Effects of Interpretation Error on User Learning in Novel Input Mechanisms

A thesis submitted to the
College of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By

Kevin Ca-Wai Lam

©Kevin Ca-Wai Lam, August 2021. All rights reserved.
Unless otherwise noted, copyright of the material in this thesis belongs to the author.

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Disclaimer

Reference in this thesis to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building, 110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C9 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

Novel input mechanisms generate signals that are interpreted as commands in computer systems. Sometimes noise from various sources can cause the system to produce errors when attempting to interpret the signal, causing a misrepresentation of the user's intention. While research has been done in understanding how these interpretation errors affect the performance of users of novel signal-based input mechanisms, such as a brain-computer interface (BCI), there is a lack of knowledge in how user learning is affected. Previous literature in command-based selection tasks has suggested that errors will have a negative impact on expertise development; however, the presence of errors could conversely improve a user's learning by demanding more attention from the user. This thesis begins by studying people's ability to use a novel input mechanism with a noisy input signal: a motor imagery BCI. By converting a user's brain signals into computer commands, a user could complete selection tasks using imagined movement. However, the high degree of interpretation errors caused by noise in the input signals made it difficult to differentiate the user's intent from the noise. As such, the results of the BCI study served as motivation to test the effects of interpretation errors on user learning. Two studies were conducted to determine how user performance and learning were affected by different rates of interpretation error in a novel input mechanism. The results from these two studies showed that interpretation errors led to slower task completion times, lower accuracy in memory recall, greater rates of user errors, and increased frustration. This new knowledge about the effects of interpretation errors can contribute to better design of input mechanisms and training programs for novel input systems.

Acknowledgements

Thank you to everyone who has supported me throughout my Master's degree and everyone who have helped me get to where I am today.

I would like to thank Carl Gutwin, one of my supervisors, for giving me an opportunity to experience the intriguing field of human-computer interaction. Your guidance and advice over the past three years have helped me become a better academic. I will be forever grateful that you took the chance with me, despite my not having a formal background in computer science. I would also like to thank Madison Klarkowski, my other supervisor, for sharing your expertise in physiological signals. Your work on psychophysiological responses in games was inspiring to read about, and I felt truly fortunate that you became one of my supervisors so soon after joining the HCI lab.

I want to thank Andy Cockburn for co-authoring the paper published in this thesis. Your comments were insightful and helped me see new perspectives in our work together.

I need to thank my family, Thanh, Phan, and Bobby, who are my biggest supporters and the main reason why I am able to succeed. Your constant love and support have always been appreciated, especially when you were driving 6+ hours between Calgary and Saskatoon every month to see me. Knowing that I could talk with you whenever I felt lonely or needed reassurance was one of the things that helped me through the toughest times. Although the pandemic brought me home a bit earlier than expected, being able to see you everyday has been the largest blessing.

I would like to thank my friends from Calgary, both the ones who are still here and the ones that have inexplicably ended up on the other side of the world. Being able to chat and play games despite our distances have helped me relieve stress and get past the more difficult parts of this degree. I could not possibly come up with enough words to say how grateful I am, but I will try. Shawn, you have always been a stoic friend to me and have shared in my hobbies ever since we met two decades ago. Getting to talk with you about anything on my mind is a boon that I appreciate. I hope we'll get another chance to travel together to Japan again, since those have been some of my best memories. Kim, even though I've known you for so long, your friendship has been one of the most invaluable things to me when I left home for the first time. Being able to share in your experience with grad school has helped me through my own struggles. I am also glad that I can be there for you when you need someone to talk to. Cindy, playing games with you has been something I looked forward to every weekend. You have always been someone that I could count on to be there when I needed it.

Finally, I want to acknowledge my idol, Miona, for your presence in my life. It was through your encouragement that I decided to take this first step into graduate studies. Your support has been comforting and motivating, and I look forward to cheering you on for your future too.

Contents

Permission to Use . . . . **i**

Abstract . . . . **ii**

Acknowledgements . . . . **iii**

Contents . . . . **iv**

List of Tables . . . . **vi**

List of Figures . . . . **vii**

List of Abbreviations . . . . **ix**

1 Introduction **1**
1.1 Problem 3
1.2 Motivation 3
1.3 Solution 4
1.4 Steps to the Solution 4
1.5 Evaluation 5
1.6 Contribution 6
1.7 Thesis Outline 6

2 Related Work **8**
2.1 Input Mechanisms 8
2.2 Learning 9
2.2.1 Transition from Novice to Expert 9
2.2.2 Retrieval Effort During Learning 12
2.2.3 Interference During Learning 13
2.3 Errors in Input Mechanisms 13
2.3.1 Noise in Input Signals 14
2.3.2 Strategies to Reduce Noise 15

3 Understanding Noise in Signal-Based Input Mechanisms **18**
3.1 Brain-Computer Interfaces 18
3.1.1 Electroencephalography 19
3.1.2 Types of BCIs 20
3.1.3 Control Strategies 21
3.1.4 Using a BCI 23
3.1.5 Interpretation Errors in BCIs 24
3.2 Development of a Motor Imagery BCI 24
3.2.1 EEG Headset Selection 25
3.2.2 EEG Headset Testing 26
3.3 BCI Study: Detection of Mu Waves Using the Ultracortex Headset 28
3.3.1 Study System 28
3.3.2 EEG Data Acquisition and Processing 30
3.3.3 Participants 30
3.3.4 Study Design 30
3.3.5 Procedure 31
3.3.6 Results 32

3.3.7 Discussion . . . 33
3.3.8 BCI Study Summary . . . 38
3.4 Issues of Noise in EEG Data . . . 39

4 Effects of Interpretation Error on Learning . . . 41
4.1 Study 1: Effects of Interpretation Error Rate on Learning a Command Set, with Immediate Feedback . . . 41
4.1.1 Study System and Menu-Selection Task Design . . . 42
4.1.2 Experimental Conditions . . . 43
4.1.3 Participants . . . 44
4.1.4 Procedure . . . 44
4.1.5 Study Design . . . 45
4.1.6 Results . . . 45
4.1.7 Study 1 Summary . . . 49
4.2 Study 2: Effects of Interpretation Error Rate on Learning a Command Set, with User Monitoring System Interpretations . . . 51
4.2.1 Study System and Menu-Selection Task Design . . . 51
4.2.2 Experimental Conditions . . . 52
4.2.3 Participants . . . 52
4.2.4 Procedure . . . 53
4.2.5 Study Design . . . 53
4.2.6 Results . . . 53
4.2.7 Study 2 Summary . . . 57
4.3 Chapter 4 Summary . . . 58

5 Discussion . . . 60
5.1 Summary of Findings . . . 60
5.2 Main Results . . . 61
5.2.1 Why did the increased IER lead to reduced learning and performance? . . . 61
5.2.2 Why was the Zero condition more accurate in Study 1, but faster in Study 2? . . . 62
5.3 Generalizing the Findings to Other Contexts . . . 63
5.4 Implications of Noise on Learning Novel Input Mechanisms . . . 65
5.5 Limitations . . . 66
5.6 Future Work . . . 69

6 Conclusion . . . 73

References . . . 75

Appendix A Consent Form . . . 84

Appendix B Post-Condition Questionnaire . . . 86

List of Tables

4.1 Study 1. Results of the ANOVAs conducted on each TLX effort question after an Aligned Rank Transform. 49
4.2 Study 2. Results of the ANOVAs conducted on each TLX effort question after an Aligned Rank Transform. 56

List of Figures

2.1 An example of the power law of practice [57]. New users begin in the cognitive stage and require the most time to complete tasks. As the users gain practice over time, they will move through the associative stage until they reach the autonomous stage, where they will have the quickest completion times. 11

3.1 The OpenBCI Ultracortex EEG Headset used to study BCIs as a novel input mechanism. . . 26
3.2 The OpenBCI GUI that was used to calibrate the Ultracortex IV Headset. On the left half of the GUI is a real-time visualization of the electrical amplitudes recorded by each electrode over time. Active electrodes show activity and the current voltage being detected by the electrode, while inactive electrodes shows a "RAILED" reading. The graph in the upper right of the GUI shows a real-time Fourier transform of all electrodes in the frequency domain. The lower right display shows a heat map over the user's scalp of the electrical activity detected by the electrodes. 27
3.3 A timeline illustrating the sequence of events taken place during a single trial. EEG activity was recorded during the events that are bounded with a solid box (from 3 seconds before stimulus onset and until 5 seconds after). The inter-trial period was randomized between every trial. 29
3.4 A layout of the International 10-10 system used for EEG electrode placements. The locations marked in green are the electrode locations used in the BCI study. 31
3.5 Raw EEG data collected from a participant over the course of the study. Each node represents one electrode attached to the participant's scalp. 33
3.6 Raw EEG data of a single non-target trial at the Cz electrode, starting at three seconds before stimulus onset and ending two seconds after hand movement. 34
3.7 Averaged EEG activity at each electrode site for a single participant. Stimulus onset occurs at 2 seconds, marked by a dotted line. Non-target trials are plotted in the top row (blue) and target trials are plotted in the second row (red). The third-row plots overlay both target and non-target trials for visual comparison. The bottom row shows the Fourier transform of the respective electrode averages. 35
3.8 Grand averages of the trials for all participant EEG recordings at each of the electrode sites. Stimulus onset occurs at 2 seconds, marked by a dotted line. Non-target trials are plotted in the top row (blue) and target trials are plotted in the middle row (red). The bottom-row plots overlay both target and non-target trials for visual comparison. 36
3.9 Grand averages at each electrode in the frequency domain. The top row shows the Fourier transform of the combined target trials (red) and non-target trials (blue). The bottom row shows the difference waves of the grand frequency averages, with non-target trials subtracted from the target trials. 37

4.1 An example of the visual guide that participants could view using the Left Shift key. The layout of the categories and items indicated the arrow key direction to press to select that category/item. 42
4.2 An example of the system interface that participants would see when completing a trial in the training blocks in Study 1. 43
4.3 Study 1. Mean selection accuracy in the memory tests, by IER and Test. 46
4.4 Study 1. Mean trial completion times in memory tests, by IER and Test. 47
4.5 Study 1. Mean user errors per trial in training blocks, by IER and Block. 48
4.6 Study 1. Mean trial completion times in the training blocks, by IER and Block. 49
4.7 Study 1. Mean NASA Task Load Index scores, by IER. 50
4.8 Study 1. Mean scores for perceived ease of learning and memorization, by IER. 51

4.9 An example of the screen layout for Study 2. Participants selected targets in order from the "Targets" list on the left; selections made by the participants were displayed in the "Selections" box on the right. 52
4.10 Study 2. Mean selection accuracy in the memory tests, by IER and Test. 54
4.11 Study 2. Mean trial completion times in memory tests, by IER and Test. 55
4.12 Study 2. Mean user errors per trial in training blocks, by IER and Block. 56
4.13 Study 2. Mean trial completion times in the training blocks, by IER and Block. 57
4.14 Study 2. Mean NASA Task Load Index scores, by IER. 58
4.15 Study 2. Mean scores for perceived ease of learning and memorization, by IER. 59

A.1 Chapter 4 Studies Consent Form, Part 1 of 2 84
A.2 Chapter 4 Studies Consent Form, Part 2 of 2 85

B.1 Chapter 4 Studies Post-Condition Questionnaire, Part 1 of 2 86
B.2 Chapter 4 Studies Post-Condition Questionnaire, Part 2 of 2 87

List of Abbreviations

ANOVA	Analysis of Varience
BCI	Brain-Computer Interface
ECoG	Electrocorticography
EEG	Electroencephalography
ERD	Event-Related Desynchronization
ERN	Error-Related Negativity
ERP	Event-Related Potential
GUI	Graphical User Interface
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
HCI	Human-Computer Interaction
HIT	Human Intelligence Task
iEEG	Intercranial Electroencephalography
IER	Interpretation Error Rate
LSL	Lab Streaming Layer
MRI	Magnetic Resonance Imaging
MTurk	Amazon Mechanical Turk
NASA	National Aeronautics and Space Administration
NMR	Nuclear Magnetic Resonance
PET	Positron Emission Topography
PCA	Principal Component Analysis
RQ	Research Question
RM	Repeated Measures
SD	Standard Deviation
SNR	Signal-to-Noise Ratio
SSVEP	Steady State Visually Evoked Potential
TLX	Task Load Index

1 Introduction

Interaction with computer systems is an integral part of daily life.[1] From the widespread use of desktop computers in administrative offices to the smartphones that allow us to access the internet from nearly anywhere, people are almost always using a computer in some way. While traditional desktop interactions were once mainly done through a mouse and keyboard, advances in technology have provided users with novel, alternative ways to interact with computer systems, such as the touch mechanisms that have become standard for modern smartphones. Along with a better understanding of how humans interact with computers, developers have now provided users with many methods of input for accessing and controlling computer systems.

The mouse and keyboard are the most widely used input mechanisms for interacting with a desktop computer for good reason. They are fast, accurate, and provide the user with fine control over their actions and input. A major benefit of using a mouse and keyboard is that they have little to no error in converting a user's intended action into a command for the computer. Many other input mechanisms, such as speech input or handwriting recognition, can be susceptible to errors in which the command that the system interprets is different from the user's input. While the mouse and keyboard have been the ubiquitous interaction methods for traditional computers, alternative input mechanisms can be vital for improving the quality of life and usability of computers for disabled users [122], despite the possibility of errors. As a result, these alternate technologies face unique obstacles in their implementation and execution, such as a noisy input signal. One of the consequences of noise in input mechanisms are *interpretation errors*. When a noisy signal is received by the computer system, such as a user's brain signals, the system may incorrectly interpret the user's intended input, leading to the wrong command being executed by the system.

An upcoming technology that has potential as an alternative input technique is the brain-computer interface (BCI). BCIs use electroencephalography (EEG) to detect changes in a user's electrical brain activity and convert their mental state into computer commands. Studies have already shown that it is possible for users with motor deficiencies (e.g., paraplegia) to use BCIs and effectively control prosthetic limbs to grab objects [88], move using motorized wheelchairs [71], and transmit visual images to the brain to allow blind people to see [99]. However, it is a technology that is still in its infancy. Various major issues have prevented BCIs from becoming more mainstream as an alternative input method: non-invasive, scalp-surface electrodes have issues with accurately detecting the electrical signals of the brain [43]; limited understanding of the

[1] Portions of this thesis appeared in the following publication: Kevin C. Lam, Carl Gutwin, Madison Klarkowski, and Andy Cockburn. The Effects of System Interpretation Errors on Learning New Input Mechanisms. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, volume 713, pages 1–13. ACM, May 2021.[68]

complexities of human brain activity prevents complex control schemes from being developed [27]; and the possibility of recording human thought can eliminate any individual privacy for the user [73]. Other issues arise from the greater time needed to process a single command when compared with a traditional input like mouse and keyboard, as well as the high rate of inaccuracy and errors that can arise from translating a user's noisy brain activity to a computer command.

One of the major obstacles to effectively using an input mechanism, such as a BCI, is the presence of noise in analog signals, typically leading to interpretation errors. When using a signal-based input mechanism (i.e., the user's input is entered into the computer as a signal that will eventually be interpreted into a command), noise can be described as any unwanted information in the signal that obscures useful information. Noise can originate from a variety of phenomena, causing the user's intent to be misrepresented. Sources of noise can come from the environment (e.g., people talking in the background), inadvertently created by the user (e.g., eye blink artifacts in EEG recordings), picked up during the transmission of a signal (e.g., interference from nearby electrical/magnetic fields), or the computer system classifying the input incorrectly (e.g., the command executed by the computer is different from the command entered by the user). Many alternative input mechanisms will have some form of noise as part of the analog input signal. The presence of this noise can make an input method unreliable at accurately determining a user's intended command. For example, speech recognition must handle the differences in speech patterns and accents between people [7], as well as acoustic noise from the surrounding environment; otherwise, the speech command that is executed by the recognizer may be different from what the user intended. Similarly, BCIs must be able to interpret specific commands from a user's brain activity, which is a naturally noisy channel. Furthermore, most BCIs are EEG-based, meaning that the input for the BCI comes from detecting minuscule changes in the electrical activity of a person's brain. In general, electrical signals are particularly susceptible to noise from various sources: noise from AC wall outlets [77], induced voltage from environmental electric/magnetic fields [77], differences in skull thickness and conductivity [5], and physiological artifacts produced from hair obstructing contact with the scalp or movement of the electrodes. Due to the sensitivity of the EEG electrodes, the readings of the brain signals are likely to be disrupted by external sources of noise. Mitigating the different sources of noise to isolate useful input information is a constant challenge for the development of any input technique.

While previous research into input mechanism and user performance has studied the impacts of interpretation errors on a user's ability to effectively use an input mechanism [129, 78], there is little understanding of how these errors can affect users during the early stages of the learning process. This thesis will provide insight into how new users of novel input mechanisms are affected by interpretation errors.

1.1 Problem

The problem addressed in this thesis is: there is a lack of understanding of the effects of system-based interpretation errors on user learning of novel input mechanisms for command selection.

1.2 Motivation

Generally, noise in input signals and the resulting interpretation errors can make it difficult for computer systems to interpret a user's intended command and lead to a variety of problems for the user. Issues may occur where a user's input is misinterpreted, such that they enter a specific command expecting a particular response, but the system unexpectedly outputs a different, wrong result. For example, a motor imagery BCI user may attempt to produce a "go" command using left arm imagery, but the system incorrectly interpreted right arm imagery from the signal, which outputs a "stop" command instead. Another interpretation error could be that the system does not even recognize the left arm imagery, resulting in no output being produced by the system (i.e., false negative error).

The problems caused by interpretation errors and the subsequent effects they have on users are known to developers and designers of novel input mechanisms. High rates of interpretation errors can lead to performance loss [78], frustrated users [98], and can make an input mechanism too unreliable to be used for safety-critical systems (e.g., using speech recognition to drive a car) [9]. However, while alternative input techniques exist for most users, some input mechanisms, like BCIs, are mainly designed for users with disabilities that prevent them from using a traditional mechanism (e.g., limited mobility). Even when techniques exist to reduce noise in an input signal (e.g., filtering [19]), these methods are not completely effective and can be difficult or time-consuming to implement. This can mean that any leftover noise in the resultant signal can be a source of interpretation errors for these types of input mechanisms. However, while current literature has studied the impact of noise (and interpretation errors as a result) on user performance [8, 28], a considerable obstacle in creating alternative input mechanisms is that there is little knowledge regarding how system interpretation errors affect user learning.

Since many novel input mechanisms are unique in their operation (e.g., motor imagery BCIs), users will always need to learn the mapping between input and command. This learning process can also entail learning to identify system interpretation errors, techniques to reduce their occurrences, and methods to correct an error. Although noise in inputs and user learning have been studied separately, few have investigated their effects together. Developing a greater understanding of how noise affects learning can aid in creating better training programs to teach users how to avoid or fix interpretation errors when they might occur.

1.3 Solution

To address the lack of understanding of how interpretation errors affect user learning and performance, three user studies were conducted. The first study investigated an EEG headset's capabilities as a brain-computer interface (BCI) input mechanism using motor imagery, as well as how noise affects an input signal. A motor imagery BCI uses electroencephalography (EEG) to detect electrical activity across a user's scalp as they imagine limb movement. A computer system then interprets the incoming electrical signals in real time and executes a computer command based on the imagined motion. Since the highly sensitive EEG measures all brain activity on the scalp during the input period, any unrelated activity, such as electrical interference from nearby power lines (e.g., 60 Hz mains hum) or movement artifacts (e.g., eye blinks), is also recorded as noise. From the results of this BCI study, it was discovered that the high level of noise in the input signal of the EEG made it difficult to isolate the brain activity that is associated with motor actions. This led to greater rates of interpretation error than expected, resulting in an inability of the system to produce distinct commands, along with participants not knowing how to perform correct inputs. Results from the BCI study were inconclusive due to these issues with noise in the EEG data, which eventually served as inspiration to further investigate the effects of interpretation errors on novel input mechanisms. The work following the BCI study focused on studying the impact of interpretation errors on the early stages of learning an input mechanism.

The outcome of the BCI study provided a foundation and motivation for studying the effects of interpretation errors on novel input mechanisms. Specifically, two studies were conducted to understand how interpretation errors affects user learning of new input mechanisms. Participants were asked to complete a series of tasks using an input mechanism with artificial interpretation errors. These interpretation errors would incorrectly translate a user's input into the wrong command or fail to produce a command. Two possible effects of interpretation errors on learning were hypothesized. The first hypothesis suggested that user learning could be negatively affected by the interference caused by the interpretation errors. Conversely, the second hypothesis postulates that more errors could require users to expend more effort to learn the input mechanism, potentially leading to improved performance. The results of these two studies provided knowledge about how user learning and performance with tasks using novel input mechanisms were affected by increasing occurrences of interpretation errors.

1.4 Steps to the Solution

Different steps were taken towards understanding how interpretation errors affect users learning novel input mechanisms:

- Investigate a novel input mechanism susceptible to noise.

    1. Select a signal-based novel input mechanism that is known for its susceptibility to input noise.

2. Determine the tasks that are commonly completed when using the novel input mechanism.
3. Determine a set of commands to be used to complete the tasks.
4. Design a study to assess user performance while completing tasks with the command set.
5. Develop study system to gather data on the effect of noise on the use of the novel input mechanism.
6. Analyze the resulting participant data to determine how noise affects user performance.

- Investigate the effects of interpretation errors on users learning new input mechanisms.
  1. Select an input mechanism with a controlled and adjustable rate of interpretation error.
  2. Establish the levels of interpretation error rate that should be studied.
  3. Determine a set of tasks with the input mechanism that users can improve at over a short period.
  4. Determine a representative command set to be used to complete the tasks.
  5. Develop a study system to gather data on the effects of interpretation error on learning to use a novel input mechanism.
  6. Analyze the resulting participant data to determine how interpretation errors affect user learning.

1.5 Evaluation

Two studies were conducted to determine the effects of interpretation error on user learning. Study 1 provided users with immediate feedback from the system when they made their selections. Study 1 showed the increasing the rate of interpretation errors led to lower accuracy during memory tests, higher rates of user errors in training, and greater perceptions of effort. Study 2 required participants to monitor and manually correct the system's interpretation of their inputs to ensure that the correct selections were being made. Study 2 showed higher rates of interpretation error caused slower completion times during memory tests, higher rates of user errors in training, and increased perceived effort. The results of these two studies also indicate a deeper relationship between interpretation errors and user learning. Completion times during training and overall learning throughout both studies were similar at all rates of error, and participants did not report much difference in how easy it was to learn the items in the tasks.

The main findings of the two studies on the effects of interpretation error on user learning are as follows:

1. Higher rates of interpretation errors generally led to slower task completion times, decreased accuracy during memory tests, higher occurrences of user errors during training, and increased perceptions of frustration. This suggests that the interference hypothesis is more applicable than the retrieval effort hypothesis to understanding how interpretation errors affect user learning of novel input mechanisms.
2. Users of Study 1 were less concerned with avoiding errors, due to the automatic feedback. During the training phase, participants were given immediate feedback regarding whether their selections were

correct, which determined if they would successfully move onto the next task or repeat the same task until they made a correct selection. These users were more accurate in making correct selections during memory tests of the Zero percent error condition.

3. Users of Study 2 spent more time and effort to carefully ensure their inputs were correctly interpreted, since they had to monitor the system's output. These users were faster in completing the tasks during memory tests in the Zero percent error condition when they had to manually check the system's interpretation output.

1.6 Contribution

The major contribution of this thesis is to provide new knowledge regarding how interpretation errors affect learning a set of commands using a novel input mechanism. The results of the studies showed that higher occurrences of interpretation error, which is often caused by noise in the input, cause problems for user learning. This new knowledge of the learning process can allow designers and developers of novel input mechanisms to understand how noisy input signals can impact a user's ability to gain proficiency with a new input mechanism. This can lead to the creation of superior training programs for both novice and expert users, new metrics for quality-assurance testing, or better practices in user-centric design of systems. With novel input mechanisms like BCIs, which some members of the population are limited to using (e.g., limited-mobility users), a better understanding of interpretation errors can further improve their quality of life.

This thesis also presents additional contributions throughout the steps towards resolving the lack of understanding of interpretation errors and their effects on learning. The BCI study provided an analysis of one BCI using common hardware for command selection and assessed the effects of noise on this device. The results of the BCI study aided in identifying explicit hypotheses about how interpretation error affects user learning. An experimental paradigm was also established for studying the effects of interpretation error on novel input mechanisms. This thesis also provides various avenues for future work regarding novel input mechanisms, interpretation errors, and their effects on learning. Furthermore, the study presented a methodological template for studying noise in HCI contexts for other similar types of input mechanisms that are vulnerable to noise and interpretation errors.

1.7 Thesis Outline

This thesis is presented in six chapters. This first chapter briefly describes and summarizes the contributions made by this thesis.

Chapter Two provides an outline of related literature on the topic of novel input mechanisms, user learning of human-computer interfaces, and the presence of noise and interpretation errors in input mechanisms. The first section presents the relevant work on signal-based input mechanisms for controlling computer systems, with a specific focus on brain-computer interfaces as the exemplar. The second section describes the various research on learning, such as the process that users undergo to gain expertise when using a new input mechanism or completing new tasks. Finally, the third section provides an overview of the literature on noise in input mechanisms and the impact on user performance, as well as the methods employed to reduce the effect of noise on the output. The program of research described in the following chapters uses this summary of related work as a foundation for the studies conducted in this thesis.

Chapter Three explores the use of brain-computer interfaces as a novel input mechanism for controlling a computer system using brain activity. In particular, the feasibility of a motor imagery BCI to provide simple commands is explored. The chapter describes a study designed for testing the use of an EEG headset as a motor imagery BCI. The results of the offline data analysis are discussed, as well as details regarding the obstacles encountered with regards to noise in the input signal. The findings of this work serve as the motivation for the following investigation into interpretation errors learning.

Chapter Four provides the methods and results of two user studies undertaken to determine the effect of interpretation errors on learning novel input mechanisms. Both studies required participants to complete selection tasks using a novel input mechanism with varying levels of artificial interpretation errors added to the input. The first study provided participants with immediate feedback regarding whether their selections were correct and if they completed their task. The second study required participants to manually monitor the system's interpretation themselves to ensure that their selections were correct. The implications of the results of these studies are considered in the next chapter.

Chapter Five discusses the findings of the two studies on interpretation errors that were performed in the previous chapter. The main findings are presented and their implications on noisy input mechanisms are examined. Limitations to the approach of the studies are considered, as well as suggestions for improvements that could be made for similar work in the future. The considerations of the results are noted for their implications on signal-based input mechanisms, like the BCI. Additional ideas are also presented as possible avenues to investigate for further research to reinforce and improve the results found in this thesis.

Chapter Six summarizes the findings and contributions made throughout this thesis towards understanding the effect of interpretation errors on input mechanisms and user learning.

2 Related Work

Noise in input mechanisms is a wide field of research that has been explored extensively. By understanding the work that has been done previously with regards to novel input mechanisms and noise, a foundation can be developed for the contributions that will be presented in this thesis. The first section presents a general overview and definition of input mechanisms. Depending on the needs of the user and the design of the computer system, input mechanisms can differ in how they transform an analog input into a digital output. The second section examines the literature on learning and expertise development with user interfaces. The current knowledge regarding how users interact with new input mechanisms provides a foundation for understanding how disruptions to the learning process can affect the user. The final section of this chapter explores relevant work on noise and interpretation errors affecting input mechanisms. Previous studies of alternative input mechanisms that are susceptible to noise are included, and the effects of noise on user performance is examined. Furthermore, a variety of methods for reducing noise in such input mechanisms are also summarized.

2.1 Input Mechanisms

Input mechanisms use some form of a real-world analog input from a user as a means of controlling an interactive, digital system. In human-computer interaction (HCI), a variety of research has been done on input mechanisms and the users that interact with them. This includes the development of novel input mechanisms (e.g., [45, 76, 139]), improvements to existing ones (e.g., [49, 56]), studies to understand the psychology of the users (e.g., [112, 151]), and methods for empirical evaluations (e.g., [145]).

Input mechanisms are categorized by a variety of properties [49], such as by the analog data being converted into input data (e.g., sound, visual, etc.), the degrees of freedom available to the user (e.g., gyroscopes in smartphones can have up to three degrees of freedom), or direct versus indirect modes of operation (i.e., pressing the “W” key on a keyboard will type the “W” letter versus dragging an item to the recycle bin on a desktop to delete the item) [49]. Furthermore, how an input is interpreted by the computer system can influence the type of interaction between the user and computer. Input information can be interpreted as: free-form movement that allows for high levels of user freedom to interact with the system (e.g., moving a mouse cursor and clicking on items), contextual information that informs the system of the user’s current state (e.g., the speed at which a smartphone user is moving or if they are stopped [55]), or command selections where the input executes an instruction from a predetermined set of commands [6, 89].

Command selection-based input mechanisms can also be separated between discrete commands (e.g., pressing Ctrl+C as a keyboard shortcut performs the "copy" function) and parameterized commands (e.g., pinching an image on a smartphone adjusts the magnification of the image based on the relative finger movements).

This thesis investigates input mechanisms that interpret inputs as command selections. Specifically, discrete command input mechanisms are the category being considered due to the wide range of input mechanisms that have been developed, studied, and used commercially with many different sensors. The most common form of a discrete command input sensor is the keyboard, which allows users to physically press keys that perform individual actions within the computer system. However, input mechanisms that use alternative sensors can also be used to execute discrete commands, such as using touch sensors for character or word input (e.g., Scriboli [49], unistrokes [20], SHARK [65]), bio-acoustic sensors that use the human body as an input surface (e.g., Skinput [45]), pressure sensors for grip gestures [6, 89], ultrasonics and cameras for in-air gestures [140], gaze recognizers for tracking eye movements [152], and EEG sensors for BCIs [101, 102, 150, 74, 25, 97].

