

ABSTRACT

Alternatives using the Leap Motion to extend Mid-Air Word-Gesture Keyboards

Garrett Benoit, M.S.

Faculty Adviser: Michael G. Poor, Ph.D.

Abstract goes here

Alternatives using the Leap Motion to extend Mid-Air Word-Gesture Keyboards

by

Garrett Benoit, B.S.

A Thesis

Approved by the Department of Computer Science

Gregory D. Speegle, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree of

Master of Science

Approved by the Thesis Committee

Michael G. Poor, Ph.D., Chairperson

Bill Poucher, Ph.D.

Brian Garner, Ph.D.

Accepted by the Graduate School

October 2015

J. Larry Lyon, Ph.D., Dean

Copyright © 2015 by Garrett Benoit

All rights reserved

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF TABLES	xiii
ACKNOWLEDGMENTS	xv
DEDICATION	xvi
1 Introduction — purpose for study, significance of study, original contribution	1
1.1 Motivation	1
1.2 Definition	3
1.3 Gestures/Text Entry: Past, Present and Future — rename this and subsections (Related Works)	3
1.3.1 A Brief History	3
1.3.2 Current Trends	3
1.3.3 Future Plans	3
1.4 Problems, Challenges, Limitations — rename this section	3
1.5 Solution Scope — rename	3
2 Literature Review — related works, describe and analyze previous works, what is known/not known, leads to hypotheses	5
2.1 3 dimensions beyond desktop — rename	5
2.2 Interaction	5
2.2.1 Gesture-based Interaction — similar to introduction materials	5
2.2.2 Barehanded Interaction/Stylus/Gloves	6
2.2.3 Gestural Manipulation — rename	6
2.2.4 Multimodal	6

2.3	Fatigue	6
2.4	Pointing Device Evaluation/Learning Effects/Motor Learning/Ergonomics and Usability	6
2.4.1	Fitts' Law — rename	6
2.4.2	Gestural Pointing — rename	6
2.5	Text Entry	6
2.5.1	Word-Gesture Keyboards	6
2.5.2	Mid-Air Text Entry	6
3	Keyboard Design	7
3.1	Word-Gesture Keyboard Implementation	7
3.1.1	Lacking Word-Recognition	7
3.1.2	Design	7
3.1.3	Display	8
3.1.4	Dictionary Creation	10
3.1.5	Calibration	11
3.2	Word-Gesture Keyboards	12
3.2.1	Touch Screen Keyboard	12
3.2.2	Leap Surface Keyboard	12
3.2.3	Leap Static-Air Keyboard	12
3.2.4	Leap Predictive-Air Keyboard	13
3.2.5	Leap Bimodal-Air Keyboard	13
3.2.6	Leap Pinch-Air Keyboard	13
4	Methodology	14
4.1	Pre-Pilot Study	14
4.1.1	Participants	14

4.1.2	Input Devices and Interaction	14
4.1.3	Task Design	16
4.1.4	Experimental Design	18
4.1.5	Procedure	18
4.1.6	Dependent Measures	19
4.2	Pilot Study	20
4.2.1	Participants	20
4.2.2	Input Devices and Interaction	21
4.2.3	Task Design	22
4.2.4	Experimental Design	24
4.2.5	Procedure	25
4.2.6	Dependent Measures	26
4.3	Final Study	31
4.3.1	Participants	32
4.3.2	Input Devices and Interaction	33
4.3.3	Task Design	35
4.3.4	Experimental Design	36
4.3.5	Procedure	37
4.3.6	Dependent Measures	39
5	Results and Analysis	44
5.1	Text-Entry Rate	44
5.1.1	Text-Entry Rate Modified-Shortest	46
5.2	Error Rate	48
5.2.1	Minimum Word Distance	48
5.2.2	Keystrokes Per Character	52

5.2.3	Minimum String Distance	55
5.2.4	Total Error Rate	57
5.3	Correctness	60
5.3.1	Fréchet Distance	60
5.4	Distance Measures	65
5.4.1	Word-Gesture Distance	65
5.4.2	Word-Gesture Velocity	68
5.4.3	Hand Velocity	70
5.5	Timing Measures	71
5.5.1	Word-Gesture Duration	71
5.5.2	Reaction Time to Errors	74
5.5.3	Reaction Time to Simulate Touch	76
5.5.4	Reaction Time for First Correct Letter	77
5.6	Quantitative Measures	79
5.6.1	Number of Touches Simulated	79
5.6.2	Number of Practice Words	81
5.7	Qualitative Measures	81
5.7.1	Level of Discomfort	81
5.7.2	Level of Fatigue	83
5.7.3	Level of Difficulty	85
5.7.4	Preference Ranking	85
6	Discussion, Future Work and Conclusion	88
6.1	Discussion	88
6.1.1	Pinching Interaction	89
6.1.2	3-Dimensional Interaction	89

6.1.3	Bimodal Interaction	90
6.1.4	Lacking Word-Recognition	90
6.2	Future Work	91
6.2.1	Implement Word-Recognition	91
6.2.2	Redesign Task	91
6.2.3	Augmented Reality	92
6.2.4	The "Skimming" Issue	92
6.2.5	Better Tracking	92
6.2.6	Alternative Interaction Plane	93
6.2.7	Multi-functional Space	93
6.2.8	Image Processing	93
6.2.9	Gaming Console Keyboards	94
6.2.10	Accessibility	94
6.3	Conclusion	94
APPENDICES		96
APPENDIX A Word-Gesturing Implementation		97
APPENDIX B Results Tables and Boxplots		98
B.1	Text-Entry Rate	99
B.1.1	Text-Entry Rate Modified-Shortest	101
B.2	Error Rates	103
B.2.1	Minimum Word Distance	103
B.2.2	Keystrokes Per Character	107
B.2.3	Minimum String Distance	111
B.2.4	Total Error Rate	113

B.3	Correctness	117
B.3.1	Fréchet Distance	117
B.4	Distance Measures	123
B.4.1	Word-Gesture Distance	123
B.4.2	Word-Gesture Velocity	127
B.4.3	Hand Velocity	129
B.5	Timing Measures	131
B.5.1	Word-Gesture Duration	131
B.5.2	Reaction Time to Errors	135
B.5.3	Reaction Time to Simulate Touch	137
B.5.4	Reaction Time for First Correct Letter	139
B.6	Quantitative Measures	141
B.6.1	Number of Touches Simulated	141
B.6.2	Number of Practice Words	143
B.7	Qualitative Measures	144
B.7.1	Level of Discomfort	144
B.7.2	Level of Fatigue	145
B.7.3	Level of Difficulty	146
B.7.4	Preference Ranking	147
	BIBLIOGRAPHY	149

LIST OF FIGURES

3.1	Display: Keyboard Layout	9
3.2	Display: Text Area	9
3.3	Display: Text Area	10
4.1	Fréchet Distance code snippet	30
5.1	Mean Text-Entry Rates	45
5.2	Mean Text-Entry Rates for Modified-Shortest	47
5.3	Mean Minimum Word Distance	49
5.4	Mean Minimum Word Distance for Modified-Shortest	51
5.5	Mean Keystrokes Per Character	53
5.6	Mean Keystrokes Per Character for Modified-Shortest	54
5.7	Mean Minimum String Distance for Modified-Shortest	56
5.8	Mean Total Error Rates	57
5.9	Mean Total Error Rate for Modified-Shortest	59
5.10	Mean Fréchet Distance	61
5.11	Mean Fréchet Distance for Modified-Shortest	63
5.12	Mean Fréchet Distance for Modified-Backspace	64
5.13	Mean Word-Gesture Distance	66
5.14	Mean Word-Gesture Distance for Modified-Shortest	68
5.15	Mean Word-Gesture Velocity	69
5.16	Mean Hand Velocity	71
5.17	Mean Word-Gesture Duration	72
5.18	Mean Word-Gesture Dur for Modified-Shortest	74
5.19	Mean Reaction Time to Errors	75

5.20	Mean Reaction Time to Simulate Touch	77
5.21	Mean Reaction Time for First Correct Letter	78
5.22	Mean Number of Touches Simulated	80
5.23	Mean Number of Practice Words	82
5.24	Level of Discomfort Boxplot	83
5.25	Level of Fatigue Boxplot	84
5.26	Level of Difficulty Boxplot	86
5.27	Mean Preference Rankings	87
B.1	Text-Entry Rate Boxplot	99
B.2	Text-Entry Rates Boxplot for Modified-Shortest	101
B.3	Minimum Word Distance Boxplot	103
B.4	Minimum Word Distance Boxplot for Modified-Shortest	105
B.5	Keystrokes Per Character Boxplot	107
B.6	Keystrokes Per Character Boxplot for Modified-Shortest	109
B.7	Minimum String Distance Boxplot for Modified-Shortest	111
B.8	Total Error Rate Boxplot	113
B.9	Total Error Rate Boxplot for Modified-Shortest	115
B.10	Fréchet Distance Boxplot	117
B.11	Fréchet Distance Boxplot for Modified-Shortest	119
B.12	Fréchet Distance Boxplot for Modified-Backspace	121
B.13	Word-Gesture Distance Boxplot	123
B.14	Word-Gesture Distance Boxplot for Modified-Shortest	125
B.15	Word-Gesture Velocity Boxplot	127
B.16	Hand Velocity Boxplot	129
B.17	Word-Gesture Duration Boxplot	131
B.18	Word-Gesture Duration Boxplot for Modified-Shortest	133
B.19	Reaction Time to Errors Boxplot	135

B.20 Reaction Time to Simulate Touch Boxplot	137
B.21 Reaction Time for First Correct Letter Boxplot	139
B.22 Number of Touches Simulated Boxplot	141
B.23 Number of Practice Words Boxplot	143
B.24 Preference Ranking Boxplot	147

LIST OF TABLES

4.1	Pre-pilot Study Details of Participants	14
4.2	Pre-pilot Study Schedule of Assessments	19
4.3	Pilot Study Details of Participants	20
4.4	Latin Squares Example	24
4.5	Pilot Study Schedule of Assessments	26
4.6	Final Study Details of Participants	33
4.7	Replicated Latin Squares Example	37
4.8	Final Study Schedule of Assessments	39
4.9	Shortest-transcribed Examples	40
4.10	Backspace-transcribed Examples	40
B.1	Text-Entry Rate Multiple Comparison	100
B.2	Text-Entry Rate Multiple Comparison Modified-Shortest	102
B.3	MWD Multiple Comparison	104
B.4	MWD Multiple Comparison Modified-Shortest	106
B.5	KSPC Multiple Comparison	108
B.6	KSPC Multiple Comparison Modified-Shortest	110
B.7	MSD Multiple Comparison Modified-Shortest	112
B.8	Total Error Rate Multiple Comparison	114
B.9	Total Error Rate Multiple Comparison Modified-Shortest	116
B.10	Fréchet Distance Multiple Comparison	118
B.11	Fréchet Distance Multiple Comparison Modified-Shortest	120
B.12	Fréchet Distance Multiple Comparison Modified-Backspace	122
B.13	Word-Gesture Distance Multiple Comparison	124
B.14	Word-Gesture Distance Multiple Comparison Modified-Shortest	126

B.15 Word-Gesture Velocity Multiple Comparison	128
B.16 Hand Velocity Multiple Comparison	130
B.17 Word-Gesture Duration Multiple Comparison	132
B.18 Word-Gesture Duration Multiple Comparison Modified-Shortest . . .	134
B.19 Reaction Time to Errors Multiple Comparison	136
B.20 Reaction Time to Simulate Touch Multiple Comparison	138
B.21 Reaction Time for First Correct Letter Multiple Comparison	140
B.22 Number of Touches Simulated Multiple Comparison	142
B.23 Level of Discomfort Multiple Comparison	144
B.24 Level of Fatigue Multiple Comparison	145
B.25 Level of Difficulty Multiple Comparison	146
B.26 Preference Ranking Multiple Comparison	148

ACKNOWLEDGMENTS

I want to express my great appreciation to several people. Acknowledgements.

Pour la rose qui fait important le temps qu'on perd

CHAPTER ONE

Introduction — purpose for study, significance of study, original contribution

In the first chapter, clearly state what the purpose of the study is and explain the study's significance. The significance is addressed by discussing how the study adds to the theoretical body of knowledge in the field and the study's practical significance for communication professionals in the field being examined.

Ph.D. students also must explain how their research makes an original contribution to the body of knowledge in their discipline. They also should address the significance of the study for mass communication education.

It is especially critical that this chapter be well developed. Without a clearly defined purpose and strong theoretical grounding, the thesis or dissertation is fundamentally flawed from the outset.

1.1 Motivation

With the increase in gesture-controlled interfaces for touch screen and other modern devices, gesture-controls have started to see a transition for use in mid-air. Mid-air, gesture-controlled content has seen its emergence in large displays (Nancel, Wagner, Pietriga, Chapuis, and Mackay 2011), smart phones (Jones, Sodhi, Forsyth, Bailey, and Maciocci 2012), augmented reality (Piumsomboon, Clark, Billinghamurst, and Cockburn 2013), and desktop computers (Sutton 2013). Mid-air pointing has been a common approach to many of these gesture-controlled interactions and is used to select and manipulate on-screen objects (Banerjee, Burstyn, Girouard, and Vertegaal 2012, Cockburn, Quinn, Gutwin, Ramos, and Looser 2011, Jota, Nacenta, Jorge, Carpendale, and Greenberg 2010, Shoemaker, Tang, and Booth 2007, Vogel and Balakrishnan 2005); however, means of reasonable mid-air text entry are fairly new. Past technologies allowed for mid-air text entry, but those approaches have

fallen short of any meaningful text entry rates (Malik and Laszlo 2004). More modern approaches of mid-air text entry have seen improved results but still low, around 13 (Markussen, Jakobsen, and Hornbæk 2013) to 18.9 (Shoemaker, Findlater, Dawson, and Booth 2009) words per minute. These approaches were limited to selection of individual characters and also lacked the multi-tap feature of touch-based entry (Markussen, Jakobsen, and Hornbæk 2013, Ni, Bowman, and North 2011, Shoemaker, Findlater, Dawson, and Booth 2009). Last year, the largest improvement was seen in mid-air text entry when Markussen et al. (Markussen, Jakobsen, and Hornbæk 2014) transitioned word-gesture keyboards for mid-air use with the development of Vulture, reaching a text entry rate of 20.6 words per minute for their initial study. They achieved an even better text entry rate of 28.1 words per minute in a second study, with training and repeated measures, indicating learning the new techniques will help bring mid-air text entry closer to touch-based text entry. Vulture reached

59

The purpose of this study is to use the Leap Motion, a new and emerging technology (Sutton 2013), for interpreting mid-air gestural inputs for text entry (Toimil, Shin, and Eisner 2013). The only previous attempt for text entry with the Leap Motion was in mid-air handwriting (Vikram, Li, and Russell 2013); however, even regular hand-writing is slow, and confined around 15 words per minute (Devoe 1967). Instead, this study aims to follow the path of Markussen et al. (Markussen, Jakobsen, and Hornbæk 2014) and use the Leap Motion to extend word-gesture keyboards to mid-air text entry. Word-gesture keyboards have garnered popularity with the advent of smart phones and tablets and have been proven to perform well on touch screens (Kristensson 2007, Zhai and Kristensson 2012, Zhai, Kristensson, Gong, Greiner, Peng, Liu, and Dunnigan 2009). This study intends to use the Leap Motion to find alternatives to mid-air, word-gesture keyboards and find a better approach than wearing a glove or detecting pinching (Markussen, Jakobsen, and Hornbæk 2014, Ni,

Bowman, and North 2011) for the mid-air equivalent of tapping and releasing for delimiters of words. This study will explore the option of using the extra degrees of freedom available in mid-air (e.g., depth) which was decided against by Markussen et al. (Markussen, Jakobsen, and Hornbæk 2014) to create virtual keyboards in mid-air, and it will also use several techniques of simulating touch for the mid-air equivalent of tapping and releasing for delimiters of words. The goal is to make it feel as similar to using a touch-based device as possible while still allowing for common gestures.

The rationale behind this research is to improve mid-air text entry using techniques that will still allow for other gesture-controls when working with gesture-interfaces (Banerjee, Burstyn, Girouard, and Vertegaal 2012, Cockburn, Quinn, Gutwin, Ramos, and Looser 2011, Jota, Nacenta, Jorge, Carpendale, and Greenberg 2010, Shoemaker, Tang, and Booth 2007, Vogel and Balakrishnan 2005). Touch-less gesture-controllers with mid-air text entry will benefit augmented reality (e.g., Google Glass, Microsoft HoloLens) as well as benefit the medical world (e.g., operating rooms) when it comes to sanitation and sterile environments to reduce the spread of pathogens.

1.2 Definition

1.3 Gestures/Text Entry: Past, Present and Future — rename this and subsections (Related Works)

1.3.1 A Brief History

1.3.2 Current Trends

1.3.3 Future Plans

1.4 Problems, Challenges, Limitations — rename this section

1.5 Solution Scope — rename

1. Explore alternatives using the Leap Motion for mid-air, word-gesture keyboards to find a better approach than pinching for the mid-air equivalent of tapping

and releasing for delimiters of words and increasing text entry rates (Markussen, Jakobsen, and Hornbæk 2014).

2. Explore alternatives using the Leap Motion to touch-based, word-gesture keyboards. Alternatives, without training or repeated measures, are not expected to surpass touch-based, word-gesture keyboards in text entry rates; however, some alternatives are expected to be similar in error rates, precision, and usability.

3. Look for evidence of a correlation between having previous word-gesture keyboard experience and the text entry rates, error rates, precision, and usability of the various word-gesture keyboards.

CHAPTER TWO

Literature Review — related works, describe and analyze previous works, what is known/not known, leads to hypotheses

The purpose of the study should suggest some theoretical framework to be explained further in this chapter. The literature review thus describes and analyzes previous research on the topic.

