes but suffer from reduced signal quality due to motion artifacts that arise from poor skin-electrode contact. In this study, a signal processing approach was evaluated as a potential solution to reduce the performance impact of artifact noise in a P300 speller system with dry electrodes. This approach utilized a Bayesian algorithm that automatically controlled the amount of EEG data collected online based on the quality of incoming data. As expected, more data were collected with the dry electrode system when compared to the wet electrode system, implying that the algorithm was able to detect and respond to the need to collect more data in low SNR scenarios. However, P300 speller performance was still significantly reduced for the dry electrode system when compared to wet electrode system, indicating that approaches other than additional data collection are likely needed for full noise mitigation. Two anomaly detection techniques revealed an increased number of low-frequency artifacts in the data collected with the dry electrode system when compared to the wet electrode system. This suggests that slow-wave motion artifacts arising from poor skin-electrode contact may have contributed to the lower SNRs in dry electrode systems, potentially resulting in decreased spelling performance. Therefore, in order to bring the performance of a P300 speller system with dry electrodes to the same level of performance

as that achieved with wet electrodes, methods to mitigate the effects of low-frequency artifact noise should be investigated in future work.

In Chapter 4, a novel experimental framework was presented that combined EEG, motion tracking, and immersive VR to investigate the psychophysiological biomarkers underlying visual-motor skill learning during a multi-day simulated marksmanship training regimen. In tasks that demand a high level of precision, such as those in sports and surgery, visual-motor skills play an important role in accurate and consistent performance. The marksmanship task used in this study, a simulation of trap shooting, required participants to physically interact with their environment using fast, precise controller movements. Over three days of training, significant changes were observed in the biomarkers extracted from the EEG and motion tracking data that correlated with improvements in marksmanship performance. Faster controller reaction times, slower peak speeds, shorter refinement phase durations, and earlier trigger responses were preceded by stronger preparatory EEG beta oscillations and more negative VEP amplitudes, indicating that practice through repetition may have improved marksmanship performance by modulating various neural and kinematic processes.

In Chapter 5, a machine learning classification algorithm was used in conjunction with the marksmanship data presented in Chapter 4 to investigate combinations of biomarkers that were most predictive of trial-by-trial performance. Two binary classifiers were trained to differentiate skilled participants (top 25%) from unskilled participants

(bottom 25%) and successful (target hit) from unsuccessful (target miss) trials. Using feature selection, subsets of preparatory phase EEG and kinematic features where shown to possess the most predictive power for classifying marksmanship performance. However, classification performance was much lower for differentiating between successful and unsuccessful trials than between skilled participants and unskilled participants, suggesting that either the differences in features did not vary on the single-trial level or that more data were needed to mitigate the effects of individual variations and noise in the data. Nevertheless, these results demonstrated ability to use machine learning to investigate biomarker interactions that provided higher classification rates than single biomarkers in isolation; however, more data are needed to ensure the stability of these results.

Finally, in Chapter 6, the performance of four time-series methods were compared for predicting single-trial marksmanship success using the raw motion tracking data to estimate the future positions of the controller given how it had moved in the past. The goal of this part of my thesis was to exploit the more complex spatiotemporal relationships in the data to potentially improve single-trial prediction performance, eliminating the need to manually design biomarkers that may not have captured the more intricate characteristics of visual-motor skill in the previous chapter. The findings from this work revealed that the flexibility of nonparametric Gaussian process regression in combination with recursive Bayes filters allowed the model to perform significantly better than the

baseline parametric Kalman filter. However, preliminary results with long short-term memory networks suggested that the more complex structure of recurrent neural networks was able to better capture the temporal dependencies in the data for longer periods of time, giving rise to improved long-term prediction performance over the Gaussian process models. Given additional training data, it may be possible to train even more advanced neural networks to further improve trajectory prediction performance.