2.2 Learning

There is a rich history of research describing how people learn to use input mechanisms to control computer systems. In the field of HCI, researchers apply aspects of knowledge from psychology and human performance to understand how users start learning to use and improve at using interactive systems. This knowledge also allows developers and designers to create better input mechanisms that are easier for users to learn to use [144], as well as produce quality of life improvements that help users with completing their tasks [15]. When users use interfaces and perform tasks with them over time, they gain proficiency and skill with their use. This gradual development can be observed as multiple stages that the user progresses through, eventually becoming experts in using an input mechanism. However, the addition of noise and interpretation errors in an input mechanism may affect how well a user learns to use that mechanism. While it can be naturally assumed that noise is likely to have negative effects on learning and task completion performance, it is also possible that errors caused by the noise could help users be more attentive with their tasks.

2.2.1 Transition from Novice to Expert

Fitts and Posner have presented three general stages in which users learn and gain expertise [33]. These three stages can be adapted with respect to completing tasks using input mechanisms. The first is the *cognitive* stage in which users have just started using and understanding the input mechanism, as well as the tasks that they can complete with it. The user's experience with using the mechanism is minimal and they mainly rely on declarative knowledge. Declarative knowledge refers to information that a person can consciously recall and be able to verbalize [121], such as knowing the steps required to shut down a computer (i.e., "click the Start button and then select the Shut Down command"). While a user can recall declarative information, it usually requires some effort and time to do so. As such, their ability to complete tasks is at its slowest, relying on their need to recall and perform each individual step explicitly. When teaching users about a new input mechanism, developers will often use tutorials or on-screen instructions to guide users through the early steps of the learning process. This acts as an aid for the user to gain the necessary declarative knowledge.

The second stage is the *associative* stage. Here, users will begin to improve upon their initial knowledge by using their experience with the input mechanism and feedback that they received while using it. This feedback can come from the system (e.g., alerts when an error occurs and suggestions for how to fix them or avoid them in the future) or from the user themselves (e.g., using the Ctrl-C hotkey rather than clicking on the Copy icon is faster and does not require the use of the mouse). As users become more familiar with the input mechanism, they begin to create associations between their actions and how the system responds. Instead of needing to think about every individual action of their task, they begin to assemble their declarative knowledge into easily retrieved procedural memory [33]. This procedural knowledge is the counterpart of declarative knowledge; information that is subconsciously known to the person and can be retrieved without expending any additional effort beyond simply thinking about it [121]. For example, the declarative knowledge "pressing a letter on the keyboard while the Shift key is held down will produce an upper-case letter" becomes procedural knowledge when users simply press the Shift key when they want an upper-case letter, without thinking about the command. Procedural memory is involved in the development of motor skills, which can sometimes be gained without the individual's explicit knowledge of how to perform that skill. At this stage, users are often practicing with using the input mechanism to complete tasks, which reinforces the correct memory pathways. Researchers have found that users of touchscreen devices are mostly capable of recalling a stroke-based gesture without thinking after having performed about 15 repetitions of the gesture [151]. Although training exercises can be performed for real-world input mechanism, these repetitions are usually done while the user uses the mechanism naturally.

Finally, the *autonomous* stage is the last step for users to become experts at using an input mechanism. When users are at this stage, they are able to quickly retrieve items from memory without needing to think about the action or command to do so. This means that they can automatically perform an operation without conscious thought and without giving any attention to the process. For example, a user familiar with keyboard hotkeys may unconsciously use the Ctrl-S shortcut to save their writing whenever they finish a sentence or take a break. Automaticity is also known as "muscle memory", which is well known for allowing people to perform complex actions without remembering each individual step, like riding a bicycle. Users who have reached the autonomous stage are able to execute commands quickly and without conscious thought, since their knowledge of the operation has been fully compiled into their procedural memory [151]. Likewise, because the process of performing a command from procedural memory has been streamlined, users require little to no feedback for these automatic actions. However, problems may arise for users in the autonomous stage when interacting with a system that is prone to interpretation errors. If an interpretation error were to occur, the user would be interrupted from their main task as they would need to make corrections and monitor the system's output for further potential errors [90].

As mentioned previously, users are able to progress through these three stages of expertise through repetitive practice. Fitts and Posner have shown that the user's gains in performance follow that of a power function, which is known as the "power law of practice" (shown in Figure 2.1). In other words, the greatest improvement to the user's performance occurs at the beginning of the learning or training process. The amount of improvement exponentially decreases over time, such that after a while, the gains are incremental over the same amount of time spent practicing. Researchers have studied different kinds of practice and their effects on learning to understand how to improve the learning process [114]. This includes exploring the different forms of feedback that a user might receive during their training period, as well as how humans process new information [104]. For example, research has found that allowing users a period to stop and rest from their practice can improve learning [142]. Similarly, it has been shown that extending the intervals between task repetitions can lead to users being more receptive to the benefits of training [69]. These types of training strategies can improve the learning process passively. Alternatively, a user's learning can also be enhanced through intentional learning or extra effort from the user themselves, which is a theory that stems from the "levels of processing" model from Craik and Lockhart [26]. This model theorizes that when the user spends more effort during their learning process, they form more long-lasting and deeper memory pathways than if they expended little effort in their learning. This theory suggests that the added effort helps to compile declarative knowledge into procedural memory, which helps people retrieve knowledge more quickly. Particularly with human-computer interaction, studies have shown that increased attention and effort in their tasks during learning lead to better memorization [23, 30]. While this idea may increase the load on the user during the training process, putting in more effort to train ultimately results in better learning in the long term.

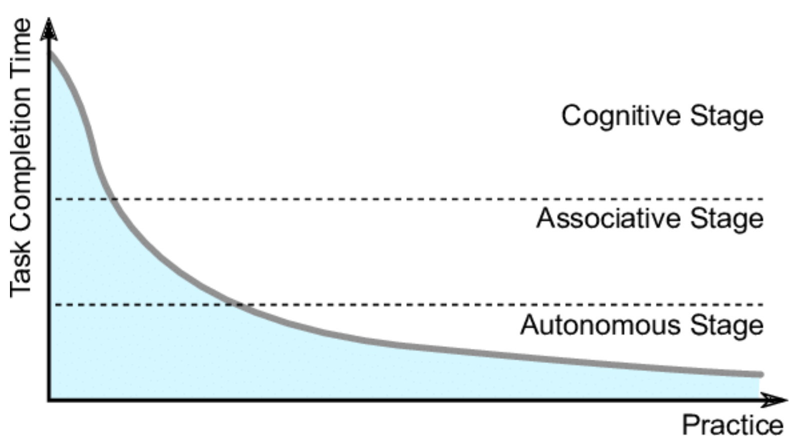

Figure 2.1: An example of the power law of practice [57]. New users begin in the cognitive stage and require the most time to complete tasks. As the users gain practice over time, they will move through the associative stage until they reach the autonomous stage, where they will have the quickest completion times.

2.2.2 Retrieval Effort During Learning

In general, it is accepted that the memory and recall of an item are improved when a person practices remembering that item. With learning to use an input mechanism, practice usually means to repetitively use the mechanism over time to develop the input action and command associations. By successfully using the input mechanism to perform commands, the user is able to reinforce the correct memory pathways. However, Pyc et. al found that users who practiced more difficult memory retrievals gained better benefits from that practice [107]. This retrieval effort hypothesis (i.e., "successful but difficult retrievals will be better for memory than successful but easy retrievals" [107]) states that increasing the difficulty of a user's training or practice will improve their learning and performance that involves recalling items from memory. Difficulty of the practice can be adjusted differently based on the task being performed. In the studies by Pyc et. al, the difficulty of the tasks was manipulated by increasing the number of items that participants had to memorize, as well as how many items had to be recalled correctly. They observed that participants who memorized more items were able to recall more items correctly during memory recall tests. Similar results were found between a group that performed the test almost immediately after memorization and a different group that underwent the recall test one week later. In the end, the increase in difficulty causes the user to put more effort into completing their task (i.e., memory recall), which aids in developing better memory pathways.

When using an input mechanism, errors in the output are likely to force users to pay more attention since they need to monitor the system's output to prevent more errors from occurring. Furthermore, when an error does occur (either a user error or system interpretation error) the user must put in additional effort to correct the error. According to Pyc et. al, the presence of interpretation errors can potentially help users in the learning process, by forcing them to be more attentive of their actions and tasks. Although the users would need to spend more time and effort making corrections due to interpretation errors, they could develop stronger memory pathways between the input action and resultant command by having to perform the action multiple times. While the retrieval effort hypothesis tested by Pyc et. al adjusted task difficulty through different number of recall targets [107], the process of fixing errors could potentially improve memory through increased repetition and error tracing (i.e., understanding what caused the error and how to correct the error).

A caveat to the approach of assuming that interpretation errors lead to higher user effort is that errors typically lead to a failed task. In other words, if a user were to attempt to recall and perform an action from muscle memory as an interpretation error occurs, they would receive feedback from the system that their memory retrieval was unsuccessful, even though it would have been correct if not for the error. From the user's perspective, it would appear as an unsuccessful retrieval until they later realized that an interpretation error had happened. The retrieval effort hypothesis only considers that effort is beneficial when retrievals from memory are successful. Some studies have shown that unsuccessful retrieval attempts can provide some advantages to learning and memory encoding [41]. However, the literature lacks examples that support the benefits of retrieval effort when the user fails to correctly recall the item. Similarly, the hypothesis has not been tested when external factors cause the result of a retrieval to be different than what the user expected or intended, such as in the case of interpretation errors.

2.2.3 Interference During Learning

When an unexpected outcome is produced by the system, the user is interrupted from their primary task and usually is required to fix the error before they can continue, such as when a wrong command is executed. Work in this area has previously examined what occurs when a user encounters such interferences during their learning process [4, 36]. As opposed to the retrieval effort hypothesis, the presence of noise and interpretation errors could lead to negative impacts to the user's learning and performance with the input mechanism. This interference hypothesis states that system errors weakens the memory association between the correct input and command (i.e., memory trace degradation) [31]. Since their impression of the correct association is being overwritten with the incorrect one, the more recent memory of the error could make it difficult for the user to recall the older, correct memory pathway. This retroactive interference is more likely to occur for memories that are similar to one another, which can often be the case when trying to recall which input action is associated to which output command. This degradation of the correct memory pathway could even affect users who are at the autonomous stage of expertise, potentially undoing any learning that they might have done [143].

Furthermore, it is also possible that interpretation errors can lead to the part-list cuing effect [11]. This effect, which is also known as cue interference, refers to the counterintuitive phenomenon that occurs when cuing an item for memory retrieval causes interruptions to the retrieval of the correct item. Normally, providing a cue usually improves a person's ability to recall, since they can use the cue as a point of reference. For example, real-world systems often use the "Enter" and "Esc" keys to confirm and cancel selections, respectively. However, if the provided cue is too similar or does not provide enough information for the person, then cue interference can occur, leading to a wrong (or false) memory being recalled instead [108]. Using the same example as above, some systems highlight the "Cancel" option by default in confirmation windows, which can cause a user to cancel their command if they press "Enter" without first changing the selected option. Regardless of the type of interference, studies with input mechanisms have shown that interpretation errors can disrupt a user's internal association of the input action and resultant command, leading them to retrieve the wrong memory pathway and suffering losses to their task performance.

2.3 Errors in Input Mechanisms

One type of error when operating an input mechanism arises from the presence of noise in the input. Noise is a particularly prominent problem in signal-based input mechanisms, and there are many sources of noise that should be considered (e.g., user errors [130], quantization errors [40], classification errors [44]). Disruptions to the signal can create issues for the computer system when processing the information, such as errors when interpreting the user's intended input. As such, much research in signal processing has been devoted to understanding how noise affects a signal. With better knowledge of how noise changes the data, methods to minimize that noise, while retaining as much useful information as possible, can be designed. Developers of novel input mechanisms implement such methods to prevent interpretation and other types of errors from occurring while using the mechanism.

2.3.1 Noise in Input Signals

One of the reasons why the keyboard as an input mechanism is ubiquitous is because it allows users to perform input commands with an extremely low error rate. Each key that they press is almost guaranteed to perform the exact operation that the user intended, barring user errors. On the other hand, input mechanisms that use sensors [34] with a less direct relationship between the user input and output command (e.g., BCIs, vision-based systems, or speech systems) requires the computer to translate the user's intent. This interpretation of the input signal is often fallible and susceptible to noise, leading to incorrect commands being executed or commands being missed entirely. Noise can arise from a variety of sources and is present in many signal-based sensors. For example, noise can come from the environment (e.g., sounds in the background, electrical interference), the user themselves (e.g., drawing a curved line instead of a straight line), or from the quantization process that converts continuous, analog signals to discrete, digital ones [16, 40]. Noise creates changes to the original signal that obscures the useful information. To reduce the amount of noise to a manageable level and obtain digital data from the analog signal, different methods are used to process the signal [105]. These including preprocessing steps (e.g., applying a low-pass filter to remove high frequency noise [19]), feature extraction (e.g., retrieving the outline of a person's face in an image with edge detection [118]), and classification (e.g., using a Hidden Markov Model to find similar words for autocorrect [86]). In particular, Casiez et. al have developed a relatively cheap low-pass filter for removing noise in user input streams [19]. This filter provides a basis for noise reduction in input mechanisms with an easy-to-implement algorithm and low computer resource cost. However, as a signal becomes noisier, it becomes more difficult to completely decouple the valuable information from the random data created from the noise. For practical purposes, most interactive systems that experience noise are designed with some flexibility towards having small amounts of noise in the system.

With signal-based inputs, the presence of noise can cause the interpretation of the user's input and intent to fail. This can occur at various points in the processing of the input signal, leading to interpretation errors that cause the system to perform a command that is different than what the user expected. These errors can occur as early as the quantization process when signal is converted. For example, the analog signal could be sampled at too low of a sampling rate, leading to a less accurate digital representation of the original signal (i.e., quantization errors [40]). Preprocessing can create errors if important bits of information are removed along with noise when filters are not carefully used [80]. Feature extraction may fail if the presence of noise creates difficulties in discovering the desired features, such as when low contrast in an image prevents edge detection from accurately predicting where an edge exists [128]. Lastly, classification of the features can suffer if the model does not provide enough distinction between similar results; for instance, stroke-based smartphone keyboards need additional context to differentiate when a user wants to type “to” versus “too” [109].

Interpretation errors can be categorized into one of three types: misinterpretations, false negatives, and false positives. Misinterpretations are when an error causes the system’s output to be different than what the user originally intended when they entered their input. False negative errors occur when the user provides an input, but no resulting command is produced by the system, either since the system did not recognize an input or that the input did not match with any of the predetermined commands. False positive interpretation errors are inversely related to false negatives: the system provides an output command despite the user not supplying an input. Most systems are designed in a manner to prevent false positives from appearing by biasing false negative errors. Generally, it is preferable that the system misses a user input than for the system to act without instruction [124, 155]. Additionally, noise and the subsequent interpretation errors create a need for the user to continually monitor the system and its output to ensure that correct interpretation of their input has occurred. This extra burden on the user is an issue that expands into Norman’s “gulf of evaluation and execution” problem [13, 100]. The gulf of evaluation can be described as the difficulty the user experiences when trying to understand the current state of the system. When the system is designed to provide meaningful feedback and information about its state, the gulf of evaluation is small. Likewise, the gulf of execution is the difference between what a user wants to do and what they system can allow them to do (e.g., one would expect that flipping a light switch upwards would turn on the lights in the room). Providing proper labeling, instructions, and training for the user allows them to understand the actions or commands that a system is capable of performing. Interpretation errors can cause the gulfs to increase, since they can obscure the true state of the system (e.g., the system performed an incorrect command leading to the wrong state) or mislead the user as to correct result of their input action.

2.3.2 Strategies to Reduce Noise

One of the main focuses of signal processing research is the study of effective noise reduction techniques [133]. An effective technique not only minimizes noise in the signal, but also does so in a timely manner with a reasonable amount of computer processing power. Different methods of noise reduction exist for different types of signals; the type of noise in a digital image (e.g., Gaussian noise, salt-and-pepper noise) can be vastly different from acoustic or electrical noise (e.g., power mains noise, thermal noise). As such, the methods that developers of input mechanisms use differ depending on the type of input signal.

Since noise can arise from a wide variety of sources, such as the environment where the input mechanism is being used, eliminating all the noise from an input signal may not be feasible. Current noise removal/reduction techniques, from simple filters to complex deep learning approaches, can remove a certain degree of noise, more so if a certain type of noise is expected to occur frequently for that input mechanism (e.g., using a band-stop or notch filter to remove the mains hum of electrical power lines). Conversely, this also means that a certain amount of noise can be expected to appear in an input signal. This may be due to the device being used in environments or states that are incompatible with the input mechanism itself (e.g., using a camera-based gesture recognizer where the lighting creates shadows that obscures the user and their gestures) or being used in such a way that designers did not intend for the input mechanism to be used (e.g., using a speech recognizer to determine the lyrics of a song). Although more advanced noise reduction techniques may aid in reducing interpretation errors made in these situations, a developer may find that the cost of implementing such a technique outweighs its benefits. Since users expect a system to interpret their input and output a system command in real time, the amount that a computer can process the input signal is limited. Many commercial input mechanisms that are susceptible to noise perform noise reduction to varying degrees, which not only ensures correct interpretations, but also improves user quality of life by reducing the need for them to take corrective actions due to errors caused noise. As such, the amount of effort and resources allocated to attenuating noise in an input mechanism needs to be balanced with how much noise is acceptable for the user to manage.

If the source of the noise is well known, such as from nearby electrical equipment [77], filters can be used as a basic method of attenuating the noisy signal [19, 38]. The function of filters can partially or fully suppress certain frequencies of a signal. Different filters (e.g., Butterworth filter, Bessel filter, etc.) change the nature of how a signal is attenuated. Furthermore, the signal that is removed by the filter depends on how the filter is designed. For example, a high-pass filter will remove low frequency signals that are sometimes tied to changes in electrode impedance caused by sweat or drift when using a BCI. Notch filters (i.e., band-stop filters) are commonly used for specifically removing the 60 Hz mains hum from high-strength power lines. Filtering is an effective technique to minimize noise when the noise occurs outside of the frequency range of the signal of interest. However, filters need to be used carefully as they can also easily remove any useful information. As such, filtering is usually done as a preprocessing step to prepare the signal for further deconstruction through feature extraction and classification.

For researchers who study specific psychological events (e.g., event-related potentials [77]), other methods are available that can help with limiting noise in the output data. Traditional uses of electroencephalography record a set of brain electrical activity over a period of time and perform analysis after all of the raw data has been gathered (i.e., offline processing). When a large number of trials are gathered for a particular time-locked phenomenon, such as the brain activity that occurs immediately after pressing a button, the different signals of all the trials can be averaged together [87]. By assuming that the noise is stationary and follows a normal distribution (e.g., Gaussian noise), the averaging process will cause the noise to cancel itself out. Another method possible with offline analysis is the manual identification and rejection of artifacts and noisy data points. As with many other studies done in HCI, noisy data points or epochs can be manually removed from the analysis. Naturally, this requires much manpower and time for the amount of data that is gathered from using an EEG, but an experienced researcher is able to identify known artifacts in the data (e.g., eye blinks, muscle movement). However, BCIs require the analysis to happen in real time as the user performs inputs (i.e., online processing), which makes it difficult to employ these techniques. There have been developments in machine learning that use training sets to learn what noise or artifacts look like and remove them in real time [136, 110]. Regression analysis has also improved the ability to extract the important features from the noise, allowing for more consistent outputs [83]. Studies have also shown success with using morphological component analysis to automatically remove eye movement and blink artifacts from both offline and online EEG data [149, 79].

While many techniques are available that help with eliminating existing noise in EEG signals, many researchers use procedures that minimize the amount of noise that can initially appear [77, 70]. For example, when a user blinks while wearing an EEG, a large artifact is produced by the rapid changing polarity caused by the eye movement. As such, users are often instructed to not blink as much as possible while using the EEG. Similarly, a user physically moving around can cause the electrodes on their scalp to shift, leading to movement artifacts in the data. Comfortable seating and relaxed posture can prevent the user from restlessly moving about [77]. In laboratory studies, the environment can be controlled to reduce sources of noise from interfering with the EEG recordings. A recording chamber that is also a Faraday cage can prevent external electrical and magnetic signals from reaching the sensitive electrodes. Lowering the temperature of the room can prevent the user from sweating such that the conductivity of the electrodes does not change over the course of the experiment [153]. Unfortunately, for a practical BCI that functions in real time, these techniques are typically not plausible for everyday use. Further work is required in this area to handle the additional noise that may be generated in uncontrolled environments, such as using a BCI while walking and performing other activities.

3 Understanding Noise in Signal-Based Input Mechanisms

When input mechanisms receive analog signals from a user's input, the computer must interpret the input in a way that reflects the user's intent. However, noise can cause incorrect interpretations of the signal to occur, which results in unexpected outputs. To better understand how interpretation errors affect these types of alternative input mechanisms, a study on a motor imagery brain-computer interface was conducted to assess the impact of noise on an input mechanism that is well known for its high rate of interpretation error. BCIs are an excellent choice for investigation, because not only do they use a noisy channel as an input signal (i.e., brain activity), they are also an input mechanism that is relatively new that will require most users to learn an entirely new control paradigm to use. As such, the process of learning to use a noisy, novel input mechanism can be studied with a BCI as a genuine real-world mechanism.

3.1 Brain-Computer Interfaces

BCIs are one of the several different forms of human-computer interfaces that use the electrical signals in a person's brain activity to control a computer system. They allow users to perform hands-free, gaze-free, and voiceless actions on a computer through interpreted commands from their brain signals. This provides BCIs with significant advantages as an input mechanism over other options; users only need to exert mental effort to execute a command. BCIs are able to overcome some of the issues that arise for other alternative input methods: speech recognition can fail in acoustically noisy environments, handwriting recognition requires the user to have enough motor control to use a writing implement, and eye tracking technologies require the user to maintain their gaze on specific areas.

BCIs also have major implications for computer users who have motor deficiencies [60, 37, 134]: they have been developed for users to control robotic prosthetic limbs that allow them to regain function from a missing limb [35]; mechanized wheelchairs can be controlled with a BCI to allow paraplegics to freely move with their own will [71]; with brain activity alone, users can write and send emails at speeds and accuracy comparable with able-bodied users [82]. Innovations and improvements to BCIs continue as the underlying EEG technology improves and more commercial EEG devices become available to the public at lower costs [32].

3.1.1 Electroencephalography

Electroencephalography (EEG) is the monitoring and recording of a brain's electrophysiological activity using electrodes. For applications when using with humans, these electrodes are typically non-invasive and are placed at standardised locations on a person's scalp [18]. There are also electrodes that can be applied directly intracranially on the brain, which is known as intracranial electroencephalography (iEEG) or electrocorticography (ECoG). While there are certain advantages of having direct electrode contact with the brain, such as a significantly better signal-to-noise ratio and a higher spatial resolution [91], the need for an invasive surgical procedure (i.e., incision into the skull) means that ECoG is often not feasible unless there was already a need to expose the patient's brain (e.g., treating intractable epilepsy [147]).

There are a variety of other techniques currently being used to study brain function, which include: functional magnetic resonance imaging (fMRI) [117], positron emission topography (PET) [154], nuclear magnetic resonance imaging (NMR), and functional near-infrared spectroscopy (fNIRS) [92]. The main reason to use EEG over any of these techniques is the extremely high temporal resolution that can measure brain activity in the order of milliseconds. This is the main focus of studying event-related potentials (ERP): measuring how the brain responds to stimulated events [77]. This has led to the discovery of the P300 component (in response to decision-making processes) and the error-related negativity component (which is elicited less than 100 ms after a mistake is made, whether the person consciously realized the mistake or not). Another advantage of EEGs is the much lower comparative cost [137]. Commercial EEGs for personal use are available for hundreds of dollars (e.g., NeuroSky offers single-electrode EEGs for approximately \$100 [95]), as well as up to tens of thousands of dollars for advanced systems with 32 or more electrodes [94]. To compare, MRI machines typically cost over hundreds of thousands of dollars [54]. The main disadvantages of EEG come from its low spatial resolution and inability to measure electrical activity below the surface of the brain. For these purposes, other techniques are desired and may sometimes be used in conjunction with EEG (e.g., fMRI and PET can be used to precisely localize regions of the brain based on changes in blood flow [117, 154]).

Scalp-based electrodes for EEG have been commonly used in clinical settings since their inception by Hans Berger in the 1930s [131]. It is contemporarily used by physicians as a tool for diagnosing epilepsy and other neurological disorders, such as encephalitis, stroke, and various sleep disorders. Furthermore, EEG is also extensively used for studying cognitive function and understanding human psychology [77]. Silver chloride (Ag/AgCl) electrodes are the main types of electrodes used for EEG as they provide potentials that do not vary with environmental variables, such as temperature and time [1]. To detect the brain's electrical activity through the scalp, most conventional clinical EEGs require a conductive gel or paste to be applied at the electrode-scalp interface to improve the conductivity of the contact. Dead skin cells and oils on the surface of the skin may also increase the electrical impedance leading to worse conductivity [59]. Thus, gentle abrasion will be performed at the contact area to lower the impedance. Recently, the development of dry electrodes (i.e., does not require using a conductive gel), have allowed EEGs to become easier to use by reducing their necessary preparation and clean-up procedures [75].

Depending on the application, the number of electrodes/channels being used will depend on the desired spatial resolution (i.e., higher spatial resolution will allow more accurate mapping of an electrical event to a specific part of the brain). When attempting to diagnose patients for epilepsy, a clinician may use 16-25 electrodes to cover the scalp as they are more interested in when an epileptic event occurs and how long it persists [85]. However, neuroscientists studying which part of the brain specific actions originate from may require 256-512 electrodes in their studies [58, 14]. As more channels are used, more computing power will be required to analyze the resulting data. As a result, BCI applications that require real-time analysis of EEG data are limited by the computer's processing power to analyze the data within a reasonable timeframe. Typically, BCIs will use 8-32 channels depending on the spatial resolution needs [8].

3.1.2 Types of BCIs

The types of BCIs available are directly based on the underlying electrophysiological monitoring technology. As such, there are both invasive and non-invasive BCIs being developed. Invasive BCIs are ones that require electrodes to be implanted on a person's brain through a surgical procedure. ECoG is mainly used for invasive BCI purposes and is often applied as an array of electrodes directly on the surface of the brain underneath the skull [115]. The use of ECoG as an interface can also sometimes be referred as a semi-invasive BCI; there are more invasive methods that exist that require the implantation of electrodes into the brain itself to get better signals from specific regions [12]. While invasive (and semi-invasive) methods provide much higher spatial resolution and signal-to-noise ratio compared to EEG (i.e., scalp-based electrodes), bodily response to foreign objects in the brain will lead to scar tissue developing over the electrodes and interfering with the signal quality. Also, the high risk and cost of a neurological surgical procedure means that implanting electrodes will only be done if a surgery is already required for a different reason (e.g., treating epilepsy [2]) or if the risk of reducing a person's quality of life from the procedure is very low (e.g., implanting optical prosthetics for blind patients [126]). As a consequence of the high risk, previous research that study invasive BCIs are often focused on 2-3 users and their experiences with using the BCI over many months or years [48].

Many invasive BCIs are developed for disabled users to help control prosthetic limbs [116]. Using BCI commands to tell a mechanical arm prosthetic to grab an object is a more natural action than issuing the same command from a traditional computer interface. It does not require the user to carry around a separate device to use the limb, and their commands can remain private (e.g., using speech commands would mean people nearby would hear the commands). Also, since the signal from an in vivo electrode is much more reliable than scalp-based electrodes, errors from interpreting the user's commands are less likely to occur. With some training, users can become proficient enough with using the BCI that they can fluidly issue commands to control their prosthetic. Passive forms of invasive BCIs can also be used to improve a person's quality of life without requiring active input from the user. For example, visual prosthetics (i.e., bionic eye

[126]) can be connected to a user's optical nerve to relay the visual information from the prosthetic directly to the brain, allowing a blind or partially blind user's vision to be restored. Although invasive BCIs provide a wide variety of benefits, they are ultimately less practical for most users, since they require a highly invasive surgery and are typically expensive to implant and maintain.

The main benefit of using a non-invasive BCI is that they are much more accessible for most users to adopt in their daily lives. Non-invasive BCIs are often used as a cheaper, safer, and less permanent alternative to invasive technologies for improving computer access to disabled or motor-impaired users. These types of BCIs can also be developed for able-bodied users as another alternative input mechanism. However, it requires that the underlying technologies that power the BCI needs to be inexpensive, portable, and, most importantly, pose no adverse health risk for the user. As such, while many different non-invasive brain monitoring technologies are available (e.g., fMRI), only EEG and fNIRS meet the criteria for use with BCI devices [63]. fNIRS technologies work by projecting near-infrared light through a person's scalp and measures the differences caused by blood flow in the brain, such as change in flow rate or oxygenation levels. This works similarly to fMRI and provides similar advantages as a brain sensing technique, namely, a high spatial resolution. However, for the purposes of BCI as alternative input mechanism for issuing computer commands quickly and accurately, EEG with its high temporal resolution is often preferred. There has been work in successfully using fNIRS for monitoring passive user information, such as emotional affect which requires the ability to locate where changes in the user's brain are occurring [63, 62].

Since EEG-based BCIs are more commonly and cheaply available, a wider range of applications have been developed for them. Video games have been created that use only brain-input controls (e.g., BrainBall sees players competing against each other to relax and increase their alpha wave amplitude [50]), virtual reality headsets can be integrated with BCI for more immersive controls (e.g., Neurable attachment for the HTC Vive [123]), aiding in meditation (e.g., Muse Headband [127]), and monitoring a user's attentiveness during various tasks (e.g., Nijholt P300 study [97]). Naturally, non-invasive BCIs can still be used for assistive purposes, such as providing an input method for writing emails or controls to drive a mechanical wheelchair [71, 21].