This chapter, however, should not merely string together what other researchers have found. Rather, you should discuss and analyze the body of knowledge with the ultimate goal of determining what is known and is not known about the topic. This determination leads to your research questions and/or hypotheses. In some cases, of course, you may determine that replicating previous research is needed.

2.1 3 dimensions beyond desktop — rename

2.2 Interaction

2.2.1 Gesture-based Interaction — similar to introduction materials

2.2.1.1 Mid-Air Gestures.

2.2.2 *Barehanded Interaction/Stylus/Gloves*

2.2.3 *Gestural Manipulation — rename*

2.2.4 *Multimodal*

2.3 *Fatigue*

2.4 *Pointing Device Evaluation/Learning Effects/Motor Learning/Ergonomics and Usability*

2.4.1 *Fitts' Law — rename*

2.4.2 *Gestural Pointing — rename*

2.4.2.1 *Mid-Air Pointing.*

2.5 *Text Entry*

2.5.1 *Word-Gesture Keyboards*

2.5.2 *Mid-Air Text Entry*

CHAPTER THREE

Keyboard Design

3.1 *Word-Gesture Keyboard Implementation*

3.1.1 *Lacking Word-Recognition*

Recreating a word-gesture keyboard from scratch with the limited, non-commercial information available was outside the scope of this study. Word-recognition had already been proven to work and benefit word-gesture keyboards immensely TODO: GET REFERENCES FOR WGK and RECOGNITION etc. In addition, the words that each participant was entering was known in advance, therefore a pseudo word-gesture keyboard implementation was used.

3.1.2 *Design*

The pseudo word-gesture implementation was created by analyzing the gesture as it was being drawn to determine which keys were being pressed. The assumptions for detecting a key press were based off of the known word being gestured and noticeable deviations made in the gesture's direction. To reduce the chance of erroneous keys being produced, the sizes of the character that was expected to be pressed next were exaggerated and the threshold for detecting deviations in the gesture were lowered between keys.

The displayed keys were 64x64 pixels in size with a gap of 10 pixels between each key. The actual size of the keys was dependent on the display device being used. The next expected key to be pressed was changed into a circular key with a radius of 76.8 pixels, 20% larger than the key widths. There was no feedback shown to the participant of the change in key size. TODO: SHOW PICTURE THAT SHOWS A BLOATING KEY.

An interpolated trail with points at a minimum of 16 pixels apart was used in determining deviations in word-gesturing. The angle of detecting a deviation was 165 degrees for all areas that were not on the expected path to the next key and was 90 degrees while on the expected path. Deviations in gesture path had to be at least 48 pixels away from each other to be detected. The expected path between two keys comprised an area from the previous expected key to the next expected key with a width 62.5% larger than key size, or 104 pixels. TODO: SHOW PICTURE OF DEVIATION OFF OF PATH VS DEVIATION ON PATH.

The specific values for detecting key presses were found using trial and error and were used to create an experience as close to a word-gesture keyboard as possible. The main deviation between between this implementation and one with word-recognition is that participants are shown updates in real-time of what the keyboard path is doing. This was determined to be an acceptable limitation.

3.1.3 Display

3.1.3.1 Keyboard Layout. The keyboard layout, seen in Figure 3.1, was laid out in a typical QWERTY keyboard fashion with key sizes of 64x64 pixels and gaps of 10 pixels. All special keys and number keys were removed to simplify the keyboard, with a backspace key added to the right of the keyboard to allow for deletion of erroneous characters.

3.1.3.2 Text Area. Figure 3.2 shows how two text areas were used to display text to participants. The top text-area showed the current word to transcribe, shown in Figure 3.2a, and the bottom area showed the currently transcribed text, shown in Figure 3.2b. The characters of both the displayed word and transcribed text were highlighted in green when the characters matched and were correct, whereas the letters in only the transcribed text would highlight in red if errors were made during transcription. The participant can then use the back space in order to correct words,

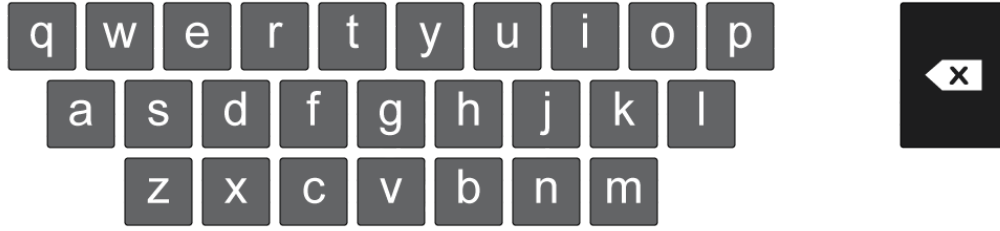


Figure 3.1. The keyboard layout used during the final study.

and they are finally highlighted in green, as seen in Figure 3.2c, when a word was completed and correct.

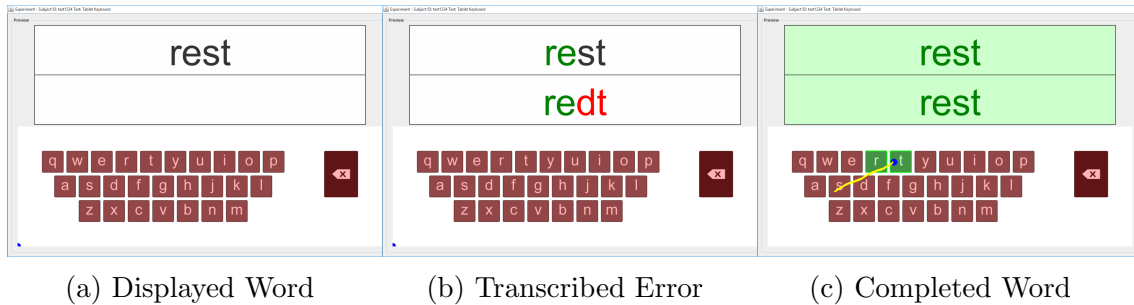


Figure 3.2. Examples of how the text areas change when showing transcribed text. **(a)** shows the word to transcribe as it first appears. **(b)** shows how correct letters are highlighted green in both text areas and transcription errors are highlighted in red. **(c)** shows a completed word, the background lights up to indicate correctness.

3.1.3.3 Real-Time Updates. As a participant is drawing the gesture-shape of a word, their progress is tracked in real time, as shown in Figure 3.3. For the keyboards that track a participant’s finger or the stylus, Figure 3.3a shows how a cylinder is displayed to indicate the direction that the finger is pointing, as well as which letter is being hovered over, indicated by a blue dot. As in Figure 3.3b, the participant is shown the path that they are traveling as well as the letter that have

been hit. The gesture-trail has a maximum length before it starts to decay as to not clutter the display.

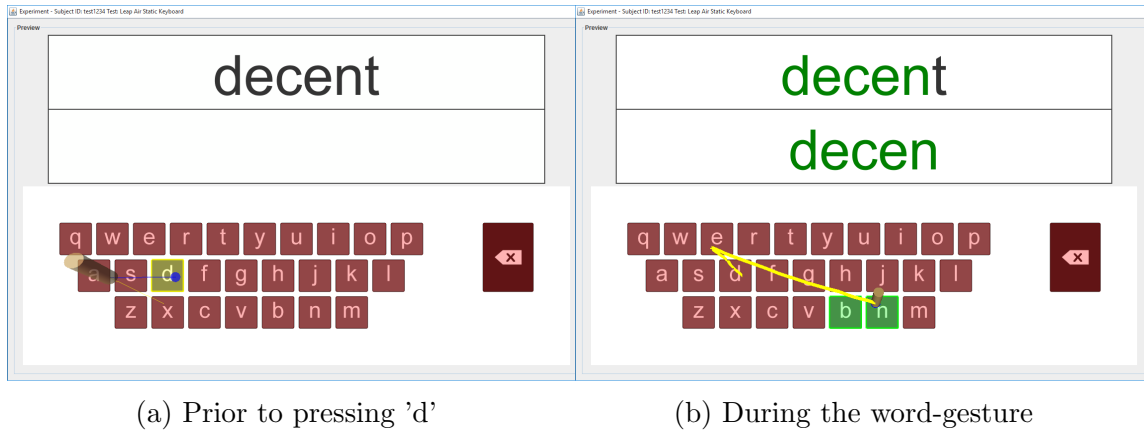


Figure 3.3. Examples the real-time display for word-gesturing. (a) shows the user just about to press the first character. (b) shows the middle of the word gesturing process for the word "decent".

3.1.4 Dictionary Creation

To make each keyboard experience as similar as possible, custom dictionaries were created for each keyboard input device containing words with similar gesture-shapes between keyboards. Dictionaries were created using a custom gesture-shape dissimilarity algorithm. The dictionaries contained words between the character lengths of 3 and 6 so that words would be rather simple, and words that most participants had seen before.

3.1.4.1 Using Similar Gesture-Shapes. The reason that custom dictionaries were chosen to be created was because it was thought that dependent measures would be more accurately represented for each keyboard if the experiences between those keyboards was nearly the same, only changing due to implementation. There was no previous research on words with similar gesture-shapes, or if this was truly necessary or beneficial.

3.1.4.2 *Gesture-Shape Dissimilarity*. Originally the Fréchet Distance was used to try to find the most similar word gesture-shapes using the sets of words with the least distance between each gesture-shape. Fréchet Distance gave acceptable results, however there were noticeable differences in some of the paths created. TODO: SHOW FIGURE OF 10 WORDS FOR FRECHET vs 10 WORDS FROM MINE — side by side comparison.

In order to achieve gesture-shapes with even closer similarity, a custom dissimilarity algorithm to rate words' gesture-shapes of the same length was created. The top results for words with the least dissimilarity between each other were then split among the dictionaries. The dissimilarity between two words is defined by the formula

$$dissimilarity(P, Q) = \frac{\sum_{i=2}^N \frac{1}{2} \left(\left(\frac{|dist(P_i, P_{i-1}) - dist(Q_i, Q_{i-1})|}{max\ distance} \right) + \left(\frac{angle(P_i - P_{i-1}, Q_i - Q_{i-1})}{\pi} \right) \right)}{N - 1} \quad (3.1)$$

where P and Q were two words of N length, P_i and Q_i were the vector locations of the characters of the words on the virtual keyboard, *max distance* was the maximum distance between any two letters on the virtual keyboard, *dist(...)* was the distance between two vector locations, and *angle(...)* was the angle between two vectors.

The dissimilarity formula forced values between the range $[0, 1]$ and treated every path between two letters with equal weight. The objective was then to find the sets of words with the lowest dissimilarity

3.1.5 Calibration

Maybe mentioned gorilla arm syndrom in Chapter 2?

— gorilla arm syndrom. Has a similar calibration as alvins Personal Space but this calibration was optional. This calibration space is also more complicated due to the nature of having to precisely interact with the generated plane.

THIS SECTION WILL GO THROUGH THE REASONS FOR CALIBRATION AND ALIVINS THING

3.1.5.1 *Separation of motor space and display space — rename (figure out ordering).*

3.1.5.2 *Size of motor space / User calibrated — rename (figure out ordering)*

.

Lack of recognition and reason for it. Simulated recognition.

3.2 *Word-Gesture Keyboards*

3.2.1 *Touch Screen Keyboard*

3.2.1.1 *Separating words — rename (figure out ordering).*

3.2.1.2 *Size of motor space / User calibrated — rename (figure out ordering)*

.

3.2.2 *Leap Surface Keyboard*

3.2.2.1 *Separating words — rename (figure out ordering).*

3.2.2.2 *Size of motor space / User calibrated — rename (figure out ordering)*

.

3.2.3 *Leap Static-Air Keyboard*

3.2.3.1 *Separating words — rename (figure out ordering).*

3.2.3.2 *Size of motor space / User calibrated — rename (figure out ordering)*

.

3.2.4 *Leap Predictive-Air Keyboard*

3.2.4.1 *Separating words — rename (figure out ordering).*

3.2.4.2 *Size of motor space / User calibrated — rename (figure out ordering)*

.

3.2.5 *Leap Bimodal-Air Keyboard*

3.2.5.1 *Separating words — rename (figure out ordering).*

3.2.5.2 *Size of motor space / User calibrated — rename (figure out ordering)*

.

3.2.6 *Leap Pinch-Air Keyboard*

NOTE: VULTURE PINCH FORCED AWKARD PINCH, LEAP ALLOWED
FOR ANY PINCH GESTURE FOR LESS FIDELITY

3.2.6.1 *Separating words — rename (figure out ordering).*

3.2.6.2 *Size of motor space / User calibrated — rename (figure out ordering)*

.

CHAPTER FOUR

Methodology

4.1 *Pre-Pilot Study*

The pre-pilot study was not a full-blown pilot. This study was used to test the conceptual feasibility of using 3 dimensions as a means of word separation for simulated touch input and to test the best-guess implementation for a word-gesture keyboard. TODO: See section SEPARATING WORDS CH3 for more information on how this process was derived.

4.1.1 *Participants*

A small sample size of 2 was used in this study. Both participants were male, ages 24 and 25. Both participants had previous experience with gesture-controllers, touch screens, and word-gesture keyboards. See TODO: TABLE BELOW for more information.

Table 4.1. Participant information including age, gender, handedness, computer usage, and previous experiences.

Subject	Gender	Age	Computer Usage per Week (hours)	Handedness	Hand Used in Experiment	Touch-device Experience	Gesture-controller Experience	Word-gesturing Experience	Impairment History
1	m	24	31 - 40	right	right	yes	yes	yes	no
2	m	25	50+	right	right	yes	yes	yes	no

4.1.2 *Input Devices and Interaction*

The interaction methods used within the study were dependent on the currently active keyboard input. All of the keyboards were simulated on a 64-bit, Windows 7 work station, with all receivers or controllers connected via USB 2.0. The participants were allowed to recalibrate the active keyboard's interaction plane for a more comfortable experience suited to their arm. Calibration was only possible if applicable to the active keyboard input. Participants were also encouraged to reposition

the gesture-controller, if applicable, and were given the option to freely rest or raise their elbows during the experiment. More information about the implementation of specific keyboard interactions, calibration, and the benefits of using a rested arm can be found in TODO: CHAPTER 3. Every keyboard was designed for use by either right or left handed participants.

4.1.2.1 Leap Motion Static-Air Keyboard. The Leap Motion Static-Air Keyboard used a Leap Motion Controller which was placed on the desk in front of the participant. The participant then used a stylus which was tracked by the Leap Motion in order to interact with a projected interaction plane. A touch was simulated by the insertion of the stylus into the interaction plane and a release was simulated upon the removal of the stylus. The interaction plane could be calibrated at any time prior to the experiment TODO: Provide picture.

4.1.2.2 Leap Motion Pinch-Air Keyboard. The Leap Motion Pinch-Air Keyboard also used a Leap Motion Controller that was positioned on the desk in front of the participant. The participant then used their bare hand, which was tracked at the center of their palm, to interact with the projected interaction plane. Touch was simulated by having the participant make a pinching gesture with their hand and a release was simulated when the participant released the pinch, opening their hand again. The interaction plane could be calibrated at any time prior to the experiment TODO: Provide picture. TODO: ALSO MAYBE REFERENCE VULTURE HERE or back in chpt

4.1.2.3 Leap Motion Surface Keyboard. Again, for the Leap Motion Surface Keyboard a Leap Motion Controller was used for tracking. Unlike the Leap Static-Air or Leap Pinch-Air keyboards, the gesture controller was placed into a custom designed holder instead of on the desk in front of the participant. The holder was attached to an inclined surface with a printed keyboard fixed on top TODO: for more details

see section BUILDING SURFACE WGK. Because placement was not guaranteed to be identical between uses, in order to accurately simulate an interaction plane projected onto the printed keyboard, the Leap Motion Surface Keyboard required a single calibration after first being inserted into the holder. The participant then used a stylus, as before, that was tracked by the Leap Motion in order to detect interaction. A touch was simulated by pressing the tip of the stylus against the printed keyboard and a release was simulated when the tip of the stylus was again removed from the printed keyboard surface.

4.1.2.4 Xbox Controller Keyboard. The Xbox Controller Keyboard used an Xbox 360 Wireless Controller that transmitted information via the Microsoft Xbox 360 Wireless Receiver for Windows. The participants were required to use any of the directional sticks or the D-pad in order to change which key was selected, and then used the 'A' button in order to select the currently highlighted key. The Xbox Controller keyboard only allowed for single-input text entry, functioning in the same way as the default Xbox 360 virtual keyboard.

4.1.3 Task Design

The design for the pre-pilot study included 4 different keyboard input devices representing each of the conditions that made up the task. The 4 conditions used were the Leap Motion Static-Air, Leap Motion Pinch-Air, Leap Motion Surface, and Xbox Controller keyboards. TODO: See section WORD GESTURE KEYBOARDS for specific details on each keyboard implementation.

The task performed by the participants consisted of 30 trials for each of the 4 keyboard input devices creating a total of 120 trials per participant. For each trial, a word was chosen at random of length 3 through 6 from the Oxford English Dictionary and displayed on the screen. A blank text-area was positioned directly below the displayed word for the transcribed, or user-generated, text TODO: GIVE FIGURE FOR

BLANK WORD DISPLAYED ON THE SCREEN. Beneath both text areas, there was a virtual representation of the keyboard the participants were using TODO: GIVE PICTURE FOR DISPLAYED KEYBOARD. The participants were then required to use the currently active keyboard input device to enter the displayed word using word-gesturing TODO: GIVE FIGURE SHOWING WORD-GESTURING. During the word-gesturing process, the participants were shown real-time updates to the displayed word and transcribed text as well as their movements within the virtual keyboard space. Detected key presses were appended directly to the transcribed text-area with correctly matching letters being highlighted in green and text entry errors being highlighted in red whereas only correctly matching letters, again highlighted in green, were applied to the displayed word TODO: REFERENCE A FIGURE HERE FOR COLOR OF TEXT. The participants were required to use the active keyboard input device's backspace key to remove errors. TODO: See section LACK OF WORD RECOGNITION for details on why a requirement to correctly enter the displayed word was enforced (best-guess nature vs word-recognition). Once a word was correctly entered, the participants were required to press the active keyboard's enter key to move onto the next word TODO: ADD PICTURE OF THE ENTER KEY.