The overall goal of the research presented in Chapters 4 - 6 was to develop various signal processing and machine learning methods to uncover psychophysiological patterns in the data that were predictive of performance during the simulated marksmanship task. These patterns reflect the markers of visual-motor skill that were learned through practice of the task. Using this information, it may be possible to develop a closed-loop feedback system that provides either offline or real-time instruction to individuals during training in order to promote greater attentional focus and more efficient motor behavior. This, along with improved methods for trajectory prediction, should be the focus of future work.

References

[1] E. W. Sellers, T. M. Vaughan, and J. R. Wolpaw, "A brain-computer interface for long-term independent home use," *Amyotroph. Lateral Scler.*, vol. 11, no. 5, pp. 449–455, 2010.

[2] Y. M. Chi, T. Jung, and G. Cauwenberghs, "Dry-contact and Non-contact Biopotential: Methodological Review," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 106–119, 2010.

[3] E. W. Sellers, P. Turner, W. a. Sarnacki, T. McManus, T. M. Vaughan, and R. Matthews, "A novel dry electrode for brain-computer interface," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5611, no. 2, pp. 623–631, 2009.

[4] T. O. Zander *et al.*, "A Dry EEG-System for Scientific Research and Brain-Computer Interfaces.," *Front. Neurosci.*, vol. 5, p. 53, 2011.

[5] C. Grozea, C. D. Voinescu, and S. Fazli, "Bristle-sensors--low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications.," *J. Neural Eng.*, vol. 8, no. 2, p. 25008, 2011.

[6] J. M. Clements, E. W. Sellers, D. B. Ryan, K. Caves, L. M. Collins, and C. S. Throckmorton, "Applying dynamic data collection to improve dry electrode system performance for a P300-based brain–computer interface," *J. Neural Eng.*, vol. 13, no. 6, p. 66018, 2016.

[7] B. Yazmir and M. Reiner, "I act, therefore I err: EEG correlates of success and failure in a virtual throwing game," *Int. J. Psychophysiol.*, vol. 122, pp. 32–41, 2017.

[8] S. Makeig, K. Gramann, T. P. Jung, T. J. Sejnowski, and H. Poizner, "Linking brain, mind and behavior," *Int. J. Psychophysiol.*, vol. 73, no. 2, pp. 95–100, 2009.

[9] J. M. Clements *et al.*, "Neurophysiology of visual-motor learning during a simulated marksmanship task in immersive virtual reality," in *Proceedings of the 25th IEEE Conference on Virtual Reality and 3D User Interfaces*, 2018.

[10] H. Berger, "Über das Elektrenkephalogramm des Menschen," *Arch. Psychiatr. Nervenkr.*, vol. 87, no. 1, pp. 527–570, 1929.

[11] S. J. Luck, "An Introduction to the Event-Related Potential Technique," *Monogr. Soc. Res. Child Dev.*, vol. 78, no. 3, p. 388, 2005.

[12] S. P. Ahlfors, J. Han, J. W. Belliveau, and M. S. Hämäläinen, "Sensitivity of MEG and EEG to source orientation," *Brain Topogr.*, vol. 23, no. 3, pp. 227–232, 2010.

[13] R. S. Menon and S. G. Kim, "Spatial and temporal limits in cognitive neuroimaging with fMRI," *Trends in Cognitive Sciences*, vol. 3, no. 6. pp. 207–216, 1999.

[14] A. Shah and S. Mittal, "Invasive electroencephalography monitoring: Indications and presurgical planning," *Ann. Indian Acad. Neurol.*, vol. 17, no. 5, p. 89, 2014.

[15] F. Sharbrough, G.-E. Chatrian, R. Lesser, H. Lüders, M. Nuwer, and T. Picton, "Guideline thirteen: guidelines for standard electrode position nomenclature. American Electroencephalographic Society.," *J. Clin. Neurophysiol.*, vol. 11, no. 1, pp. 111–113, 1994.

[16] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, a Review," *Sensors*, vol. 12, no. 12, pp. 1211–1279, 2012.