3.1.3 Control Strategies

There are a few major control paradigms that currently exist for people to use a BCI: steady-state visually evoked potentials (SSVEP), P300, and event-related desynchronization (ERD). While other control methods also exist, these are the most popular for their ease of implementation and the better overall understanding of the underlying brain functions that drive them. In existing literature, most BCI applications use either SSVEP or P300 control strategies as they are reported to have lower rates of errors due to factors such as noise.

SSVEP BCIs are based on the phenomenon that when a person sees an oscillation, the brain produces a response at the same frequency of the stimulating oscillation [101]. In lab settings, this oscillation is usually a stimulus in the form of a light, shape, or image that flickers on and off at a specific frequency. When a user focuses their attention on that stimulus, a relatively strong signal at the same frequency (and its harmonic frequencies) can be observed near the participant's visual cortex (i.e., the back of the head by the occipital lobe). This phenomenon holds true even when multiple stimuli are presented at varying frequencies that do not interfere with each other (e.g., frequencies at 11 Hz and 15 Hz do not share many harmonic frequencies and are relatively far apart). While the participant focuses on a particular stimulus, that frequency will appear as the strongest and the other peripheral stimuli will elicit weaker responses. As a computer interface, SSVEP allows users to select a command by simply looking at their intended target. The BCI will then be able to determine the difference in strength of the brain's responses to determine which item should be selected. The number of possible items/stimuli that can be displayed at once is limited by how different the oscillation frequencies are from one another, as wrong interpretations are more likely to occur as frequencies are closer together.

P300 BCIs are similar to SSVEP BCIs in that they rely on a user's visual attention towards a particular stimulus. The major difference between the two paradigms is that P300 stimuli flash instead of oscillating [102]. P300 spellers are a common application that essentially provide a method of text entry using a P300 BCI [66]. A grid of letters, numbers, and symbols are ordinally laid out on a black background. A flash stimulus is performed across each item, one at a time, in the grid by highlighting the item with a white background for a short duration (e.g., 100 ms). While the user is focused on an item as the flash occurs for that specific item, a response called the P300 (i.e., a positive peak in the brainwave occurring approximately 300 ms after the stimulus onset) is elicited from the user's brain. Like with SSVEP, peripheral targets that are not being attended to will have weaker P300 responses. Since the flashes occur quickly throughout the grid, the BCI can wait and use multiple flashes to confirm a selection and reduce the likelihood of interpretation errors. This has led P300 spellers to becoming a popular BCI tool for performing text-entry tasks in a relatively quick and accurate manner (e.g., writing emails [71]).

ERD BCIs use imagined movements of the user's body to elicit responses from different parts of their brain; hence, these types of interfaces are sometimes known as motor imagery BCIs [61, 25]. µ waves (mu waves) are the sensorimotor rhythms of brain activity that are related to voluntary movement. These waves typically exhibit a frequency from between 7-13 Hz and are localized around the motor cortex (i.e., top of the head). When a person performs a motor action (e.g., bending an arm), a suppression of the mu wave occurs, which can be observed in a frequency-domain analysis as a decrease in amplitude at the 7-13 Hz range. This response occurs in different areas of the brain depending on the action (e.g., bending of the left arm will cause an ERD on the right side of the motor cortex). This suppression occurs because the affected neurons are normally firing in synchrony when at rest but become desynchronized during an action as the neurons fire in different and complex patterns. The ERD phenomenon also occurs if the person views someone else performing an action; some researchers believe this is related to mirror neurons (i.e., a neuron that activates when the person or animal sees another performing the same action). This also applies when a motor action is imagined (e.g., thinking about bending an arm), albeit to a weaker degree. By mapping different imagined movements and their respective ERDs to different computer commands, the basis of motor imagery BCIs is formed. With sufficient training, a user can effectively send mental commands to a computer without the need for physical interaction and unlike methods that use visually-evoked potentials (e.g., SSVEP, P300), the user does not need to visually attend to a stimulus.

There also exists another category of EEG-based BCI devices that monitor a user's passive state, known generally as passive BCIs [62, 150]. They monitor various underlying neurological processes as a user performs tasks (either related with the BCI or not) and the BCI provides the computer with ways to improve the user's performance or quality of life. For example, a passive BCI monitoring the error-related negativity (ERN) component [148] of a user's brainwaves would be able to detect when a user makes a mistake and help them make corrections faster than they normally could. The advantage of these passive BCIs is that they function without the need for the user to be aware of them, meaning that they can be integrated with other control paradigms as a hybrid BCI to aid and improve their usability (e.g., reducing command interpretation errors using ERN detection [113]). While further understanding of the brain's processes is still required, future passive BCI devices may be able to guide users with planning and decision-making by being able to directly look at the underlying processes as they occur.

3.1.4 Using a BCI

The idea that BCIs will allow users with motor deficiencies to be able to perform motor tasks is one of the major draws of the technology. Being able to control a computer without the need for physical interaction means that almost anyone can use it as an interface. Along with cheaper EEGs being made commercially available, widespread adoption by able-bodied users should also be possible soon. However, the problem of BCI illiteracy is a major obstacle to creating a single one-size-fits-all BCI.

To start, being able to effectively use a BCI (i.e., issuing commands with relatively high interpretation accuracy) often requires a long and extensive training period. With most current EEG electrodes, slight displacements of the electrodes can cause noise to appear in the input signal, which may interfere with a BCI's interpretation of the user command [81]. Another major source of noise comes from blinking (often classified as an artifact), caused by the rapid muscle movement of the eyelid. As such, users need to be as motionless as possible when attached to an EEG device and using a BCI to perform computer commands. This means that even BCI control paradigms that are reacting to a user's attended state, such as SSVEP or P300, users must learn to remain as still as possible while the BCI is determining a command selection. For BCIs that use ERD (e.g., motor imagery), most users will need to train themselves to produce stronger mu wave suppressions when imagining movement, since imagined motor actions naturally produce weaker ERDs than actual movement. An effective method of generating reliable signals for the BCI differs from person to person, which means potential users will likely have to try different imagined motor actions to produce the strongest ERD for them [72]. Repeatedly imagining and training the same motor action can also benefit from neuroplasticity; neuron pathways that facilitate mu wave suppression will be reinforced and become easier to access at will.

While these pose initial problems to effectively using a BCI, they can be eventually resolved through training and practice. However, even with practice, not all users are able to use BCIs. This anomaly that affects approximately 20% of users is known as BCI illiteracy [3], which is defined as the inability to use one or more BCI control paradigms. Even though brain functions are roughly located in the same area for most people (e.g., the visual cortex near the back of the head controls sight), there are differences between people's brain structures; the individual pathways and neurons will vary. Since the most commonly used non-invasive BCIs are based on EEGs with scalp-based electrodes, a person whose relevant brain activity that occurs deeper in their brain may not produce a strong enough signal at the scalp surface to be able to use certain BCIs. As such, there is no single, universal BCI paradigm that work for all users. For some users who are illiterate with one paradigm, they could potentially use a different one, especially if the signals originate from different areas of the brain (e.g., SSVEP signals at the visual cortex/back of the head; ERD signals at the motor cortex/top of the head). However, there still exists a small amount of users that are completely unable to use any currently available BCI paradigm. While researchers have attempted to provide both software and hardware solutions to these issues, the greatest success thus far has been attempts at using machine learning to make improvements for BCI-illiterate users [136].

3.1.5 Interpretation Errors in BCIs

Errors that occur during the interpretation of the input signal can result in wrong commands being executed by the computer. For BCIs, one of the main sources of interpretation errors is caused by noisy brain activity being recorded. This noise is primarily in the form of electrical noise and interference that is received through the EEG electrodes, and often arises from nearby power supply lines in the wall, AC-powered lights, and other nearby electronic equipment (e.g., computer monitors, signal amplifiers, Wi-Fi routers etc.) [77]. Similarly, movement artifacts produced from the user can also be a cause of poor input interpretations. Any noise, interference or artifacts affecting an electrical signal will modify the properties of the signal's information, which can make it extremely difficult to discern the effects of the modification if the original signal is unknown. Strategies to reduce noise in EEG signals are discussed in Section 2.3.2.

3.2 Development of a Motor Imagery BCI

To determine the feasibility of using the brain as an alternative input for controlling computer systems, motor imagery/ERD BCIs were studied. Motor imagery was mainly chosen as the BCI control paradigm due to the input method being primarily neurological. As opposed to the SSVEP and P300 paradigms which require the user to visually attend to a stimulus, motor imagery can be commanded with only the user's thoughts and intent. In practice, this means that motor imagery BCIs have several advantages over other BCI control paradigms, as well as alternative input mechanisms. Motor imagery does not require an external stimulus to activate (e.g., the P300 component is detected when a change occurs to a visual stimulus that the person is looking at), allows for complete input privacy from their physical surroundings (e.g., speech recognition needs the user to speak their commands out loud, usually with a clear tone), and can be used continuously (e.g., as long as the imagery is maintained in the user's mind, an input signal will be continuously streamed to the computer).

The current issues with using motor imagery as a control paradigm, and BCIs as an alternate input mechanism, are also acknowledged. Since motor imagery is derived from the ERD signals that appear from the imagination of physical movement, actual muscle activation located elsewhere in the body can cause stronger ERD signals to appear (i.e., real motor actions typically produce stronger ERD events than imagined or mirrored ones). For example, if a two-class motor imagery BCI (i.e., a BCI that uses two distinct EEG events to differentiate inputs) is waiting for an input of either left arm or right arm imagery, physical movement from the legs could be misconstrued by the system as an imagined arm motion and potentially causing the wrong command to be executed. Thus, when using a motor imagery BCI, users are often discouraged from physical movements, including those performed as a result of subconscious behaviour, such as blinking or moving their tongue inside their mouth.

3.2.1 EEG Headset Selection

When considering the device to use to study BCI input, several important factors were considered. Currently, there are no ready-to-use motor imagery BCIs that can be operated immediately out of the box. Most research groups that study BCIs develop their own in-house system, since the requirements of the device and software greatly vary depending on the interaction being studied. To start, different EEG devices were researched for their suitability for studying how noise and interpretation errors affect the use and learning of BCIs as an input mechanism. After comparing various commercial devices, the OpenBCI Ultracortex Mark IV Headset was selected (shown in Figure 3.1). One of the major factors in deciding on the device was the cost. While more expensive EEG systems could provide more accurate readings, these systems would be unreasonable for most potential BCI users to afford. The Ultracortex headset allows for up to 16 channels whose locations can be easily adjusted depending on the needs of the researcher or user. This means that electrodes can be concentrated on the scalp over specific areas of interest (e.g., centrally over the motor cortex for detecting ERD events). The Ultracortex headset is portable through its wireless connection through Bluetooth, which also eliminates the possibility of interference from connecting wires. Also, the source code for the device is entirely open source, allowing for a high degree of customization with its firmware to suit difference needs, as well as the availability of online resources to aid in development and troubleshooting.

The Ultracortex electrodes are dry silver chloride (Ag-AgCl) electrodes. Using dry electrodes removes the need for that conductive gel to be applied onto the user's scalp, allowing for quicker preparation times, less clean-up, and faster transitions from one user to another when used for consecutive participants in user studies. Duvinage et al. found that, while a dry electrode system is less accurate than the gel-based systems (i.e., wet electrodes) often used in medical applications, it still performed at a level useful for non-critical applications (e.g., BCIs for gaming) [29]. Although the ability in recording accurate signals is lower, dry electrode EEG systems are still usable for non-critical applications (e.g., BCIs for gaming). When compared with using a gel-based system that are commonly used for medical or research purposes, dry electrodes are more likely to be used for BCI systems, since they can be used with minimal preparation (e.g., does not require the application of a conductive gel). Reducing the time and effort necessary to prepare an EEG device for BCI use is necessary for the input mechanism to reach higher rates of adoption by the general population.

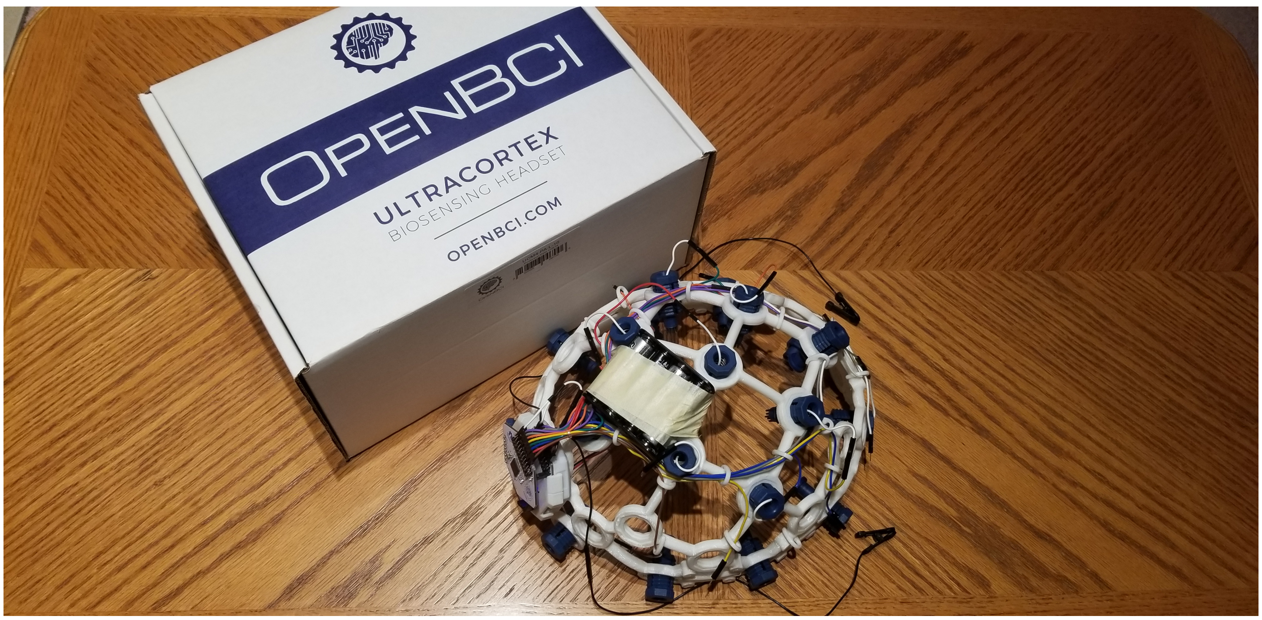

Figure 3.1: The OpenBCI Ultracortex EEG Headset used to study BCIs as a novel input mechanism.

3.2.2 EEG Headset Testing

Initial tests to understand the Ultracortex headset were performed by observing the elicitation of alpha waves in several testers. This was done using OpenBCI's own software, the OpenBCI Graphical User Interface (GUI), designed for basic monitoring of raw EEG data with the Ultracortex and their other devices. The GUI allows for real-time observation and recording of the electrical signals received at each electrode, in both time and frequency domains. A heat map of the scalp is also available that shows activity based on the location of the electrodes. An example of the GUI can be found in Figure 3.2. The Ultracortex was placed on users in accordance with set-up instructions provided by the OpenBCI website. The set-up steps are similar to the procedure used in the BCI study and is detailed in Section 3.2.4. Calibration is primarily done by asking the participants to perform an action that elicits a strong response that can be seen throughout all electrodes (e.g., blinking, clenching the jaw). In Figure 3.2, a blink artifact can be seen in all active electrodes at approximately -2.4 seconds.

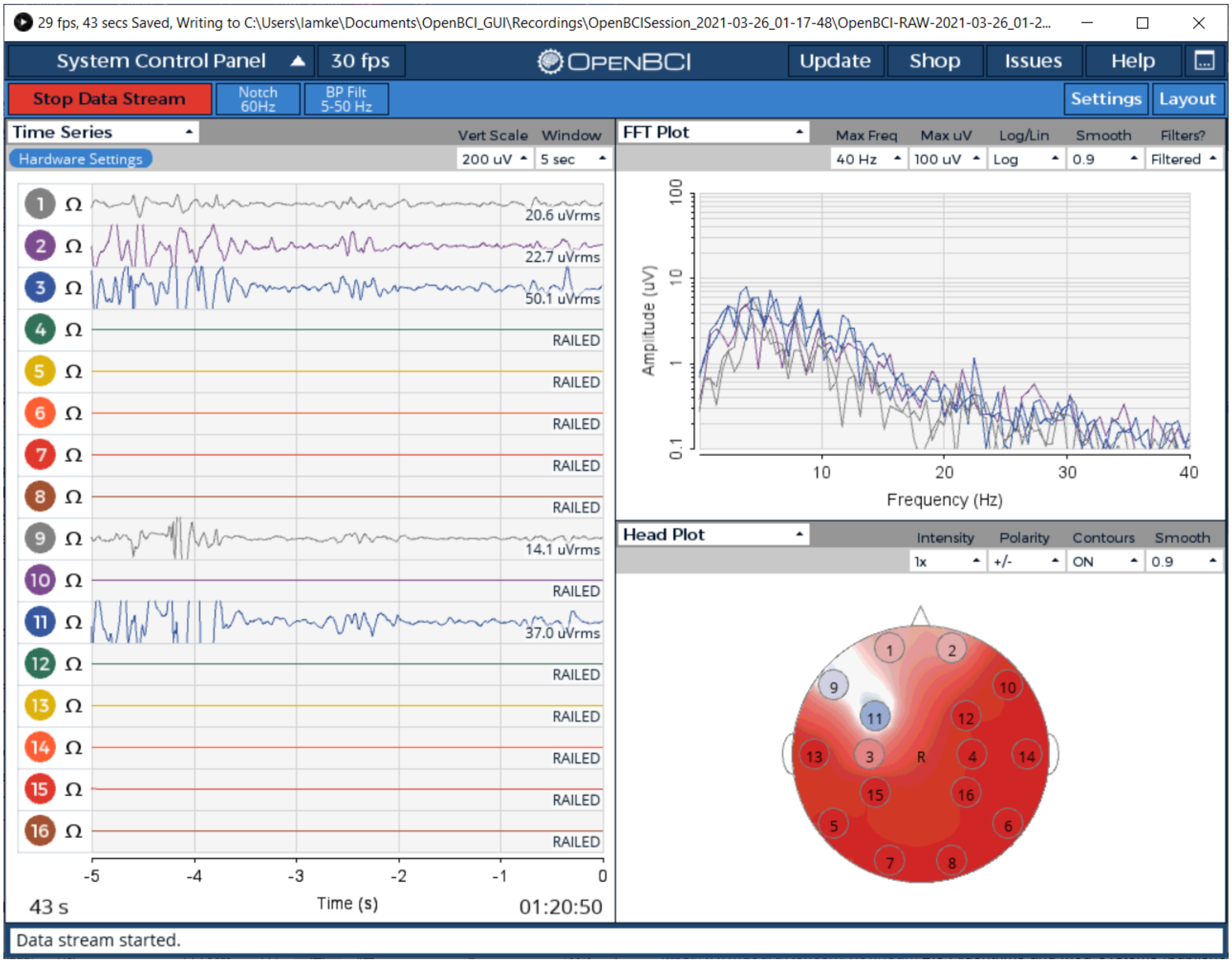

Figure 3.2: The OpenBCI GUI that was used to calibrate the Ultracortex IV Headset. On the left half of the GUI is a real-time visualization of the electrical amplitudes recorded by each electrode over time. Active electrodes show activity and the current voltage being detected by the electrode, while inactive electrodes shows a "RAILED" reading. The graph in the upper right of the GUI shows a real-time Fourier transform of all electrodes in the frequency domain. The lower right display shows a heat map over the user's scalp of the electrical activity detected by the electrodes.

The testers were then directed to perform a few different actions that elicit strong responses in the EEG (e.g., blinking, clenching of the jaw, etc.), to test proper adherence of the electrodes. Finally, to detect their alpha waves, the users were instructed to close their eyes and relax as much as possible. In the state of relaxation, alpha waves are typically detected as increased oscillations across the occipital lobe (i.e., the back of the head) in the 8-12 Hz range. Successful users would see an increase in their EEG amplitudes over this range. After confirming the usability of the Ultracortex with the basic tests, a study was designed to gather motor imagery data from a user study using an oddball paradigm (i.e., participants are presented with repeating stimuli that are rarely interrupted by a non-conforming stimulus) [120].

3.3 BCI Study: Detection of Mu Waves Using the Ultracortex Headset

The BCI study was designed to test the capabilities of the Ultracortex headset with detecting and recording motor imagery. Specifically, the brain activity of interest are mu waves that become supressed when an event related synchronization event occurs, typically evoked through real or imagined muscle movement. A secondary objective of the study was to test the responsiveness of system by gathering data in timed trials and to test the offline analysis procedures.

Participants were asked to complete a series of tasks by moving their fingers when prompted by an on-screen stimulus, with an occasional "oddball" stimulus [120]. Real movement was desired as opposed to imagined motor actions, because the suppression of mu waves is more prominent (i.e., easier to detect) when the movement actually occurs [17, 51]. To fulfill the secondary objective, this study used an oddball paradigm to evoke a neurological event that can occurs after a short period of time from when the stimulus is received. The oddball paradigm is commonly used for studying ERPs that are time-locked to a stimulus (i.e., brain activity that occurs after a period from when the person receives a stimulus). Research into ERPs show that there is a greater neurological reaction, specifically from the P300 component, during an oddball stimulus when compared with normal stimuli. A previous study used similar methods to successfully locate both P300 and ERD events in participants with drug-resistant epilepsy [119]. After data gathering was completed, offline analysis of the EEG data was done in an attempt to isolate the ERD events.

3.3.1 Study System

The software used in this study was developed in Java and Python. To gather EEG data from the OpenBCI Ultracortex headset, an open-source interface called the Lab Streaming Layer (LSL) was used to synchronize the handling of data collection and operation of the study software. The headset would send EEG data over the Bluetooth connection via the LSL, which could provide the data output to several different programming/scripting languages for further analysis or manipulation, including Java, Python, and MATLAB. However, initial testing of the output data found problems with some of the languages. For example, the LSL Java output was lacking timestamps that would have allowed for determining the time a specific data point occurred after the trial began. Some outputs also showed significant delay between the actual time that the brain activity occurred and when the signal was received by the computer, where the delay was visibly apparent during real-time monitoring of the EEG recording. As such, the LSL Python output was selected for being able to provide the most accurate and detailed EEG recordings in real time.

The user study's GUI was developed in Java using JavaFX. Since the incoming EEG data was being received by a Python script, the data needed to be redirected into the Java program for synchronizing with the ongoing study. A socket was implemented within both the Python script and Java program that allowed two-way, real-time communication between the two languages. This allowed continuous streaming of the EEG data into the program controlling the operation of the study, as well as any other programs for monitoring the neurological state of the participants.

The tasks involved performing a motor action (i.e., finger and hand movement of both hands) when an on-screen prompt was received during a single trial. A fixation cross was present on the screen for 3 seconds. A high-pitch sound cue followed for 0.5 seconds that informed the participant of imminent start of a trial, signalling for the participant to refrain from any movement, including blinking. Participants would then receive one of four stimuli that instructed them with which hand to perform the motor actions. They continuously moved their fingers on both hands in a chording motion for the duration that the stimulus was visible on the screen, which lasted for 3 seconds. A 0.5 second low-pitch tone denoted the end of the stimulus and for the participants to stop all movement again. Lastly, the fixation cross reappeared that marked the end of a trial, although data recording continued for another 2 seconds. An inter-trial period was added between the end of a trial and the start of the next trial to prevent participants from recognizing a pattern and anticipating the start of the next trial. This period was chosen randomly from between 0.1 seconds and 0.8 seconds, as described by previous work [77, 132]. Participants were shown a solid black screen during the inter-block periods, indicating that EEG data recording had been paused and were free to move. A diagram of the procedure timeline is shown in Figure 3.3.

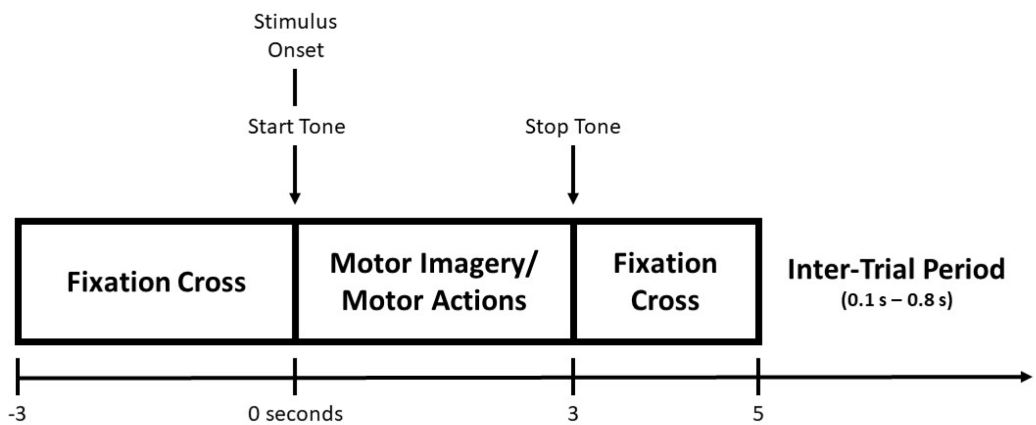

Figure 3.3: A timeline illustrating the sequence of events taken place during a single trial. EEG activity was recorded during the events that are bounded with a solid box (from 3 seconds before stimulus onset and until 5 seconds after). The inter-trial period was randomized between every trial.

Two different stimuli were used to instruct participants on whether to perform the motor actions during the trial. The target stimulus was a solid green circle in the middle of the screen, which informed the participant to move their fingers in the prescribed chording motions. The non-target stimulus was a solid red circle, which indicated that the participants should not be moving at all during the trial.

The study involved 10 blocks of 10 trials each. Participants started and ended their session with a baseline EEG test that recorded their neutral brain activity while staring at a black screen. The baselines allow for zeroing of the EEG data by eliminating individual differences between the participants. Participants were allowed to blink and make small adjustments to their posture during the baseline tests but were still asked to minimize such movements as much as possible. Within each block, 3 of the 10 trials were targets. These oddball targets did not immediately follow with another target trial.

3.3.2 EEG Data Acquisition and Processing

EEG data was recorded from 7 silver chloride (Ag-AgCl) electrodes. The electrodes were distributed across the scalp using a modified version of the International 10-20 system, with the electrodes being located at: C3, C4, Cz, FC1, FC2, CP1, and CP2. Electrode locations are shown in Figure 3.4. The 10-20 system refers to the standardized distribution of electrodes over the scalp [18]. The locations are divided into distances that are at 10% intervals from the front to the back of the skull (i.e., from the nasion to the inion) and at 20% intervals from the left ear to the right ear. The electrode placements used in this study is from the 10-10 system, which is a higher resolution system that allows for more concentrated data gathering. The letters denote the region of the brain that the electrode is reading signals from: central lobe (C), frontal lobe (F), and parietal lobe (P). FC denotes that the electrode is between the frontal and central locations, and CP is between the central and parietal locations. Odd numbers denote the site being on the left half of the head and even numbers are used for the right half. A lower-case “z” is used for the centre line on the scalp that connects the nasion and the inion. Reference nodes, used to ground the electrical signal, were located at both left and right mastoids (i.e., ear lobes). The electrode locations for this study were selected such that a higher concentration of data could be recorded from the area of the scalp best associated with motor imagery (i.e., top of the head). Electrical impedances at each electrode were less than 5 kΩ for all participants. The sampling rate of the EEG was 125 Hz.

The electrical signals were recorded with a sampling frequency of 125 Hz. MATLAB was used to perform analysis after all participant data had been gathered. During offline analysis, the EEG data was filtered using a 1 Hz high-pass filter and a 45 Hz low-pass filter.

3.3.3 Participants

Four participants (one man and three women) were recruited for the BCI study from the University of Saskatchewan’s Human Computer Interaction Lab. Each participant was a student working at the lab and were aged 21-29. None of the participants had previously used a BCI or EEG before, and had never participated in any other EEG-related studies.