Small deviations in the above task were required for how the participants interacted with the Xbox Controller keyboard. For this keyboard, there was no word-gesturing feature implemented, instead the participant was asked to use single-input text entry using a standard Xbox 360 Controller. The Xbox Controller keyboard was implemented in the same way as a conventional console keyboard, TODO: see section WGK XBOX CONTROLLER for more information. The participants had to press the 'A' button to select a letter, 'Y' button to delete, and the start button to move to the next word.

4.1.4 *Experimental Design*

The initial pre-pilot study used a within-subjects design without any counterbalancing. Both participants used every keyboard input device.

4.1.5 *Procedure*

There was only a single study visit for each participant in which all tasks were performed. The study visit took between 30 and 45 minutes to complete depending on how many calibrations were performed. The full set of tasks, and their expected durations, are detailed in TODO: SCHED OF ASSESSMENTS the pre-pilot schedule of assessments. The participants followed the same process for each of the 4 keyboard input device’s tasks.

First, the participants were given a brief explanation and demonstration of the active keyboard input. The explanation dialog contained the name of the active keyboard, how it was interacted with, and whether or not it was a word-gesture keyboard. The researcher then demonstrated how to enter the word ”test” using the active keyboard. The participants were then given control and handed the stylus or used their bare hand to interact with the keyboard to get a sense of how it worked.

Participants were then instructed to use the keyboard to perform practice words which were randomly chosen from the Oxford English Dictionary with lengths between 3 and 6 characters long. No practice words were duplicated and the dictionary was filtered for offensive words. There was no limit placed on how many words could be performed while practicing. The participants were told to continue until they felt they were able to efficiently and comfortably type each word with minimal errors. During the practice phase, the opportunity to optionally recalibrate the interaction-space any number of times was given if it was applicable to the active keyboard.

Next, the participants performed the task itself. As detailed in TODO: SECTION TASKS, the participants were instructed to enter a total of 30 words for the

current keyboard input device. None of the experiment words were duplicated between themselves or the words used during practice and again were pulled at random from the Oxford English Dictionary with lengths between 3 and 6 characters, filtering for swear words. Participants were not allowed to recalibrate the interaction space during the task.

After all tasks were completed for each of the 4 input devices, the participants were asked to fill out an exit survey TODO: SEE APPENDIX. The exit survey asked the participants for their age, gender, major, and handedness as well as several questions detailing any prior touch, gesture-controller, or word-gesturing experience or impairments that might relate to the study. Finally the participants were required to fill out the Likert scale relating to difficulty, discomfort and fatigue experienced when using the devices as well as rank each device on a numerical scale from best to worst.

Table 4.2. Schedule of Assessments for a single study visit (in minutes).

	Controller	Leap Motion Surface	Leap Motion Static-Air	Leap Motion Pinch-Air	Exit Survey	total
explain	.5	.5	.5	.5	0	2
calibrate	0	2	2	2	0	6
practice	5	5	5	5	0	20
task	3	3	3	3	0	12
survey	0	0	0	0	5	5
total	8.5	10.5	10.5	10.5	5	45

4.1.6 *Dependent Measures*

The pre-pilot study only collected the playback data of participants. The playback data included detected key presses, the calibrated interaction plane, and the tracking location data.

4.2 Pilot Study

The pilot study expanded on what was learned from the pre-pilot. Additional Mid-Air keyboard interactions were added and the displayed virtual keyboard was redesigned to be simpler and remove obtrusive features.

4.2.1 Participants

A sample size of 7 was used for the pilot study. There were 3 male and 4 female participants, ages ranging from 21 to 24 with a median age of 22. Participants' computer usage ranged from 6 to more than 50 hours per week with a median usage of 21 to 30 hours per week. All of the participants described their right hand as being dominant and all participants used their right hand during the experiment except for one participant who switched back and forth. All of the participants had previous experience with touch devices, whereas all but one participant had previous experience with gesture-controllers and only 57% having previous experience with word-gesture keyboards. No participants had any impairment that affected their ability to enter text with computers. TODO: LINK TO THE TABLE BELOW FOR MORE DETAILS

Table 4.3. Participant information including age, gender, handedness, computer usage, and previous experiences.

Subject	Gender	Age	Computer Usage per Week (hours)	Handedness	Hand Used in Experiment	Touch-device Experience	Gesture-controller Experience	Word-gesturing Experience	Impairment History
1	male	21	21 - 30	right	right	yes	yes	no	no
2	male	24	41 - 50	right	right	yes	yes	yes	no
3	female	22	50+	right	right	yes	yes	yes	no
4	female	23	50+	right	right	yes	yes	yes	no
5	female	21	6 - 10	right	both	yes	yes	no	no
6	female	24	6 - 10	right	right	yes	no	no	no
7	male	21	21 - 30	right	right	yes	yes	yes	no

4.2.2 *Input Devices and Interaction*

The pilot study saw the introduction of three additional keyboard inputs, the Touch Screen Keyboard, the Leap Motion Bimodal-Air Keyboard, and the Leap Motion Predictive-Air Keyboard. The Touch Screen Keyboard was added because it is the de facto interaction for modern word-gesture keyboards TODO: FIND A REFERENCE TO THIS. The other two keyboards were added as alternative implementations to mid-air, word-gesture keyboards.

As in the pre-pilot, interaction with the different keyboard devices were dependent on whatever device was active at the time. All of the keyboards except for the Touch Screen Keyboard were simulated on the same 64-bit, Windows 7 work station as before. The Touch Screen Keyboard was simulated on the Ideum Multi-touch Table Platform tabletop, running 64-bit Windows 8. Again, all receivers or controllers were connected through USB 2.0. The participants were again allowed to recalibrate the active keyboard’s interaction plane. Again, participants were encouraged to reposition the gesture-controller and were given the option to use either hand and rest or raise their arms during the experiment.

4.2.2.1 Touch Screen Keyboard. The Touch Screen Keyboard was used on a large tabletop touch screen. The participant then used their finger to interact with the virtual keyboard on the screen in the same way as typical touch devices. Touch was simulated when the participants finger touched the screen and release was simulated when the finger was lifted from the surface TODO: Provide a picture.

4.2.2.2 Leap Motion Bimodal-Air Keyboard. The Leap Motion Bimodal-Air Keyboard used a Leap Motion Controller which was placed on the desk in front of the participant. The participant then used a stylus which was tracked by the Leap Motion in order to determine the location over the projected virtual keyboard. A touch was simulated by pressing the space bar key on a standard QWERTY keyboard and a

touch release was simulated upon the release of the 'Space Bar.' The interaction plane could be calibrated at any time prior to the experiment TODO: Provide picture.

4.2.2.3 Leap Motion Predictive-Air Keyboard. The Leap Motion Predictive-Air Keyboard used a Leap Motion Controller which was placed on the desk in front of the participant. The participant then used a stylus which was tracked by the Leap Motion in order to interact with a projected interaction plane. A touch was simulated by recognizing and predicting a forward gesture of the stylus toward the interaction plane and a release was simulated by recognizing a backward gesture away from the interaction plane. As before, the interaction plane could be calibrated at any time prior to the experiment TODO: Provide picture.

4.2.3 Task Design

As in the pre-pilot, the conditions of the task were represented by the 7 different keyboard input devices. The 7 conditions used were the Leap Motion Static-Air, Leap Motion Pinch-Air, Leap Motion Surface, and Xbox Controller keyboards as before, with the addition of the Leap Motion Predictive-Air, Leap Motion Bimodal-Air, and Touch Screen keyboards. TODO: See section Word GESTURE KEYBOARDS for specific details of the new keyboard implementations.

Task profiles were created for each of the 7 keyboard input devices. Each task profile consisted of 10 separate trials for a total of 70 trials per participant. The reduction in trials from 30 words to 10 words for each device was due to a complaint of fatigue during the pre-pilot study, one of the participants was unable to finish. The addition of task profiles were in an attempt to standardize the data collected rather than using random, changing words between uses of the same keyboard. Instead of choosing 10 words at random for each and every keyboard and participant, the task profiles insured that the same 10 words were used across each unique keyboard device for all participants. This was handled by generating static, unchanging dictionaries

for each keyboard, guaranteeing a total of 70 unique words rather than a total of 490 unique words for the 7 participants. The 10 words selected for each dictionary were generated by a custom dissimilarity algorithm that produced the top 10 least dissimilar gesture-shapes across all words in the Oxford English Dictionary for words of length 3 through 6 characters. This meant that only 10 different gesture-shapes were used by each participant across all input devices, ensuring that all participants' experiences with each keyboards are as similar as possible to each other and other participants TODO: DICTIONARY ALG.

For each trial, a word was chosen at random from the active keyboard's previously constructed dictionary and displayed on the screen. A blank text-area was positioned directly below the displayed word for the participants' transcribed text TODO: GIVE FIGURE FOR BLANK WORD DISPLAYED ON THE SCREEN. Beneath both text areas, the virtual representation of the keyboard that was previously displayed was updated and simplified TODO: GIVE PICTURE FOR DISPLAYED KEYBOARD. The shift, enter, and number keys were all removed, and the backspace key readjusted. The participants were then required to use the currently active keyboard input device to enter the displayed word using word-gesturing as before TODO: GIVE FIGURE SHOWING WORD-GESTURING. During the word-gesturing process, participants were still shown real-time updates to the displayed word and transcribed text as well as their movements within the virtual keyboard space TODO: REFERENCE A FIGURE HERE FOR COLOR OF TEXT. The participants were required to use the active keyboard device's backspace key to remove errors, however already correct transcribed characters were protected from being deleted. The change to protect the correctly transcribed characters was because of the high sensitivity and precision required to only delete the erroneous characters. Once a word was correctly entered, the participants were to release the simulated touch by the

appropriate means of the active keyboard to move to the next word instead of hitting the enter key.

As before, deviations in the above task were required for how the participants interacted with the Xbox Controller keyboard. For this keyboard, there was still no word-gesturing feature implemented, instead the participant was asked to use single-input text entry using a standard Xbox 360 Controller. The participants had to press the 'A' button to select a letter, 'Y' button to delete, and the start button to move to the next word.

4.2.4 Experimental Design

A within-subjects design was again used for this study. The strength of a within-subjects design is that the overall power will increase and there will be a reduction in error variance associated with individual differences. The weakness of using the within-subjects design is that it suffers from carryover effects between each keyboard input device. The participation in one condition may affect performance in other conditions. To account for this weakness, the study was supplemented with a Latin Squares design for counterbalancing. Table 4.4 shows how the Latin Squares design was utilized for a sample size of 7 with an equal number of different keyboard inputs. TODO: NEED REFERENCE FOR WITHIN SUBJECT AND LATIN SQUARES.

Table 4.4. Latin Squares design for 7 participants and 7 conditions.

subject	conditions						
1	A	B	C	D	E	F	G
2	B	C	D	E	F	G	A
3	C	D	E	F	G	A	B
4	D	E	F	G	A	B	C
5	E	F	G	A	B	C	D
6	F	G	A	B	C	D	E
7	G	A	B	C	D	E	F

4.2.5 Procedure

Again, each subject participated in a single study visit which took between 30 and 70 minutes to complete depending on how many calibrations were performed. The full set of tasks, and their expected durations, are detailed in Figure 4.5, the pilot study schedule of assessments. The participants followed the same process for each of the 7 keyboard input devices' tasks.

First, the participants were given a brief explanation and demonstration of the active keyboard input. The explanation dialog contained the name of the active keyboard, how it was interacted with, and whether or not it was a word-gesture keyboard. The researcher then demonstrated how to enter the word "test" using the active keyboard. The participants were then given control and handed the stylus or used their bare hand to interact with the keyboard to get a sense of how it worked.

Participants were then instructed to use the keyboard to perform practice words which were randomly selected from the Oxford English Dictionary with lengths between 3 and 6 characters long. The practice words were filtered for offensive words and against the previously constructed experiment dictionaries so that no participants would be able to see any of the experiment words in advance. There was no limit placed on how many words could be performed while practicing. The participants were told to continue until they felt they were able to efficiently and comfortably type each word with minimal errors. During the practice phase, the opportunity to recalibrate the interaction-space any number of times was given if it was applicable to the active keyboard.

Next, the participants performed the task itself. As detailed in Section 4.2.3, the participants were instructed to enter a total of 10 words for the current keyboard input device. The words selected were pulled at random from the active keyboard's previously constructed dictionary until all 10 words in the dictionary were used. Participants were not allowed to recalibrate the interaction space during the task.

After all tasks were completed for each of the 7 input devices, the participants were asked to fill out an exit survey TODO: SEE APPENDIX. The exit survey asked the participants for their age, gender, major, and handedness as well as several questions detailing any prior touch, gesture-controller, or word-gesturing experience or impairments that might relate to the study. Finally the participants were required to fill out the a Likert scale section relating to difficulty, discomfort and fatigue experienced when using the devices as well as rank each device on a numerical scale from best to worst.

Table 4.5. Schedule of Assessments for a single study visit (in minutes).

	Controller	Touch Screen	Leap Motion Surface	Leap Motion Static-Air	Leap Motion Pinch-Air	Leap Motion Predictive-Air	Leap Motion Bimodal-Air	Exit Survey	total
explain	.5	.5	.5	.5	.5	.5	.5	0	3.5
calibrate	0	0	2	2	2	2	2	0	10
practice	5	5	5	5	5	5	5	0	35
task	2	2	2	2	2	2	2	0	14
survey	0	0	0	0	0	0	0	5	5
total	7.5	7.5	9.5	9.5	9.5	9.5	9.5	5	67.5

4.2.6 Dependent Measures

The design choice to not fully realize word-recognition, for the implemented word-gesture keyboard, heavily influenced the design of the task and therefore also affected the dependent measures. Each individual trial was designed as a single word rather than a phrase that included many words, as in Vulture TODO: REFERENCE VULTURE, therefore most dependent measures were analyzed on the word-gesture level. It’s possible to analyze these values at the phrase level if every the combined trials are viewed as a single phrase.

4.2.6.1 Text-Entry Rate. Typically, text-entry rates are calculated using the standard Words Per Minute formula

$$WPM = \frac{|T - 1|}{S} \times 60 \times \frac{1}{5} \quad (4.1)$$

where $|T - 1|$ is the length of the transcribed string and S is the amount of time, in seconds, that was taken to transcribe the word from the time the first character was produced TODO: GET VULTURE-3 REFERENCE. When dealing with timing word-gestures, however, the formula must be modified to:

$$WPM = \frac{|T|}{S} \times 60 \times \frac{1}{5} \quad (4.2)$$

where $|T - 1|$ is replaced with $|T|$ and S represents the amount of time, in seconds, that was taken to transcribe the word including the time that it took to produce the first character. The modification is required because the time it takes to produce the first character must be included when timing word-gestures TODO: VULTURE-17.

4.2.6.2 Error Rates. There were several techniques used to measure error rates and find the best representation of keyboard performance and account for the task design.

The first error rate was modeled after the techniques in Vulture TODO: REFER VULTURE and uses the Minimum Word Distance which was calculated in the same way was the Minimum String Distance TODO: VULTURE-15, MACKENZIE 2003:

$$MWD \text{ error rate} = \frac{MWD(P, T)}{\overline{S_P}} \times 100\% \quad (4.3)$$

Minimum Word Distance differentiated from Minimum String Distance in that it was calculated on a per-word level than on a per-character level where P and T were the sets of words in the presented and transcribed strings, and $\overline{S_P}$ was the mean size of the optimal alignments calculated on a per-word level TODO: REFERENCE VULTURE. It is important to note that because participants were forced to correctly type in words and there were no errors present in the final transcribed text, words that were considered erroneous had to be defined differently. For this study to use Minimum Word Distance, if the participant made any errors at all regardless of being

forced to correct them, then the word was counted as incorrect. This gives the formula

$$MWD \text{ error rate} = \frac{IW}{CW + IW} \times 100\% \quad (4.4)$$

where IW are words where the participant made any mistakes at all regardless of corrections and CW are words where the participant got the word correct with one attempt.

The next error rate used was the Keystrokes Per Character method, otherwise known as KSPC TODO: MACKENZIE 2003. The Keystrokes Per Character formula

$$KSPC \approx \frac{C + INF + IF + F}{C + INF} \quad (4.5)$$

used Soukoreff and MacKenzie’s keystroke taxonomy, where C represented the correct characters in the transcribed text, INF were the incorrect, not fixed characters in the transcribed text, F was used to show the keystrokes which were editing functions, namely backspace, and IF were the incorrect but fixed characters in the input stream, but not editing keys. The Keystrokes Per Character method was less than ideal and has many limitations as an error metric TODO: REF MACKENZIE 2003, yet it was still beneficial since the task required each word to be typed correctly before moving on to the next and the word-gesture keyboards were designed with a lack of true word-recognition TODO: REFER TO WGK IMPLEMENTATION. It’s important to note that INF always equated to zero, reducing Formula 4.5 to

$$KSPC \approx \frac{C + IF + F}{C} \quad (4.6)$$

because participants were required to correctly transcribe each word.

The final error rate used was the Total Error Rate TODO: MACKENZIE 2003. The Total Error Rate was also described by the previous keystroke taxonomy used in

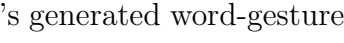
KSPC, giving the formula

$$Total\ Error\ Rate = \frac{INF + IF}{C + INF + F} \times 100\% \quad (4.7)$$

where, again, C represented the correct characters in the transcribed text, INF were the incorrect, not fixed characters in the transcribed text, F was used to show the keystrokes which were editing functions, namely backspace, and IF were the incorrect but fixed characters in the input stream, but not editing keys. Again, Formula 4.7 can be reduced to

$$Total\ Error\ Rate = \frac{IF}{C + F} \times 100\% \quad (4.8)$$

because participants were required to correctly transcribe words.