[17] R. S. Khandpur, *Handbook of biomedical instrumentation*. Tata McGraw-Hill Education, 1992.

[18] E. S. Kappenman and S. J. Luck, "The effects of electrode impedance on data quality and statistical significance in ERP recordings.," *Psychophysiology*, vol. 47, no. 5, pp. 888–904, 2010.

[19] P. Griss, P. Enoksson, H. K. Tolvanen-Laakso, P. Merilainen, S. Ollmar, and G. Stemme, "Micromachined electrodes for biopotential measurements," *J. Microelectromechanical Syst.*, vol. 10, no. 1, pp. 10–16, Mar. 2001.

[20] Y. M. Chi, Y. Te Wang, Y. Wang, C. Maier, T. P. Jung, and G. Cauwenberghs, "Dry and noncontact EEG sensors for mobile brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 2, pp. 228–235, 2012.

[21] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. 2012.

[22] G. F. Woodman, "A brief introduction to the use of event-related potentials in studies of perception and attention.," *Atten. Percept. Psychophys.*, vol. 72, no. 8, pp. 2031–46, Nov. 2010.

[23] L. A. Farwell and E. Donchin, "The Truth Will Out: Interrogative Polygraphy ('Lie Detection') With Event-Related Brain Potentials," *Psychophysiology*, vol. 28, no. 5.

pp. 531–547, 1991.

[24] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.

[25] J. Polich, C. L. Ehlers, S. Otis, A. J. Mandell, and F. E. Bloom, "P300 latency reflects the degree of cognitive decline in dementing illness," *Electroencephalogr. Clin. Neurophysiol.*, vol. 63, no. 2, pp. 138–144, 1986.

[26] S. J. Luck, *An Introduction to the Event-Related Potential Technique*, 2nd ed., vol. 78, no. 3, 2005.

[27] E. K. Vogel and S. J. Luck, "The visual N1 component as an index of a discrimination process," *Psychophysiology*, vol. 37, no. 2, pp. 190–203, Mar. 2000.

[28] G. R. Mangun, "Modulations of Sensory-Evoked Brain Potentials Indicate Changes in Perceptual Processing During Visual-Spatial Priming," 2014.

[29] B. Van Den Berg, L. G. Appelbaum, K. Clark, M. M. Lorist, and M. G. Woldorff, "Visual search performance is predicted by both prestimulus and poststimulus electrical brain activity," *Sci. Rep.*, vol. 6, pp. 1–13, 2016.

[30] K. Clark, L. G. Appelbaum, B. van den Berg, S. R. Mitroff, and M. G. Woldorff, "Improvement in Visual Search with Practice: Mapping Learning-Related Changes in Neurocognitive Stages of Processing," *J. Neurosci.*, vol. 35, no. 13, pp. 5351–5359, 2015.

[31] C. M. Hamamé, D. Cosmelli, R. Henriquez, and F. Aboitiz, "Neural mechanisms of human perceptual learning:electrophysiological evidence for a two-stage process," *PLoS One*, vol. 6, no. 4, pp. e19221–e19221, 2011.

[32] B. Fisch, *Fisch and Spehlmann's EEG Primer Basic Principles of Digital and Analog EEG*. Elsevier Health Sciences, 2000.

[33] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal—state-of-the-art and guidelines," *J. Neural Eng.*, vol. 12, no. 3, p. 31001, 2015.

[34] K. T. Sweeney, T. E. Ward, and S. F. McLoone, "Artifact removal in physiological signals-practices and possibilities," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 488–500, 2012.

[35] B. W. McMenamin *et al.*, "Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG," *Neuroimage*, vol. 49, no. 3, pp. 2416–2432, 2010.

[36] B. Z. Allison and J. A. Pineda, "ERPs evoked by different matrix sizes: Implications for a brain computer interface (BCI) system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 110–113, 2003.