3.3.4 Study Design

This study had two main goals:

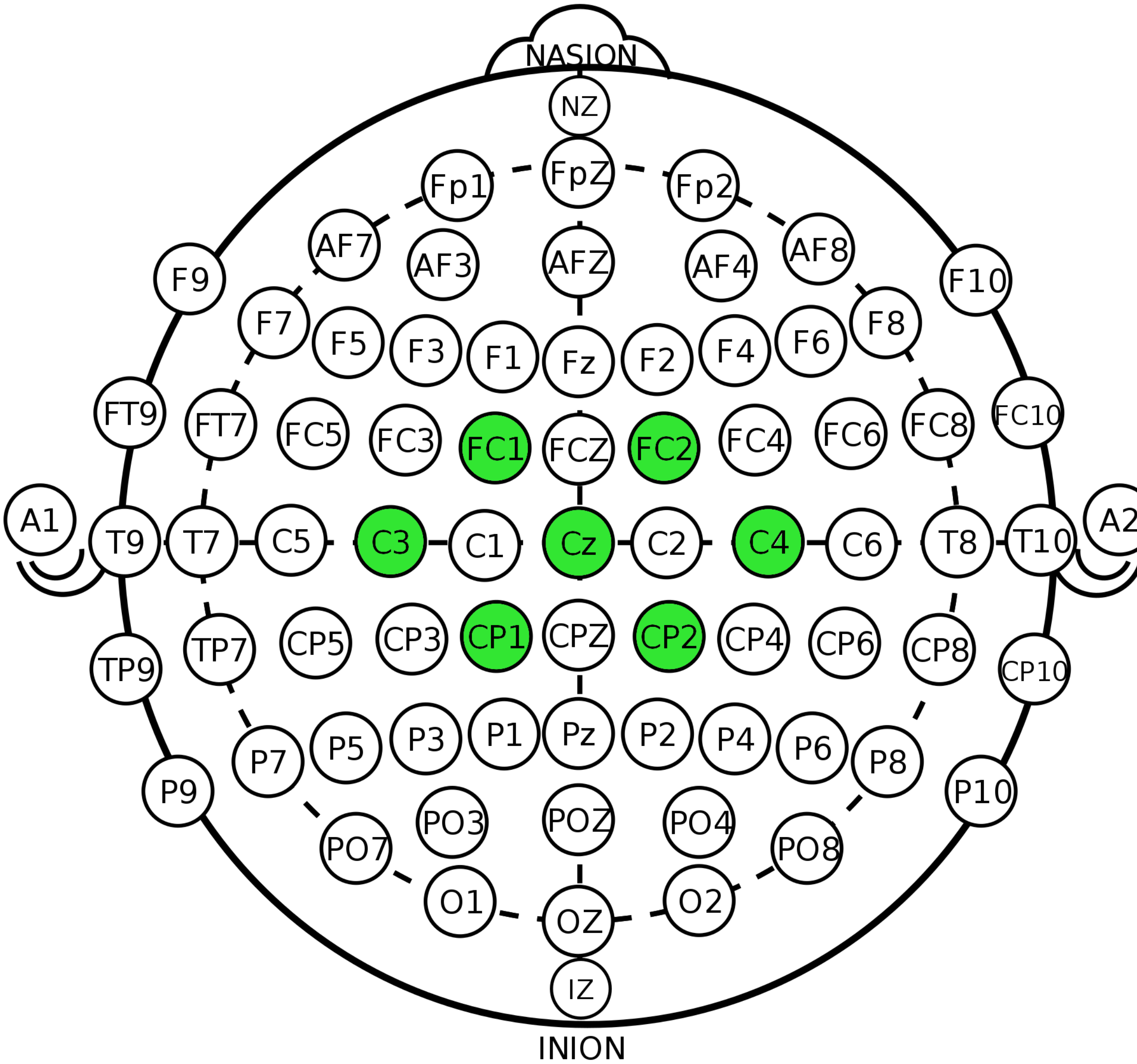

Figure 3.4: A layout of the International 10-10 system used for EEG electrode placements. The locations marked in green are the electrode locations used in the BCI study.

1. Understand the capabilities of the OpenBCI Ultracortex IV Headset to detect mu suppression associated with ERD events.

2. Determine the LSL and study system's responsiveness in recording EEG data.

The BCI study was conducted as a within-participants design, where each participant performed the same set of tasks, under the same conditions.

3.3.5 Procedure

Participants provided their vocal consent before the start of the study. They were instructed the day before their session to clean their head with shampoo and comb their hair. They were also asked to not use any hair products on the day of the study. The electrodes of the Ultracortex headset were used as a guide to find their eventual resting locations on the scalp of the participant. After determining their locations, a small amount of Nuprep EEG skin prep gel was applied to each location. Using a cotton swab, the gel was used to prepare the electrode contact locations by abrading the skin to remove dead skin cells and any other contaminants. The gel was cleaned off the skin before the headset was placed onto the participant's head.

Adjustments to the headset and calibrations to the EEG headset were made using the OpenBCI GUI. After the calibration, participants were given instructions for how to complete the study's tasks. Participants then underwent a two-minute baseline test prior to completing the tasks. Following the baseline test, the participants completed 10 blocks of tasks, with 10 trials per block. After all blocks were completed, participants finished one final baseline test before being disconnected from the headset. Lastly, participants were asked a series of subjective questions about their experience with wearing the Ultracortex headset and completing the study's tasks.

3.3.6 Results

Trials with eye blink artifacts, movement artifacts, and/or EEG amplitudes that exceed 5 *SDs* above or below the trial's mean were removed from offline data analysis as outliers. The EEG data recorded before and after a trial was kept in analysis, as it could provide useful information to events that precede or follow ERD and ERP events.

Participant Results

Participant EEG data was gathered from 7 electrodes as they completed the study. An example of the raw EEG collected is shown in Figure 3.5. Individual trials are divided between target trials (oddball stimulus) and non-target trials. An example of the raw EEG of an individual trial is shown in Figure 3.6. In both plots, the y-axis served as a relative scale.

The participant data was filtered during offline analysis using a 1 Hz high-pass filter and a 45 Hz low-pass filter. Outliers were removed through visual inspection of each trial. Individual participant data were averaged, separated between target trials and non-target trials. Artifacts (i.e., blink, movement, etc.) that occurred after the end of the instructed hand movements at the 4-second mark, were not considered to be causes of outliers, since the data in this period was not used in the frequency domain analysis. Trials with such artifacts were not removed from analysis. An example of a single participant's trial averages is shown in Figure 3.7.

Grand Averages

Grand averages of all trials, divided between target and non-target, were obtained by averaging the filtered EEG data from each of the participants, over each of the seven electrodes. The grand averages were only

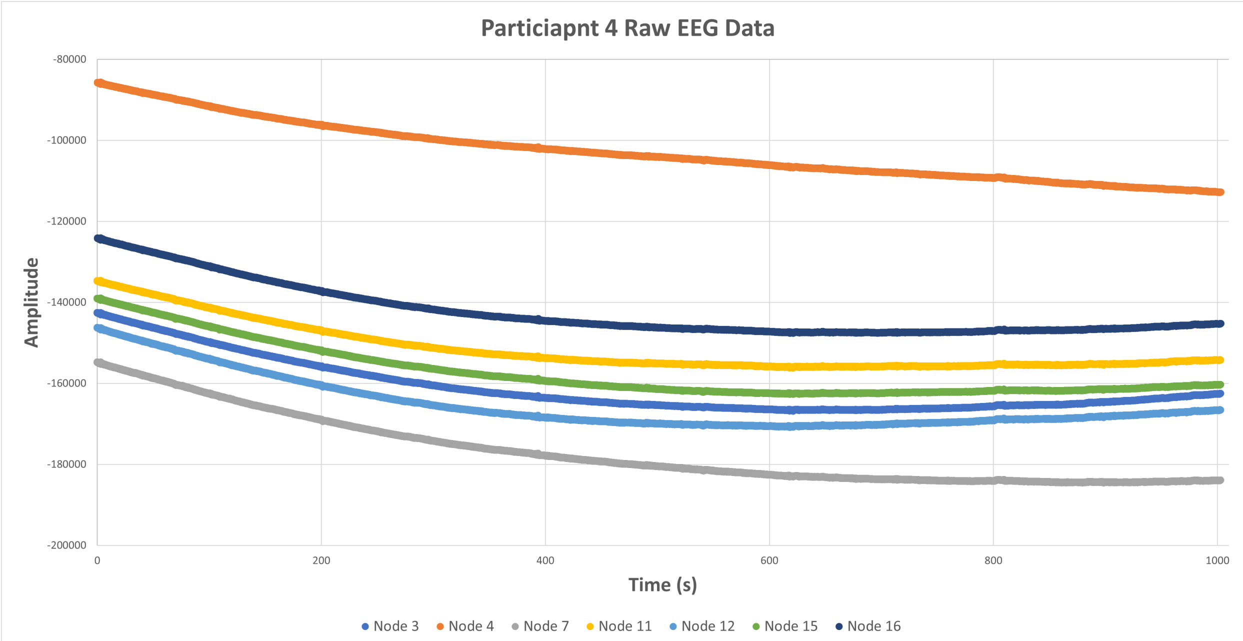

Figure 3.5: Raw EEG data collected from a participant over the course of the study. Each node represents one electrode attached to the participant's scalp.

calculated for a period of 6.5 seconds, where the stimulus was evoked at the two-second mark. The time domain plot of the grand averages is shown in Figure 3.8.

Fourier transforms were performed to convert the grand averages into the frequency domain at each electrode site. The data used started at the time of stimulus onset and ended two seconds afterwards, when the visual stimulus disappeared from the screen and the stop tone occurred. Difference waves were calculated by subtracting non-target trial frequencies from target frequencies. The grand averages in the frequency domain and the difference waves are shown in Figure 3.9.

3.3.7 Discussion

The results of the BCI study provided some insight into the EEG data being recorded by the Ultracortex IV Headset. Although there were only four participants in the study, the trials collected enough data to form grand averages that allow for review of the study system and methods used to gather motor imagery data.

Summary of Results

The results of the study can be summarized from the plots presented in the previous section.

1. Grand averages in the time domain showed small differences in responses at all electrode sites between target trials and non-target trials approximately one second after stimulus onset.

2. Grand averages in the frequency domain showed inconsistent difference waves throughout all frequencies, with the largest variance at lower frequencies.

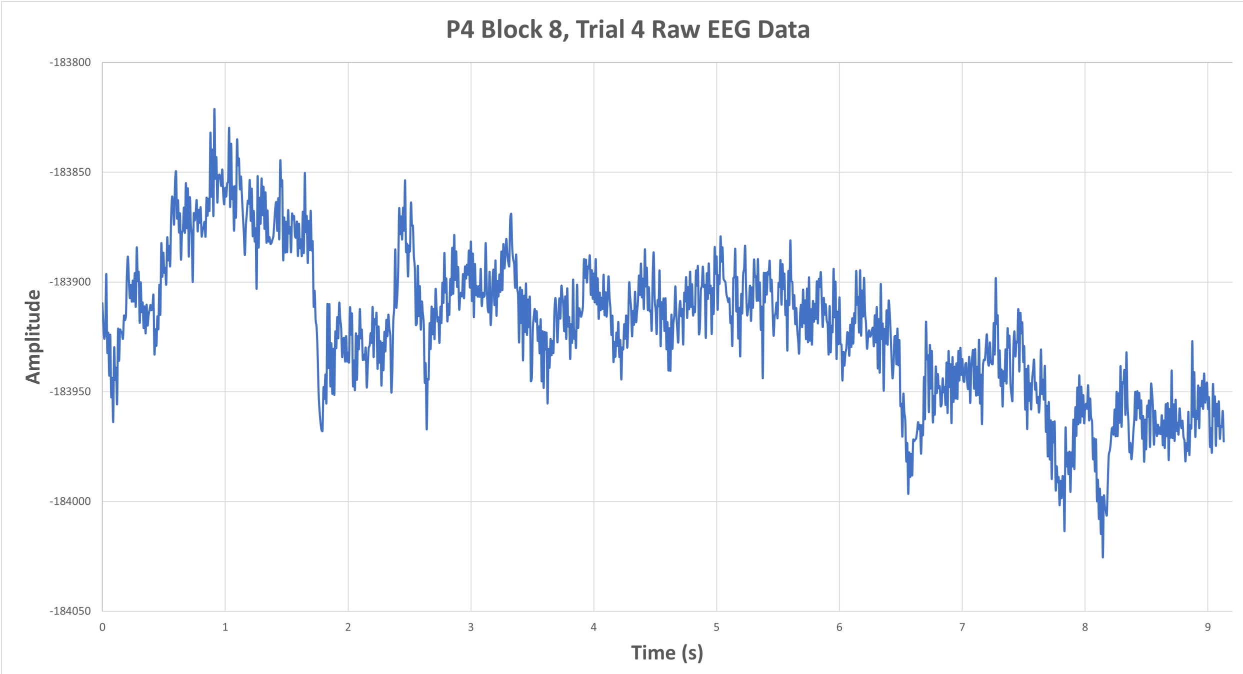

Figure 3.6: Raw EEG data of a single non-target trial at the Cz electrode, starting at three seconds before stimulus onset and ending two seconds after hand movement.

3. All participants experienced signal drift (primarily downwards) at all electrode sites over the course of the study.

Time Domain Results

The plots in Figure 3.8 show the time domain grand averages of the trials over the course of 6.5 seconds. Before the stimulus appears on the screen, most of the centrally located electrodes (i.e., C3, C4, CP1, and CP2) show a steady-state signal, which may point towards a relaxed user state as they wait for the stimuli to appear. Small oscillations prior to the stimulus are seen in most of the electrodes, which can be attributed to an anticipatory response as the participants were waiting for the stimuli to appear. Almost immediate responses are seen in the small peaks near the stimulus onset (approximately a few hundred milliseconds). These are likely the automatic responses to the stimulus, such as the P300 ERP component. One second after the stimulus is presented, a large reaction can be observed, represented by a sharp positive peak. The great change in brain activity is likely caused by a combination of the auditory tone that precedes the stimulus and the visual occurrence of the coloured circle on the screen, both of which are being consciously noted by the participant. Following the initial peak, most electrodes experience a few smaller negative and positive peaks (e.g., Cz) before eventually settling towards a neutral state. The sharp decline at the end of the trial, observed in all electrodes, is likely due to movement artifacts produced after the end of the trial (i.e., four seconds after the start of the trial) in the inter-trial period.

The grand averages in the time domain showed several differences in the EEG between the target trials

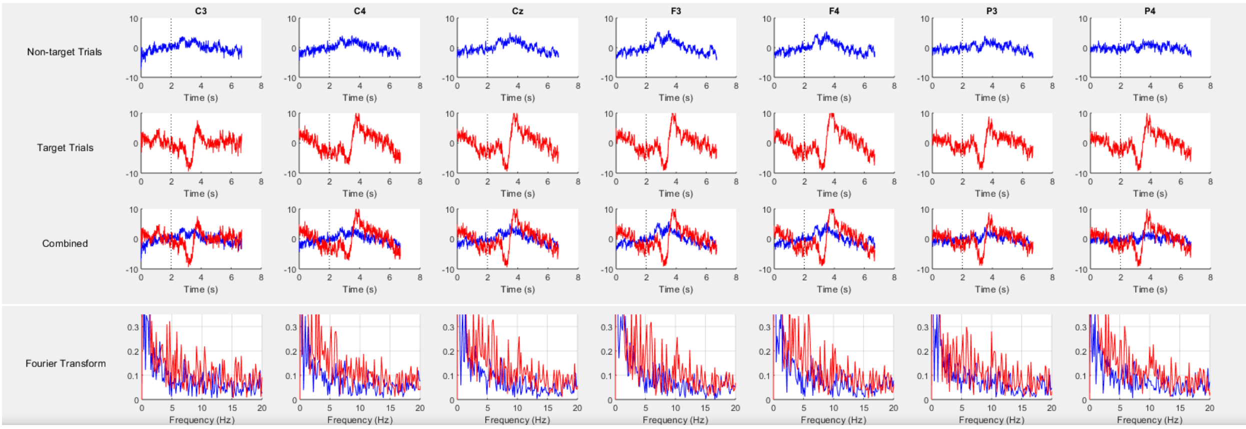

Figure 3.7: Averaged EEG activity at each electrode site for a single participant. Stimulus onset occurs at 2 seconds, marked by a dotted line. Non-target trials are plotted in the top row (blue) and target trials are plotted in the second row (red). The third-row plots overlay both target and non-target trials for visual comparison. The bottom row shows the Fourier transform of the respective electrode averages.

and non-target trials. The main difference is observed in higher, sharper peaks that occur in the target trials approximately one second after the stimulus is evoked, followed immediately by a more negative-going peak. This difference in the target trials is a common response seen in ERP studies as a reaction to oddball stimuli [119], such as the less frequent targets that instruct the participant to perform the motor actions with their hands. Infrequent stimuli elicit stronger EEG responses that can be used to isolate ERP components that react to the stimulus. The target trials generally show more EEG activity than in non-target trials, particularly noticeable when comparing the responses preceding the onset of a stimulus. This additional activity could be potential brain activity from the participants as they anticipate an upcoming stimulus. However, since the order of the trials were randomized without repetition, the participants should not have known beforehand whether the upcoming stimulus was an oddball target. It is more likely that the extra peaks observed before stimulus onset was the result of noise in the EEG data.

Similar plots were observed throughout the different electrode sites. Since both hands were used for the motor actions during the target trials, it is expected that the EEGs from both sides of the head are about equal. In both target and non-target trials, the electrode at Cz experienced more erratic peaks following the stimulus onset. This could possibly be evidence of motor actions in the target trials and suppression of movement in non-target trials. However, it is also possible that due to the position of the Cz electrode (i.e., directly on top of the head), it is more sensitive to small amounts of head movement, and the activity seen in the results are simply movement artifacts.

Frequency Domain Results

Figure 3.9 shows the Fourier transform of the time domain plots for the two second period when the visual stimulus was on screen. During this period, participants would have been performing the prescribed motor

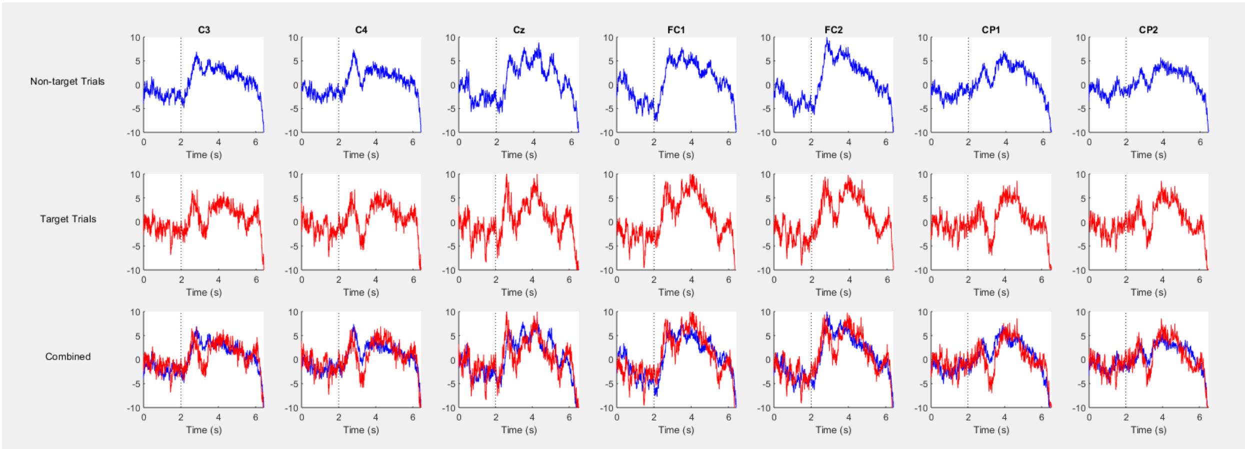

Figure 3.8: Grand averages of the trials for all participant EEG recordings at each of the electrode sites. Stimulus onset occurs at 2 seconds, marked by a dotted line. Non-target trials are plotted in the top row (blue) and target trials are plotted in the middle row (red). The bottom-row plots overlay both target and non-target trials for visual comparison.

actions with their hands in target trials. The frequency range of the plots was limited to a maximum of 20 Hz, since the primary objective of the BCI study was to observe mu suppression (i.e., ERD events) related to motor actions. Mu suppression typically occurs in the 8-13 Hz range [103] during a motor action, which is usually observed as a decrease in amplitude at that range. However, from the difference waves, there is inconclusive evidence to suggest mu suppression is being detected. Rather, the difference waves show a lot of variance between target and non-target trials, which appears more like random noise. This variance is greater in the lower frequencies, particularly at less than 1 Hz. Because the data was previously filtered with a high-pass filter at 1 Hz, this is likely noise created from the filtering process and design of the filters.

Generally, the target trials had frequencies with much greater variance as indicated by their thin, sharp peaks. This alternatively means that non-target trials experienced less activity during the stimulus. As such, it can be assumed that the higher variance associated with the target trials is caused by the hand movements. Although the difference waves do not to show the mu suppression that is usually correlated with motor actions, the movements during target stimuli could have been detected by the electrodes on the scalp. Similarly, slight head movements that the participant is unaware of could also lead to movement artifacts. This idea is supported by the movement artifacts that can be potentially seen in the time domain plots.

The difference waves show some slight differences between electrodes. The frequencies above 10 Hz in the C3, C4, CP1, and CP2 electrodes are slightly greater in target trials. This suggests some increased activity in the central and parietal regions of the brain during motor actions of the hands. Since the hand movements during target trials were conducted with both hands, there were no differences found in the electrodes located on different sides of the head.

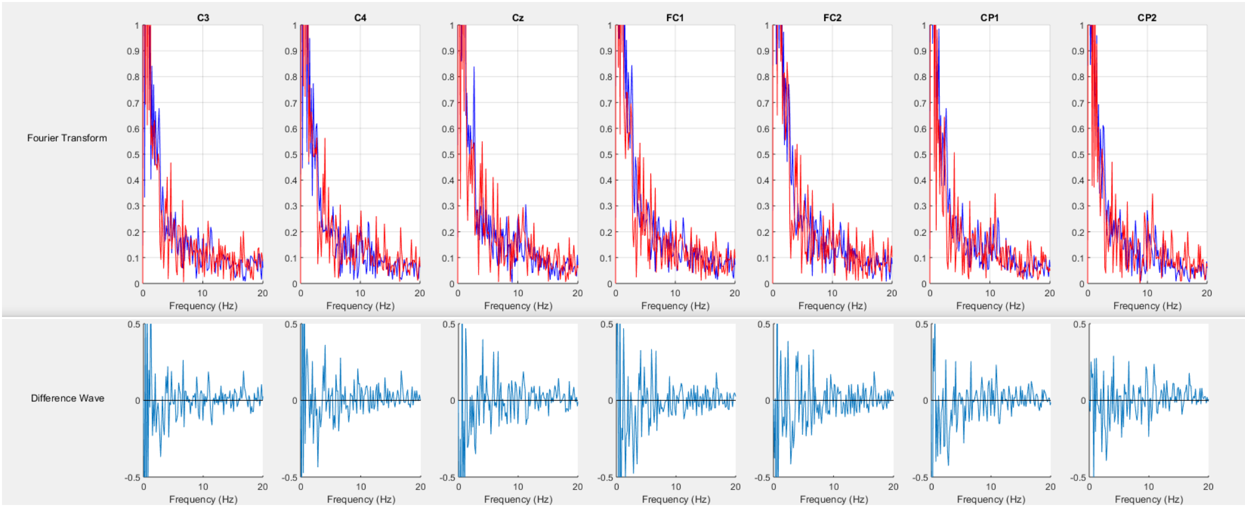

Figure 3.9: Grand averages at each electrode in the frequency domain. The top row shows the Fourier transform of the combined target trials (red) and non-target trials (blue). The bottom row shows the difference waves of the grand frequency averages, with non-target trials subtracted from the target trials.

Signal Drift

A downwards signal drift was experienced by all participants throughout the study. The effects were greatest near the beginning of the study, but eventually settled by the end of the 30-minute session, with some upwards drift starting to appear. A two-minute baseline was performed before the start of the study and another was done following the study's completion. These baselines were averaged and used to subtract from the study's data to minimize the effects of the signal drift. Furthermore, a high-pass filter was used to eliminate slow changes to the EEG data, which is a technique that has previously helped in eliminating drift [42]. However, these techniques of eliminating drift are usually done in offline analysis.

Offline Data Analysis

This study used offline analysis to procure the results shown in the previous section, as it provides several advantages. The main benefit of processing the entire dataset after it has been fully collected is that it allows for the calculation of grand averages. By averaging the data, some stationary noise can automatically be eliminated as it cancels itself out over the numerous trials. Another benefit of offline data processing is that it allows the researcher to spend time to carefully inspect the raw EEG data to remove outlying trials that contain too much noise or artifacts. For example, when trained in how to visually detect eye blink artifacts, a researcher can manually tag a trial when they find such an artifact. Furthermore, they can adjust the frequency of the filters by varying amounts to gain higher fidelity of their desired frequency range and observe the changes to the data that different filters produce. Various automatic techniques that extract and classify EEG features, such as principal component analysis (PCA), can be tested to find the technique that produces the best results for the desired feature [125]. These advantages of offline processing have been used extensively in traditional EEG research that study brain function; however, they do not typically apply for BCIs, which require real-time, online processing to function as an input mechanism for a computer system.

Real-time analysis requires a sophisticated system that can perform the tasks of preprocessing, feature extraction, and classification of incoming EEG data as quickly as possible. While other studies have shown that this is possible with motor imagery data [25, 138], the incoming data must be free from noise to a reasonable degree. Real-time filtering of frequencies that fall outside of the desired range can be easily done, but it is more difficult to detect and remove artifacts. Artifacts produced through eye blinks typically follow a usual pattern but shifts to the position of EEG electrodes can cause noise that appear as a feature. When this kind of noise occurs during an input action for a BCI, the system may accidentally interpret the noise as a feature, leading to the wrong command being produced. In cases where machine learning is used for advanced feature extraction and classification, the system can be trained with offline BCI data to teach it to recognize these artifacts. However, this would also require additional sessions to procure the training set that would be unique and relevant for a single user.

Presence of Noise

From the time domain grand averages, noise was observed before the stimulus was evoked in most electrodes during the target trials. There is not enough information with the current data to uncover the source of this noise, but the effect of the noise is not simply isolated to before stimulus onset. Since the noise was present before the stimulus appeared, there is no reason to suggest that it did not continue into data recorded after stimulus onset during target trials. It is possible that this noise was part of the differences seen between target and non-target trials. It is unclear why this noise was more apparent in target trials than non-target ones.

The frequency domain shows the noise more clearly. Across the entire range, the frequency varies greatly with sharp peaks. The highest amplitudes are found in the lowest frequencies (i.e., less than 1 Hz), which is where the largest variations are found as well. The difference waves also exhibit a high degree of noise. When looking at only the relevant frequency range for mu waves (i.e., 8-13 Hz), there appears to be no consistent difference between target and non-target trials across all electrodes. The difference varies in a noise-like fashion, with a mean around zero. With these noisy results, it is difficult to visually extract features or patterns from the plots. The amplitudes of the noisy peaks in the difference waves appear greater than any potential feature that could be discovered, even if a technique like PCA was used.

3.3.8 BCI Study Summary

The purpose of the BCI study was to understand how well the Ultracortex headset can detect mu suppression events and its performance in recording EEG data. Offline analysis was used since the data processing techniques are simpler and quicker to implement (i.e., an experienced researcher manually removing noisy trials versus developing a system for real-time processing). From the observed results, it was difficult to find evidence of mu suppression occurring due to motor imagery. Rather, the noise that remained in the grand averages after filtering and removing outlying trials showed that either the methods used were flawed, the suppression was too weak to detect, or a combination of both.

The Ultracortex headset was capable of recording EEG data that showed differences when the user was presented with an oddball stimulus. However, unknown sources of noise prior to stimulus onset meant that this data was also subject to continuous noise in the target trials. Furthermore, when a Fourier transformed was performed on the EEG recorded after a stimulus was evoked, the frequencies were too noisy to make out any features or patterns in the results. The difference waves were also too noisy to find substantial differences between the target and non-target trials.

3.4 Issues of Noise in EEG Data

The grand averages calculated from the EEG data recorded during the study were permeated with noise. This noise made it difficult to visually appraise the information gathered from the electrodes and discover any useful features or patterns from the results. Although a more aggressive approach to eliminating noise could be used (e.g., filtering closer to the frequency range of interest), it is possible that useful information can also be removed by this process. Similarly, additional noise could be generated from these methods if used improperly. While some findings from the study can be speculated, substantial conclusions cannot be drawn while the noise remains. Furthermore, developing a system that can automatically interpret these results is difficult when the uncertainty caused by the noise is so prevalent.

For BCIs to function as an input mechanism, they need to be able to interpret the user's input, in real time, that comes in the form of EEG signals recorded from electrodes on the user's scalp. These input signals need to be correctly interpreted by the system such that the user's intended command is executed when the system has completed processing of the incoming data. To improve the user experience with using the BCI, the processing of the input must occur as quickly as possible, meaning that the system must be able to conduct the steps of preprocessing, feature extraction, and classification without any human oversight. As such, any noise that interferes with this process can cause the system to misinterpret the input.

Since brain activity at the scalp is usually weak, the electrodes can detect the smallest changes in electrical activity. Normally, this means that more underlying signals from the brain can be detected. However, this also means that greater activity that is also detected by the electrodes will mask the desired signal. As shown in the previous sections, EEG electrodes are highly sensitive to many sources of noise. This noise can come from other nearby electrical signals in the surrounding environment, artifacts created from the user moving, or even unrelated brain activity. When unrelated signals are introduced into the input, the data received by the system becomes noisy.

While the EEG input remains highly susceptible to noise and artifacts, it becomes difficult to develop a motor imagery BCI that can feasibly function as a novel, alternative input mechanism. Examples from previous literature have shown successful implementations that allow users to control a computer with imagined movement alone [25]. However, many studies also show a low degree of accuracy in the computer system's ability to correctly, and consistently, interpret the user's input and execute the appropriate command [106, 135]. It can be difficult to remove noise from the input signal, since many current solutions for offline noise processing of EEG data (e.g., manual artifact rejection) cannot be realistically applied for real-time BCI use. The ecological viability of BCIs as an alternative human-computer interface will continue to remain limited until a universal method for easily and quickly removing noise from the EEG input is developed. Until then, users who rely on BCIs as an input mechanism will need to be able to handle errors caused by noisy input. This is even more important for new users of the device, as the effects of noise on learning is still an underdeveloped field of study. These issues of noise encountered in interpreting the results of the BCI study served as motivation for the following chapter and the work to investigate the effects of interpretation errors in learning to use novel input mechanisms.

4 Effects of Interpretation Error on Learning

The study of a brain-computer interface presented many issues regarding the noise found in the EEG input. Due to the high rate of interpretation errors occurring because of the noise, it was unfeasible to continue studying the BCI as a novel input mechanism. While further efforts of reduce the noise could have produced better results (e.g., using morphological component analysis to remove artifacts [149]), its presence would have remained. Interpretation errors that are the result of noise can affect a user's ability to successfully use an input mechanism regardless of their experience with using that mechanism. As such, it is important to understand how noise and interpretation errors can impact a user.

While there has been previous research that has studied how computer system interpretation errors affect a user's performance and satisfaction with that system [129, 78], the effects of these errors have not been assessed with their impact to learning how to use the system. Although the assumption that the occurrence of interpretation errors may negatively affect the user's learning, there have also been studies in psychology and, more specifically, HCI that show a difficult learning process may improve their performance through harder work. Two hypotheses were drawn from these ideas: an interference hypothesis, where more interpretation errors will lead to worse learning of the input mechanism (and negatively impacting the user's ability to associate input actions with computer commands); and a retrieval effort hypothesis, where increased interpretation errors will lead to users putting in more effort to better learn the system.