4.2.6.3 Correctness. Correctness of a single word-gesture was determined by calculating the Fréchet Distance between the expected word-gesture path and the participant's generated word-gesture path. Figure  **TODO: PUT A PICTURE HERE OF THE** shows both the expected and user-generated paths of the word 'TODO: CHOOSE WORD'. The Fréchet Distance between two curves P and Q , or in this case gesture-shapes, is defined as the minimum length leash needed to walk a dog when the person walks along P and the dog walks along Q (Har-Peled and Raichel 2014). Figure 4.1 shows the recursive implementation of Fréchet Distance used in this study where P and Q were the two paths being walked, CA was the matrix that contains all possible distance values for each comparison, and i and j are the indices that were being examined for that particular recursive phase.

4.2.6.4 Distance Measures. Distance measures were used to evaluate the movements of the participants' hands in the interaction plane. The two primary distance measures were the distance traveled to complete a word's gesture-shape,

```

private float frechetRecursive(Vector [] P, Vector [] Q,
    Float [][] CA, int i, int j) {
    float CAij = 0;
    if(CA[i][j] > -1) {
        CAij = CA[i][j];
    } else if(i == 0 && j == 0) {
        CA[i][j] = distance(P[0], Q[0]);
        CAij = CA[i][j];
    } else if(i > 0 && j == 0) {
        CA[i][j] = Math.max(frechetRecursive(P, Q, CA, i - 1, 0),
            distance(P[i], Q[0]));
        CAij = CA[i][j];
    } else if(i == 0 && j > 0) {
        CA[i][j] = Math.max(frechetRecursive(P, Q, CA, 0, j - 1),
            distance(P[0], Q[j]));
        CAij = CA[i][j];
    } else if(i > 0 && j > 0) {
        float min = Math.min(frechetRecursive(P, Q, CA, i - 1, j),
            frechetRecursive(P, Q, CA, i - 1, j - 1));
        min = Math.min(min, frechetRecursive(P, Q, CA, i, j - 1));
        CA[i][j] = Math.max(min, distance(P[i], Q[j]));
        CAij = CA[i][j];
    } else {
        CA[i][j] = Float.POSITIVE_INFINITY;
    }
    return CAij;
}

```

Figure 4.1. A snippet of code showing the recursive implementation of Fréchet distance.

recorded in centimeters, and the average velocity of the participant's hand recorded in centimeters per second.

4.2.6.5 Timing Measures. Timing measures were used to calculate text-entry rates as well as attempt to evaluate the participant's level of focus. The primary time measure taken was the duration required to complete a word's gesture-shape in seconds. The time measure attempting to determine a participant's level of focus was by recording the reaction time to respond to errors. Finally, the duration it took

for participants to first simulate a touch and to correctly enter the first letter were recorded.

4.2.6.6 Quantitative Measures. There were two quantitative measures recorded, the number of practice words completed for each input per participant, and the number of times a touch was simulated for each subject per input.

The number of practice words the participant completed aims to rate the complexity of each keyboard, comparing it to the recorded qualitative measures. Tracking the number of times a touch was simulated should be linked to error rates. This helped to try to determine if there were detection errors with the device itself rather than errors generated by the participants.

4.2.6.7 Qualitative Measures. The qualitative measures in this study were recorded by utilizing an exit survey once the task for all of the keyboards had been completed. The participants were asked to rate each keyboard that they used in terms of discomfort, difficulty, and fatigue using a Likert scale with 5 options. Discomfort was defined as the awkwardness of the keyboard and whether it required an uncomfortable position or gesture to use. Difficulty evaluated whether the keyboard input device was confusing to understand how to use, and fatigue asked the participant if they had experienced tiredness, soreness, or fatigue from the keyboard they had just used. Lastly, participants ranked the keyboards from 1, most preferred, to 7, the least preferred. TODO: SEE APPENDIX FOR EXIT SURVEY.

4.3 Final Study

In the final study, the Xbox Controller Keyboard was removed due to its irrelevance, see Section TODO: XBOX WGK FUTURE WORK SECTION for more information on how gaming consoles' virtual keyboards can be improved by moving away from single-input text entry and instead implementing a word-gesture keyboard. In addition, the use of a stylus was removed as an interaction tool from all mid-air

keyboards, allowing participants to interact barehanded with the mid-air keyboard inputs.

4.3.1 Participants

A sample size of 18 was used in the final study. The justification for this sample size comes from the formula to calculate the sample size for two independent group means using a pooled standard deviation: TODO: GET REFERENCE FOR THIS FORMULA

$$N = \frac{2(z_{\frac{\alpha}{2}} + z_{1-\beta})^2}{\left(\frac{\mu_1 - \mu_2}{\sigma_{pooled}}\right)^2} \quad (4.9)$$

where $z_{\frac{\alpha}{2}}$ and $z_{1-\beta}$ were the z-scores for the α and β , respectively, μ_1 and μ_2 were the means of the two populations being compared and σ_{pooled} is the pooled standard deviation of the two populations. A power of $1 - \beta = 0.80$ and a significance level of $\alpha = 0.05$ were used when calculating the sample size. The derived sample size was the average sample size for all relevant variable comparisons based on the study objectives. Outliers requiring a sample size greater than 100 were removed. Furthermore, a sample size of 18 justifies a Replicated Latin Squares design for 6 input methods. The Latin Squares design was chosen for counterbalancing the experimental design and to reduce the effect of participation in one condition affecting performance of other conditions. Further details are explained in Section 4.3.4.

There were 13 male and 5 female participants, ages ranging from 18 to 24 with a median age of 21. Participants' computer usage ranged from 1 to greater than 50 hours per week with a median usage between 31 to 50 hours per week. All but two of the participants described their right hand as being dominant with one participant describing their left hand and the other claiming to be ambidextrous. Correspondingly, all participants used their right hand during the experiment except for one participant who used their left hand and another who switched back and forth. All of the participants had previous experience with touch devices, whereas

83% of participants had previous experience with gesture-controllers and only 56% having previous experience with word-gesture keyboards. No participants had any impairment that affected their ability to enter text with computers. Refer to Table 4.6 for more specific details of each participant.

Table 4.6. Participant information including age, gender, handedness, computer usage, and previous experiences.

Subject	Gender	Age	Computer Usage per Week (hours)	Handedness	Hand Used in Experiment	Touch-device Experience	Gesture-controller Experience	Word-gesturing Experience	Impairment History
1	female	20	41 - 50	right	right	yes	yes	no	no
2	male	24	31 - 40	right	right	yes	yes	yes	no
3	male	19	1 - 5	both	both	yes	yes	no	no
4	male	23	41 - 50	right	right	yes	yes	yes	no
5	male	21	21 - 30	right	both	yes	yes	no	no
6	female	22	50+	right	right	yes	yes	yes	no
7	male	18	41 - 50	right	right	yes	yes	yes	no
8	female	21	11 - 20	right	right	yes	no	yes	no
9	male	22	50+	left	left	yes	yes	yes	no
10	male	18	21 - 30	right	right	yes	yes	yes	no
11	female	21	50+	right	right	yes	no	no	no
12	male	22	11 - 20	right	right	yes	yes	yes	no
13	male	22	50+	right	right	yes	yes	no	no
14	female	18	50+	right	right	yes	yes	yes	no
15	male	23	50+	right	right	yes	yes	no	no
16	male	18	6 - 10	right	right	yes	no	yes	no
17	male	19	11 - 20	right	right	yes	yes	no	no
18	male	18	31 - 40	right	right	yes	yes	no	no

4.3.2 Input Devices and Interaction

In order to focus the study on only word-gesture keyboards, the final study saw the removal of the Xbox Controller Keyboard. Again, all of the keyboards except for the Touch Screen Keyboard were simulated on the same 64-bit, Windows 7 work station as before. The Touch Screen Keyboard was simulated on the Ideum Multi-touch Table Platform tabletop, running 64-bit Windows 8. All receivers or controllers were connected through USB 2.0. The participants were again allowed to recalibrate the active keyboard’s interaction plane, however a default interaction space was provided that was unlikely to need to be recalibrated. Participants were encouraged to use the default-calibrated interaction space. Again, the participants were allowed to reposition the gesture-controller and were told to use the keyboards in whatever way they felt they could perform best.

A major change to all of the mid-air keyboards is that the stylus was removed. The absence of the stylus was aimed to remove the barrier between the participant and the keyboard input. The goal was to make it feel more natural and more like a touch screen.

4.3.2.1 Touch Screen Keyboard. The Touch Screen Keyboard was used on a large tabletop touch screen. The participant then used their finger to interact with the virtual keyboard on the screen in the same way as typical touch devices. Touch was simulated when the participants finger touched the screen and release was simulated when the finger was lifted from the surface TODO: Provide a picture.

4.3.2.2 Leap Motion Surface Keyboard. Again, for the Leap Motion Surface Keyboard a Leap Motion Controller was used for tracking. Unlike the other Leap-based keyboards, the gesture controller was placed into a custom designed holder instead of on the desk in front of the participant. The holder was attached to an inclined surface with a printed keyboard fixed on top TODO: for more details see section BUILDING SURFACE WGK. Placement, this time, was insured to be similar between uses, and therefore the Leap Motion Surface Keyboard required only a single calibration for all participants. The participant then used a stylus, as before, that was tracked by the Leap Motion in order to detect interaction. A touch was simulated by pressing the tip of the stylus against the printed keyboard and a release was simulated when the tip of the stylus was again removed from the printed keyboard surface.

4.3.2.3 Leap Motion Static-Air Keyboard. The Leap Motion Static-Air Keyboard used a Leap Motion Controller which was placed on the desk in front of the participant. The participant then used the pointer finger of their dominant hand, which was tracked by the Leap Motion, to interact with a projected interaction plane. A touch was simulated by the insertion of their finger into the interaction plane and a release was simulated upon the removal of their finger. TODO: Provide picture.

4.3.2.4 Leap Motion Pinch-Air Keyboard. The Leap Motion Pinch-Air Keyboard also used a Leap Motion Controller that was positioned on the desk in front of the participant. The participant then used their bare hand, which was tracked at the center of their palm, to interact with the projected interaction plane. Touch was simulated by having the participant make a pinching gesture with their hand and a release was simulated when the participant released the pinch, opening their hand again. TODO: Provide picture. TODO: ALSO MAYBE REFERENCE VULTURE HERE or back in chpt

4.3.2.5 Leap Motion Bimodal-Air Keyboard. The Leap Motion Bimodal-Air Keyboard used a Leap Motion Controller positioned on the desk in front of the participant as before. The participant then used the pointer finger of their dominant hand, which was tracked by the Leap Motion, to determine the location over the projected virtual keyboard. A touch was simulated by pressing the space bar key on a standard QWERTY keyboard and a touch release was simulated upon the release of the 'Space Bar.' TODO: Provide picture.

4.3.2.6 Leap Motion Predictive-Air Keyboard. As the others, the Leap Motion Predictive-Air Keyboard saw a Leap Motion Controller placed on the desk in front of the participant. The participant then used the pointer finger of their dominant hand, which was tracked by the Leap Motion, to interact with a projected interaction plane. A touch was simulated by recognizing and predicting a forward gesture of the participant's finger toward the interaction plane and a release was simulated by recognizing a backward gesture away from the interaction plane. TODO: Provide picture.

4.3.3 Task Design

As in the pilot, the conditions of the task were represented by the 6 different keyboard input devices. The 6 conditions used were the Leap Motion Static-Air, Leap

Motion Pinch-Air, Leap Motion Surface, Leap Motion Predictive-Air, Leap Motion Bimodal-Air, and Touch Screen keyboards.

Task profiles were created for each of the 6 keyboard input devices. Each task profile consisted of 15 separate trials for a total of 90 trials per participant. The increase in trials from 10 words to 15 words for each device was due to the removal of the Xbox Controller Keyboard. The creation of the task profiles was handled by generating static, unchanging dictionaries for each keyboard, guaranteeing a total of 90 unique words across all participants. The 15 words selected for each dictionary were again generated by the same custom dissimilarity algorithm, producing the top 15 least dissimilar gesture-shapes across all words in the Oxford English Dictionary for words of length 3 through 6 characters. TODO: DICTIONARY ALG.

For each trial, a word was chosen at random from the active keyboard’s previously constructed dictionary and displayed on the screen. A blank text-area was positioned directly below the displayed word for the participants’ transcribed text TODO: GIVE FIGURE FOR BLANK WORD DISPLAYED ON THE SCREEN. Beneath both text areas, the keyboard input devices were virtually represented. The participants were then required to use the currently active keyboard input device to enter the displayed word using word-gesturing TODO: GIVE FIGURE SHOWING WORD-GESTURING. During the word-gesturing process, participants were again shown real-time updates TODO: REFERENCE A FIGURE HERE FOR COLOR OF TEXT. The participants were required to use the active keyboard device’s backspace key to remove errors, however already correct transcribed characters were still protected from being deleted. Once a word was correctly entered, the participants were to release the simulated touch to move to the next word.

4.3.4 *Experimental Design*

As justified in the pilot, a within-subjects design was used for the final study. To minimize carryover effects, the study was supplemented with a Replicated Latin

Squares design for counterbalancing. Table 4.7 shows how the Replicated Latin Squares design was utilized for 6 different keyboard inputs with a sample size of 18.

Table 4.7. The three replications requires for a Replicated Latin Squares design for 18 participants and 6 conditions.

First Replication						
participants	conditions					
1	A	B	C	D	E	F
2	B	C	D	E	F	A
3	C	D	E	F	A	B
4	D	E	F	A	B	C
5	E	F	A	B	C	D
6	F	A	B	C	D	E

Second Replication						
participants	conditions					
7	F	A	B	C	D	E
8	A	B	C	D	E	F
9	B	C	D	E	F	A
10	C	D	E	F	A	B
11	D	E	F	A	B	C
12	E	F	A	B	C	D

Third Replication						
participants	conditions					
13	E	F	A	B	C	D
14	F	A	B	C	D	E
15	A	B	C	D	E	F
16	B	C	D	E	F	A
17	C	D	E	F	A	B
18	D	E	F	A	B	C

4.3.5 Procedure

Each subject participated in a single study visit which took between 30 and 45 minutes to complete. For this study, almost no calibrations were performed. The full set of tasks, and their expected durations, are detailed in Table 4.8, the final study

schedule of assessments. The participants followed the same process for each of the 6 keyboard input devices' tasks.

First, the participants were given a brief explanation and demonstration of the active keyboard input. The explanation dialog contained the name of the active keyboard, how it was interacted with. The researcher then demonstrated how to enter the word "test" using the active keyboard. The participants were then given control and interacted with the keyboard to get a sense of how it worked.

Participants were then instructed to use the keyboard to perform practice words which were randomly selected from the Oxford English Dictionary with lengths between 3 and 6 characters long. The practice words were again filtered for offensive words and against the previously constructed experiment dictionaries. There was no limit placed on how many words could be performed while practicing. The participants were told to continue until they felt they were able to efficiently and comfortably type each word with minimal errors. The participants were told that the keyboards could be recalibrated if absolutely necessary but were encouraged to learn to use the default calibration. This change was brought about because in the pilot study, participants had a hard time finding calibrations that worked, sometimes taking upwards of 30 minutes with poor results.

Next, the participants performed the task itself. As detailed in Section 4.3.3, the participants were instructed to enter a total of 15 words for the current keyboard input device. The words selected were pulled at random from the active keyboard's previously constructed dictionary until all 15 words in the dictionary were used.

After the task for the active keyboard was completed, the participants were asked to fill out a small survey section. The survey asked participants to use the Likert scale to rate each keyboard in terms of difficulty, discomfort and fatigue experienced when using the devices. TODO: SEE APPENDIX FOR INTERMITTENT SURVEY OR FIGURE

Finally, after all tasks were completed for each of the 6 input devices, the participants were asked to fill out an exit survey TODO: SEE APPENDIX. The exit survey, as in the pilot, asked the participants for their age, gender, major, and handedness as well as several questions detailing any prior touch, gesture-controller, or word-gesturing experience or impairments that might relate to the study. Lastly, the participants were asked to rank each device on a numerical scale from best to worst.

Table 4.8. Schedule of Assessments for a single study visit (in minutes).

	Touch Screen	Leap Motion Surface	Leap Motion Static-Air	Leap Motion Pinch-Air	Leap Motion Predictive-Air	Leap Motion Bimodal-Air	Exit Survey	total
explain	.5	.5	.5	.5	.5	.5	0	3
calibrate	0	0	.5	0	.5	.5	0	1.5
practice	3	3	3	3	3	3	0	18
task	1.5	1.5	1.5	1.5	1.5	1.5	0	9
survey	.5	.5	.5	.5	.5	.5	3	6
total	5.5	5.5	6	5.5	6	6	3	37.5

4.3.6 *Dependent Measures*

Again because each individual trial was designed as a single word rather than a phrase that included many words, most dependent measures were analyzed on the word-gesture level. It's possible to analyze these values at the phrase level if every the combined trials are viewed as a single phrase. In addition, due to the lack of word-recognition in the design of the word-gesture keyboard, the forcing of participants to make corrections to transcribed words, and to help accommodate for device detection errors that were out of the participants' control, some of the dependent measures have had modified forms added. The modifications that were used directly effect the data processing for the transcribed word.

The first of the modifications was by using the shortest form of the transcribed word, Table 4.9 shows how this worked. The shortest-transcribed modification helps to account for device detection errors as well as the fact that participants were forced

to correct words. The drawback, however, can be seen in Example 4 of Table 4.9, the participant’s intention seemed to be the word ”fired” instead of ”fire” but this information was lost with the shortest-transcribed modification.

Table 4.9. Examples of the shortest-transcribed modification.

	Presented text:	Input Stream:	Transcribed text:
Example 1	quick	wquickl←←←←←←←quick	wquickl
Example 2	dot	fdot←←←←di←ot	fdot
Example 3	burn	burnm←	burn
Example 4	fire	fuired←←←←←ired←	fuire

The second modification was by including backspaces as part of the presented word when a participant made mistakes, since participants were required to correct errors. With the modification to the presented text, the transcribed string was now represented by the entire input stream. Table 4.10 shows how this worked. The main motivation behind this modification was to mirror the participants’ requirement to use backspaces until a word was correctly entered, especially for the correctness measure, Fréchet Distance.