[37] N. K. Squires, K. C. Squires, and S. A. Hillyard, "Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man," *Electroencephalogr. Clin. Neurophysiol.*, vol. 38, no. 4, pp. 387–401, 1975.

[38] S. Silvoni *et al.*, "P300-Based Brain-Computer Interface Communication: Evaluation and Follow-up in Amyotrophic Lateral Sclerosis.," *Front. Neurosci.*, vol. 3, p. 60, 2009.

[39] D. J. McFarland and J. R. Wolpaw, "Brain-Computer Interfaces for Communication and Control.," *Commun. ACM*, vol. 54, no. 5, pp. 60–66, 2011.

[40] J. E. Huggins, P. A. Wren, and K. L. Gruis, "What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis.," *Amyotroph. Lateral Scler.*, vol. 12, no. 5, pp. 318–24, 2011.

[41] J. N. Mak *et al.*, "Optimizing the P300-based brain-computer interface: current status, limitations and future directions.," *J. Neural Eng.*, vol. 8, no. 2, p. 25003, 2011.

[42] A. Kübler, E. Holz, T. Kaufmann, and C. Zickler, "A User Centred Approach for Bringing BCI Controlled Applications to End-Users," *Brain-Computer Interface Syst. - Recent Prog. Futur. Prospect.*, pp. 1–20, 2013.

[43] C. Guger, G. Krausz, B. Z. Allison, and G. Edlinger, "Comparison of dry and gel based electrodes for p300 brain-computer interfaces.," *Front. Neurosci.*, vol. 6, p. 60, 2012.

[44] C. Guger *et al.*, "How many people are able to control a P300-based brain–computer interface (BCI)?," 2009.

[45] H. Serby, E. Yom-Tov, and G. F. Inbar, "An Improved P300-Based Brain-Computer Interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 1, pp. 89–98, 2005.

[46] A. Lenhardt, M. Kaper, and H. J. Ritter, "An adaptive P300-based online brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 2, pp. 121–130, 2008.

[47] J. Jin *et al.*, "An adaptive P300-based control system," *J. Neural Eng.*, vol. 8, no. 3, 2011.

[48] C. S. Throckmorton, K. A. Colwell, D. B. Ryan, E. W. Sellers, and L. M. Collins, "Bayesian approach to dynamically controlling data collection in P300 spellers," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 3, pp. 508–517, 2013.

[49] B. O. Mainsah *et al.*, "Increasing BCI communication rates with dynamic stopping towards more practical use: an ALS study," *J. Neural Eng.*, vol. 12, no. 1, p. 16013, 2015.

[50] J. T. Gwin, K. Gramann, S. Makeig, and D. P. Ferris, "Removal of movement artifact from high-density EEG recorded during walking and running," *J. Neurophysiol.*, vol. 103, no. 6, pp. 3526–34, 2010.

[51] J. T. Gwin, K. Gramann, S. Makeig, and D. P. Ferris, "Electrocortical activity is coupled to gait cycle phase during treadmill walking," *Neuroimage*, vol. 54, no. 2, pp. 1289–1296, 2011.

[52] P. De Sanctis, J. S. Butler, J. M. Green, A. C. Snyder, and J. J. Foxe, "Mobile brain/body imaging (MoBI): High-density electrical mapping of inhibitory processes during walking," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2012, pp. 1542–1545.

[53] J. Wagner, S. Makeig, M. Gola, C. Neuper, and G. Muller-Putz, "Distinct Band Oscillatory Networks Subserving Motor and Cognitive Control during Gait Adaptation," *J. Neurosci.*, vol. 36, no. 7, pp. 2212–2226, 2016.

[54] M. Seeber, R. Scherer, and G. R. Müller-Putz, "EEG Oscillations Are Modulated in Different Behavior-Related Networks during Rhythmic Finger Movements," *J. Neurosci.*, vol. 36, no. 46, pp. 11671–11681, 2016.