To better understand these hypotheses, two user studies were designed to determine the effects of interpretation error on a user's ability to learn how to use an input mechanism to perform simple selection tasks.

4.1 Study 1: Effects of Interpretation Error Rate on Learning a Command Set, with Immediate Feedback

The first study tasked participants with learning to use a set of commands to select a target item from hierarchical menu. Menu-based selection was used as many input mechanisms use menus to navigate a computer system (e.g., on-screen gestures). Participants were asked to repeatedly select twelve items, organized into four categories, at different rates of interpretation error in their input. To study the effects of varying error rates, artificial errors were added to keyboard arrow key inputs, an input mechanism that normally has no interpretation errors. Participant's performance was recorded through their task completion times and correct selection accuracy throughout training blocks (where they were provided a visual guide) and memory tests (where they had to select the target based on recall).

4.1.1 Study System and Menu-Selection Task Design

A web-based application was developed in HTML, CSS, and JavaScript. A target item was displayed in the middle of the screen, asking users to select that item using the arrow keys from a hierarchal menu. The menu was organized into four categories of three items each with a layout matching the cardinal directions of the keyboard's arrow keys. An example of the menu layout is shown in Figure 4.1.

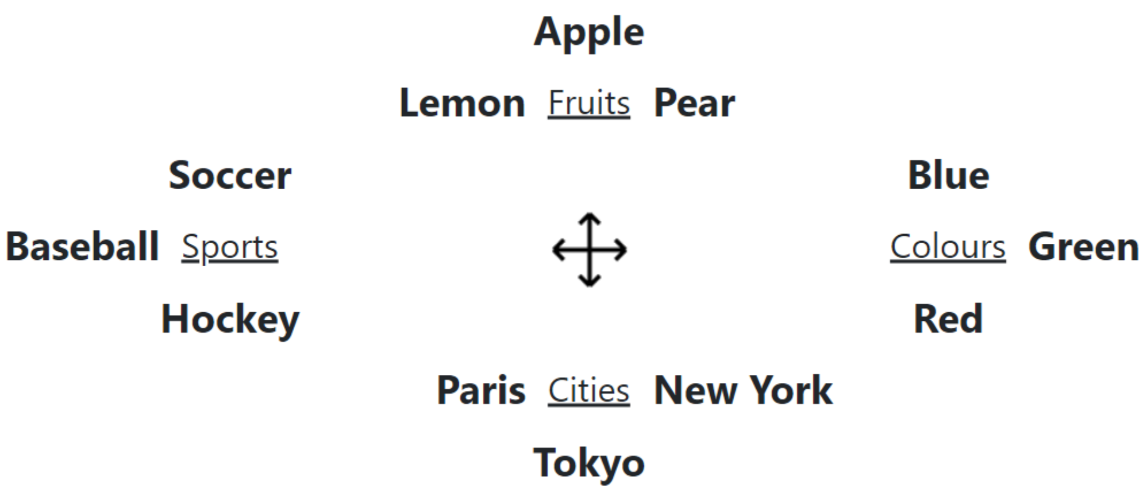

Figure 4.1: An example of the visual guide that participants could view using the Left Shift key. The layout of the categories and items indicated the arrow key direction to press to select that category/item.

To select an item, participants would first press the direction of the category that the item belonged to, and then press the direction associated with the item (e.g., pressing Up in Figure 4.1 would select the Fruits category and then pressing Left would select the Lemon item). Participants were given 500 ms following an item selection to make a correction if either they or the system made an error. The duration for error correction was determined through pilot testing of the system.

The study involved a total of nine training blocks and three memory tests; a memory test was completed after every third training block. Both training blocks and memory tests required participants to select each of the twelve items from the menu in a random order (sampling without replacement). During the training blocks, participants were provided with a visual guide (e.g., Figure 4.1) that they could access at any time (using the Shift key). While visible, the guide showed the participant's current location in the menu by highlighting their selection on the guide, as well as showing the item/category in a display box. When a selection was resolved by the system, a correct/incorrect response was provided by the display box (i.e., a green or red border, respectively). A screenshot of the system during the training blocks is provided in Figure 4.2.

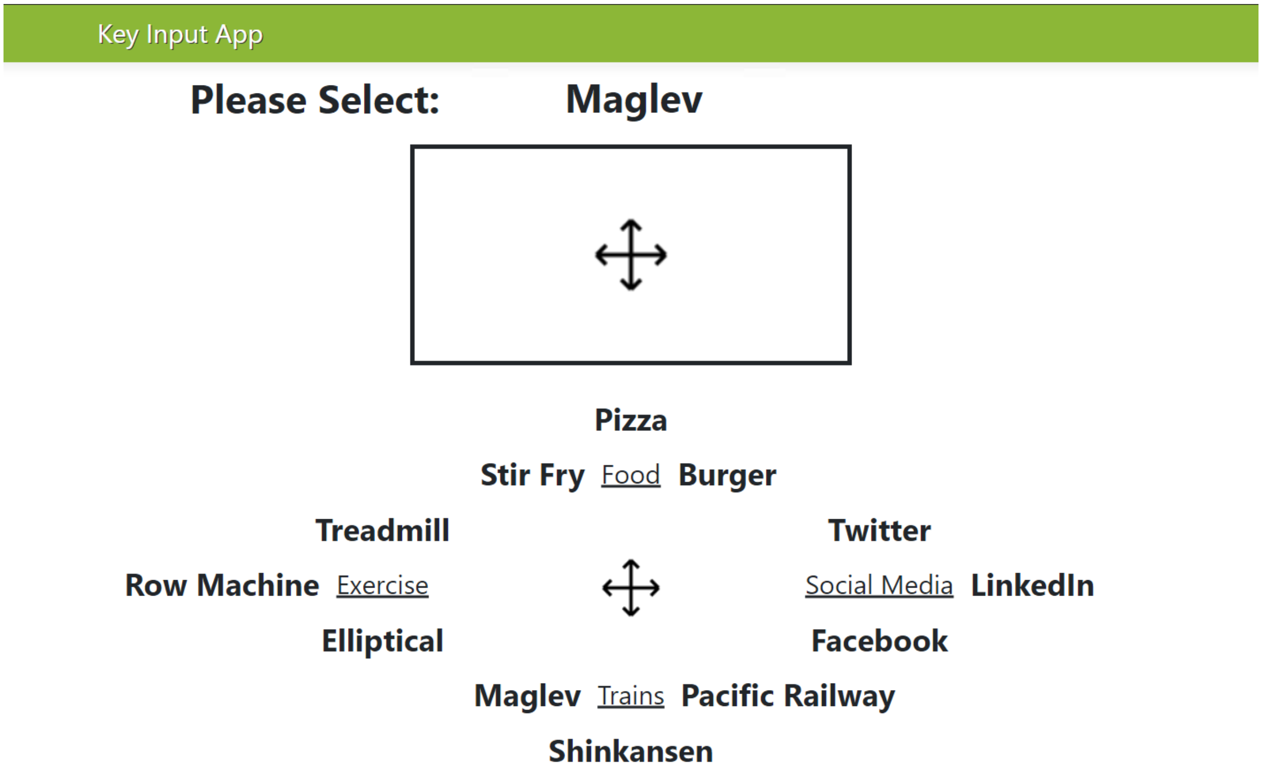

Figure 4.2: An example of the system interface that participants would see when completing a trial in the training blocks in Study 1.

The memory tests were similar to the training blocks, except participants were not shown their location in the menu, could not view the guide, and were given only one opportunity to select the correct item within two keypress per target. To accurately measure participants' performance, no interpretation errors were included throughout the memory tests.

4.1.2 Experimental Conditions

Interpretation error rate (IER) is the rate at which the system added artificial errors to the participant's input. Each artificial error had an equally likely possibility of being either a misinterpretation or a false negative. Misinterpretations would cause an input to be interpreted as a different one (e.g., pressing Left resulted in the system detecting an Up press); the new input was chosen at random from one of the directions not pressed. False negative errors would cause the input to be ignored (e.g., pressing Down resulted in no change). The IER would emulate the rate at which incorrect interpretations occurred within the system. Participants were all informed prior to the start of the tasks that there was a possibility of system-based errors occurring during their inputs.

The study used three levels of interpretation error rate:

- *Zero percent error*: no interpretation errors were added to the participant's input.
- *Five percent error*: each input towards selecting an item (i.e., arrow key press) had a 5% chance of

being interpreted incorrectly as either a misinterpretation or false negative.

- *Twenty percent error*: probability of an interpretation error occurring was 20%.

In all conditions, user-borne errors (e.g., inadvertently pressing the wrong key, or incorrectly remember the location of an item in the menu) were possible and were not manipulated by the study system. False positive interpretation errors (e.g., pressing the wrong key, but the system selected the correct key; or the system making a selection when no key was pressed) were not included in our study as a possible type of error.

4.1.3 Participants

Participants were recruited through Amazon's Mechanical Turk (MTurk) service, which is an online platform where users (MTurk workers) can opt-in to complete Human Intelligence Tasks (HITs). Previous studies in HCI have also used data gathered from MTurk [47, 64], but causes of outliers due to bots and negligent workers had to be accounted for during data analysis. Only workers from the United States and with a HIT acceptance rate of 90% or greater (a measure of work quality based on previous tasks) were accepted into the study. Survey response times were recorded and checked to ensure that they were not completed too quickly without reading the questions, or that the same answer was used for every response.

75 participants were recruited (51 men, 22 women, 2 non-binary) aged 20-72 (mean = 35.7, SD = 11.3). Participants were all paid $4 for completing the study, which took approximately 20 minutes. Two participants reported not having a standard arrow key layout: one reported that his Up and Down keys as being fitted into the space between the Left and Right keys. Self-reported estimates of weekly computer usage among participants averaged 39.9 hours (SD = 20.5 hours). Regarding their experience with alternative computer input mechanisms, 20 participants reported having used swipe typing, 20 had used gesture commands, 28 had used speech commands, and 12 had used handwriting recognition. No participants were removed from the study upon review of their survey responses.

Ethics approval for this study was provided by the ethics review board of the University of Saskatchewan.

4.1.4 Procedure

Participants completed an informed consent form before starting the study. They were then asked to complete a demographics questionnaire and shown an instructional video that described the study's tasks and how to perform them. After finishing the video, they were directed to a one-minute timed practice session where they used the arrow key controls to select items. Following the end of the practice session, participants completed 12 blocks (9 training blocks; 3 memory tests) of tasks, with 12 trials (i.e., target items) per block, for a single IER condition. After completing all blocks, participants were asked to complete a final questionnaire that included questions regarding their perceived effort (i.e., NASA Task Load Index [46]), and their perceived learning and memorization of the items.

4.1.5 Study Design

This study attempted to address three main research questions (RQ):

- RQ1-1: Does increasing interpretation error affect user learning (i.e., accuracy and completion time of memory tests)?
- RQ1-2: Does increasing interpretation error affect performance in training (i.e., error rate and completion time of training blocks)?
- RQ1-3: Does increasing interpretation error affect user perception of effort or ease of learning (i.e., subjective questionnaire responses)?

The study used a two-factor design, 3x9 for the analysis of training blocks, and 3x3 for the analysis of the memory testing conditions:

- Interpretation Error Rate (IER), between-subjects: Zero, Five, or Twenty percent error
- Block/Test, within-subjects, 1-9 for training blocks, and 1-3 for memory tests

The primary dependent variables considered were: accuracy and completion of memory tests, user error count and completion time in training blocks, and subjective ratings of effort from post-session questionnaires. Other measures were also gathered (e.g., the rate at which the guide was used during the training blocks, time when the guide was accessed, time of the participant's first keypress within a trial) to explore the participant's performance in further detail.

4.1.6 Results

The effect size for significant repeated-measures (RM) ANOVA results are reported as generalized eta-squared for both studies: η^2 (with values less than .01 considered small, approximately .06 considered medium, and greater than .14 considered large [24]).

Outlier data points were determined as any trial with a completion time or first keypress time (i.e., the duration between the start of a trial and when the participant first pressed an arrow key) greater than 3 *SDs* above the block's mean completion time. In total, 404 trials out of 10800 total trials were removed from the data analysis of Study 1. In training blocks, any trials with artificial interpretation errors (i.e., misinterpretations or false negatives) were also discarded from the analysis to prevent the average block completion time from being inaccurately inflated. No artificial errors were added to any of the trials during the memory tests.

All pairwise t-tests were corrected using the Holm-Bonferroni method.

Effects of IER on User Learning

Overall accuracy in the memory tests was calculated as the number of correct selections made out of the 12 trials in each test; a correct selection required the participant to correctly choose both the correct category and correct item in their first attempt. Overall, accuracy was 82.7% in the Zero condition, 72.0% in the Five condition, and 70.0% in the Twenty condition. No trials were discarded in this analysis beyond the standard outlier detection (i.e., completion time greater than 3 *SDs*). The accuracies over the three memory tests, including overall accuracies and separate accuracies in category and item selection, are shown in Figure 4.3.

A two-way RM-ANOVA (IER×Test) showed significant effects of both IER and Test on accuracy (IER: $F_{2,216} = 5.4, p < .001, \eta^2 = 0.047$; Test: $F_{2,216} = 14, p < .0001, \eta^2 = 0.11$). There was no interaction between IER and Test ($F_{4,216} = 0.0078, p = .99$). Post-hoc pairwise t-tests showed significant differences between the Zero condition and both Five and Twenty conditions (both $p < .05$), but not between the Five and Twenty conditions ($p = .65$). Post-hoc testing between the memory tests also showed that the 2nd and 3rd tests had significantly higher accuracy than in the 1st test (both $p < .001$), but no difference between the 2nd and 3rd tests ($p = .21$).

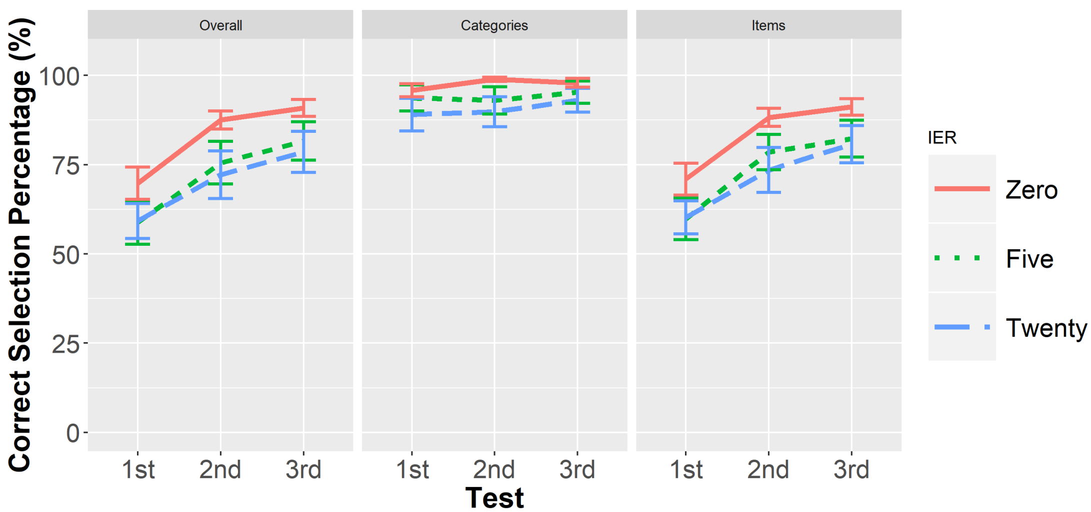

Figure 4.3: Study 1. Mean selection accuracy in the memory tests, by IER and Test.

Completion time of the memory tests was measured as the time from when the target item appeared on the screen until the time that the system registered an item selection (regardless if the selection was correct or not). No trials were discarded in this analysis beyond the standard outlier detection. The average completion time of the Zero condition was 2.51 s, of the Five condition was 2.32 s, and of the Twenty condition was 2.45 s. The mean completion times of each condition across the three memory tests are shown in Figure 4.4.

RM-ANOVA (IER×Test) showed a significant effect of Test on completion time ($F_{2,216} = 10, p < .0001, \eta^2 = 0.088$), but no effect of IER ($F_{2,216} = 0.89, p = .41$) and no interaction ($F_{4,216} = 0.0021, p = .999$). Post-hoc t-tests showed significantly difference completion times between the 1st test and both 2nd and 3rd tests (both $p < .0001$), but not between the 2nd and 3rd tests ($p = .077$).

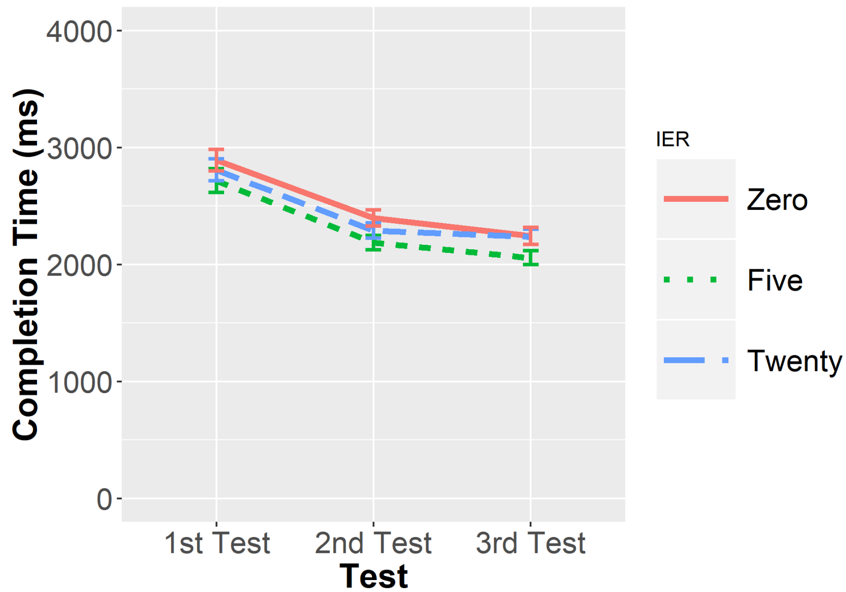

Figure 4.4: Study 1. Mean trial completion times in memory tests, by IER and Test.

Effects of IER on User Performance During Training

User errors during the training blocks were measured as the number of arrow keypresses made that were incorrect, when the participant was in a position to perform the correct keypress (e.g., if the participant was at the start of a trial, and the correct category was in the Left direction, but the participant pressed the Down arrow key, one error was counted in this case). Errors made as the result of an artificial system interpretation error were not considered as user errors. Any trials with an artificial error added to the participant's input were discarded from the analysis along with the standard outlier removal method.

The average user error rate of the Zero condition was 0.26 errors per trial, of the Five condition was 0.39 errors per trial, and of the Twenty condition was 0.83 errors per trial. User error rate of each condition across the training blocks are shown in Figure 4.5.

RM-ANOVA (IER×Block) showed significant effects of both IER and Block on user errors (IER: $F_{2,648} = 21, p < .0001, \eta^2 = 0.060$; Block: $F_{8,648} = 6.2, p < .0001, \eta^2 = 0.071$). There was no interaction between IER and Block ($F_{16,648} = 0.91, p = .56$). Post-hoc pairwise t-tests showed significant differences between all IER conditions (all $p < .01$). Post-hoc tests on neighbouring training blocks only showed differences between blocks 1 and 2 ($p < .0001$).

Trial completion time in the training blocks was measured from the time when the target item appeared until the time that the system registered the correct item selection. Trials with artificial system interpretation errors were discarded from analysis along with the standard outlier removal method. The average completion time in the Zero condition was 2.81 s, in the Five condition was 2.66 s, and in the Twenty condition was 2.86 s. Completion times for each condition across the nine training blocks are shown in Figure 4.6.

RM-ANOVA (IER×Block) showed a significant effect of Block on completion time ($F_{8,641} = 18, p <$

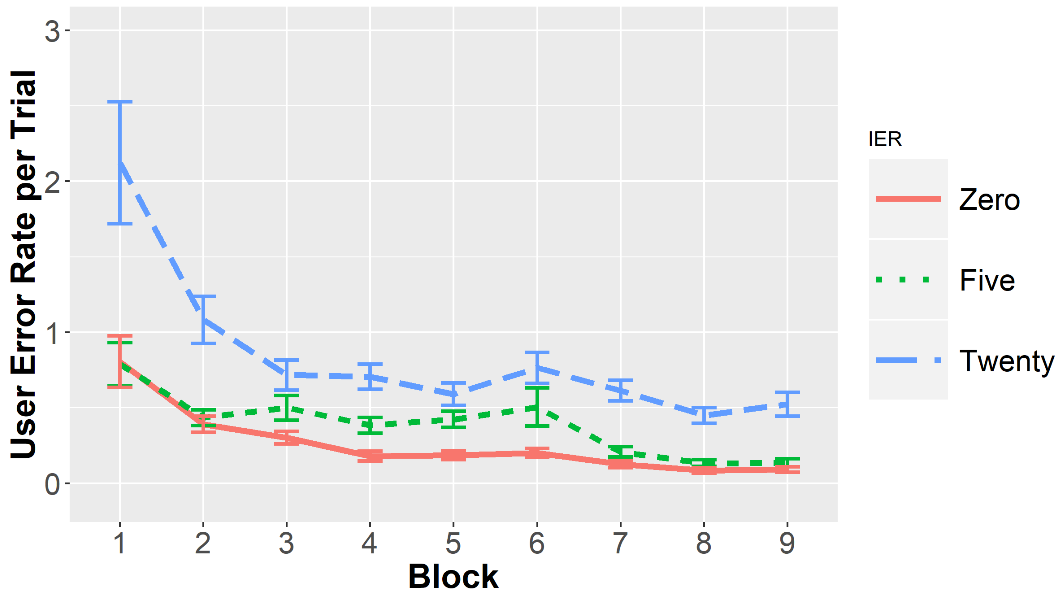

Figure 4.5: Study 1. Mean user errors per trial in training blocks, by IER and Block.

$.0001, \eta^2 = 0.19$), but no effect of IER ($F_{2,641} = 2.1, p = .13$) and no interaction ($F_{16,641} = 0.92, p = .54$). Post-hoc pairwise t-tests on neighbouring blocks showed significant differences between blocks 1-2 and blocks 2-3 (both $p < .05$).

The mean total training time of each condition was measured to determine whether differences in the overall time spent in training would affect the differences seen in the various IER conditions. No trials were discarded for this analysis. The overall completion time of the training blocks in the Zero condition was 308 s, in the Five condition was 336 s, and in the Twenty condition was 468 s. A one-way ANOVA of the overall completion time of the training blocks showed a significant effect of IER ($F_{2,72} = 11, p < .0001, \eta^2 = 0.23$). Post-hoc pairwise t-tests found significant differences between the Twenty condition and both Zero and Five conditions (both $p < .01$).

Subjective Measures

The perceived effort of participants was recorded using the NASA Task Load Index (TLX). The mean scores are shown in Figure 4.7. Aligned Rank Transform (ART) were performed on the aggregated TLX responses [146]. An ART was used to enable the calculation of ANOVAs for non-parametric data by transforming the data to be suitable for an ANOVA. ARTs are often used in HCI studies to allow factorial analysis for examining interaction effects of multiple experiment variables. Using ARTs, one-way ANOVAs were performed on each of the TLX questions with the factor of IER. Significant effects were found in the responses for physical effort, temporal effort, and frustration (all $p < .05$). The full results of the ANOVAs are tabulated in Table 4.1.

Post-hoc pairwise t-tests were performed on the questions that had significant effects. For both physical and temporal effort, significant differences were only found between the Zero and Twenty conditions ($p =$

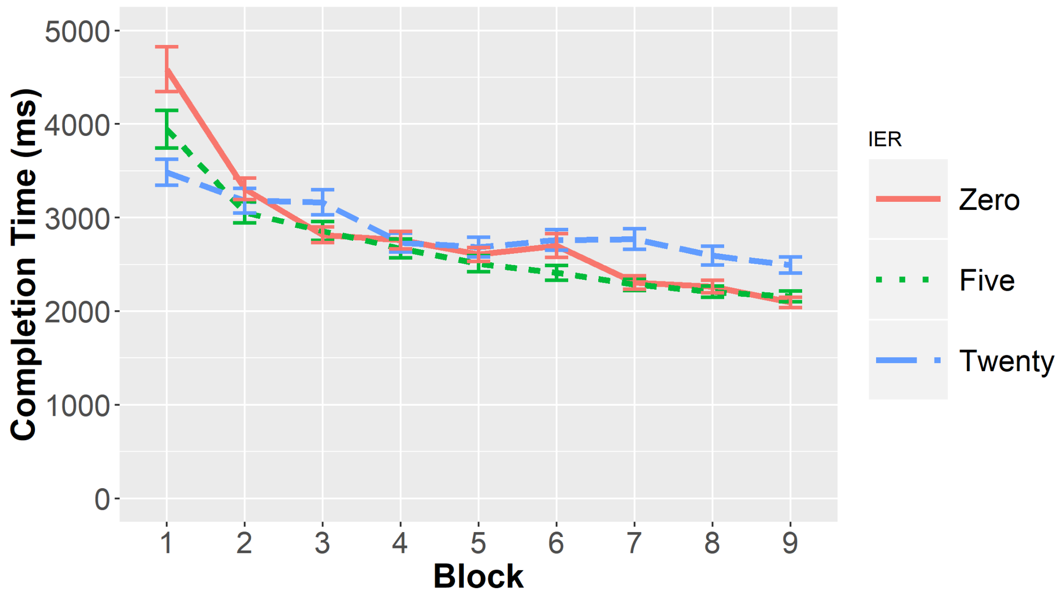

Figure 4.6: Study 1. Mean trial completion times in the training blocks, by IER and Block.

Table 4.1: Study 1. Results of the ANOVAs conducted on each TLX effort question after an Aligned Rank Transform.

TLX Question	$F_{2,72}$	p
Mental	3.0	.055
Physical	5.8	$<$ **.05**
Temporal	5.2	$<$ **.05**
Performance	0.068	.93
Effort	0.62	.54
Frustration	4.9	$<$ **.05**

.0035 and $p = .0057$, respectively). For frustration, significant differences were found between the Twenty condition and both Zero and Five conditions (both $p < .05$).

Participants were also asked to rate their perceived ease of learning and memorization on two separate 7-point Likert scales. The mean scores are shown in Figure 4.8. Using Aligned Rank Transform on the aggregated responses, one-way ANOVAs were performed on each of the questions with the factor of IER. No significant effects were found in the responses of either of the questions for ease of learning ($F_{2,72} = 0.011, p = .99$) and ease of memorization ($F_{2,72} = 1.2, p = .29$).

4.1.7 Study 1 Summary

The main goal of Study 1 was to determine if interpretation errors would affect a user's learning and transition to expertise for a menu selection task, as well as their performance during training. Participant data was

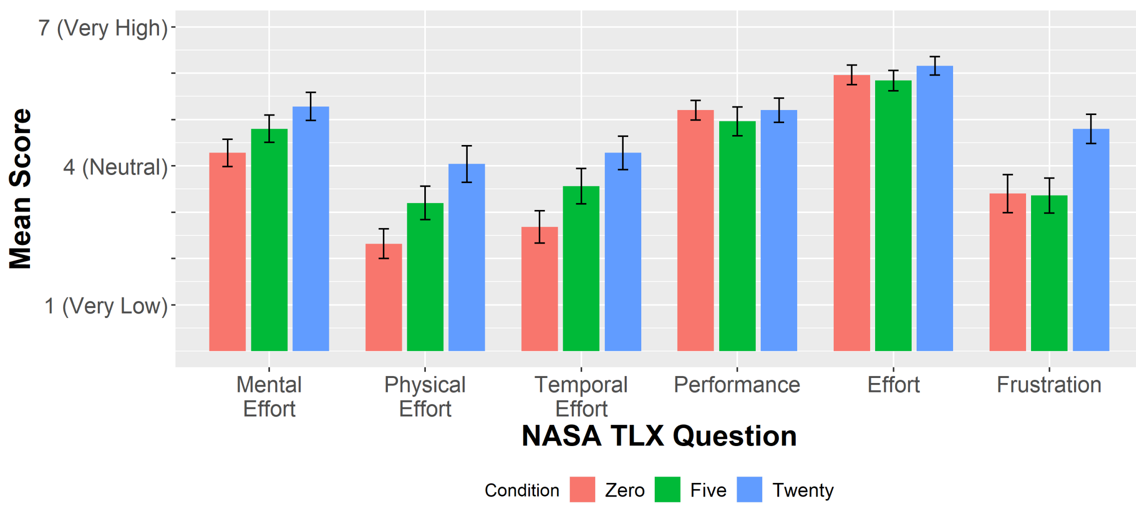

Figure 4.7: Study 1. Mean NASA Task Load Index scores, by IER.

gathered with a user study and analyzed their data using metrics such as accuracy and completion time during the memory tests, and error rate and completion time during the training blocks. Participants were also asked to rate their perceived effort, learning, and memorization of the study's tasks.

During the memory tests, the Zero IER condition was significantly more accurate than both the Five and Twenty conditions, but there was no difference in the completion times. Accuracy also increased significantly after the first test, while completion time decreased in a similar manner, showing the greatest improvement after having completed a memory test once.

During the training blocks, the Zero condition also had significantly lower user errors per trial, as well as lower total completion time of the training blocks. There was no difference between conditions for the average completion times of the individual trials. User errors were the highest in the first training block but showed a significant decrease in the second block. Similarly, completion times also significantly decreased within the first three blocks, showing that participants had the largest performance gains near the beginning of the study.

Subjective responses for perceived effort (i.e., physical effort, temporal effort, and frustration) showed a generally increasing trend as the rate of artificial errors increased, with the Zero condition requiring less effort and the Twenty condition requiring the most effort. The different rates of interpretation error did not affect participant's perception towards how easy it was to learn and memorize the item set.