Table 4.10. Examples of the backspace-transcribed modification.

	Presented text:	Input Stream:	Modified Presented text:	Transcribed text:
Example 1	quick	wquickl←←←←←←←quick	←←←←←←←quick	wquickl←←←←←←←quick
Example 2	dot	fdot←←←←di←ot	←←←←←d←ot	fdot←←←←di←ot
Example 3	burn	burnm←	burn←	burnm←
Example 4	fire	fuired←←←←←ired←	f←←←←←ire←	fuired←←←←←ired←

4.3.6.1 Text-Entry Rate. Again, text-entry rates are calculated using modified Words Per Minute, Formula 4.2, where $|T-1|$ is replaced with $|T|$ and S represents the amount of time, in seconds, that was taken to transcribe the word including the time that it took to produce the first character. The text-entry rate was calculated with and without the shortest-transcribed modification.

4.3.6.2 *Error Rates.* There were several techniques used to measure error rates and find the best representation of keyboard performance and account for the task design.

The first error rate, again, uses modified Minimum Word Distance, Formula 4.4, where IW are words where the participant made any mistakes at all regardless of corrections and CW are words where the participant got the word correct with one attempt. The shortest-transcribed modification, though, allows for the original Minimum Word Distance formula, Formula 4.3 from Vulture TODO: VULTURE REF, where P and T were the sets of words in the presented and transcribed strings, and $\overline{S_P}$ was the mean size of the optimal alignments calculated on a per-word level.

Next, due to the addition of the shortest-transcribed modification and because of the lack of word-recognition implementation of the word-gesture keyboards, the Minimum String Distance error rate was able to be used TODO: REF MAKENZIE 2003. Using the simplified keystroke taxonomy that was presented before, Minimum String Distance can be defined as the formula

$$MSD \text{ error rate} = \frac{INF}{C + INF} \times 100\% \quad (4.10)$$

where C represented the correct characters in the transcribed text and INF were the incorrect, not fixed characters in the transcribed text.

As before, the simplified Keystrokes Per Character formula, Formula 4.6, was used where C represented the correct characters in the transcribed text, INF were the incorrect, not fixed characters in the transcribed text, F was used to show the keystrokes which were editing functions, namely backspace, and IF were the incorrect but fixed characters in the input stream, but not editing keys. Additionally, Formula 4.5 was used with the shortest-transcribed modification.

TODO: MENTION THAT TOTAL ERROR RATE SIMPLIFIED IS ACTUALLY ALMOST EXACTLY LIKE THE CORRECTED ERROR RATE The final

error rate, again, was the simplified Total Error Rate from the Formula 4.8 where, again, C represented the correct characters in the transcribed text, INF were the incorrect, not fixed characters in the transcribed text, F was used to show the keystrokes which were editing functions, namely backspace, and IF were the incorrect but fixed characters in the input stream, but not editing keys. With the edition of the shortest-transcribed modification, Formula 4.7 was also utilized.

4.3.6.3 Correctness. As in the pilot study, correctness of a single word-gesture was determined by calculating the Fréchet Distance between the expected word-gesture path and the participant's generated word-gesture path. Figure 4.1 shows the recursive implementation of Fréchet Distance used in this study where P and Q were the two paths being walked, CA was the matrix that contains all possible distance values for each comparison, and i and j are the indices that were being examined for that particular recursive phase. The Fréchet Distance was also calculated for both the shortest-transcribed and backspace-transcribed modifications.

4.3.6.4 Distance Measures. The two primary distance measures were the distance traveled to complete a word's gesture-shape, recorded in centimeters, and the average velocity of the participant's hand recorded in centimeters per second.

4.3.6.5 Timing Measures. As before in the pilot, the primary time measure taken was the duration required to complete a word's gesture-shape in seconds. Additionally, the participant's reaction time to errors, the duration it took for participants to first simulate a touch, and the time it took to correctly enter the first letter were recorded.

4.3.6.6 Quantitative Measures. Again, there were two quantitative measures recorded, the number of practice words completed for each input per participant, and the number of times a touch was simulated for each subject per input.

4.3.6.7 *Qualitative Measures.* The qualitative measures, in the final study, were gathered from the intermittent surveys and the final exit survey. In the intermittent surveys, the participants were asked to rate each keyboard that they used in terms of discomfort, difficulty, and fatigue using a Likert scale with 5 options. In the final exit survey, participants ranked the keyboards from 1, most preferred, to 6, the least preferred. TODO: SEE APPENDIX FOR EXIT SURVEY.

CHAPTER FIVE

Results and Analysis

The statistical methods used consisted of One-Way ANOVAs for each set of the dependent measures, followed by using Tukey’s Honest Significant Difference (HSD) for multiple-compare for the post-hoc analysis. The ranking system from the exit survey used the Friedman’s test for analysis in conjunction with Tukey’s HSD for a post-hoc analysis.

In addition, an N-Way ANOVA was used on each set of variables for each keyboard with the participants’ previous word-gesture experience as a factor TODO: USE MATLAB REFERENCES FOR TUKEYS AND ANOVAS. Surprisingly, there were no significant effects with word-gesture experience as a factor, meaning, participants who had word-gesture experience performed about the same as participants who didn’t.

5.1 *Text-Entry Rate*

Figure 5.1 shows the mean text-entry rates for each keyboard input device. Participants reached a mean text-entry rate of 19.5 WPM ($SD = 3.0$) for the Touch Screen Keyboard, 17.1 WPM ($SD = 3.3$) for the Leap Surface Keyboard, 8.6 WPM ($SD = 1.9$) for the Leap Static-Air Keyboard, 11.3 WPM ($SD = 2.0$) for the Leap Pinch-Air Keyboard, 9.6 WPM ($SD = 2.3$) for the Predictive-Air Keyboard, and 15.8 WPM ($SD = 2.2$) for the Leap Bimodal-Air Keyboard. It should be noted that the participant with the highest Bimodal-Air text-entry rate, 20.7 WPM, was on par with the Touch Screen Keyboard text-entry rates. The Pinch-Air keyboard, with a mean text-entry rate of 11.3 WPM, is consistent with the results from Vulture ($M = 11.8$) for participants’ first session TODO: VULTURE REFERENCE. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means,

$F(5, 102) = 55.5017$, $p\text{-value} = 1.4206e-27$, ($SD_{pooled} = 2.5$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

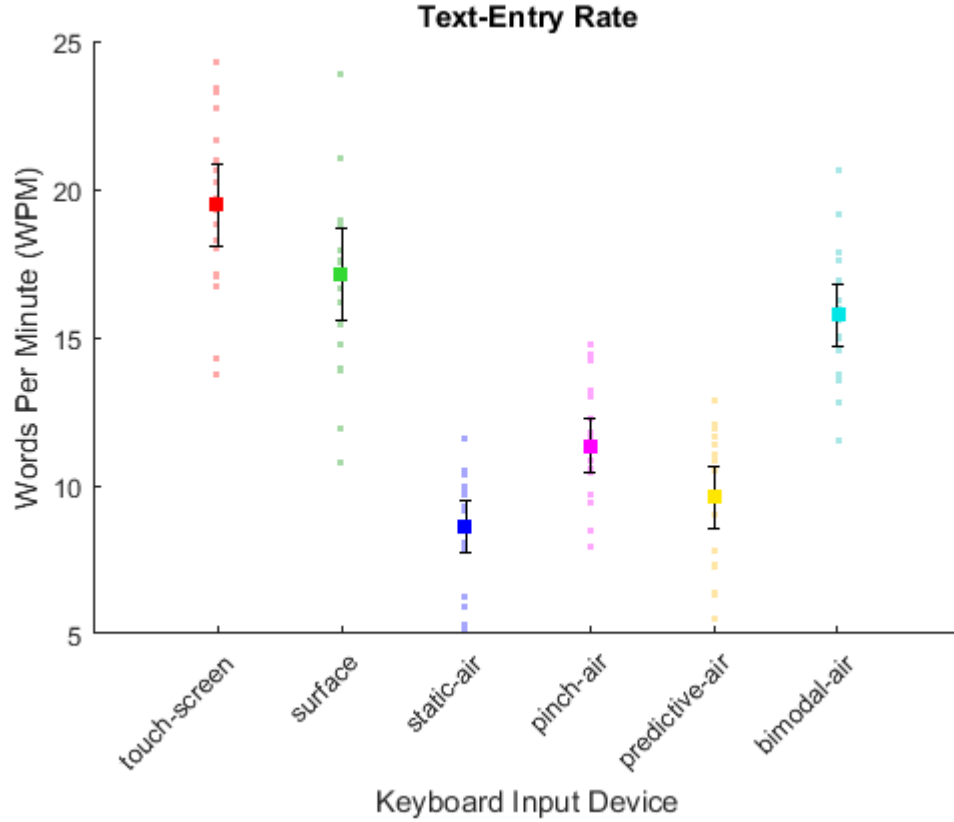


Figure 5.1. Mean Text-Entry Rates for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.1, revealed significant differences in text-entry rates between the Touch Screen Keyboard and all mid-air keyboards with $p\text{-values} < 0.001$. There were significant differences between the Leap Surface Keyboard and the Static-Air, Pinch-Air, and Predictive-Air keyboards all with $p\text{-values} < 0.0001$. There were also significant differences found between the Leap Bimodal-Air and all other mid-air keyboards, all with $p\text{-values} < 0.0001$. Finally, there was a significant difference found between the Pinch-Air and Static-Air keyboards with a $p\text{-value} = 0.0170$.

Unfortunately, no mid-air keyboards were able to achieve speeds as fast as the Touch Screen Keyboard and were all significantly slower, however the results are promising with the Bimodal-Air keyboard reaching a mean text-entry rate of 15.8 WPM without repeated sessions. With training, as in Vulture TODO REF VULTURE, the Bimodal Keyboard is expected to increase in performance by TODO: VUL PERCENT , reaching speeds sufficiently high enough to compete the with the Touch Screen keyboard, and to surpass pinching as a means of mid-air word-gesturing.

The Static-Air keyboard underperformed with the Predictive-Air keyboard achieving marginally better results, however these results were expected. Adding the 3rd-Dimension as a means of interaction further increased keyboard complexity, and increased the mental coupling required to move from the gestures in the motor-space to the feedback on the display TODO: VULTURE 24. It is, however, interesting to see how well the Leap Surface Keyboard performed since it had the same implementation as the Leap Static-Air Keyboard, just instead being projected onto a piece of paper representing the keyboard. This implies that visually representing the Leap Static-Air keyboard in 3D space using augmented reality might have the same results due to the decreased coupling between the motor-space and display.

5.1.1 *Text-Entry Rate Modified-Shortest*

Figure 5.2 shows the mean text-entry rates for each keyboard input device using the shortest-transcribed modification. Participants reached a mean text-entry rate of 19.7 WPM ($SD = 2.9$) for the Touch Screen Keyboard, 17.3 WPM ($SD = 3.3$) for the Leap Surface Keyboard, 8.8 WPM ($SD = 1.9$) for the Leap Static-Air Keyboard, 11.6 WPM ($SD = 2.1$) for the Leap Pinch-Air Keyboard, 9.9 WPM ($SD = 2.3$) for the Predictive-Air Keyboard, and 15.9 WPM ($SD = 2.2$) for the Leap

Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 54.6430$, $p\text{-value} = 2.5223e-27$, ($SD_{pooled} = 2.5$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

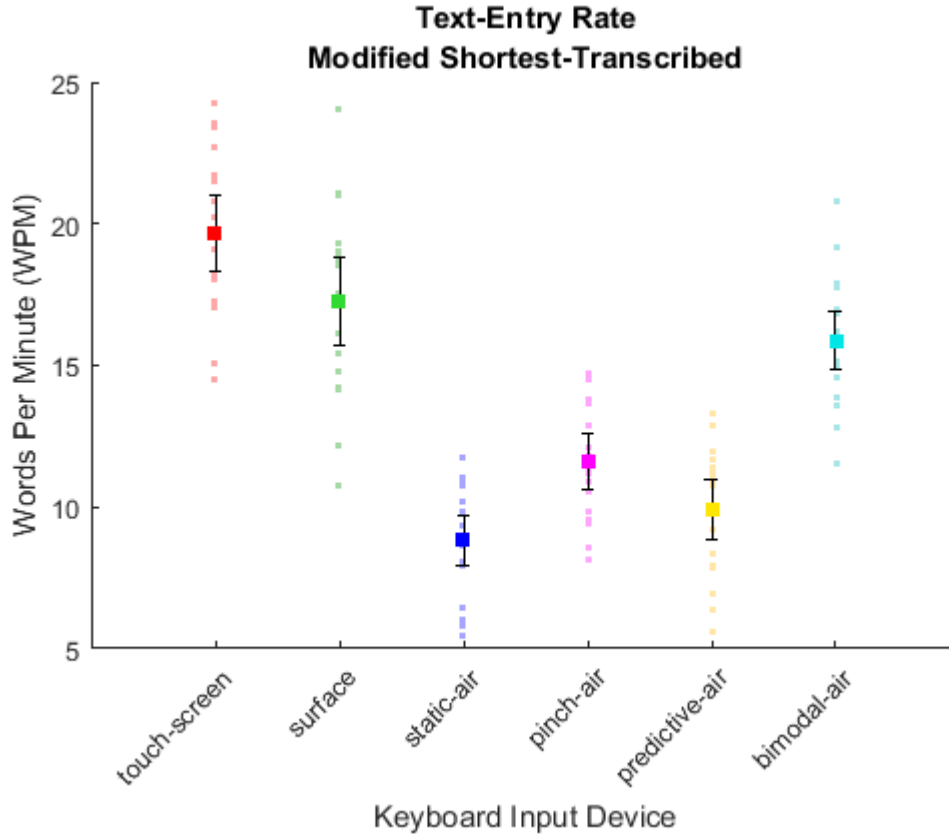


Figure 5.2. Mean Text-Entry Rates using the shortest-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.2, revealed significant differences in text-entry rates between the Touch Screen Keyboard and all mid-air keyboards with $p\text{-values} < 0.001$. There were significant differences between the Leap Surface Keyboard and the Static-Air, Pinch-Air, and Predictive-Air keyboards all with $p\text{-values} < 0.0001$. There were also significant differences found between the Leap Bimodal-Air and all other mid-air keyboards, all with $p\text{-values} < 0.0001$. The highest Bimodal-Air

text-entry rates were on par with the Touch Screen Keyboard. Finally, there was a significant difference found between the Pinch-Air and Static-Air keyboards with a p -value = 0.0148.

The results from using the shortest-transcribed modification are marginally better than those without, implying that errors were more likely corrected throughout typing the words rather than at the end. The cases where errors were made while still hitting all required letters seem to have very little impact on the duration it takes to complete words.

5.2 Error Rate

5.2.1 Minimum Word Distance

Figure 5.3 shows the mean Minimum Word Distance for each keyboard input device. Participants reached a mean MWD of 13.0% ($SD = 13.4$) for the Touch Screen Keyboard, 18.9% ($SD = 14.0$) for the Leap Surface Keyboard, 56.7% ($SD = 19.4$) for the Leap Static-Air Keyboard, 34.4% ($SD = 19.4$) for the Leap Pinch-Air Keyboard, 48.9% ($SD = 25.2$) for the Predictive-Air Keyboard, and 20.4% ($SD = 16.9$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 16.5496$, p -value = $6.1467e-12$, ($SD_{pooled} = 18.5$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.3, revealed significant differences in Minimum Word Distance between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards with p -values < 0.01 . There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards both with p -values < 0.0001 . There were also significant

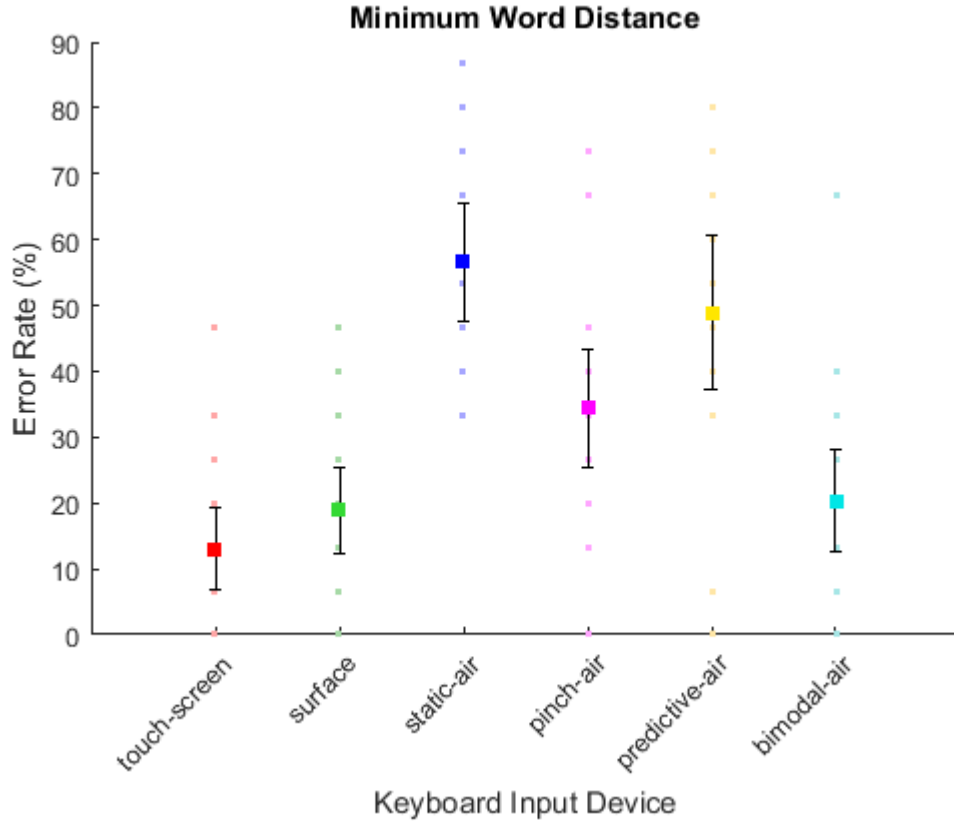


Figure 5.3. Mean Minimum Word Distance for each keyboard with error bars showing 95% confidence intervals.

differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, both with p -values < 0.001 . There was finally a significant difference found between the Leap Pinch-Air and Static-Air keyboards with p -value $= 0.0062$.