[55] H. Enders, F. Cortese, C. Maurer, J. Baltich, A. B. Protzner, and B. M. Nigg, "Changes in cortical activity measured with EEG during a high-intensity cycling exercise," *J. Neurophysiol.*, vol. 115, no. 1, pp. 379–388, 2016.

[56] E. Jungnickel and K. Gramann, "Mobile Brain/Body Imaging (MoBI) of Physical Interaction with Dynamically Moving Objects," *Front. Hum. Neurosci.*, vol. 10, p.

306, 2016.

[57] C. Berka, A. Behneman, N. Kintz, R. Johnson, and G. Raphael, "Accelerating Training Using Interactive Neuro- Educational Technologies: Applications to Archery, Golf and Rifle Marksmanship," *Int. J. Sport Soc.*, vol. 1, no. 4, pp. 87–104, 2010.

[58] C. H. Hillman, R. J. Apparies, C. M. Janelle, and B. D. Hatfield, "An electrocortical comparison of executed and rejected shots in skilled marksmen," *Biol. Psychol.*, vol. 52, no. 1, pp. 71–83, 2000.

[59] A. J. Haufler, T. W. Spalding, D. L. Santa Maria, and B. D. Hatfield, "Neuro-cognitive activity during a self-paced visuospatial task: comparative EEG profiles in marksmen and novice shooters," *Biol. Psychol.*, vol. 53, no. 2, pp. 131–160, 2000.

[60] C. M. Janelle and B. D. Hatfield, "Visual Attention and Brain Processes That Underlie Expert Performance: Implications for Sport and Military Psychology," *Mil. Psychol.*, vol. 20, no. sup1, pp. S39–S69, 2008.

[61] G. M. Loze, D. Collins, and P. S. Holmes, "Pre-shot EEG alpha-power reactivity during expert air-pistol shooting: A comparison of best and worst shots," *J. Sports Sci.*, vol. 19, no. 9, pp. 727–733, 2001.

[62] D. J. Crews and D. M. Landers, "Electroencephalographic measures of attentional patterns prior to the golf putt," *Med.Sci.Sports Exerc.*, vol. 25, no. 1, pp. 116–126, Jan. 1993.

[63] C. Babiloni *et al.*, "Golf putt outcomes are predicted by sensorimotor cerebral EEG rhythms," *J. Physiol.*, vol. 586, no. 1, pp. 131–139, 2008.

[64] W. Salazar, D. M. Landers, S. J. Petruzzello, M. Han, D. J. Crews, and K. A. Kubitz, "Hemispheric asymmetry, cardiac response, and performance in elite archers.," *Res. Q. Exerc. Sport*, vol. 61, no. 4, pp. 351–9, 1990.

[65] D. M. Landers, M. Han, W. Salazar, and S. J. Petruzzello, "Effects of learning on electroencephalographic and electrocardiographic patterns in novice archers.," *Int. J. Sport Psychol.*, vol. 25, no. 3, pp. 313–330, 1994.

[66] J. Seo *et al.*, "Stronger activation and deactivation in archery experts for differential cognitive strategy in visuospatial working memory processing," *Behav. Brain Res.*, vol. 229, no. 1, pp. 185–193, 2012.

[67] F. A. dos S. Mendes *et al.*, "Motor learning, retention and transfer after virtual-reality-based training in Parkinson's disease - effect of motor and cognitive demands of games: A longitudinal, controlled clinical study," *Physiotherapy*, vol. 98, no. 3, pp. 217–223, 2012.

[68] B. Abernethy and R. J. Neal, "Visual characteristics of clay target shooters," *J. Sci. Med. Sport*, vol. 2, no. 1, pp. 1–19, 1999.

[69] J. Causer, S. J. Bennett, P. S. Holmes, C. M. Janelle, and A. M. Williams, "Quiet eye duration and gun motion in elite shotgun shooting," *Med. Sci. Sports Exerc.*, vol. 42, no. 8, pp. 1599–1608, 2010.