In general, this study showed that increasing IER would lead to a reduction in user performance and their learning of item placements in a menu. This is represented by the decreasing accuracy in memory tests and higher completion time in the training blocks (both overall time and average trial times) as IER increased. This first study provided immediate feedback to whether a participant's selection was correct. To determine the ecological validity of these results in a real-world system, a second study was carried out that would require participants to monitor their input and check if any errors had occurred.

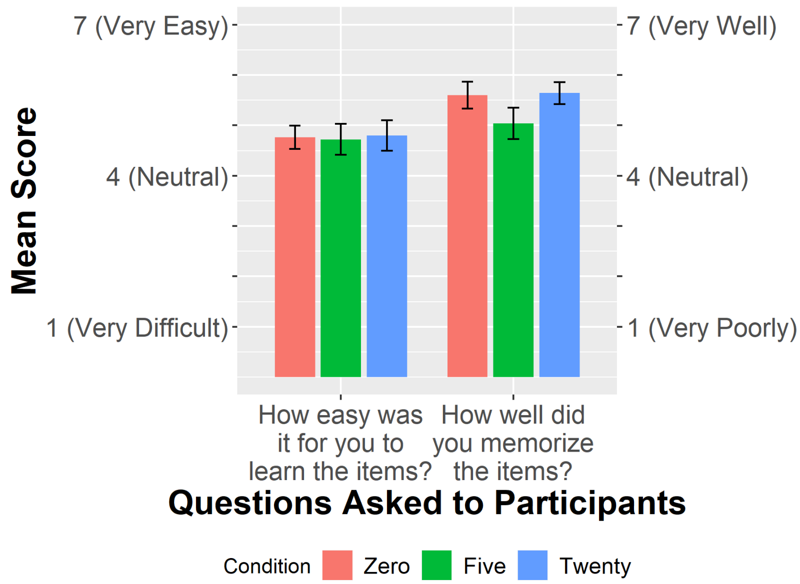

Figure 4.8: Study 1. Mean scores for perceived ease of learning and memorization, by IER.

4.2 Study 2: Effects of Interpretation Error Rate on Learning a Command Set, with User Monitoring System Interpretations

The second study expands on the methods of the first study by requiring participants to manually check the system's interpretation of their selections. By doing so, this would mirror a real-world system more closely as users often need to check their input for correctness (e.g., autocorrection on smartphones can sometimes misinterpret a user's word or even change a correct word into an incorrect one). Similar to the previous study, this study asked participants to repeatedly select twelve items from a menu with differing rates of interpretation error between participants. Instead of providing immediate feedback whether their interpreted selection was correct, participants were required to monitor the output and compare the selections with the targets themselves.

4.2.1 Study System and Menu-Selection Task Design

The system used in this second study was similar to the one used in Study 1. However, the main difference was that instead of being shown a single target item at a time (and allowing participants to select an item from the hierarchical menu to complete the trial), participants were provided with an ordered list of all of the target items for the entire block at once. The target list was provided in a box on the left edge of the screen, and the order of the items were randomized between each block. Participants were required to correctly select each of the target items in the same order as presented in the list. Once a selection was made, it would appear in a box on the right edge of the screen, mirroring the box for target items. Furthermore, unlike the previous study, the display box in the middle of the screen would not indicate whether a select was correct. An example of the study screen is provided in Figure 4.9.

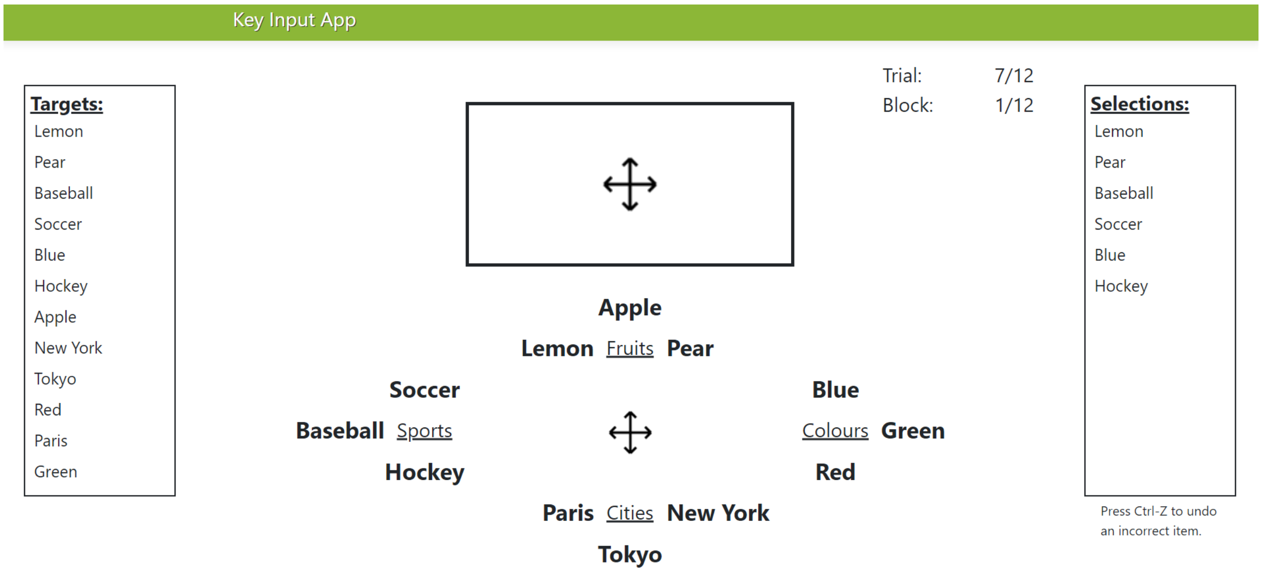

Figure 4.9: An example of the screen layout for Study 2. Participants selected targets in order from the "Targets" list on the left; selections made by the participants were displayed in the "Selections" box on the right.

If a selection were incorrect, the participant would not be able to proceed with the next trial until they cleared (i.e., Ctrl-Z) the incorrect item. This would require participants to monitor their selection to determine if any errors had occurred during their input, whether they were user errors or system interpretation errors. Although participants were informed when they were not allowed to proceed with the next trial, the feedback only occurred after the participant proceeded with the next trial without making corrections first. This feedback was with regards to being unable to proceed with the study, as opposed to giving the participant feedback on the correctness of their selections.

4.2.2 Experimental Conditions

The factors of this study were the same as in Study 1. Three levels were used for IER, between participants: Zero, Five, and Twenty percent error. Each participant performed the tasks throughout nine training blocks and three memory tests.

4.2.3 Participants

Using the same criteria from the previous study, 75 participants (54 men, 20 women, 1 did not disclosed) aged 19-70 (mean = 37.5, *SD* = 11.5) through MTurk. Workers who had already participated in Study 1 (whether they completed the study or not) were disqualified from participating in this study. One participant reported not having a standard arrow key layout (e.g., Up and Down keys were fitted between the Left and

Right keys). Self-reported estimates of weekly computer usage among participants averaged 41.6 hours per week (*SD* = 21.0). 29 participants reported they had previously used swipe typing, 23 had used gesture commands, 20 had used voice commands, and 11 had used handwriting recognition. No participants were removed from this study upon review of their survey responses and response times.

4.2.4 Procedure

The procedure for this study was the same as that for Study 1. Participants first completed the consent form and demographics questionnaire. They watched an instructional video that taught them how to complete the upcoming tasks and were provided a one-minute practice session. Afterwards, they completed 12 blocks of the selection tasks (9 training blocks; 3 memory tests) for a single IER condition. After all blocks were finished, the answered a final subjective questionnaire about their experience with the study.

4.2.5 Study Design

This second study addressed three different research questions (RQ):

- RQ2-1: Does increasing interpretation error affect user learning (i.e., accuracy and completion time of memory tests) when users need to monitor their selections?
- RQ2-2: Does increasing interpretation error affect performance in training (i.e., error rate and completion time of training blocks) when users need to monitor their selections?
- RQ2-3: Does increasing interpretation error affect user perception of effort or ease of learning (i.e., subjective questionnaire responses) when users need to monitor their selections?

Study 2 used the same two-factor design as Study 1: 3x9 for the analysis of training blocks, and 3x3 for the analysis of the memory testing conditions:

- Interpretation Error Rate (IER), between-subjects: Zero, Five, or Twenty percent error
- Block/Test, within-subjects, 1-9 for training blocks, and 1-3 for memory tests

The primary dependent variables considered were: accuracy and completion of memory tests, user error count and completion time in training blocks, and subjective ratings of effort from post-session questionnaires. Other measures were also gathered (e.g., the rate at which the guide was used during the training blocks, time when the guide was accessed, time of the participant's first keypress within a trial) to assess participant performance.

4.2.6 Results

Similar to the previous study, outlier data points were determined as any trial with a completion time or first keypress time greater than 3 *SDs* above the block's mean completion time. In total, 343 trials out of

10800 total trials were removed from the data analysis of Study 2. Trials during the training blocks that had artificial interpretation errors (i.e., misinterpretations or false negatives) were also discarded from the analysis. Again, no artificial errors were added to any of the trials during the memory tests.

All pairwise t-tests were corrected using the Holm-Bonferroni method.

Effects of IER on User Learning

Overall accuracy in the memory tests was 71.2% in the Zero condition, 69.5% in the Five condition, and 70.6% in the Twenty condition. No trials were discarded in this analysis beyond the standard outlier detection. The accuracies over the three memory tests, including overall accuracies and separate accuracies in category and item selection, are shown in Figure 4.10.

A two-way RM-ANOVA (IER×Test) showed a significant effect of Test on accuracy ($F_{2,216} = 17, p < .0001, \eta^2 = 0.13$), but no effect of IER ($F_{2,216} = 0.059, p = .94$) and no interaction between IER and Test ($F_{4,216} = 0.25, p = .91$). Post-hoc pairwise t-tests showed significant differences between all pairs of Test levels (all $p < .5$).

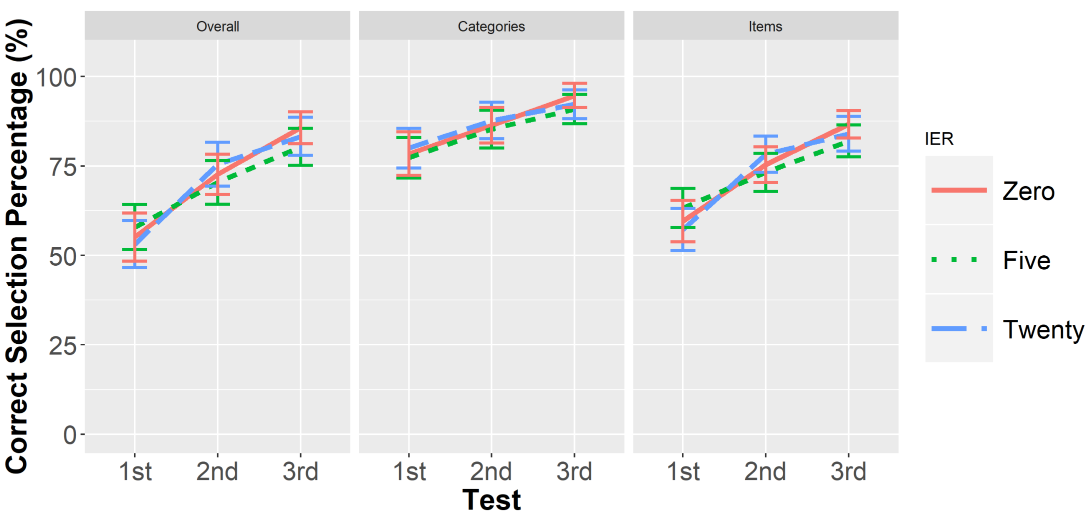

Figure 4.10: Study 2. Mean selection accuracy in the memory tests, by IER and Test.

The average completion time of the Zero condition was 2.12 s, of the Five condition was 2.50 s, and of the Twenty condition was 2.80 s. No trials were discarded in this analysis due to the standard outlier detection. The mean completion times of each condition across the three memory tests are shown in Figure 4.11.

RM-ANOVA (IER×Test) showed significant effects of both IER and Test on completion time (IER: $F_{2,216} = 9.1, p < .001, \eta^2 = 0.078$; Test: $F_{2,216} = 14, p < .0001, \eta^2 = 0.11$). There was no interaction between IER and Test ($F_{4,216} = 0.29, p = .88$). Post-hoc t-tests on completion time showed significant difference between all levels of IER (all $p < .001$) and all levels of Test (all $p < .0001$).

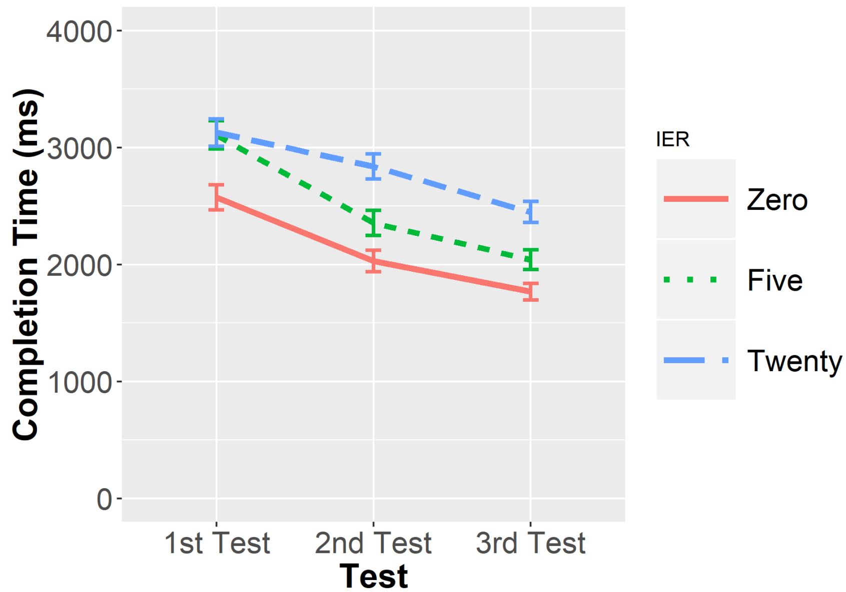

Figure 4.11: Study 2. Mean trial completion times in memory tests, by IER and Test.

Effects of IER on User Performance During Training

The average user error rate of the Zero condition was 0.19 errors per trial, of the Five condition was 0.25 errors per trial, and of the Twenty condition was 0.51 errors per trial. Any trials with an artificial error added to the participant's input were discarded from the analysis along with the standard outlier removal method. User error rate of each condition across the training blocks are shown in Figure 4.12.

RM-ANOVA (IER×Block) showed significant effects of both IER and Block on user errors (IER: $F_{2,648} = 9.1, p < .001, \eta^2 = 0.027$; Block: $F_{8,648} = 2.5, p = .0104, \eta^2 = 0.030$). There was no interaction between IER and Block ($F_{16,648} = 0.41, p = .98$). Post-hoc pairwise t-tests showed significant differences between all IER conditions (all $p < .05$). Post-hoc tests on neighbouring training blocks only showed differences between blocks 1 and 2 ($p < .001$).

The average completion time in the Zero condition was 2.57 s, in the Five condition was 2.68 s, and in the Twenty condition was 2.85 s. Trials with artificial system interpretation errors were discarded from analysis along with the standard outlier removal method. Completion times for each condition across the nine training blocks are shown in Figure 4.13.

RM-ANOVA (IER×Block) showed a significant effect of Block on completion time ($F_{8,645} = 4.1, p < .0001, \eta^2 = 0.048$), but no effect of IER ($F_{2,645} = 0.49, p = .62$) and no interaction ($F_{16,645} = 0.43, p = .97$). Post-hoc pairwise t-tests on neighbouring blocks only showed significant differences between blocks 1-2 (both $p < .0001$).

The mean total completion time of the training blocks in the Zero condition was 280 s, in the Five condition was 333 s, and in the Twenty condition was 497 s. No trials were discarded for this analysis. A one-way ANOVA of the overall completion time of the training blocks showed a significant effect of IER

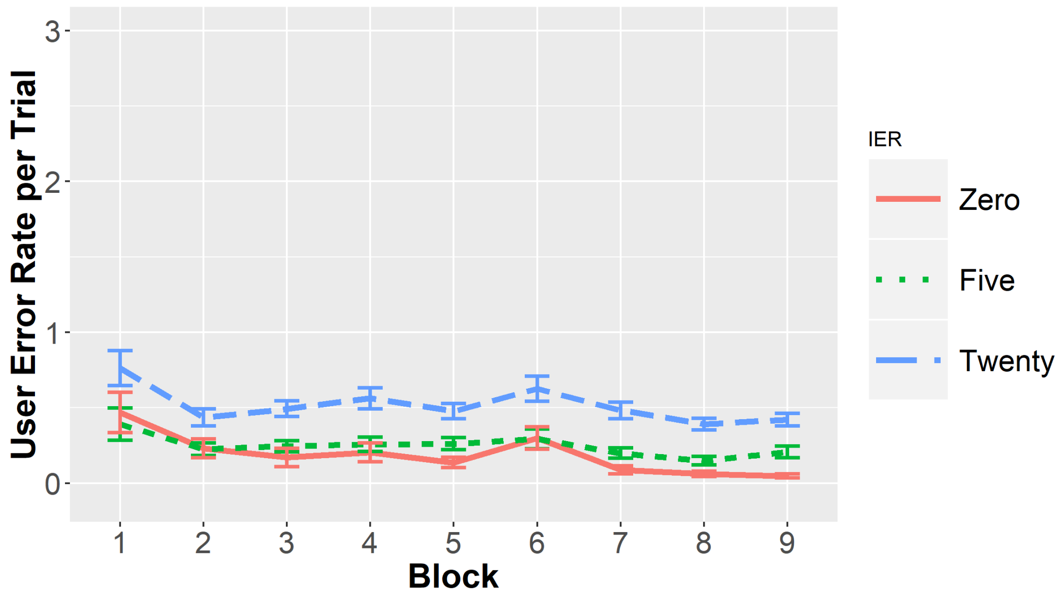

Figure 4.12: Study 2. Mean user errors per trial in training blocks, by IER and Block.

($F_{2,72} = 16, p < .0001, \eta^2 = 0.31$). Post-hoc pairwise t-tests found significant differences between the Twenty condition and both Zero and Five conditions (both $p < .001$).

Subjective Measures

The mean scores of the TLX responses from Study 2 are shown in Figure 4.14. Using Aligned Rank Transform on the aggregated responses, one-way ANOVAs were performed on each of the TLX questions with the factor of IER. Significant effects were only found in the responses for frustration ($F_{2,72} = 3.8, p < .05$) and in none of the other questions. Post-hoc pairwise t-tests performed on the responses for frustration only found significant differences between the Zero and Twenty conditions ($p < .05$). Results of the ANOVAs are tabulated in 4.2.

Table 4.2: Study 2. Results of the ANOVAs conducted on each TLX effort question after an Aligned Rank Transform.

TLX Question	$F_{2,72}$	p
Mental	1.1	.34
Physical	1.4	.26
Temporal	1.6	.22
Performance	1.8	.17
Effort	1.2	.32
Frustration	3.7	$<$ **.05**

The mean scores for participants' perceived ease of learning and memorization in Study 2 are shown in Figure 4.15. Using Aligned Rank Transform on the aggregated responses, one-way ANOVAs found a

Figure 4.13: Study 2. Mean trial completion times in the training blocks, by IER and Block.

significant effect on the perceived ease of memorization ($F_{2,72} = 4.3, p < .05$), but no effect on the perceived ease of learning ($F_{2,72} = 0.20, p = .82$). Post-hoc t-tests on the responses of perceived ease of memorization only found significant differences between the Five and Twenty conditions ($p < .05$).

4.2.7 Study 2 Summary

The goal of Study 2 was to determine if the results of Study 1 would be replicated in a setting that more closely mirrored how users need to ensure that a system's output matches their input. In other words, if users needed to constantly check whether a computer system would interpret their input correctly, would their learning of a system and performance with it be affected when system interpretation errors were present? Using a similar procedure as Study 1, participant data was gathered through training blocks and memory tests, and subjective responses were gathered through questionnaires.

During the memory tests, the Zero IER condition was significantly faster than both the Five and Twenty conditions, but there was no difference in accuracy. These results contrast with the ones from Study 1, where the Zero condition was more accurate, but not faster than the other conditions. As in Study 1, accuracy increased throughout all three levels of Test for each of the IER conditions, while completion time decreased similarly.

During the training blocks, the Zero condition had lower user errors per trial while also having a lower total completion time. There were also no differences between IER conditions for the average completion times of each trial. These differences between the conditions reflect the ones from Study 1 similarly. Likewise, user errors were highest in the first training block for all IER conditions with a significant decrease by the second block. However, when compared with the average rate of errors across all conditions, there were

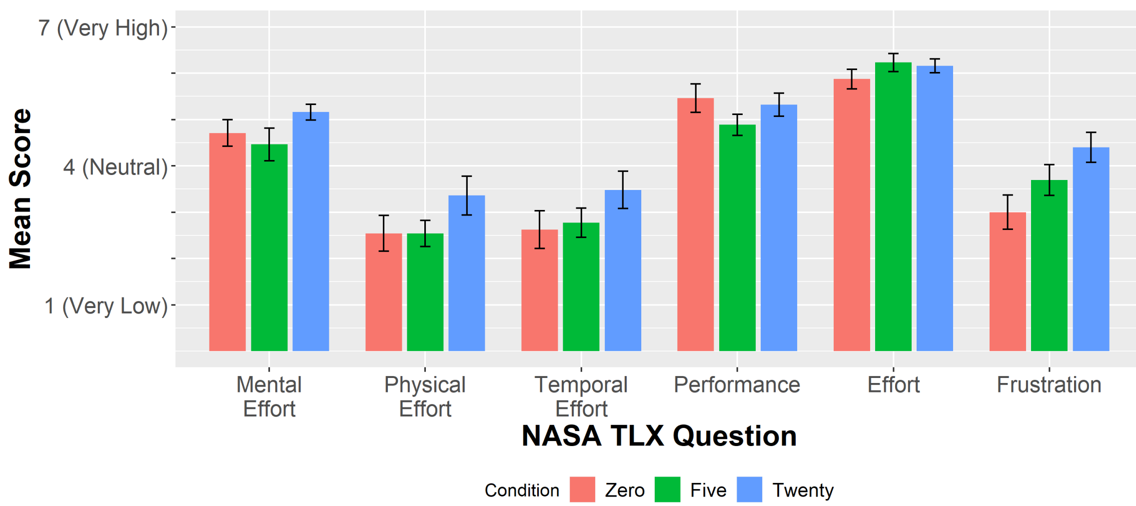

Figure 4.14: Study 2. Mean NASA Task Load Index scores, by IER.

generally less errors per trial occurring in Study 2 than in Study 1. Completion times followed a similar trend with the highest times in the first block but decreased significantly in the second block.

In this second study, subjective responses for perceived effort only showed an increase in frustration as the rate of artificial errors increased. Unlike the first study, participants in the Twenty condition felt that the items were easier to memorize than the participants in the Five condition (where there were no differences in the perceived ease of memorization in the first study).

This second study showed that when users had to monitor the system's interpretation of their own inputs (as opposed to receiving immediate feedback regarding their selection), there were some differences in how they performed the tasks. The main difference was that in a no-error condition, users that needed to monitor the system's interpretations were faster during the memory tests, while users who were provided immediate feedback were more accurate during the same tests. Users who monitored the interpretations also performed less user errors overall during the training blocks. These differences show that the act of monitoring for system interpretation errors could affect how a user performs a menu selection task.

4.3 Chapter 4 Summary

The two studies presented in this chapter provide evidence towards two competing ideas towards how interpretation errors by the computer system could affect user learning. In one case, it is possible that interpretation errors could negatively affect a user's development and learning of how an input action translates to a computer command. Alternatively, interpretation errors may cause users to work harder to better learn the input-to-command associations, resulting in faster learning and improved performance.

The results of the investigation show that increasing rates of interpretation error has detrimental effects on the accuracy of item selection and completion times of tasks during memory tests, as well as increased

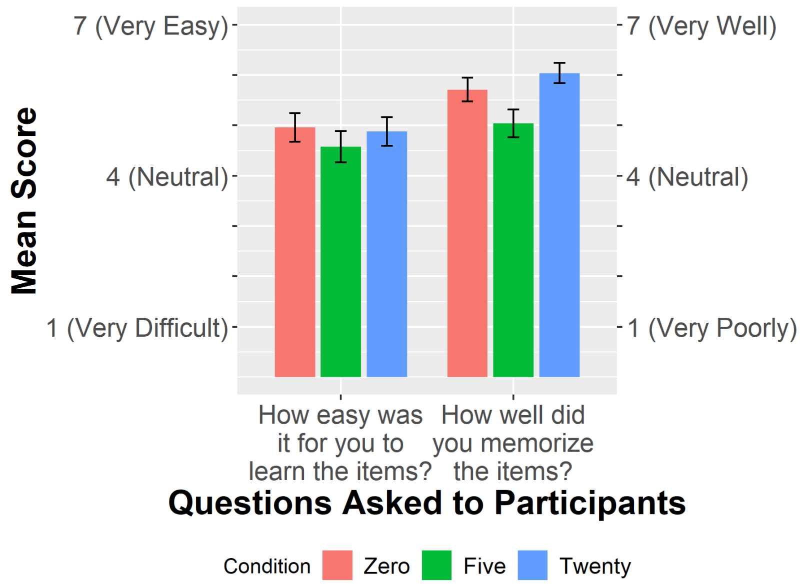

Figure 4.15: Study 2. Mean scores for perceived ease of learning and memorization, by IER.

user errors and perceived effort during training. When participants are forced to monitor their selections, they tend to be faster at completing memory tests, as opposed to participants who were provided automatic feedback by the system, who were more accurate. In both cases, participants found that completing tasks with the presence of interpretation errors required more perceived effort; in particular, higher rates of error made the tasks more frustrating.

The following chapter provides a more detailed explanation of the results, as well as how they generalize to real-world environments. In more general terms, the implications of these results will also be discussed regarding how noise (which can lead to system interpretation errors) can affect a user's learning ability of an alternative input mechanism, such as a BCI.

5 Discussion

The problem of noise in input interpretation of novel user input techniques, specifically with BCIs, were presented in Chapter 3. In arguing that noise in input will lead to interpretation errors by the computer system, two studies in Chapter 4 provided evidence that noise can have an impact on how users are able to learn and effectively use non-traditional input techniques. In this chapter, the results of the two interpretation error studies will be explored in further detail and explanations for the results will be considered. The implications of noise on learning novel input mechanisms, as well as limitations to the studies' approach and potential future work, will be discussed. Finally, the consequences of the interpretation error findings and how they can affect learning to use novel input mechanisms, such as BCIs, are examined.

5.1 Summary of Findings

The investigation into the effects of system interpretation errors on learning input mechanisms found several significant results between the two studies performed in Chapter 4:

- Learning occurred in all conditions for both studies: Accuracy in the memory tests showed improvement as the tests progressed, user errors in the training blocks decreased, and completion times showed significant decrease in both testing and training.
- The Zero-error condition had less user error in both studies: When compared with the conditions that had artificial interpretation errors, the zero-error condition had significantly less user error occurring on average. This occurred despite participants in the Zero-error condition spending less time in training overall (i.e., lower total time to complete the training blocks), but having similar average completion times for each trial.
- Study 1: The Zero-error condition was more accurate on memory tests: When provided with immediate feedback regarding the correctness of their item selection, participants were significantly more accurate during the memory tests than conditions that had artificial interpretation errors. However, there was no difference in the completion time of memory tests between the different conditions.
- Study 2: The Zero-error condition was faster during memory tests: When participants were required to monitor the system's interpretation of their input, they were significantly faster at completing the memory tests. Conversely, there was no difference in the accuracy of the memory tests between conditions.

- Interpretation errors affected the perception of effort for both studies: Participants from Study 1 felt that tasks involving artificial interpretation errors required more physical and temporal effort, and were more frustrating. Participants from Study 2 similarly found the higher IER conditions were more frustrating.

5.2 Main Results

5.2.1 Why did the increased IER lead to reduced learning and performance?

Throughout both studies, the presence of system interpretation errors caused reductions in accuracy, longer completion times, more frequent user errors, and higher perceptions of effort. To reflect on these results, the two hypotheses presented at the start of Chapter 4 are used as a basis. The interference hypothesis states that system interpretation errors will have a negative impact on user learning of new input mechanisms, while the retrieval effort hypothesis conversely states that the increased effort and attention required due to interpretation errors will cause users to learn new systems faster and have better retention. It is possible that these two effects are occurring simultaneously at various levels and may even be competing. However, the results primarily show that the negative interference from the interpretation errors outweigh any benefit that would be gained from the increased effort placed into completing the tasks.

In both studies, the Five and Twenty IER conditions had significantly greater user errors on average (i.e., errors not caused artificially by the study system) when completing training block trials compared to the Zero condition. This could be due to the occurrence of cue interference [11, 108]. The reinforcement of memory pathways that associate a system command and its input action is disrupted by the presence of interpretation errors, leading the participant to retrieve the wrong input for a certain target command due to confusion with similar pathways [108]. It is also possible that higher rates of IER hinder the memorization of the correct associations between command and input action. While participants could memorize an input action for a command, they may not be able to recall if that action was correct. The addition of interpretation errors meant that participants may have had greater trouble determining whether their input action-to-command association was accurate, especially if the system's feedback informed them otherwise. Since the study system never changed an incorrect item selection to a correct selection due to an interpretation error, participants only experienced perceived correct actions resulting in incorrect selections. Another explanation could be that participants in higher IER conditions may incorrectly attribute a system interpretation error as a user error caused by themselves. In other words, if an interpretation error occurred as a participant entered the correct input action, causing them to select the wrong item, it would be more difficult for them to accept that their previous action was supposed to be correct. This might lead them to second-guess their inputs in the future and cause doubts about items that they may have already memorized. Participants may also become more frustrated at their own perceived inability to perform the correct input actions, which can be observed in the increased frustration perceived by participants in the higher error conditions (e.g., Figures

4.7 and 4.14).