The error rates here were much higher than those seen in Vulture for Pinch or Touch TODO: REF VULTURE because this is a modified version of MWD. Since participants were required to fix errors, the MWD from Vulture would have seen error rates of 0% so instead if any errors were detected during the typing process, whether it be from device detection issues or the participant going for the wrong letter, the word is counted as erroneous. This MWD is showing us how many words needed correcting for any reason.

The Static-Air and Predictive-Air keyboards underperformed, as expected, due to the increased coupling and the added complexity of working with the 3rd-Dimension. The Leap Bimodal-Air Keyboard was significantly less erroneous than the Static-Air and Predictive-Air keyboards and was not significantly different from the Touch Screen Keyboard, implying that removing the 3rd-Dimension significantly reduces coupling and hand-eye coordination efforts. Again, the Leap Surface was on par with the Touch Screen keyboard, suggesting that with augmented reality, the Static-Air Keyboard could see a significant improvement in performance in all areas.

5.2.1.1 MWD Modified-Shortest. Figure 5.4 shows the mean Minimum Word Distance with the shortest-transcribed modification for each keyboard input device. Participants reached a mean MWD of 6.3% ($SD = 9.0$) for the Touch Screen Keyboard, 6.3% ($SD = 6.3$) for the Leap Surface Keyboard, 25.6% ($SD = 20.7$) for the Leap Static-Air Keyboard, 20.0% ($SD = 11.2$) for the Leap Pinch-Air Keyboard, 26.0% ($SD = 16.6$) for the Predictive-Air Keyboard, and 10.0% ($SD = 11.7$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 8.5172$, $p\text{-value} = 9.0987e-07$, ($SD_{pooled} = 13.5$), prompting the use of Tukey’s HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.4, revealed significant differences in Minimum Word Distance with the shortest-transcribed modification between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards with $p\text{-values} < 0.05$. There were significant differences between the Leap Surface Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards all with $p\text{-values} < 0.05$. There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, both with $p\text{-values} < 0.01$.

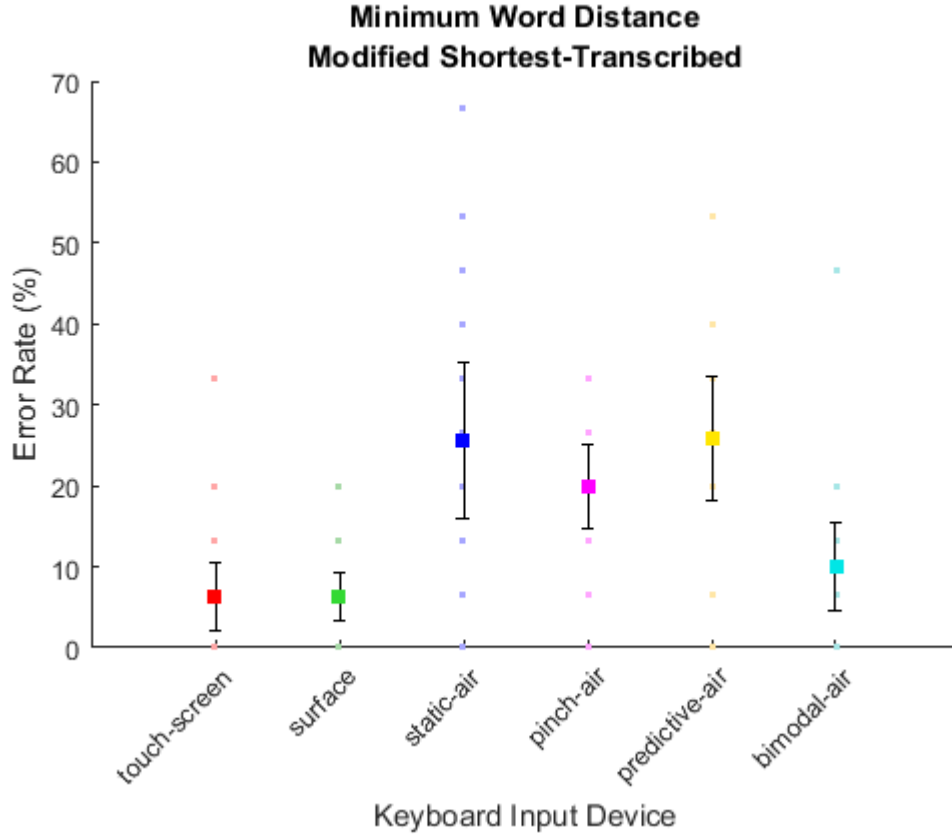


Figure 5.4. Mean Minimum Word Distance using the shortest-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

The huge reductions in MWD error rates, near 50% reduction for all keyboards, when using the shortest-transcribed modification indicates that many participants seem to be making errors after hitting all of the required letters for a word that they are entering. This may imply situations like Example 4 from Table 4.9, where participants are typing the wrong word, such as "fired" compared to "fire". More likely the case, for all of the Leap Motion-based keyboards it implies that many participants made errors on exiting the interaction plane. This kind of error, when exiting the interaction plane, was observed often by the researcher and is thought to be heavily influenced by participants pulling their hands away in an arcing motion, especially for participants who were resting their arm TODO: GET PICTURE OF ARCING MOTION. It is expected that a word-recognition implementation of the

word-gesture keyboards would see even lower error rates closer to those found in Vulture TODO: REF VULTURE, implying that the best-guess implementation was more sensitive and less robust than word-recognition to deviations in a word’s gesture-shape.

5.2.2 Keystrokes Per Character

Figure 5.5 shows the mean Keystrokes Per Character for each keyboard input device. Participants reached a mean KSPC of 1.2 keystrokes per character ($SD = 0.18$) for the Touch Screen Keyboard, 1.4 keystrokes per character ($SD = 0.38$) for the Leap Surface Keyboard, 2.0 keystrokes per character ($SD = 0.79$) for the Leap Static-Air Keyboard, 1.6 keystrokes per character ($SD = 0.42$) for the Leap Pinch-Air Keyboard, 1.9 keystrokes per character ($SD = 0.64$) for the Predictive-Air Keyboard, and 1.2 keystrokes per character ($SD = 0.18$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 9.8827$, $p\text{-value} = 9.9960e-08$, ($SD_{pooled} = 0.49$), prompting the use of Tukey’s HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.5, revealed significant differences in Keystrokes Per Character between the Touch Screen Keyboard and the Leap Static-Air and Predictive-Air keyboards, both with $p\text{-values} < 0.001$. There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards both with $p\text{-values} < 0.05$. There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, both with $p\text{-values} < 0.01$.

Keystrokes Per Character was useful because with the best-guess implementation, words are produced at the character level rather than at the whole word level as with word-recognition. With KSPC, if words were input without any detected errors, then we would expect a KSPC of 1.0 keystrokes. This metric helps show the rate at which extra erroneous characters were produced for each keyboard, for example, the

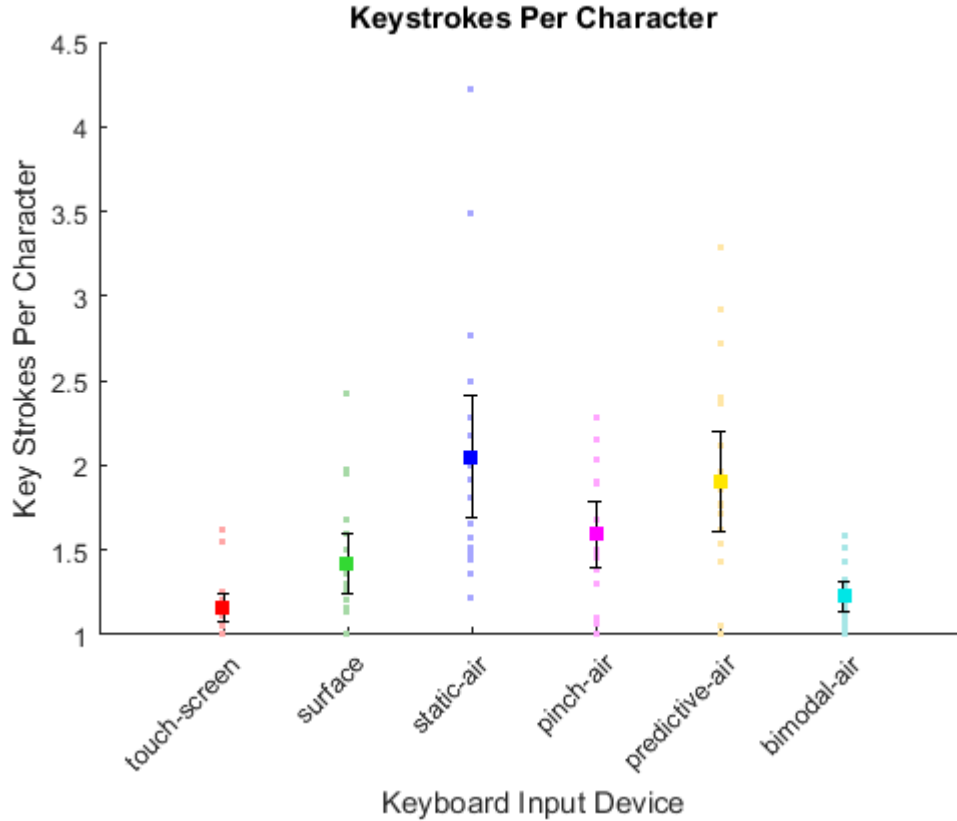


Figure 5.5. Mean Keystrokes Per Character for each keyboard with error bars showing 95% confidence intervals.

KSPC of the Static-Air keyboard was 2.0 keystrokes, meaning 100% more interactions were produced than required to type the words successfully. As before the Static-Air and Predictive-Air keyboard performances were underwhelming and the Bimodal-Air performed on par with the Touch Screen and Leap Surface keyboards. The Pinch-Air performed somewhere in between the other Leap Motion-based keyboards. As can be seen, a trend is developing in the dependent measures. This trend is expected to hold for all other variables.

5.2.2.1 KSPC Modified-Shortest. Figure 5.6 shows the mean Keystrokes Per Character with the shortest-transcribed modification for each keyboard input device. Participants reached a mean KSPC of 1.1 keystrokes per character ($SD = 0.14$) for the

Touch Screen Keyboard, 1.3 keystrokes per character ($SD = 0.27$) for the Leap Surface Keyboard, 1.7 keystrokes per character ($SD = 0.48$) for the Leap Static-Air Keyboard, 1.4 keystrokes per character ($SD = 0.31$) for the Leap Pinch-Air Keyboard, 1.7 keystrokes per character ($SD = 0.43$) for the Predictive-Air Keyboard, and 1.1 keystrokes per character ($SD = 0.15$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 11.5949$, $p\text{-value} = 7.0164e-09$, ($SD_{pooled} = 0.32$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

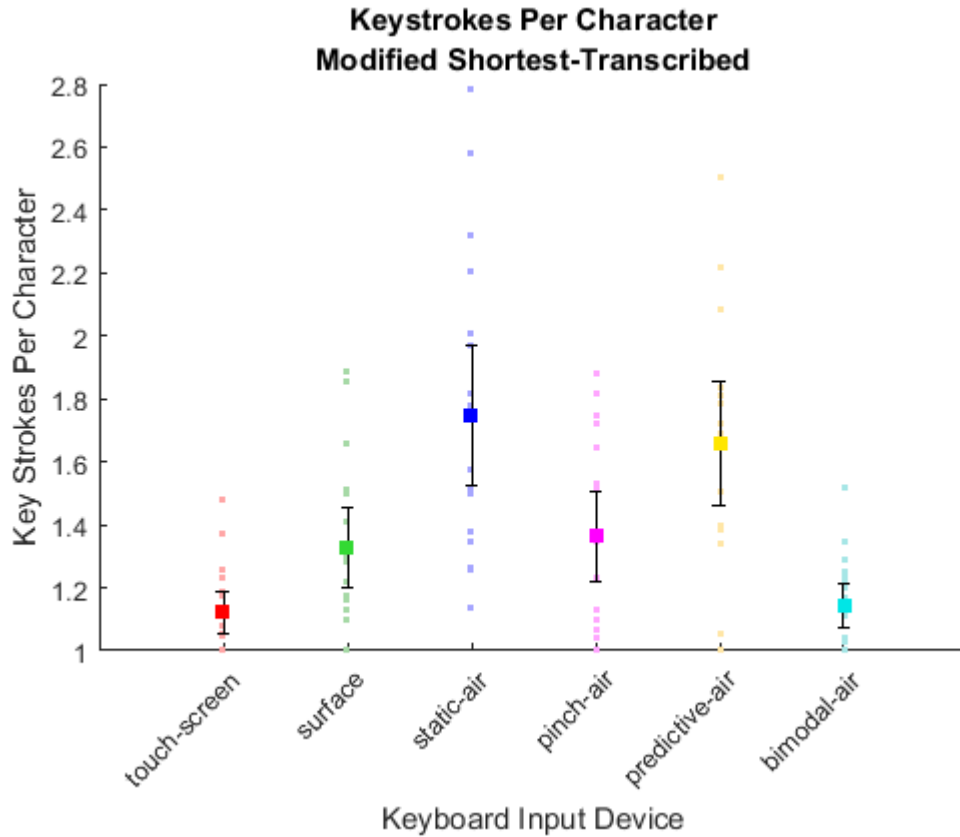


Figure 5.6. Mean Keystrokes Per Character using the shortest-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.6, revealed significant differences in Keystrokes Per Character with the shortest-transcribed modification between the

Touch Screen Keyboard and the Leap Static-Air and Predictive-Air keyboards with p -values < 0.0001 . There were significant differences between the Leap Surface Keyboard and the Leap Static-Air and Predictive-Air keyboards both with p -values < 0.05 . There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, both with p -values < 0.001 . Finally there was a significant difference between the Pinch-Air and Static-Air keyboards with a p -value $= 0.0072$.

As before, like with MWD, KSPC saw huge improvements with the shortest-transcribed modification, especially for Leap Motion-based keyboards. This reaffirms that a large number of errors are produced during exiting the interaction plane.

5.2.3 Minimum String Distance

5.2.3.1 MSD Modified-Shortest. Figure 5.7 shows the mean Minimum String Distance with the shortest-transcribed modification for each keyboard input device. Participants reached a mean MSD of 1.6% ($SD = 2.4$) for the Touch Screen Keyboard, 2.2% ($SD = 2.4$) for the Leap Surface Keyboard, 6.4% ($SD = 5.0$) for the Leap Static-Air Keyboard, 5.1% ($SD = 3.1$) for the Leap Pinch-Air Keyboard, 5.4% ($SD = 3.3$) for the Predictive-Air Keyboard, and 2.5% ($SD = 2.6$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 6.8688$, p -value $= 1.4591e-05$, ($SD_{pooled} = 3.3$), prompting the use of Tukey’s HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.7, revealed significant differences in Minimum String Distance with the shortest-transcribed modification between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards with p -values < 0.05 . There were significant differences between the Leap Surface Keyboard and the Leap Static-Air and Predictive-Air keyboards both with

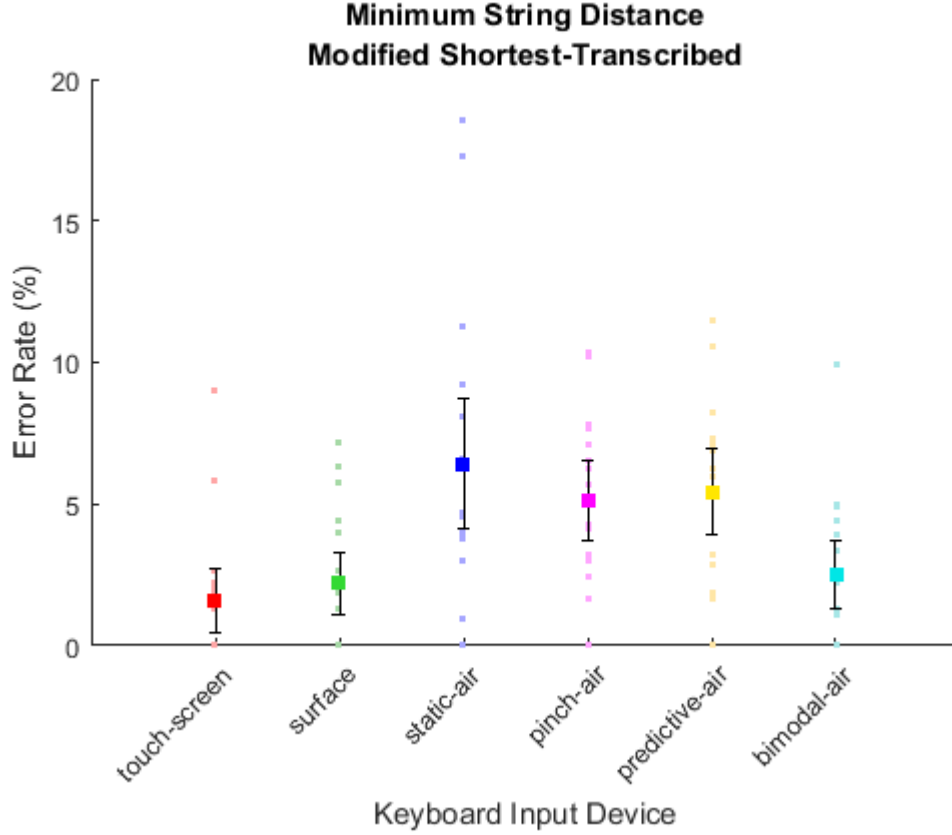


Figure 5.7. Mean Minimum String Distance using the shortest-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

p -values < 0.05 . There were also significant differences found between the Leap Bimodal-Air and the Static-Air keyboards with a p -value = 0.0067.