[70] M. P. Paulus *et al.*, "A neuroscience approach to optimizing brain resources for human performance in extreme environments.," *Neurosci. Biobehav. Rev.*, vol. 33, no. 7, pp. 1080–8, 2009.

[71] D. Elliott, L. E. M. Grierson, S. J. Hayes, and J. Lyons, "Action representations in perception, motor control and learning: Implications for medical education," *Med. Educ.*, vol. 45, no. 2, pp. 119–131, 2011.

[72] F. Vidal, C. Meckler, and T. Hasbroucq, "Basics for sensorimotor information processing: Some implications for learning," *Frontiers in Psychology*, vol. 6, no. FEB. Frontiers Media SA, p. 33, 2015.

[73] K. Krasich, B. Ramger, L. Holton, L. Wang, S. R. Mitroff, and L. Gregory Appelbaum, "Sensorimotor Learning in a Computerized Athletic Training Battery," *J. Mot. Behav.*, vol. 48, no. 5, pp. 401–412, 2016.

[74] J. L. Park, M. M. Fairweather, and D. I. Donaldson, "Making the case for mobile cognition: EEG and sports performance," *Neurosci. Biobehav. Rev.*, vol. 52, pp. 117–130, 2015.

[75] H. M. Rao *et al.*, "Sensorimotor learning during a marksmanship task in immersive virtual reality," *Front. Psychol.*, vol. 9, no. FEB, pp. 1–14, 2018.

[76] E. C. Lee, M. S. Fragala, S. A. Kavouras, R. M. Queen, J. L. Pryor, and D. J. Casa, "Biomarkers in Sports and Exercise: Tracking Health, Performance, and Recovery in Athletes.," *J. strength Cond. Res.*, vol. 31, no. 10, pp. 2920–2937, 2017.

[77] G. Cheron *et al.*, "Brain Oscillations in Sport: Toward EEG Biomarkers of Performance," *Front. Psychol.*, vol. 7, p. 246, 2016.

[78] P. Pudil, J. Novovieova ', and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 1, pp. 119–1, 1994.

[79] H. Weiming, X. Xuejuan, X. Dan, and T. Tieniu, "Traffic accident prediction using vehicle tracking and trajectory analysis," *Intell. Transp. Syst. 2003. Proceedings. 2003 IEEE*, vol. 1, pp. 220–225, 2003.

[80] K. Kim, D. Lee, and I. Essa, "Gaussian process regression flow for analysis of motion trajectories," *Comput. Vis. (ICCV), 2011 IEEE Int. Conf.*, no. 1, pp. 1164–1171, 2011.

[81] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-Sequence Prediction of Vehicle Trajectory via LSTM Encoder-Decoder Architecture."

[82] F. Altché and A. de La Fortelle, "An LSTM Network for Highway Trajectory Prediction," 2018.

[83] D. Thrun, Sebastian and Burgard, Wolfram and Fox, *Probabilistic Robotics*. MIT Press, 2005.

[84] J. Ko and D. Fox, "GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models," *2008 IEEE/RSJ Int. Conf. Intell. Robot. Syst. IROS*, pp. 3471–3476, 2008.

[85] J. Chen, S. Shen, and H. Y. K. Lau, "Hitting flying objects with learning from demonstration," in *2017 18th International Conference on Advanced Robotics, ICAR 2017*, 2017, pp. 55–60.

[86] D. Ellis, E. Sommerlade, and I. Reid, "Modelling pedestrian trajectories with gaussian processes," *Ninth Int. Work. Vis. Surveill.*, pp. 27110–27110, 2009.

[87] S. Kim *et al.*, "BRVO: Predicting pedestrian trajectories using velocity-space reasoning," *Int. J. Rob. Res.*, vol. 34, no. 2, pp. 201–217, 2015.

[88] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 961–971, 2016.

[89] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks."

[90] P. T. Pasqual and J. O. Wobbrock, "Mouse pointing endpoint prediction using kinematic template matching," *Proc. 32nd Annu. ACM Conf. Hum. factors Comput. Syst.*, pp. 743–752, 2014.