Finally, the higher IER conditions were perceived as more frustrating to complete in both studies. The increased effort as a result may have led to participants feeling discouraged by the faulty system interpretation, causing them to be less willing to try to memorize the input actions. This idea is also supported by the increased time it took for participants to complete the training blocks (i.e., higher overall completion times) in the Five and Twenty conditions, which could have led to greater fatigue.

5.2.2 Why was the Zero condition more accurate in Study 1, but faster in Study 2?

While both studies saw a decrease in learning and performance as the rate of interpretation error increased, there were differences between the two studies in how the performance suffered, particularly during the memory tests. Study 1 showed lower accuracy during the tests as IER increased without any difference in completion times, while Study 2 tests were completed slower at higher IER with similar accuracy between conditions. This can possibly be explained by how the studies' tasks were presented to the participants. Generally, command selection and memorization tasks, like the ones performed in the Chapter 4 studies, have well-known trade-offs between completion speed and accuracy. That is, a faster completion time for this type of task would require the user to sacrifice how precise they are in inputting correct commands, and vice-versa. Between the two studies, the target stimuli were presented differently: single, discrete targets in Study 1 (Figure 4.2), and an ordered list of all targets in Study 2 (Figure 4.9). As such, it can be assumed that different stimuli presentation between the two studies caused participants in the Zero condition of Study 1 to value accuracy more than their counterparts in Study 2, who were more interested in speed. This assumption is supported by the total training times for the Zero conditions: participants averaged 280 s to complete all training blocks in Study 2, but participants in Study 1 spent more time with an average of 308 s. This contrasts with participants in the Twenty condition, who spent more time in Study 2 (497 s) than Study 1 (468 s).

Another possible explanation could arise from how the two study systems provided feedback to the participants regarding the correctness of their selections. In Study 1, participants received immediate feedback after their selection was registered by the system. If they made an incorrect selection, they were prompted again with the same target until they successfully selected the correct item. In Study 2, selections were added to a list on the opposite edge of the screen of the target list. If an item was added to the selections list that did not match with the targets list, the participants were prevented from continuing onto the next trial until they fixed the incorrect item. These differences in the two studies' systems meant that participants in Study 1 were given immediate feedback to quickly correct any errors (user or system interpreted), while participants in Study 2 expended more effort to make careful, accurate selections. As a result, Study 1 participants may have felt that errors were less important since any errors they made were automatically resolved by the system (i.e., they did not need to perform a separate action to correct the errors as Study 2 participants needed to do).

This meant that in the higher IER conditions, Study 2 participants were always aware of the cost of making an error, resulting in similar accuracies when compared with the error-free condition. This has been seen in other work, which shows that increased effort can lead to better memorization [30, 22]. Additionally, the immediate feedback and error correction may be a digital equivalent to the concept of mechanical guidance, which is a form of supervision during the learning process that sets physical restrictions on the user (i.e., the system or instructor forcing a particular response to a stimulus) [96]. Although mechanical guidance is helpful for learning rote actions, it can also reduce a learner's autonomy and their motivation to learn [52]. That may explain why participants in Study 1 were less motivated to learn to use the interface when immediate feedback, regarding whether they needed to repeat the task, was available.

Conversely, the increased completion time as IER increased in Study 2 may have been caused by requiring participants to monitor the correctness of their selections more carefully. While the immediate feedback from Study 1 allowed participants to quickly correct both user and interpretation errors, the system in Study 2 provided no clear feedback and participants needed to recognize and correct errors themselves. Thus, the combination of higher IER and the need to identify errors themselves may have led participants to lose confidence and second-guess their inputs, resulting in the higher observed completion times as the rate of interpretation error increased.

In both studies, the negative effects of the interpretation errors were present despite participants in the higher-error conditions having spent significantly more time in training overall. The results showed that they affected both memory (more than 10% lower accuracy on average) and completion time of the tasks (half a second longer per trial). While the interpretation errors may have led to participants putting more effort towards completing the tasks, the errors were ultimately detrimental to the user.

5.3 Generalizing the Findings to Other Contexts

The two studies examining system interpretation errors and the effects on learning were performed using a system that was designed to test their respective research questions. This meant that to reduce the possibility of external factors affecting the results, the system made participants complete learning tasks using an artificial input mechanism and commands not found in real-world computer systems. Arrow keys were chosen for the input mechanism since it is a mechanism with no inherent interpretation error. This allowed for the rate of interpretation errors to be precisely controlled for the purposes of the study. However, there are generalizations that can be made based on the studies' design to apply the results to real-world contexts.

The user's learning of the input mechanism and tasks followed Fitts and Posner's stages of expertise development [33] in a similar fashion to many real-world input mechanisms. This occurred despite the artificial input mechanism (i.e., using arrow keys to perform menu selections) being specifically designed for the study. The study system allows for novice users (i.e., in the cognitive stages) to use visual guidance to navigate the hierarchal menus while they were still inexperienced. As they became more familiar with the input mechanism, users would eventually develop procedural memory routines (i.e., in the associative and autonomous stages). This would allow these expert users to retrieve the whole, single unit of an input sequence (e.g., the sequence needed to select the "Apple" item) from memory without needing to think about which individual key to press. This type of learning process has been observed for other forms of inputs, such as Appert and Zhai's gestural stroke-based commands [6]. Moreover, the effective use of many input techniques often relies on the user's procedural memory for fast command execution (e.g., in-air and above-the-surface gestures [76], marking menus [67], finger tap sequences [10], or pressure gestures [89, 141]).

Throughout each of the studies, participants were trained to use the input mechanism to complete the tasks with a training process that reflects some methods used in real life. Flashcard-style systems, that utilize a similar target-and-response system used to provide immediate feedback in Study 1, have been used to teach new input mechanisms to users (e.g., using the Quizlet smartphone app to memorize command shortcuts [111]). Although this memorization-based training approach can help with learning, its effectiveness can be improved upon, such as by applying the technique of spaced practice [69]. Likewise, the learning process that occurred in Study 2 was similar to how users would naturally use and learn a novel input mechanism (i.e., the user is required to monitor the system's interpretation of their input and manually correct any errors). These considerations are especially important when developing training programs for input mechanisms that are likely to have system interpretation errors.

The scenarios experienced by the participants were artificially created to suit the purposes of the study and have some differences with how users would typically interact with a real-world input mechanism. The participants were MTurk workers being paid a fixed amount to complete a series of tasks using the instructions provided to them. As such, they had little intrinsic motivation or interest in learning to use the artificial input mechanism to complete the studies' tasks. While it may have been in their interest to complete the tasks as quickly as possible (to sooner move onto other MTurk tasks), the participants may not have put in much effort into memorizing the items and their respective input commands that were only useful for these studies. However, conducting studies on MTurk can provide advantages over a typical lab study, such as better ecological validity. MTurk allows for vaster and quicker recruitment of participants from a more diverse demographic range than normally seen in HCI studies. Between the two studies conducted in Chapter 4, a total of 150 participants were recruited with an age range of 19 – 72. As such, the results can be interpreted as being more realistically reflective of the population [47].

During the completion of the studies, participants engaged in a form of learning that shares similarities with memory-based learning, but also carries aspects of cognitive-based learning. With this "memory plus cognitive" form of learning, users are able to apply their previous experience with errors (from both the user themselves and incorrect system interpretations) to make corrections when an error occurs, as well as adjust their own behaviour to avoid similar errors in the future. This type of learning is often seen in real-world situations where users are continuously improving their own behaviour by relying on previously gained knowledge about how a system works, whether it is a computer system or otherwise. When learning how to use new input mechanisms, the system does not usually provide feedback when an error occurs that the user can easily use to learn from or make future improvements. In general, input mechanisms and their commands are predominantly learned using a "memory-only" (or "memory-mostly") approach where the user themself needs to recall how they successfully performed a specific command. For example, when using speech recognition, an incorrect interpretation of the user's speech command does not often provide helpful feedback as to why the interpretation failed (e.g., the user's accent, spoke too quietly, etc.). Similarly, the occurrence of errors as a result of incorrect system interpretation could be seemingly random and erratic due to the presence of environmental factors. For instance, visual recognition software (e.g., eye-tracking) can sometimes fail due to lighting conditions that the user is unaware of, such as glare from bright light sources. Since the user is not made aware of these situations, it can be difficult for them to fix the current mistake, as well as change their actions to prevent similar errors from happening again.

Overall, the studies provide a strong foundation for future work, including additional experiments to replicate the above results with real input mechanisms and systems with intrinsic interpretation error. However, the current findings can already provide designers with new, practical information for creating novel input mechanisms and improving existing ones. Particularly, designers using novel input mechanisms should be aware that the presence of interpretation errors may have detrimental effects on a user's ability to effectively learn those new mechanisms. While there are potential benefits of using a novel system, the cost to the user must be noted, such as the time and effort it would take to learn to use something new when compared with using a more well-known and established mechanism (e.g., using the mouse for item selection). These studies regarding the effects on learning also add to the previous literature that have identified other problems when system interpretation errors occur; for example, autocorrect software making wrong (or unnecessary) corrections cause users to distrust the software in making accurate corrections in the future [84].

5.4 Implications of Noise on Learning Novel Input Mechanisms

As discussed in Chapter 3, the presence of noise when using a novel input mechanism can often lead to the system making interpretation errors. These errors can be a variety of different types: misinterpretations, false negatives, or false positives. While the former two were used in the Chapter 4 studies as potential results of artificial interpretation errors, false positives (i.e., errors where the system performs a command when the user did not intend to provide an input) were not tested since their occurrence would be difficult to control in a testing environment. However, this phenomenon can occur in real-world input mechanisms. For example, a smartphone that can be accessed through voice activation (e.g., Google Assistant) may pick up environmental sounds or another person's conversation and incorrectly try to interpret it as a command. As the complexity of the process increases, the number of interpretations that the system can generate also rises, which can often lead to the system providing the wrong result. Using a BCI, which can detect small amounts of electrical activity over a person's scalp, is often inaccurate as the range of possible inputs and outputs is large, and the system is unable determine the user's intent. This is because in input mechanisms susceptible to noise, the additional useless information of the noise disrupts the system's ability to accurately interpret the input to provide the user's intended output command.

By extrapolating the effects on input mechanisms caused by interpretation errors as a consequence of noise, the results found in Chapter 4 can be applied to noisy inputs devices like the BCI. That is, as the noisiness of an input stream increases, it can be predicted that a user's performance will decrease. The noise will also lower users' ability to effectively learn a new input mechanism, requiring them to spend more time and effort to gain expertise with that mechanism when compared with one that has little or no inherent noise (e.g., keyboard and mouse). Since noise can cause unpredictable interpretation errors, it is likely that a user's learning process for a noisy input will follow a similar process as those of participants of the second Chapter 4 study. Users will need to monitor the system's output more closely to ensure that a correct interpretation of their input has been made. For BCI devices that suffer from a high degree of noise, it becomes much more important and necessary that effective strategies are developed and used to reduce the amount of noise. In some cases, this could even mean possibly reducing the number of recognized commands, such that the remaining commands are more frequently interpreted accurately. With a limited number of commands, users may be able to learn to use the device faster, as well as handle errors more efficiently.

Furthermore, the results showed that higher rates of interpretation errors lead to users making more errors themselves. Needing to repeatedly fix these errors can cause the user to lose confidence in their ability to differentiate between an error caused by the user and an error caused by noise, which may have consequences when the user is attempting to associate input actions with their respective commands (i.e., cue interference). This process is made much more difficult for BCI users, as the input (i.e., brain activity) is not something that they can typically control at will. There have been examples from the literature that suggest training can help improve command selection accuracy with BCIs [53, 93], but it can still be difficult to train users to avoid or minimize errors, whether they come from the system or the user. Without proper feedback from the system as to the cause of the error (e.g., specific details about what caused a command interpretation to fail), users may not be able to effectively identify when an error might occur, which could prevent them from changing their behaviour to experience errors less often. As such, a noisier input mechanism can be much more difficult to learn and use to the same level of other alternatives.

5.5 Limitations

The studies conducted in Chapter 4 were performed in a controlled environment where the system interpretation errors could be set at a desired rate, which led to some limitations in the approach and design of the experiments. First, the arrow key input mechanism and menu selection tasks were specifically designed for the purposes of the studies. Using the arrow keys as the input mechanism (i.e., one that inherently exhibits no interpretation errors) allowed for a system to be created that would introduce artificial interpretation errors into the users' input at controlled rates. The menu selection tasks allowed for participant performance and their efforts to be observed, as well as comparisons to be made due to the different levels of IER. However, participants were simply asked to complete the tasks with no context as to why they were performing them. Although the tasks were constructed to simulate learning and using a noisy novel input mechanism, participants were not explicitly provided this information. The lack of context for completing the tasks may have deprived motivation from the participants, leading them to put less effort into completing the study. Similarly, the commands that participants were asked to select were unrealistic. The command set used in the studies were selected to prevent knowledge transfer effects, resulting in categories such as "Cities" and "Fruits". Nevertheless, these categories do not reflect what users of a real-world input mechanism might encounter. Furthermore, successful selections that users made only led to further selection tasks. With a real-world mechanism, a command selection would typically result in a command being executed or function being performed by the system. Without such tangible feedback, it may have been more difficult for the participants of the studies to learn and memorize the items, and they may have been less motivated to complete the tasks.

Second, the rates of interpretation error that were chosen as levels in the studies (i.e., 0%, 5%, and 20% error) were based on previously published work regarding novel input mechanisms and their respective intrinsic error rates. Depending on the mechanism, interpretation errors can occur at rates from as little as 2% up to 40% at the worst [140]. While these rates are based on novel input mechanisms that might not reflect established mechanisms that are used widely by the general population, they provide a realistic example of how often a system might misinterpret a user's input. However, these published error rates are often measured in controlled lab settings that typically do not change from user to user. In real-world use, input mechanisms are frequently used in differing environments that may affect the system's ability to correctly interpret a user's input. For example, a smartphone's speech recognition software may suffer more errors in a noisy area, or an eye tracking device may fail to detect the user's eyes in a dimly lit room. In these cases, the error rates and the effects of the different environmental conditions on successful input interpretation are not well known.

Third, the pair of studies made observations of the participants after a relatively short training period with the input mechanism and the selection tasks. Although the participants showed improved recall of the items towards the end of the testing, only the short-term retention and immediate effects of the training could be observed. Between the two studies, the longest average time to complete the training blocks was 497 s for the Twenty percent error condition, where the higher rate of artificial interpretation errors would expectedly require users to spend more time to fix those errors. Real-world learning of input mechanisms often involves repeated practice over a much longer period (e.g., on the order of weeks to months). Practice with a real-world input mechanism also typically involves using it to perform real tasks (e.g., using a speech command to set an alarm), as opposed to simply completing a task for the sake of practice. The studies did not examine how longer or continued training would benefit memory or how well the input commands would be retained over a longer period (e.g., hours or days after training). Memory tests were only conducted during the duration of the studies, and no follow-up was performed. This meant that the long-term effects of the learning were not observed for these two studies. The effects of continued practice over longer periods of time and long-term retention were not tested. Likewise, the transfer effects of the learning were not investigated. The input mechanism and system were designed such that participants would not have an advantage in completing the tasks if they had previous experience with a similar mechanism. On the other hand, this also meant that participants could not rely on their previous knowledge to improve their learning process with this novel input mechanism. The participants would have a limited ability to transfer the things they learned during the experiment to other mechanisms as well. These conditions used in the studies were necessary to develop an initial baseline to understand how learning in the early training stages is affected by interpretation errors.

Fourth, of the three stages of learning described by Fitts and Posner [33], the participants of the two studies were only trained until the second, associative stage (i.e., participants were beginning to make the connections between an input sequence and the item it selects). Since the user studies only examined the participants ability to retain the item associations in the short term, they were not able to reach the final, autonomous stage of the learning process, which would require a much greater amount of time and practice with the input mechanism and system. Reaching this last stage would mean the users would have been able to simply think of the item they wanted to select, then be able to perform its respective command without needing to recall the input sequence. There is a possibility that once users reach automaticity, or are familiar with a similar input mechanism, the presence of interpretation errors would cause greater problems towards effectively using the new mechanism. This may be due to the added need to monitor the results of the system's interpretation; thus, preventing a user from simply inputting a command without remembering the individual key presses.

Fifth, the difficulty of tasks throughout the studies were increased by adding artificial errors to the participant's input. Since participants needed to spend more effort and time paying attention to the system's interpretation, their ability to quickly and accurately complete tasks would decrease. This meant that the participants in the Twenty percent error conditions would spend more time to complete the experiment overall, while participants of the Zero percent error conditions could simply finish one task and move onto the next without needing to check the interpretation or make subsequent corrections. This method of increasing difficulty fits the principles described by Schmidt and Bjork for improving training techniques [114]. However, the adjustment of difficulty is unlike the methods used by previous studies in HCI that investigate the same phenomenon. For example, increasing the number of items or making items more similar to each other are more typical approaches that researchers have tested different levels of difficulty. Although the studies in Chapter 4 have varying levels of task completion difficulty, they are not directly comparable with difficulty adjustments done for most other user studies.

Sixth, the two types of artificial system interpretation errors used in the studies (i.e., misinterpretations and false negatives) were added through random distribution. Both types of errors had an equally likely probability of occurring. However, real-world recognition systems, such as in-air gesture recognition, often have better recognition of certain inputs over others. They are typically biased to one type of error over the others, which could influence how users approach avoiding or correcting of these errors. This phenomenon could be further affected by how the user performs the input, such as how they pronounce a word for a speech recognizer. Similarly, as users learn to use an input mechanism and become experts through natural use, they may become more aware of interpretation errors occurring during specific commands and how they affect their inputs, which may change how careful the users are when performing those commands. The random distribution used in the studies provide a basic method to introducing artificial errors and allow for an initial understanding of how interpretation errors affect the learning process. However, the artificial errors of the studies did not include false positive errors (i.e., a selection is made by the system without user input). Only misinterpretations and false negatives were included, as they were the more common type of interpretation errors, but false positives are also possible in some real-world input mechanisms.

Finally, the data used for the studies were gathered using Amazon's Mechanical Turk. Although previous work have successfully used MTurk to carry out HCI studies [47, 64], the issues with conducting studies remotely need to be considered. Since the participants completed their tasks online, it was difficult to directly observe their interaction with the study system. For example, given that both studies had multiple memory tests that disabled the guide during the duration of the tests, it was possible that some participants cheated the tests by taking a screenshot or picture of the guide when it was available during the training blocks. Participants were also in uncontrolled environments, which could have caused some participants to become distracted by outside factors, such as other people in the same room. Furthermore, MTurk users are not an accurate representation of the general population. Goodman et. al [39] discusses some weaknesses of MTurk participants, including having different view of money and less likely to pay attention to experimental materials. It was possible that some participants attempted to complete the tasks and questionnaires as quickly as possible with little care for providing genuine responses to their experience, since MTurk pays upon completion of the each study rather than time spent in the study.

5.6 Future Work

The limitations described in the previous section show possible improvements that can be made for future studies in the same area of research. They also outline potential points of interest for further study as a result of the studies' findings.

Conducting the same studies from Chapter 4 using a more established input mechanism will allow the results to more realistically reflect how user learning of a novel input mechanism is affected by interpretation errors. One of the limitations of the studies was that they were performed using an input mechanism and set of selection tasks that were created specifically for use with the study. By using an input mechanism with established real-world applications, such as stroke gestures used in marking menus [67], further understandings can be gained by observing how interpretation errors in an input mechanism with high, intrinsic command selection accuracy, using a similar method to add artificial interpretation errors into the user's input stream. Similarly, testing the input mechanism in various environmental conditions can provide more realistic probabilities of an interpretation error occurring. For example, performing similar studies where users are learning to use a new input mechanism, such as camera-based gesture trackers or speech recognizers, in an adverse environment for the system (e.g., low lighting for cameras, white noise in the surroundings of a microphone, etc.) may bring about poorer user performance and learning. This further work can provide knowledge regarding the magnitude of the adverse effects that system interpretation errors cause in real-world input mechanisms (i.e., how much worse are the interpretations when environmental noise is also a factor).

Examining the long-term effects of learning with the continual presence of interpretation errors in the user's input would allow designers of input mechanisms to have a better idea of how to improve their training procedures. Since the Chapter 4 studies only looked at the short-term retention and performance, performing longitudinal studies that observe the long-term effects can reveal new findings about how interpretation errors affect the development of expertise. The participants of the studies only managed to reach the associative stage of learning the input sequences and their respective commands. However, further studies that examine longer training over days and weeks may find users reaching the autonomous stage where they will be able to perform a command without needing to recall the individual actions (i.e., using muscle memory). The length of time and practice required to reach automaticity may increase if the negative effects shown in short-term continue throughout the whole training process. With interpretation errors occasionally disrupting this learning process, it may even be possible that users will never reach the autonomous stage. The problems caused by interpretation errors are increased even further if feedback from the system is not provided and the user needs to monitor the system's interpretation for potential errors. Users may find that the effort required to manage a novel input mechanism with an unreliable output is too high for the benefits it may provide.

By adjusting the difficulty of the tasks imposed on the user, a relationship may be found between system interpretation errors and the difficulty. The difficulty of the tasks in the studies were adjusted through varying rates of interpretation errors. While this may force the participants to spend more time to complete the tasks (by requiring them to correct the errors or retry failed tasks), this form of difficulty adjustment is not typical of those done by other similar HCI studies. Making the tasks harder, such as increasing the number of menu levels to traverse or increasing the total number of items, could cause the effects of interpretation errors on learning to become more profound. Studying this possibility in a future study may provide better knowledge of how using input mechanisms for complex tasks may affect the user's ability to execute their tasks effectively.

In a future work, it may also be interesting to examine how users can adapt to system interpretation errors if they frequently experience the same errors. Although the artificial interpretation errors in the studies were randomly distributed between misinterpretations and false negatives, the errors of a real-world input mechanism would typically be non-uniformly distributed. In other words, some errors would occur more often than others, depending on the environmental conditions and how the user interacts with the mechanism. As users become more familiar with an input mechanism and the system's interpretation of their inputs, they may find subtle techniques to avoid such errors by interacting with the system more carefully. For example, users may feel less inclined to use a particular speech command if they find that the recognition system is having trouble understanding their pronunciation of a certain word. While this could eventually reduce the overall number of wrong interpretations, it is another facet of the system that the user will have to learn and practice. Users who become experts with using such an input mechanism will also be able to navigate the system with a much lower rate of errors than a user operating the mechanism for the first time. Furthermore, developing a system that can provide feedback to users that help them adjust to minimize the number of potential errors that could occur. Systems, such as autocorrect for smartphone keyboards, already exist that aid users by suggesting similar commands or outputs that the user may be attempting to execute. By observing how users adapt to unreliable systems, knowledge regarding how the user experience and user's learning processes for noisy inputs can also be established.

Determining an acceptable level of noise in an input is important such that users can learn how to use the input mechanism without suffering the consequences of interpretation errors. While there exists much previous work on the acceptable signal-to-noise ratio (SNR) for a variety of input streams, there are few that consider the effects of the SNR on learning. By extending the results from Chapter 4, the users will have diminished learning capabilities for novel input mechanisms that suffer from a high degree of noise in the input. However, since it is unlikely that most input mechanisms will employ methods to completely remove all noise from an input, the system must be subject to a small remainder of noise. In other words, these systems will experience occasional interpretation errors as a result. A future study that examines how frequent interpretation errors need to be before users start to experience learning and performance deficits, as well as when users start to notice errors not borne from themselves, can provide insight towards how much noise suppression is needed for the input stream. Significant effects were already occurring in the Five percent error condition in the Chapter 4 studies (i.e., one interpretation error in every 20 key inputs), which mean that the learning process is particularly sensitive to disruptions. Furthermore, the amount of noise in an input varies with the input mechanism (e.g., BCIs have noisier sources of data when compared with speech recognizers) but can also be proportional to the amount of incoming input data. Input mechanisms like BCIs require much more data than other input mechanisms, which can be accidentally discarded as noise if the noise reduction techniques are not carefully used. Since removing noise also potentially removes useful information, the purity of the incoming data can be preserved by only removing as much noise as necessary. Finding the upper limit of interpretation errors before these effects start happening might provide designers with a guideline towards how much noise reduction should be applied to a user's input.

While MTurk studies can provide benefits in recruiting a large number of participants and carrying out multiple studies in a short amount of time, the limitations of online, remote studies should be addressed in future work. Many of the problems of MTurk studies, as described in the previous section, can be alleviated by conducting the studies in a traditional, controlled lab setting. That is, being able to observe participants directly while they perform the studies' tasks can prevent participants from cheating during memory tests and minimize distractions from drawing attention away from the tasks. Alternatively, different methods of motivation could help improve attentiveness towards the tasks (e.g., using a payment model that is based on the quality of work). Another benefit of lab studies is that participants are often recruited from an environment (e.g., a university) that has people who are more likely to be interested or invested in the findings of the studies that they are participating in, leading to them putting more care into their work.

6 Conclusion

Users interact with computer systems by using input mechanisms as an interface for command selection. Input mechanisms allows a user to translate their intent into a command that the computer executes. There are a wide range of novel input mechanisms that allow different types of input, such as speech or hand gestures. Brain-computer interfaces are one such input mechanism, allowing users to issue commands through their thoughts and brain activity. Although alternative mechanisms can offer certain advantages over the traditional inputs, an additional factor of interpreting the user's input is required of the computer system. Signals used by input mechanisms are often susceptible to noise that can arise from various sources, such as the environment or from artifacts of the input. Noise can interfere and make it difficult to discern the useful information in the signal. As a result, the system can incorrectly interpret a user's input due to the noise, causing an interpretation error to occur. This thesis explored the issue of noise in input signals of novel input mechanisms and how it affects the interpretation of a user's intended command.

An initial study was conducted with an EEG headset to understand its capabilities of this device as a motor-imagery BCI input mechanism. The results were analyzed offline after data collection was completed and found that there was too much noise in the EEG signals for the mu suppression associated with motor actions to be conclusively detected. The noise persisted despite efforts to minimize noise through filtering and remove artifacts through visual inspection of the raw EEG data. The findings of this study provided the motivation for the following work on investigating the effects of noise on learning with novel input mechanisms.

Two studies were performed using an input mechanism with simulated noise to explore how interpretation errors affect user learning. Both studies added artificial interpretation errors into the user's tasks but differed in how feedback was provided back to the user. The results of the two studies showed that interpretation errors have a significant negative impact on a user's ability to learn a novel input mechanism, as well as on their perception of the effort required to complete the command selection tasks. The findings also indicate that a user may prioritize speed and accuracy differently depending on whether the system provides feedback when an error occurs. Despite any potential gains from increased effort in completing the tasks, interpretation errors create adverse effects for users attempting to learn to use a novel input mechanism.

This thesis provided new understanding of how interpretation errors can affect users as they use novel input mechanisms, particularly when learning to perform a set of commands. Users of novel input mechanisms that use analog input signals may be susceptible to noise, leading to system interpretation errors. Although previous literature had suggested two competing effects on learning while handling interpretation errors, the negative effects of the errors were more substantial. By gaining a better understanding of the negative impact of noise and interpretation errors on learning, designers and developers of novel input mechanisms can create higher quality training programs to teach both new and experienced users how to use input mechanisms that are prone to interpretation errors. The findings can also be used as a foundation by usability researchers to understand how well a signal-based input mechanism is perceived by the user base. Furthermore, the knowledge of the effects of interpretation errors on learning can lead to more user-centred designs that account for the possibility of errors from both the user and the system. Finally, the findings of this work also provided directions for further investigations of novel input mechanisms, noise, and user learning. The findings of this thesis can help to improve the design and learning of new input mechanisms for human-computer interaction.

References

[1] Anas Albulbul. Evaluating major electrode types for idle biological signal measurements for modern medical technology. *Bioengineering*, 3(3):20, August 2016.

[2] Rafeed Alkawadri. Brain–computer interface (BCI) applications in mapping of epileptic brain networks based on intracranial-EEG: An update. *Frontiers in Neuroscience*, 13:191, March 2019.

[3] Brendan Z. Allison and Christa Neuper. Could anyone use a BCI? In *Brain-Computer Interfaces*, pages 35–54. Springer London, 2010.

[4] Michael C. Anderson and James H. Neely. Interference and inhibition in memory retrieval. In *Memory*, pages 237–313. Elsevier, 1996.

[5] Marios Antonakakis, Sophie Schrader, Ümit Aydin, Asad Khan, Joachim Gross, Michalis Zervakis, Stefan Rampp, and Carsten H. Wolters. Inter-subject variability of skull conductivity and thickness in calibrated realistic head models. *NeuroImage*, 223:117353, December 2020.

[6] Caroline Appert and Shumin Zhai. Using strokes as command shortcuts. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, pages 2289–2298. ACM Press, 2009.

[7] Levent M Arslan and John HL Hansen. Language accent classification in american english. *Speech Communication*, 18(4):353–368, 1996.

[8] Mahnaz Arvaneh, Cuntai Guan, Kai Keng Ang, and Chai Quek. Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Transactions on Biomedical Engineering*, 58(6):1865–1873, June 2011.

[9] Omar Bagdadi and András Várhelyi. Development of a method for detecting jerks in safety critical events. *Accident Analysis & Prevention*, 50:83–91, January 2013.

[10] Gilles Bailly, Eric Lecolinet, and Yves Guiard. Finger-count & radial-stroke shortcuts. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, pages 591–594. ACM Press, 2010.

[11] David R. Basden and Barbara H. Basden. Some tests of the strategy disruption interpretation of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6):1656–1669, 1995.