The error rates seen for the Pinch-Air and Touch Screen keyboards for MSD with the shortest-transcribed modification are similar to those found in Vulture for their MWD TODO: VULTURE REF. It is expected that with practice or repeated sessions, these error rates would become even less. The keyboards again follow the same pattern as other measures, providing more evidence of the effects of coupling and complexity of 3-Dimensions versus a tactile surface. Notably, the Leap Bimodal-Air seems to be unaffected by coupling and complexity, so far having never been significantly different than the Touch Screen or the Leap Surface keyboards.

5.2.4 Total Error Rate

Figure 5.8 shows the mean Total Error Rate for each keyboard input device. Participants reached a mean Total Error Rate of 4.9% ($SD = 5.5$) for the Touch Screen Keyboard, 9.3% ($SD = 7.1$) for the Leap Surface Keyboard, 23.6% ($SD = 12.1$) for the Leap Static-Air Keyboard, 14.1% ($SD = 9.0$) for the Leap Pinch-Air Keyboard, 20.1% ($SD = 11.4$) for the Predictive-Air Keyboard, and 6.8% ($SD = 5.5$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 12.9381$, $p\text{-value} = 9.4983e-10$, ($SD_{pooled} = 8.8$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

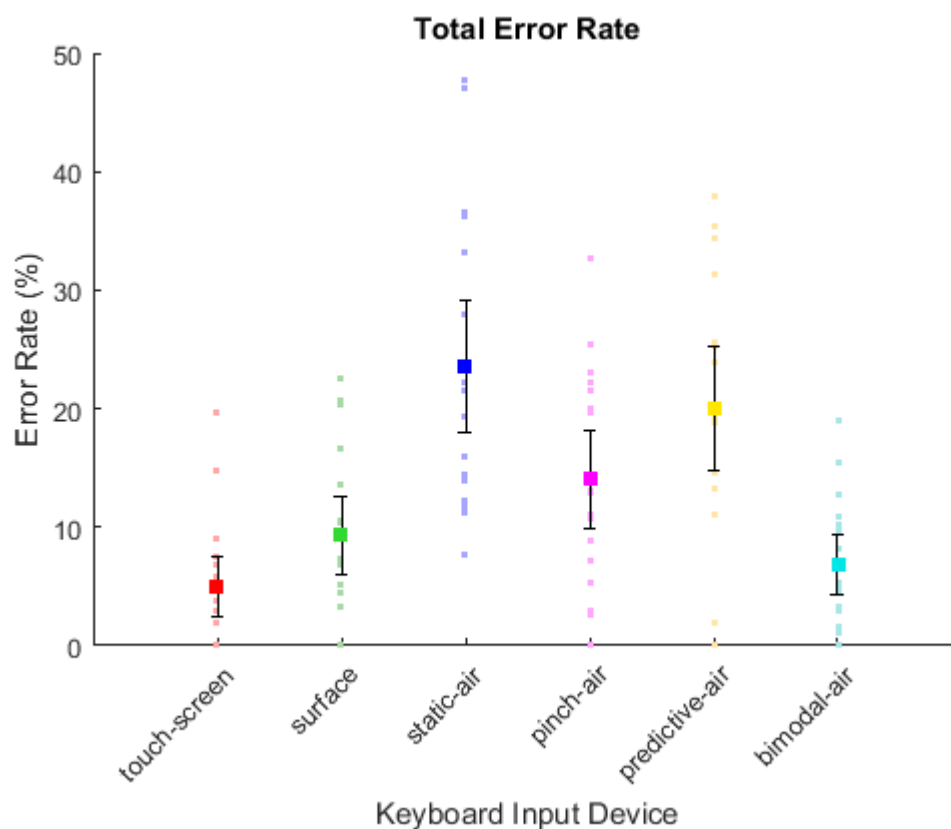


Figure 5.8. Mean Total Error Rates for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.8, revealed significant differences in Total Error Rate between the Touch Screen Keyboard and the Leap Static-Air and Predictive-Air keyboards, both with p -values < 0.0001 . There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards both with p -values < 0.01 . There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, both with p -values < 0.01 . Finally, there was a significant difference between the Pinch-Air and the Static-Air keyboards with a p -value $= 0.0191$.

Again, we see the keyboards following the same pattern. The Bimodal-Air and Leap Surface keyboards perform about as well as the Touch Screen Keyboard while the Static-Air and Predictive-Air keyboards continue to fall short. The Pinch-Air keyboard falls somewhere in the middle for Total Error Rate, not significantly different from any of them.

5.2.4.1 Total Error Rate Modified-Shortest. Figure 5.9 shows the mean Total Error Rate with the shortest-transcribed modification for each keyboard input device. Participants reached a mean Total Error Rate of 4.9% ($SD = 5.4$) for the Touch Screen Keyboard, 9.2% ($SD = 7.0$) for the Leap Surface Keyboard, 23.3% ($SD = 11.8$) for the Leap Static-Air Keyboard, 13.5% ($SD = 8.9$) for the Leap Pinch-Air Keyboard, 19.4% ($SD = 10.7$) for the Predictive-Air Keyboard, and 6.7% ($SD = 5.3$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 13.0717$, p -value $= 7.8147e-10$, ($SD_{pooled} = 8.6$), prompting the use of Tukey’s HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.9, revealed significant differences in Total Error Rate with the shortest-transcribed modification between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards with p -values < 0.05 . There were significant differences between the Leap Surface

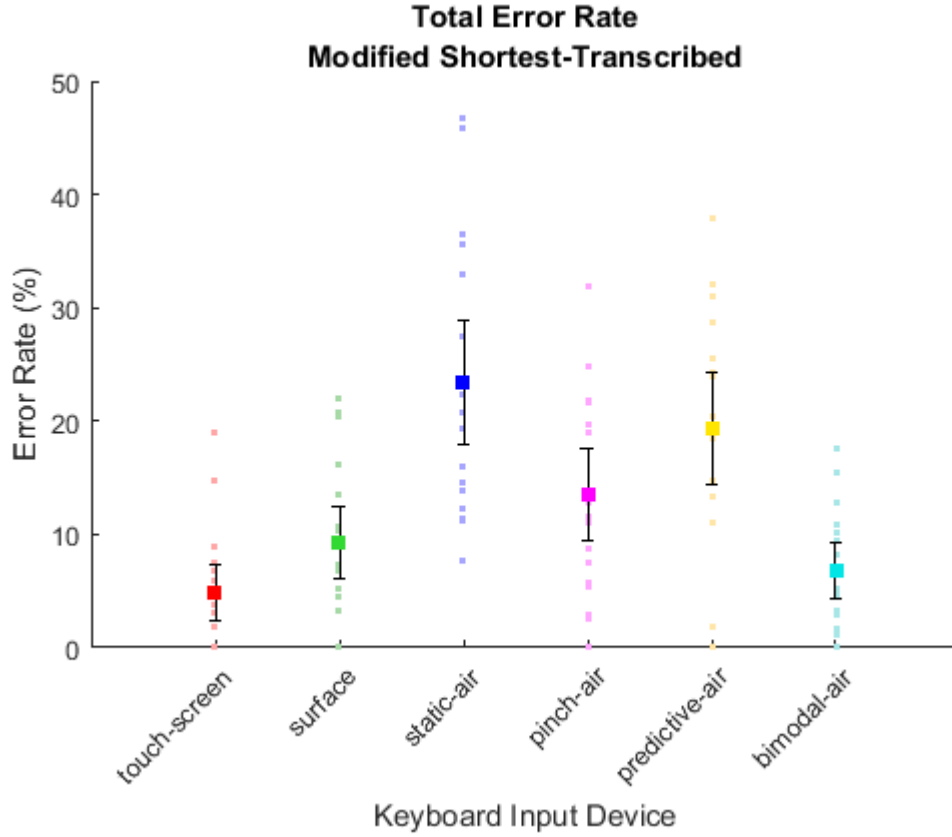


Figure 5.9. Mean Total Error Rate using the shortest-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

Keyboard and the Leap Static-Air and Predictive-Air keyboards both with p -values < 0.01 . There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards with a p -value < 0.001 . Finally, there is a significant difference between the Pinch-Air and Static-Air keyboards with a p -value = 0.0101.

Surprisingly, unlike KSPC and MWD, the Total Error Rate didn't follow suit and show large decreases in error rate with the shortest-transcribed modification. These observations do still however support what was seen text-entry rates also being relatively unaffected by the shortest-transcribed modification, implying that errors made by participants were more likely corrected throughout typing the words rather

than at the end. This is a major downside of lacking word-recognition because gestures are interrupted and corrected rather than being followed through completely as is natural with word-gesture keyboards.

5.3 Correctness

5.3.1 Fréchet Distance

Figure 5.10 shows the mean Fréchet Distance for each keyboard input device. Participants reached a mean Fréchet Distance of 231.2 pixels ($SD = 26.7$) for the Touch Screen Keyboard, 242.8 pixels ($SD = 30.4$) for the Leap Surface Keyboard, 364.3 pixels ($SD = 86.5$) for the Leap Static-Air Keyboard, 501.2 pixels ($SD = 42.8$) for the Leap Pinch-Air Keyboard, 350.6 pixels ($SD = 86.8$) for the Predictive-Air Keyboard, and 396.5 pixels ($SD = 102.1$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 37.9416$, $p\text{-value} = 8.0562e-22$, ($SD_{pooled} = 69.4$), prompting the use of Tukey’s HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.10, revealed significant differences in Fréchet Distance between the Touch Screen Keyboard and all of the mid-air keyboards with $p\text{-values} < 0.0001$. There were significant differences between the Leap Surface Keyboard and all of the mid-air keyboards with $p\text{-values} < 0.001$. There were also significant differences found between the Leap Pinch-Air and all of the other mid-air keyboards with $p\text{-values} < 0.001$.

The Fréchet Distance is the only place where the trends are broken and the Bimodal-Air keyboard performs significantly worse than the Touch Screen and Leap Surface keyboards. The Pinch-Air Keyboard also performed significant worse than all other keyboards. These results are expected to be caused by a combination of the word-gesture velocity seen in Section 5.4.2 and the affects of increase or decreased coupling or complexity from working with the 3rd-Dimension.

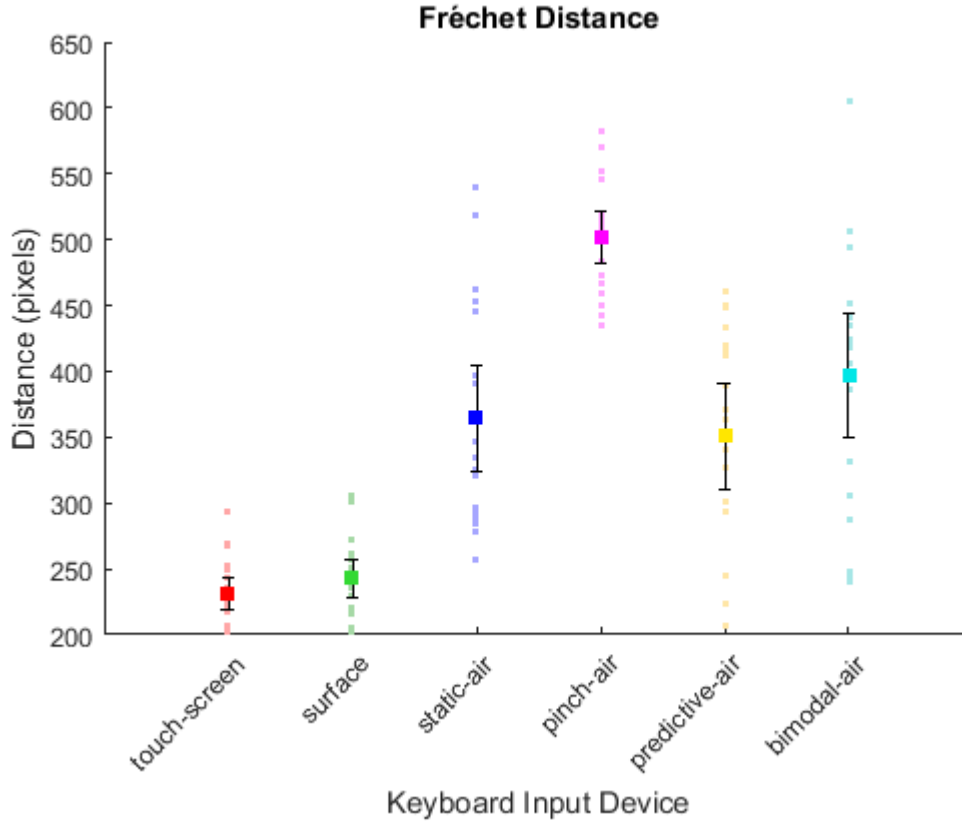


Figure 5.10. Mean Fréchet Distance for each keyboard with error bars showing 95% confidence intervals.

The Leap Bimodal-Air seems to suffer in correctness due to its very high word-gesture velocities whereas the Static-Air and Predictive-Air benefit in correctness due to their very slow word-gesture velocities. The Leap Surface and Touch Screen keyboards are relatively unaffected in correctness by their high word-gesturing velocities due to their decreased coupling from the word-gesture motor space to the displayed screen.

The Pinch-Air Keyboard suffers from the combination of high word-gesture velocities and effects of increased coupling, however it also seems to be affected heavily by its deviation in its tracking method increasing its overall distance traveled for each word-gesture, see Section 5.4.1. All other mid-air keyboards track the participant's pointer finger whereas the Pinch-Air keyboard tracks the participant's palm

since their fingers move so sharply when creating a pinching gesture. This difference in tracking combined with high word-gesture velocities are the expected factors for this decrease in performance. TODO: STATISTICALLY CHECK IF VELOCITY AFFECTS CORRECTNESS AND IF TRACKING AFFECTS CORRECTNESS (stylus, palm, finger, etc)

5.3.1.1 Fréchet Distance Modified-Shortest. Figure 5.11 shows the mean Fréchet Distance with the shortest-transcribed modification for each keyboard input device. Participants reached a mean Fréchet Distance of 221.0 pixels ($SD = 18.0$) for the Touch Screen Keyboard, 231.0 pixels ($SD = 22.7$) for the Leap Surface Keyboard, 306.4 pixels ($SD = 53.2$) for the Leap Static-Air Keyboard, 459.2 pixels ($SD = 20.1$) for the Leap Pinch-Air Keyboard, 281.6 pixels ($SD = 49.9$) for the Predictive-Air Keyboard, and 368.4 pixels ($SD = 82.4$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 65.7440$, $p\text{-value} = 2.3767e-30$, ($SD_{pooled} = 47.2$), prompting the use of Tukey’s HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.11, revealed significant differences in Fréchet Distance with the shortest-transcribed modification between the Touch Screen Keyboard and all of the mid-air keyboards with $p\text{-values} < 0.01$. There were significant differences between the Leap Surface Keyboard and all of the mid-air keyboards with $p\text{-values} < 0.05$. There were also significant differences found between the Leap Bimodal-Air Keyboard and the other mid-air keyboards with $p\text{-values} < 0.01$. Finally, there were significant differences found between the Leap Pinch-Air Keyboard and the Static-Air and Predictive-Air keyboards with $p\text{-values} < 0.0001$.

Again, the increase in correctness for all keyboards, namely the mid-air keyboards, when using the shortest-transcribed modification implies that many errors are being produced when exiting the interaction plane. This is especially evident

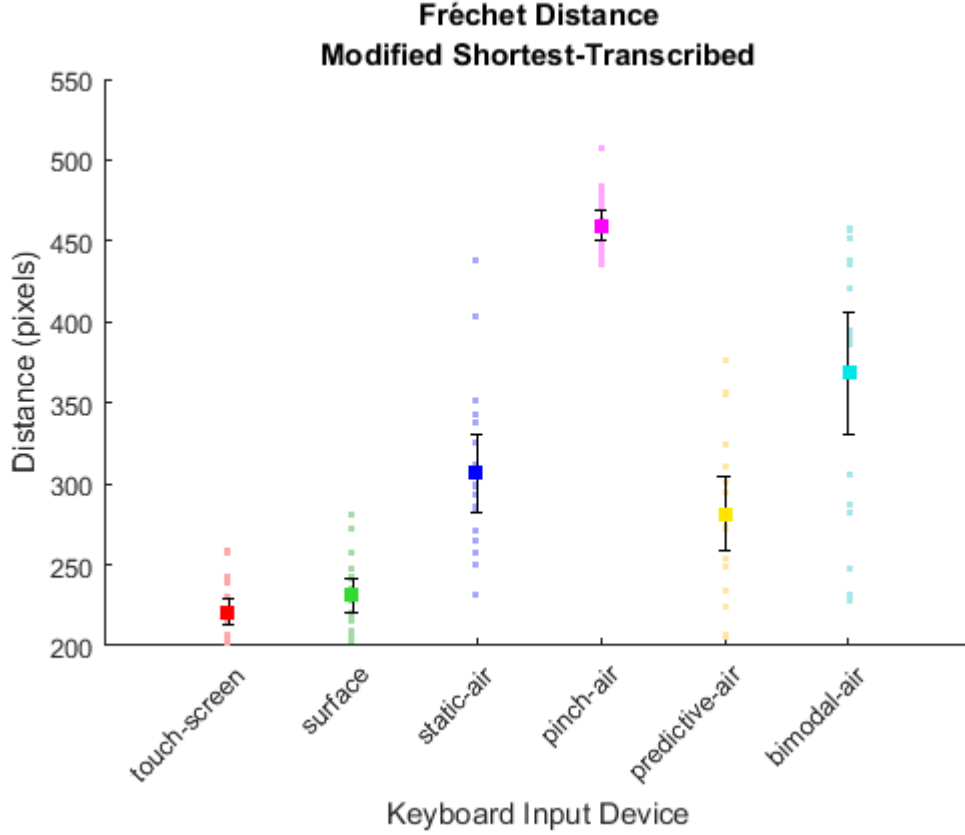


Figure 5.11. Mean Fréchet Distance using the shortest-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

for the Static-Air, Pinch-Air, and Predictive-Air keyboards all seeing the largest improvements. For the Static-Air and Predictive-Air keyboards, these errors are caused by the addition of the 3rd-Dimension and participants having trouble entering and exiting the interaction plane accurately. The Pinch-Air Keyboard, again, sees a combination of high word-gesture velocity and a difference in tracking that is expected to produce these kinds of errors.