[91] G. E. Cox, G. Kachergis, and R. M. Shiffrin, "Gaussian Process Regression for Trajectory Analysis," *Proc. Cogn. Sci.*, 2012.

[92] P. Biswas, G. A. Aydemir, P. Langdon, and S. Godsill, "Intent Recognition Using Neural Networks and Kalman Filters," *Lncs*, vol. 7947, pp. 112–123, 2013.

[93] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *J. Basic Eng.*, vol. 82, no. 1, p. 35, 1960.

[94] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning.*, vol. 14, no. 2. 2004.

[95] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Johns Hopkins University Press, 2013.

[96] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer, 2012.

[97] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[98] P. J. Werbos, "Backpropagation Through Time: What It Does and How to Do It," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[99] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," 2014.

[100] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance.," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15–21, 2008.

[101] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, 2004.

[102] G. Townsend *et al.*, "A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns," *Clin. Neurophysiol.*, vol. 121, no. 7, pp. 1109–1120, 2010.

[103] D. J. Krusienski *et al.*, "A comparison of classification techniques for the P300 Speller.," *J. Neural Eng.*, vol. 3, no. 4, pp. 299–305, 2006.

[104] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing*. Boston: McGraw-Hill, 2000.

[105] L. Bougrain, C. Saavedra, and R. Ranta, "Finally, what is the best filter for P300 detection?," *TOBI Work. lll-Tools …*, vol. 2012, pp. 2–3, 2012.

[106] F. H. Lopes Da Silva, K. Van Hulten, and J. G. Lommen, "Automatic detection and localization of epileptic foci," *Progr.Rep.Med.Fys.Inst.Tno*, no.4, 1976.

[107] I. Clark, R. Biscay, M. Echeverría, and T. Virués, "Multiresolution decomposition of non-stationary eeg signals: A preliminary study," *Comput. Biol. Med.*, vol. 25, no. 4, pp. 373–382, 1995.

[108] K. P. Indiradevi, E. Elias, P. S. Sathidevi, S. Dinesh Nayak, and K. Radhakrishnan, "A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram," *Comput. Biol. Med.*, vol. 38, no. 7, pp. 805–816, 2008.

[109] Y. Zhang, X. Ji, and Y. Zhang, "Classification of EEG signals based on AR model and approximate entropy," *Int. Jt. Conf. Neural Networks*, pp. 1–6, 2015.

[110] I. Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[111] M. Adib and E. Cretu, "Wavelet-based artifact identification and separation technique for EEG signals during galvanic vestibular stimulation.," *Comput. Math. Methods Med.*, vol. 2013, p. 167069, Jun. 2013.

[112] İ. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *J. Neurosci. Methods*, vol. 148, no. 2, pp. 113–121, Oct. 2005.

[113] I. International Shooting Union, "Official Statutes, Rules and Regulations of the International Shooting Union," vol. 2013, no. January 2013, pp. 401–470, 1996.

[114] D. J. Zielinski, H. Rao, N. Potter, L. G. Appelbaum, and R. Kopper, "Evaluating the effects of image persistence on dynamic target acquisition in low frame rate virtual environments," 2016.

[115] B. D. Hatfield, A. J. Haufler, T.-M. Hung, and T. W. Spalding, "Electroencephalographic studies of skilled psychomotor performance.," *J. Clin. Neurophysiol.*, vol. 21, no. 3, pp. 144–156, 2004.

[116] A. B. Nelson *et al.*, "Beta Oscillatory Changes and Retention of Motor Skills during Practice in Healthy Subjects and in Patients with Parkinson's Disease," *Front. Hum. Neurosci.*, vol. 11, p. 104, Mar. 2017.

[117] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. 2006.

[118] A. Y. Ng and M. I. Jordan, "On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes."

[119] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

[120] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997.