[12] André Beer-Furlan, Stephan A Munich, and Michael Chen. Augmenting superior sagittal sinus functionality. commentary: Motor neuroprosthesis implanted with neurointerventional surgery improves capacity for activities of daily living tasks in severe paralysis—first in human experience. *Journal of NeuroInterventional Surgery*, 13(2):100–101, December 2020.

[13] Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. Making sense of sensing systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*, pages 415–422. ACM Press, 2002.

[14] Antal Berényi, Zoltán Somogyvári, Anett J. Nagy, Lisa Roux, John D. Long, Shigeyoshi Fujisawa, Eran Stark, Anthony Leonardo, Timothy D. Harris, and György Buzsáki. Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *Journal of Neurophysiology*, 111(5):1132–1149, March 2014.

[15] Nigel Bevan. International standards for HCI and usability. *International Journal of Human-Computer Studies*, 55(4):533–552, October 2001.

[16] F. Bordoni and A. D'Amico. Noise in sensors. *Sensors and Actuators A: Physical*, 21(1-3):17–24, February 1990.

[17] Lieke Braadbaart, Justin H.G. Williams, and Gordon D. Waiter. Do mirror neuron areas mediate mu rhythm suppression during imitation and action observation? *International Journal of Psychophysiology*, 89(1):99–105, July 2013.

[18] John T Cacioppo, Louis G Tassinary, and Gary Berntson. *Handbook of psychophysiology*. Cambridge university press, 2007.

[19] Géry Casiez, Nicolas Roussel, and Daniel Vogel. 1 € filter. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, 2012.

[20] Steven J. Castellucci and I. Scott MacKenzie. Graffiti vs. unistrokes. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, 2008.

[21] Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao. High-speed spelling with a noninvasive brain–computer interface. *Proceedings of the National Academy of Sciences*, 112(44):E6058–E6067, October 2015.

[22] Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys*, 47(2):1–36, January 2015.

[23] Andy Cockburn, Per Ola Kristensson, Jason Alexander, and Shumin Zhai. Hard lessons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*. ACM Press, 2007.

[24] Jacob Cohen. Eta-squared and partial eta-squared in fixed factor anova designs. *Educational and Psychological Measurement*, 33(1):107–112, April 1973.

[25] Damien Coyle, Jhonatan Garcia, Abdul R Satti, and T Martin McGinnity. EEG-based continuous control of a game using a 3 channel motor imagery BCI: BCI game. In *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*. IEEE, April 2011.

[26] Fergus I.M. Craik and Robert S. Lockhart. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6):671–684, December 1972.

[27] E Curran. Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems. *Brain and Cognition*, 51(3):326–336, April 2003.

[28] Paul Doliotis, Alexandra Stefan, Christopher McMurrough, David Eckhard, and Vassilis Athitsos. Comparing gesture recognition accuracy using color and depth information. In *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '11*. ACM Press, 2011.

[29] Matthieu Duvinage, Thierry Castermans, Mathieu Petieau, Thomas Hoellinger, Guy Cheron, and Thierry Dutoit. Performance of the emotiv epoc headset for p300-based applications. *BioMedical Engineering OnLine*, 12(1):56, 2013.

[30] Brian D. Ehret. Learning where to look. In *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*. ACM Press, 2002.

[31] Anthony B. Fallon, Kim Groves, and Gerald Tehan. Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology*, 34(5-6):301–307, October 1999.

[32] Bryn Farnsworth. Eeg headset prices – an overview of 15+ eeg devices - imotions, Jul 2019.

[33] Paul M Fitts and Michael I Posner. *Human performance.* Brooks/Cole Publishing Company, 1967.

[34] B. J. Fogg, Gregory Cueller, and David Danielson. Motivating, influencing, and persuading users: An introduction to captology. In *Human-Computer Interaction Fundamentals*, pages 127–140. CRC Press, March 2009.

[35] Antonio Frisoli, Claudio Loconsole, Daniele Leonardis, Filippo Banno, Michele Barsotti, Carmelo Chisari, and Massimo Bergamasco. A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):1169–1179, November 2012.

[36] Ella Gabitov, Arnaud Boutin, Basile Pinsard, Nitzan Censor, Stuart M. Fogel, Geneviève Albouy, Bradley R. King, Julie Carrier, Leonardo G. Cohen, Avi Karni, and Julien Doyon. Susceptibility of consolidated procedural memory to interference is independent of its active task-based retrieval. *PLOS ONE*, 14(1):e0210876, January 2019.

[37] Xiaorong Gao, Dingfeng Xu, Ming Cheng, and Shangkai Gao. A BCI-based environmental controller for the motion-disabled. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):137–140, June 2003.

[38] Pramod Gaur, Ram Bilas Pachori, Hui Wang, and Girijesh Prasad. A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and riemannian geometry. *Expert Systems with Applications*, 95:201–211, April 2018.

[39] Joseph K Goodman, Cynthia E Cryder, and Amar Cheema. Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, 26(3):213–224, 2013.

[40] R.M. Gray. Quantization noise spectra. *IEEE Transactions on Information Theory*, 36(6):1220–1244, 1990.

[41] Phillip J. Grimaldi and Jeffrey D. Karpicke. When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, 40(4):505–513, January 2012.

[42] S. Gupta and H. Singh. Preprocessing EEG signals for direct human-system interface. In *Proceedings IEEE International Joint Symposia on Intelligence and Systems*. IEEE Comput. Soc. Press, 1996.

[43] Gido Hakvoort, Boris Reuderink, and Michel Obbink. *Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system.* Number TR-CTIT-11-03 in CTIT Technical Report Series. Centre for Telematics and Information Technology (CTIT), Netherlands, February 2011.

[44] Levi J. Hargrove, Guanglin Li, Kevin B. Englehart, and Bernard S. Hudgins. Principal components analysis preprocessing for improved classification accuracies in pattern-recognition-based myoelectric control. *IEEE Transactions on Biomedical Engineering*, 56(5):1407–1414, May 2009.

[45] Chris Harrison, Desney Tan, and Dan Morris. Skinput. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, 2010.

[46] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology*, pages 139–183. Elsevier, 1988.

[47] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, 2010.

[48] Lea Hehenberger, Reinmar J. Kobler, Catarina Lopes-Dias, Nitikorn Srisrisawang, Peter Tumfart, John B. Uroko, Paul R. Torke, and Gernot R. Müller-Putz. Long-term mutual training for the CYBATHLON BCI race with a tetraplegic pilot: A case study on inter-session transfer and intra-session adaptation. *Frontiers in Human Neuroscience*, 15:635777, February 2021.

[49] Ken Hinckley, Patrick Baudisch, Gonzalo Ramos, and Francois Guimbretiere. Design and analysis of delimiters for selection-action pen gesture phrases in scriboli. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*. ACM Press, 2005.

[50] Sara Ilstedt Hjelm and Carolina Browall. Brainball-using brain activity for cool competition. In *Proceedings of NordiCHI*, volume 7, 2000.

[51] Hannah M. Hobson and Dorothy V.M. Bishop. Mu suppression – a good measure of the human mirror neuron system? *Cortex*, 82:290–310, September 2016.

[52] John Honeybourne, Michael Hill, and Helen Moors. *Advanced PE and Sport*. Nelson Thornes, 2004.

[53] Han-Jeong Hwang, Kiwoon Kwon, and Chang-Hwang Im. Neurofeedback-based motor imagery training for brain–computer interface (BCI). *Journal of Neuroscience Methods*, 179(1):150–156, April 2009.

[54] Block Imaging. Mri machine price guide, 2019.

[55] Toshiki Iso and Kenichi Yamazaki. Gait analyzer based on a cell phone with a single three-axis accelerometer. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services - MobileHCI '06*. ACM Press, 2006.

[56] Robert J. K. Jacob. Human-computer interaction. *ACM Computing Surveys*, 28(1):177–179, March 1996.

[57] Colby Johanson, Carl Gutwin, Jason T. Bowey, and Regan L. Mandryk. Press pause when you play. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, October 2019.

[58] Markus Junghofer, Thomas Elbert, Don M. Tucker, and Brigitte Rockstroh. Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, 37(4):523–532, July 2000.

[59] Emily S. Kappenman and Steven J. Luck. The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology*, 47(5):888–904, March 2010.

[60] Laura Kauhanen, Pasi Jylänki, Janne Lehtonen, Pekka Rantanen, Hannu Alaranta, and Mikko Sams. EEG-based brain-computer interface for tetraplegics. *Computational Intelligence and Neuroscience*, 2007:1–11, 2007.

[61] Jasmin Kevric and Abdulhamit Subasi. Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomedical Signal Processing and Control*, 31:398–406, January 2017.

[62] M. Jawad Khan and Keum-Shik Hong. Passive BCI based on drowsiness detection: an fNIRS study. *Biomedical Optics Express*, 6(10):4063, September 2015.

[63] Muhammad Jawad Khan and Keum-Shik Hong. Hybrid EEG–fNIRS-based eight-command decoding for BCI: Application to quadcopter control. *Frontiers in Neurorobotics*, 11:6, February 2017.

[64] Aniket Kittur, Ed H. Chi, and Bongwon Suh. Crowdsourcing user studies with mechanical turk. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, 2008.

[65] Per-Ola Kristensson and Shumin Zhai. SHARK2. In *Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST '04*. ACM Press, 2004.

[66] D.J. Krusienski, E.W. Sellers, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw. Toward enhanced p300 speller performance. *Journal of Neuroscience Methods*, 167(1):15–21, January 2008.

[67] Gordon Kurtenbach, Abigail Sellen, and William Buxton. An empirical evaluation of some articulatory and cognitive aspects of marking menus. *Human-Computer Interaction*, 8(1):1–23, March 1993.

[68] Kevin C. Lam, Carl Gutwin, Madison Klarkowski, and Andy Cockburn. The effects of system interpretation errors on learning new input mechanisms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, volume 713, pages 1–13. ACM, May 2021.

[69] Thomas K Landauer. Optimum rehearsal patterns and name learning. *Practical aspects of memory*, pages 625–632, 1978.

[70] Patrick Ledwidge, Jeremy Foust, and Adam Ramsey. Recommendations for developing an eeg laboratory at a primarily undergraduate institution. *Journal of Undergraduate Neuroscience Education*, 17(1):A10, 2018.

[71] Robert Leeb, Doron Friedman, Gernot R. Müller-Putz, Reinhold Scherer, Mel Slater, and Gert Pfurtscheller. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: A case study with a tetraplegic. *Computational Intelligence and Neuroscience*, 2007:1–8, 2007.

[72] Junhua Li and Liqing Zhang. Active training paradigm for motor imagery BCI. *Experimental Brain Research*, 219(2):245–254, April 2012.

[73] QianQian Li, Ding Ding, and Mauro Conti. Brain-computer interface applications: Security and privacy challenges. In *2015 IEEE Conference on Communications and Network Security (CNS)*. IEEE, September 2015.

[74] Jinyi Long, Yuanqing Li, Hongtao Wang, Tianyou Yu, Jiahui Pan, and Feng Li. A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(5):720–729, September 2012.

[75] M. Lopez-Gordo, D. Sanchez-Morillo, and F. Valle. Dry EEG electrodes. *Sensors*, 14(7):12847–12870, July 2014.

[76] Wei Lu, Zheng Tong, and Jinghui Chu. Dynamic hand gesture recognition with leap motion controller. *IEEE Signal Processing Letters*, 23(9):1188–1192, September 2016.

[77] Steven J Luck. *An introduction to the event-related potential technique*. MIT press, 2014.

[78] B O Mainsah, L M Collins, K A Colwell, E W Sellers, D B Ryan, K Caves, and C S Throckmorton. Increasing BCI communication rates with dynamic stopping towards more practical use: an ALS study. *Journal of Neural Engineering*, 12(1):016013, January 2015.

[79] Joseph W. Matiko, Stephen Beeby, and John Tudor. Real time eye blink noise removal from EEG signals using morphological component analysis. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, July 2013.

[80] Dennis J. McFarland, Lynn M. McCane, Stephen V. David, and Jonathan R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalography and Clinical Neurophysiology*, 103(3):386–394, September 1997.

[81] Dennis J. McFarland, William A. Sarnacki, Theresa M. Vaughan, and Jonathan R. Wolpaw. Brain-computer interface (BCI) operation: signal and noise during early training sessions. *Clinical Neurophysiology*, 116(1):56–62, January 2005.

[82] Dennis J. McFarland and Jonathan R. Wolpaw. Brain-computer interfaces for communication and control. *Communications of the ACM*, 54(5):60–66, May 2011.

[83] D.J. McFarland and J.R. Wolpaw. Sensorimotor rhythm-based brain-computer interface (BCI): feature selection by regression improves performance. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(3):372–379, September 2005.

[84] Michelle A McSweeney. *The pragmatics of text messaging: making meaning in messages*. Routledge, 2018.

[85] John Hopkins Medicine. Electroencephalogram (eeg), 2019.

[86] John V. Monaco and Charles C. Tappert. The partially observable hidden markov model and its application to keystroke dynamics. *Pattern Recognition*, 76:449–462, April 2018.

[87] A. Mouraux and G.D. Iannetti. Across-trial averaging of event-related EEG responses and beyond. *Magnetic Resonance Imaging*, 26(7):1041–1054, September 2008.

[88] G.R. Muller-Putz and G. Pfurtscheller. Control of an electrical prosthesis with an SSVEP-based BCI. *IEEE Transactions on Biomedical Engineering*, 55(1):361–364, January 2008.

[89] Kazuya Murao, Hayami Tobise, Tsutomu Terada, Toshiki Iso, Masahiko Tsukamoto, and Tsutomu Horikoshi. Mobile phone user authentication with grip gestures using pressure sensors. *International Journal of Pervasive Computing and Communications*, 11(3):288–301, September 2015.

[90] Brad A Myers. A user acceptance equation for intelligent assistants. In *AAAI 2007 Spring Symposium on Interaction Challenges for Intelligent Assistants*, 2007.

[91] Naresh Nagabushan, Taber Fisher, Giovanni Malaty, Mark Witcher, and Sujith Vijayan. A comparative study of motor imagery based BCI classifiers on EEG and iEEG data. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, November 2019.

[92] Noman Naseer and Keum-Shik Hong. fNIRS-based brain-computer interfaces: a review. *Frontiers in Human Neuroscience*, 9:3, January 2015.

[93] Christa Neuper and Gert Pfurtscheller. Neurofeedback training for BCI control. In *Brain-Computer Interfaces*, pages 65–78. Springer Berlin Heidelberg, 2009.

[94] ANT Neuro. eego mylab ant neuro, 2019.

[95] NeuroSky. Mindwave mobile 2 headset, 2015.

[96] Allen Newell and Paul S Rosenbloom. Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, 1(1981):1–55, 1981.

[97] Anton Nijholt, Danny Plass-Oude Bos, and Boris Reuderink. Turning shortcomings into challenges: Brain–computer interfaces for games. *Entertainment Computing*, 1(2):85–94, April 2009.

[98] Anton Nijholt, Desney Tan, Brendan Allison, Jose del R. Milan, and Bernhard Graimann. Brain-computer interfaces for hci and games. In *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*. ACM Press, 2008.

[99] Soroush Niketeghad and Nader Pouratian. Brain machine interfaces for vision restoration: The current state of cortical visual prosthetics. *Neurotherapeutics*, 16(1):134–143, September 2018.

[100] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[101] Rupert Ortner, Brendan Z. Allison, Gerd Korisek, Herbert Gaggl, and Gert Pfurtscheller. An SSVEP BCI to control a hand orthosis for persons with tetraplegia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(1):1–5, February 2011.

[102] R C Panicker, S Puthusserypady, and Ying Sun. An asynchronous p300 BCI with SSVEP-based control state detection. *IEEE Transactions on Biomedical Engineering*, 58(6):1781–1788, June 2011.

[103] Cheolsoo Park, David Looney, Naveed ur Rehman, Alireza Ahrabian, and Danilo P. Mandic. Classification of motor imagery BCI using multivariate empirical mode decomposition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(1):10–22, January 2013.

[104] Harold Pashler, Nicholas J. Cepeda, John T. Wixted, and Doug Rohrer. When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1):3–8, 2005.

[105] G. Pfurtscheller, G.R. Muller-Putz, A. Schlogl, B. Graimann, R. Scherer, R. Leeb, C. Brunner, C. Keinrath, F. Lee, G. Townsend, C. Vidaurre, and C. Neuper. 15 years of BCI research at graz university of technology: current projects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):205–210, June 2006.

[106] Gert Pfurtscheller. The hybrid BCI. *Frontiers in Neuroscience*, 4:30, 2010.

[107] Mary A. Pyc and Katherine A. Rawson. Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4):437–447, May 2009.

[108] Matthew B. Reysen and James S. Nairne. Part-set cuing of false memories. *Psychonomic Bulletin & Review*, 9(2):389–393, June 2002.

[109] Jochen Rick. Performance optimizations of virtual keyboards for stroke-based text entry on a touch-based tabletop. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*. ACM Press, 2010.

[110] Claudia Sannelli, Mikio Braun, Michael Tangermann, and Klaus-Robert Müller. Estimating noise and dimensionality in bci data sets: towards illiteracy comprehension. In *International Brain-Computer Interface Workshop and Training Course*, volume 1, pages 26–31. Citeseer, 2008.

[111] Abdulaziz B Sanosi. The effect of quizlet on vocabulary acquisition. *Asian Journal of Education and e-learning*, 6(4):2321–2454, 2018.

[112] Joey Scarr, Andy Cockburn, Carl Gutwin, and Philip Quinn. Dips and ceilings. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, 2011.

[113] Gerwin Schalk, Jonathan R Wolpaw, Dennis J McFarland, and Gert Pfurtscheller. EEG-based communication: presence of an error potential. *Clinical Neurophysiology*, 111(12):2138–2144, December 2000.

[114] Richard A. Schmidt and Robert A. Bjork. New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4):207–218, July 1992.

[115] P. Shenoy, K.J. Miller, J.G. Ojemann, and R.P.N. Rao. Generalized features for electrocorticographic BCIs. *IEEE Transactions on Biomedical Engineering*, 55(1):273–280, January 2008.

[116] Jerry J. Shih, Dean J. Krusienski, and Jonathan R. Wolpaw. Brain-computer interfaces in medicine. *Mayo Clinic Proceedings*, 87(3):268–279, March 2012.

[117] Ranganatha Sitaram, Andrea Caria, Ralf Veit, Tilman Gaber, Giuseppina Rota, Andrea Kuebler, and Niels Birbaumer. fMRI brain-computer interface: A tool for neuroscientific research and treatment. *Computational Intelligence and Neuroscience*, 2007:1–10, 2007.

[118] K. Sobottka and I. Pitas. Face localization and facial feature extraction based on shape and color information. In *Proceedings of 3rd IEEE International Conference on Image Processing*. IEEE, 1996.

[119] Daniela Sochurková, Milan Brázdil, Pavel Jurák, and Ivan Rektor. P3 and ERD/ERS in a visual oddball paradigm. *Journal of Psychophysiology*, 20(1):32–39, January 2006.

[120] Nancy K Squires, Kenneth C Squires, and Steven A Hillyard. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38(4):387–401, April 1975.

[121] Jason Stanley and Timothy Willlamson and. Knowing how. *Journal of Philosophy*, 98(8):411–444, 2001.

[122] Constantine Stephanidis and Anthony Savidis. Universal access in the information society: Methods, tools, and interaction technologies. *Universal Access in the Information Society*, 1(1):40–55, June 2001.

[123] Eliza Strickland. Startup neurable unveils the world's first brain-controlled vr game, Aug 2017.

[124] Jeffrey Stylos, Brad A. Myers, and Andrew Faulring. Citrine: providing intelligent copy-and-paste. In *Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST '04*. ACM Press, 2004.

[125] Abdulhamit Subasi and M. Ismail Gursoy. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications*, 37(12):8659–8666, December 2010.

[126] Xiaohong Sui, Liming Li, Xinyu Chai, Kaijie Wu, Chuanqing Zhou, Xiaodong Sun, Xun Xu, Xiaoxin Li, and Qiushi Ren. Visual prosthesis for optic nerve stimulation. In *Biological and Medical Physics, Biomedical Engineering*, pages 43–83. Springer US, 2009.

[127] Decho Surangsrirat and Apichart Intarapanich. Analysis of the meditation brainwave from consumer EEG device. In *SoutheastCon 2015*. IEEE, April 2015.

[128] Vasileios Syrris, Stefano Ferri, Daniele Ehrlich, and Martino Pesaresi. Image enhancement and feature extraction based on low-resolution satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5):1986–1995, May 2015.

[129] R. Tandra and A. Sahai. Fundamental limits on detection in low SNR under noise uncertainty. In *2005 International Conference on Wireless Networks, Communications and Mobile Computing*. IEEE, 2005.

[130] Carol Tenopir. Common end user errors. *Library Journal*, pages 31–32, 1997.

[131] Mario Tudor, Lorainne Tudor, and Katarina Ivana Tudor. Hans berger (1873-1941)–the history of electroencephalography. *Acta medica Croatica: casopis Hravatske akademije medicinskih znanosti*, 59(4):307–313, 2005.

[132] R. VanRullen, N. A. Busch, J. Drewes, and Julien Dubois. Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. *Frontiers in Psychology*, 2:1–9, 2011.

[133] Saeed V Vaseghi. *Advanced digital signal processing and noise reduction*. John Wiley & Sons, 2008.

[134] T.M. Vaughan, D.J. McFarland, G. Schalk, W.A. Sarnacki, D.J. Krusienski, E.W. Sellers, and J.R. Wolpaw. The wadsworth BCI research and development program: at home with BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):229–233, June 2006.

[135] L. Vega-Escobar, A.E. Castro-Ospina, and L. Duque-Munoz. Feature extraction schemes for BCI systems. In *2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)*. IEEE, September 2015.

[136] Carmen Vidaurre and Benjamin Blankertz. Towards a cure for BCI illiteracy. *Brain Topography*, 23(2):194–198, November 2009.

[137] Athanasios Vourvopoulos and Sergi Bermudez i Badia. Usability and cost-effectiveness in brain-computer interaction. In *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, February 2016.

[138] Hongtao Wang and Anastasios Bezerianos. Brain-controlled wheelchair controlled by sustained and brief motor imagery BCIs. *Electronics Letters*, 53(17):1178–1180, August 2017.

[139] Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. Interacting with soli. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, October 2016.

[140] Zhengjie Wang, Yushan Hou, Kangkang Jiang, Wenwen Dou, Chengming Zhang, Zehua Huang, and Yinjing Guo. Hand gesture recognition based on active ultrasonic sensing of smartphone: A survey. *IEEE Access*, 7:111897–111922, 2019.

[141] Graham Wilson, Stephen A Brewster, and Martin Halvey. Augmenting tactile interaction with pressure-based input. In *Digital Engagement '11*, 2011.

[142] Carolee J. Winstein and Richard A. Schmidt. Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4):677–691, 1990.

[143] John T. Wixted. The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55(1):235–269, February 2004.

[144] Jacob O Wobbrock. The future of mobile device research in hci. In *CHI 2006 workshop proceedings: what is the next generation of human-computer interaction*, pages 131–134. Citeseer, 2006.

[145] Jacob O. Wobbrock, Edward Cutrell, Susumu Harada, and I. Scott MacKenzie. An error model for pointing based on fitts' law. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, 2008.

[146] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, 2011.

[147] Tong Yang, Shahin Hakimian, and Theodore H. Schwartz. Intraoperative ElectroCorticoGraphy (ECog): indications, techniques, and utility in epilepsy surgery. *Epileptic Disorders*, 16(3):271–279, September 2014.

[148] Nick Yeung, Matthew M. Botvinick, and Jonathan D. Cohen. The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111(4):931–959, October 2004.

[149] Xinyi Yong, Rabab K. Ward, and Gary E. Birch. Artifact removal in EEG using morphological component analysis. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, April 2009.

[150] Thorsten O Zander and Christian Kothe. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering*, 8(2):025005, March 2011.

[151] Shumin Zhai. Foundational issues in touch-surface stroke gesture design — an integrative review. *Foundations and Trends® in Human–Computer Interaction*, 5(2):97–205, 2012.

[152] Guangtao Zhang, John Paulin Hansen, Katsumi Minakata, Alexandre Alapetite, and Zhongyu Wang. Eye-gaze-controlled telepresence robots for people with motor disabilities. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, March 2019.

[153] Minghui Zhu, Weiwei Liu, and Pawel Wargocki. Changes in EEG signals during the cognitive activity at varying air temperature and relative humidity. *Journal of Exposure Science & Environmental Epidemiology*, 30(2):285–298, June 2019.

[154] Yunqi Zhu, Kedi Xu, Caiyun Xu, Jiacheng Zhang, Jianfeng Ji, Xiaoxiang Zheng, Hong Zhang, and Mei Tian. PET mapping for brain–computer interface stimulation of the ventroposterior medial nucleus of the thalamus in rats with implanted electrodes. *Journal of Nuclear Medicine*, 57(7):1141–1145, February 2016.

[155] John Zimmerman, Anthony Tomasic, Isaac Simmons, Ian Hargraves, Ken Mohnkern, Jason Cornwell, and Robert Martin McGuire. Vio: a mixed-initiative approach to learning and automating procedural update tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*. ACM Press, 2007.

Appendix A

Consent Form

Key Input App

Before proceeding, please read the following. You must give your consent to continue.

Consent Form

Title: Determining the effects of error on user inputs

Researcher(s):

- Kevin Lam, Researcher, Department of Computer Science, University of Saskatchewan, kevin.lam@usask.ca
- Dr. Carl Gutwin, Supervisor, Department of Computer Science, University of Saskatchewan, +1 306 966-8646, gutwin@cs.usask.ca
- Dr. Madison Klarkowski, Supervisor, Department of Computer Science, University of Saskatchewan, +1 306 966-8940, madison@cs.usask.ca

Purpose(s) and Objective(s) of the Research: This study will examine the effects of error on how client-side user input is processed.

Procedures: After completing a demographics questionnaire, participants will perform a series of item selection tasks from different categories. These selection tasks will be performed using the arrow keys on the keyboard. The arrow key input data will be logged by the system, and the participants will be asked to complete an effort questionnaire following the tasks.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks:

- There are no known or anticipated risks to you by participating in this specific research project.
- There is always a risk of minor anxiety or stress associated with participating in an experiment.

Potential Benefits: Your participation will contribute to a better understanding of how client-side user errors can be improved. This knowledge will aid designers and developers in creating better input methods.

Compensation:

- To thank you for participating, we will provide you with a $5 honorarium.
- The entire study (including questionnaires) should take approximately 30 minutes to complete.

Confidentiality:

- Confidentiality will be maintained throughout the study, and all data gathered during the study will be anonymized. Any identifying information, such as MTurk ID number, will be stored separately from the data. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation within academic venues.
- Only the researchers listed above will have access to the raw data.
- **Storage of Data:**
  - Data (including survey and interview responses, logs of computer use) will be stored on a secure password-protected server for 5 years after data collection.
  - After 5 years, the data will be destroyed. All digital data will be wiped from hard disks beyond any possibility of data recovery.

Figure A.1: Chapter 4 Studies Consent Form, Part 1 of 2

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Deciding to withdraw from the study will not affect your status or standing (including ratings on any online services such as MTurk).
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Withdrawal requests can be made by contacting us directly with the methods listed above, or by entering "WITHDRAW" as the MTurk completion code.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up: To obtain a summary of results from the study, please contact Kevin Lam (kevin.lam@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top of this form.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office (ethics.office@usask.ca, (306) 966-2975). Out of town participants may call toll free at: (888) 966-2975.

Copies:

- If you would like to keep a copy of this consent form for your records, right-click this web page, click "Save Page As..." and follow the prompts provided by your web browser.

By choosing "I give my consent" below, you are indicating that you...

- Have read and understand the description provided.
- Consent to participate in the research project.
- Understand that a copy of this consent form is available to you for your records.

Do you give your consent?

○ I give my consent

○ I do not give my consent

Continue

Figure A.2: Chapter 4 Studies Consent Form, Part 2 of 2

Appendix B

Post-Condition Questionnaire

Key Input App

Please rate your experience with the tasks that you have just completed.

Please provide a rating on the previous set of tasks:

	Very Low	Low	Somewhat Low	Neutral	Somewhat High	High	Very High
Mental Demand	○	○	○	○	○	○	○
Physical Demand	○	○	○	○	○	○	○
Time Pressure	○	○	○	○	○	○	○
Your Performance	○	○	○	○	○	○	○
Your Effort	○	○	○	○	○	○	○
Frustration Level	○	○	○	○	○	○	○

Did you encounter any errors in the system (not caused by you) while completing the tasks?

○ Yes

○ No

If you answered 'Yes' to the question above, please specify the type of errors that occurred:

Please list the errors you encountered.

If you answered 'Yes' to the question above, how often did the errors occur (%)?

Figure B.1: Chapter 4 Studies Post-Condition Questionnaire, Part 1 of 2

How easy was it for you to learn the items in this set of tasks?

- ○ Very Difficult
- ○ Difficult
- ○ Somewhat Difficult
- ○ Neutral
- ○ Somewhat Easy
- ○ Easy
- ○ Very Easy

How well did you memorize the items by the end of the previous set of tasks?

- ○ Very Poorly
- ○ Poorly
- ○ Somewhat Poorly
- ○ Neutral
- ○ Somewhat Well
- ○ Well
- ○ Very Well

Was there anything in the previous set of tasks that hindered your learning of the items?

- ○ Yes
- ○ No

If you answered 'Yes' to the question above, please specify the hindrances that occurred:

Please list the hindrances you encountered.

Did you use any external aids in the memory tests (e.g. taking a screenshot or picture of the guide)?
Please answer honestly. Your compensation will not be affected by your response.

- ○ Yes
- ○ No

If you answered 'Yes' to the question above, please specify the reason why you used an external aid.

Please provide your reason(s).

Are there any additional comments or suggestions that you would like to share?

Additional comments

Continue

Figure B.2: Chapter 4 Studies Post-Condition Questionnaire, Part 2 of 2