[121] A. Genz, "Numerical computation of rectangular bivariate and trivariate normal and t probabilities," Kluwer Academic Publishers, 2004.

[122] D. Iter, J. Kuck, and P. Zhuang, "Target Tracking with Kalman Filtering, KNN and LSTMs," 2016.

[123] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2014.

[124] T. DeVries and G. W. Taylor, "Dataset Augmentation in Feature Space," 2017.

[125] M. Al-Shedivat, A. G. Wilson, Y. Saatchi, Z. Hu, and E. P. Xing, "Learning Scalable Deep Kernels with Recurrent Structure," vol. 18, pp. 1–37, 2016.

Biography

Jillian Clements was born in Raleigh, North Carolina on November 17, 1987. She received her Bachelor of Science degree in Electrical Engineering from North Carolina State University in 2011. While at North Carolina State University, she received a grant from the Office of Undergraduate Research to conduct research under the guidance of Dr. Michael Escuti and was inducted into the IEEE-Eta Kappa Nu Electrical and Computer Engineering honor society. Following a brief career in industry, Jillian returned to academia and obtained a Master of Science degree in Electrical and Computer Engineering from Duke University in 2015 under the guidance of Dr. Leslie Collins. She received her Doctor of Philosophy degree in Electrical and Computer Engineering from Duke University in August of 2018. During her time as a Ph.D. student, she was the recipient of the National Science Foundation WISeNet IGERT fellowship in 2015 - 2017 and authored the following conference and journal publications:

1. **Clements, J. M.**, Kopper, R., Zielinski, D. J., Rao, H., Sommer, M. A., Kirsch, E., Mainsah, B. O., Collins, L. M., and Appelbaum, L. G. (2018). Neurophysiology of visual-motor learning during a simulated marksmanship task in immersive virtual reality. In *Proceedings of the 25th IEEE Conference on Virtual Reality and 3D User Interfaces, March 18 - 22, 2018, Reutlingen, Germany.*

2. Rao, H. M., Khanna, R., Zielinski, D. J., Lu, Y., **Clements, J. M.,** Potter, N. D., Sommer, M. A., Kopper, R., and Appelbaum, L. G. (2018)**.** Sensorimotor learning during a marksmanship task in immersive virtual reality. *Frontiers in Psychology: Movement Science and Sport Psychology*. doi: 10.3389/fpsyg.2018.00058

3. Soriano, M. C., Niso, G., **Clements, J. M**., Ortín, S., Carrasco, S., Gudín, M., Mirasso, C.R. and Pereda, E. (2017). Automated detection of epileptic biomarkers in

resting-state interictal MEG data. *Frontiers in Neuroinformatics*, 11. doi: 10.3389/fninf.2017.00043

4. **Clements, J. M.,** Sellers, E. W., Ryan, D., Caves, K., Collins, L. M., and Throckmorton, C. S. (2016)**.** Applying dynamic data collection to improve dry electrode system performance for a P300-based brain-computer interface. *Journal of Neural Engineering,* 13(6). doi: 10.1088/1741-2560/13/6/066018

5. **Clements, J. M.**, Sellers, E.W., Ryan, D., Caves, K., Collins, L. M., and Throckmorton, C. S. (2016). Analyzing the performance of dry electrodes for p300-based brain-computer interfaces in participants with amyotrophic lateral sclerosis. In *Proceedings of the International Brain-Computer Interface Meeting, May 30 - June 3, 2016, Asilomar, California.* doi: 10.3217/978-3-85125-467-9-45

6. **Clements, J. M**., Collins, L. M., Sellers, E.W., Laughlin, S., and Throckmorton, C. S. (2014). Improving the performance of dry electrodes for p300 brain-computer interfaces. In *Proceedings of the 6th International Brain-Computer Interface Conferenc*e, *September 16 - 19, 2014, Graz, Austria*. Graz University of Technology Publishing House. doi:10.3217/978-3-85125-378-8-89. Poster presentation.