5.3.1.2 Fréchet Distance Modified-Backspace. Figure 5.12 shows the mean Fréchet Distance with the backspace-transcribed modification for each keyboard input device. Participants reached a mean Fréchet Distance of 212.3 pixels ($SD = 9.3$) for the Touch Screen Keyboard, 223.5 pixels ($SD = 18.5$) for the Leap Surface Keyboard,

258.9 pixels ($SD = 33.5$) for the Leap Static-Air Keyboard, 444.7 pixels ($SD = 8.0$) for the Leap Pinch-Air Keyboard, 237.7 pixels ($SD = 24.8$) for the Predictive-Air Keyboard, and 356.8 pixels ($SD = 85.0$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 97.1196$, $p\text{-value} = 3.5271e-37$, ($SD_{pooled} = 39.7$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

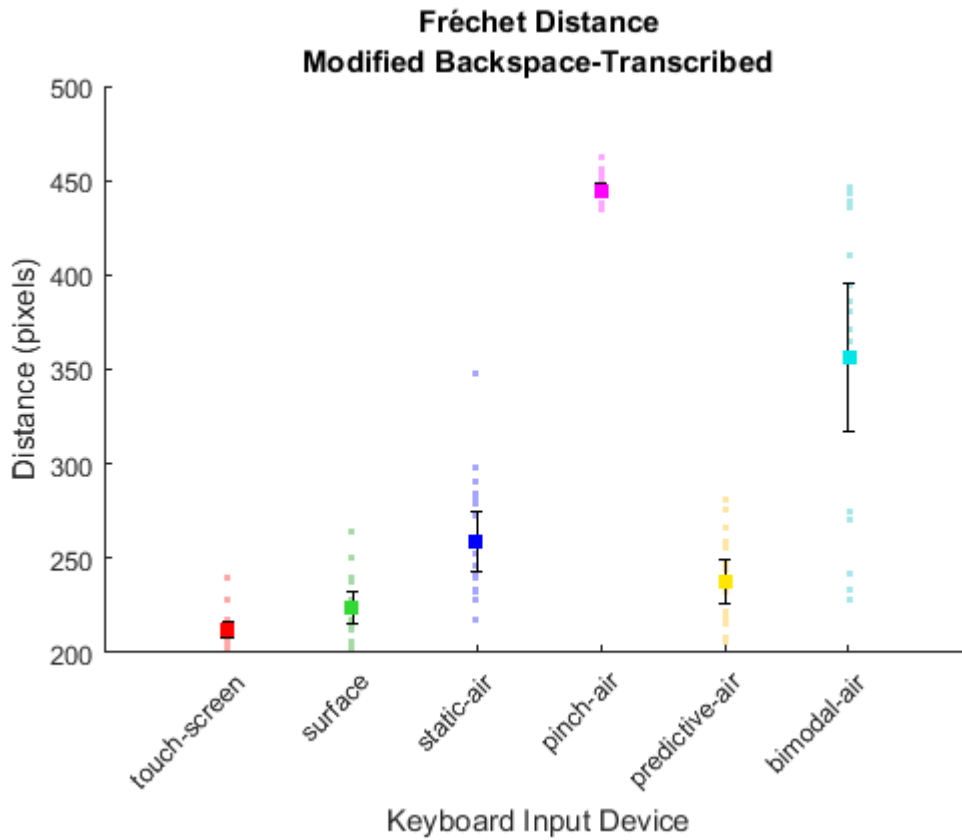


Figure 5.12. Mean Fréchet Distance using the backspace-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.12, revealed significant differences in Fréchet Distance with the backspace-transcribed modification between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Bimodal-Air keyboards with $p\text{-values} < 0.01$. There were significant differences between the Leap Surface

Keyboard and the Leap Pinch-Air and Bimodal-Air keyboards both with p -values < 0.0001 . There were also significant differences found between the Leap Bimodal-Air and all other mid-air keyboards with p -values < 0.0001 . Finally, there were significant differences between the Leap Pinch-Air Keyboard and the Static-Air and Predictive-Air keyboards with p -values < 0.0001 .

If the backspace key is included in the path considered to be correct for Fréchet Distance, since participants were required to correct words, a boost in correctness is seen across the board. Figure TODO: GET PICTURE OF PATHS FOR EXPECTED WITH BACKSPACE AND PATH TAKEN shows a modified expected path to include the backspace key. The Bimodal-Air and Pinch-Air keyboards still suffer from the previously mentioned problems in Section 5.3.1. The Static-Air and Predictive-Air, however, see huge improvements in correctness because of the low word-gesture velocities and high error rates. The Predictive-Air Keyboard, in this case, is not significantly different than the Touch Screen and Leap Surface keyboards, implying that participants were relatively attentive to correcting errors for it.

5.4 Distance Measures

5.4.1 Word-Gesture Distance

Figure 5.13 shows the mean Word-Gesture Distance for each keyboard input device. Participants reached a mean Word-Gesture Distance of 1087.5 pixels ($SD = 85.9$) for the Touch Screen Keyboard, 1423.9 pixels ($SD = 246.0$) for the Leap Surface Keyboard, 1590.9 pixels ($SD = 508.2$) for the Leap Static-Air Keyboard, 2165.3 pixels ($SD = 540.5$) for the Leap Pinch-Air Keyboard, 1634.7 pixels ($SD = 481.4$) for the Predictive-Air Keyboard, and 1627.7 pixels ($SD = 321.3$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 13.9225$, p -value

$= 2.2937e-10$, ($SD_{pooled} = 398.6$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

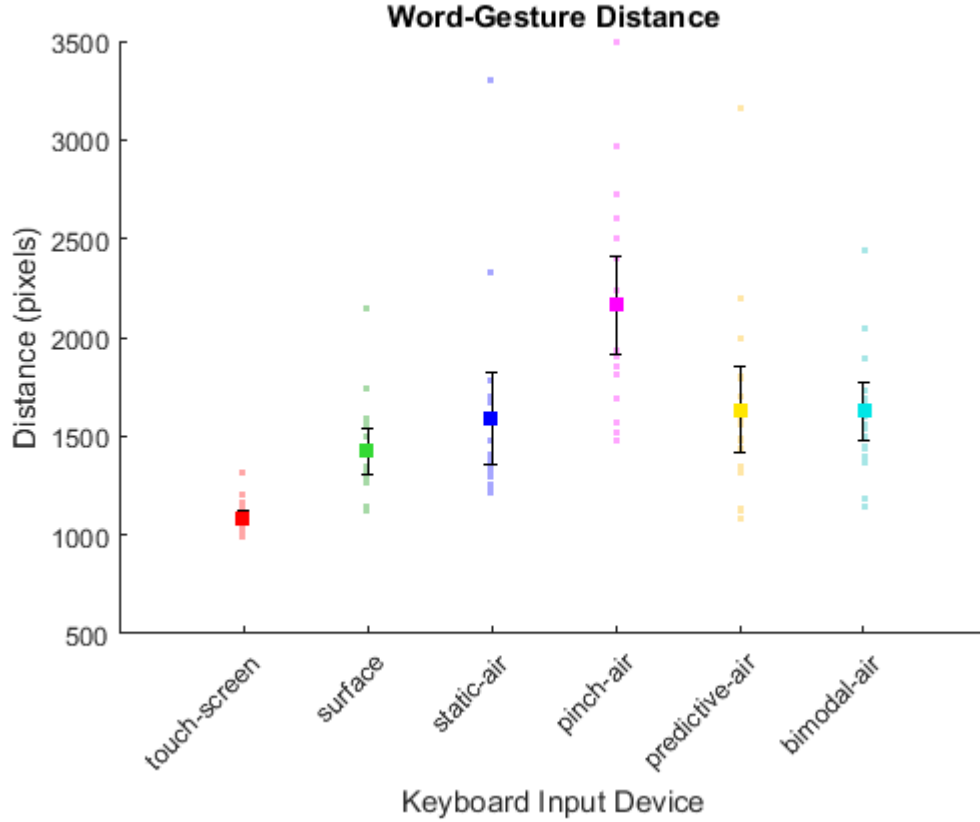


Figure 5.13. Mean Word-Gesture Distance for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.13, revealed significant differences in Word-Gesture Distance between the Touch Screen Keyboard and all of the mid-air keyboards with p -values < 0.01 . There was a significant difference between the Leap Pinch-Air and all other Leap Motion keyboards with p -values < 0.01 .

As expected, all mid-air keyboards required longer word-gesture distances than the Touch Screen Keyboard due to the effect coupling between the word-gesture motor space and the keyboard display. It is surprising, however, that the Leap Surface Keyboard seems to have required significantly more word-gesture distance than the

Touch Screen Keyboard. This could be because of the requirement to use a stylus rather than the participants bare hand.

The Pinch-Air keyboard required a significantly longer word-gesture distance than all other keyboards. This is expected to be caused by the difference in tracking method as previously mentioned. Tracking the palm requires larger, less controlled movements than with tracking fingers or a stylus.

5.4.1.1 Word-Gesture Distance Modified-Shortest. Figure 5.14 shows the mean Word-Gesture Distance with the shortest-transcribed modification for each keyboard input device. Participants reached a mean Word-Gesture Distance of 1081.3 ($SD = 84.6$) for the Touch Screen Keyboard, 1396.5 ($SD = 226.5$) for the Leap Surface Keyboard, 1524.8 ($SD = 412.5$) for the Leap Static-Air Keyboard, 1981.4 ($SD = 470.6$) for the Leap Pinch-Air Keyboard, 1536.7 ($SD = 346.1$) for the Predictive-Air Keyboard, and 1566.5 ($SD = 253.6$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 14.403$, $p\text{-value} = 1.1613e-10$, ($SD_{pooled} = 325.1$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.14, revealed significant differences in Word-Gesture Distance with the shortest-transcribed modification between the Touch Screen Keyboard and all other keyboards with $p\text{-values} < 0.05$. There were significant differences between the Leap Surface and Pinch-Air keyboards and all other Leap Motion keyboards with $p\text{-values} < 0.01$.

Again, the small difference between the word-gesture distance and the word-gesture distance using the shortest-transcribed modification, implies that most errors were corrected by participants whilst in the middle of typing words.

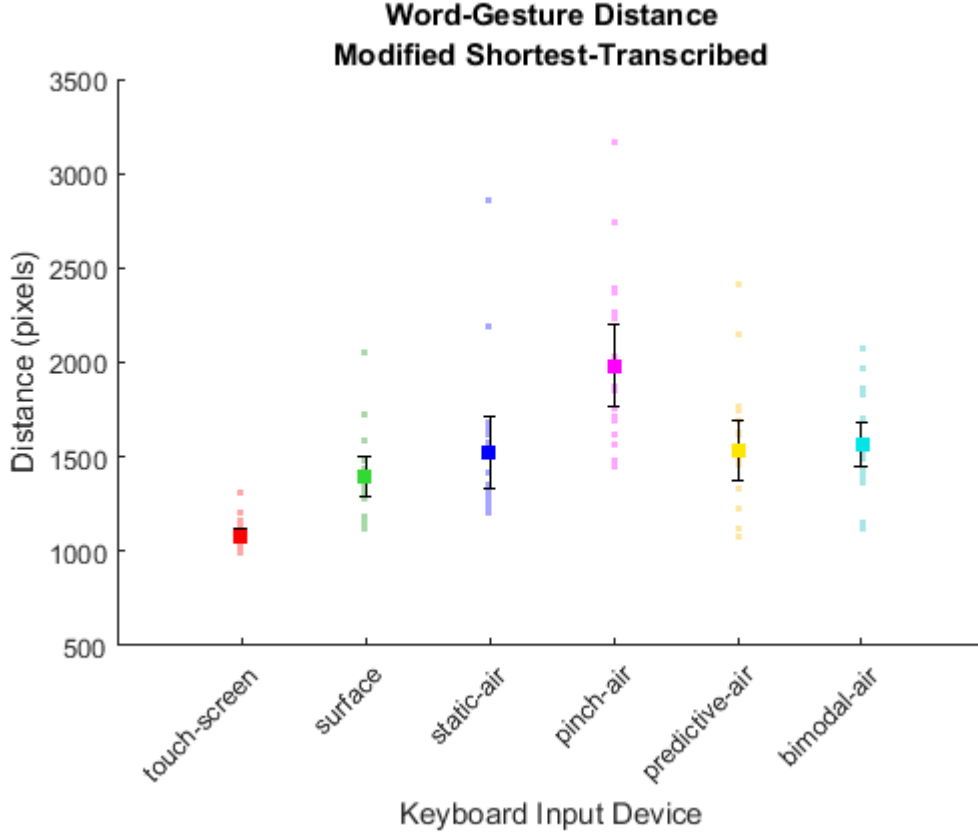


Figure 5.14. Mean Word-Gesture Distance using the shortest-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

5.4.2 Word-Gesture Velocity

Figure 5.15 shows the mean Word-Gesture Velocity for each keyboard input device. Participants reached a mean Word-Gesture Velocity of 673.3 *pixels/s* ($SD = 126.4$) for the Touch Screen Keyboard, 641.1 *pixels/s* ($SD = 165.5$) for the Leap Surface Keyboard, 380.9 *pixels/s* ($SD = 78.2$) for the Leap Static-Air Keyboard, 643.8 *pixels/s* ($SD = 204.7$) for the Leap Pinch-Air Keyboard, 405.6 *pixels/s* ($SD = 87.6$) for the Predictive-Air Keyboard, and 777.6 *pixels/s* ($SD = 240.9$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 17.233$, $p\text{-value} = 2.4922e-12$, ($SD_{pooled} = 161.8$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

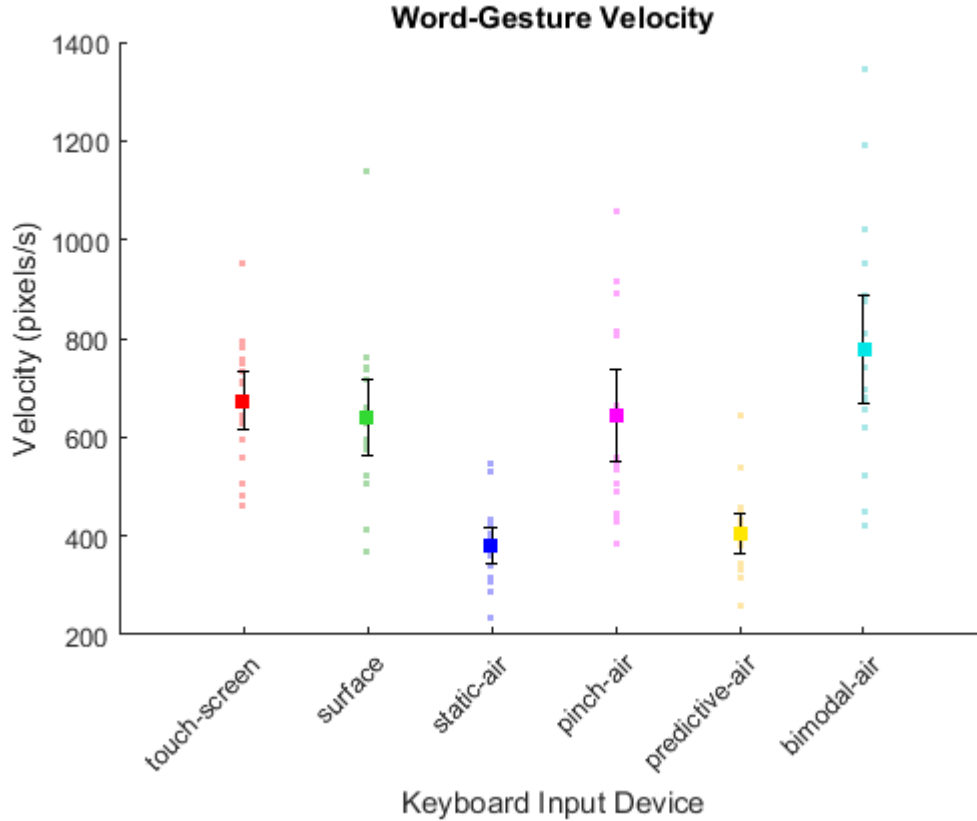


Figure 5.15. Mean Word-Gesture Velocity for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.15, revealed significant differences in Word-Gesture Velocity between the Touch Screen Keyboard and the Leap Static-Air and Predictive-Air keyboards with p -values < 0.0001 . There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards with p -values < 0.001 . There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards with p -values < 0.0001 . Finally, there were significant differences found between the Leap Pinch-Air and the Static-Air and Predictive Air keyboards with p -values < 0.001 .

The Static-Air and Predictive-Air keyboards were significantly slower than all of the other keyboards for word-gesture velocity. This was expected due to the

enhanced precision required to working with a 3rd-Dimension in mid-air. It is interesting, though, that these two keyboards also had the highest error rates across all error measures, implying a very high degree of difficulty, even with slower movements.

5.4.3 Hand Velocity

Figure 5.16 shows the mean Hand Velocity for each keyboard input device. Participants reached a mean Hand Velocity of 35.7 *cm/s* ($SD = 6.7$) for the Touch Screen Keyboard, 15.0 *cm/s* ($SD = 3.9$) for the Leap Surface Keyboard, 6.9 *cm/s* ($SD = 1.4$) for the Leap Static-Air Keyboard, 10.5 *cm/s* ($SD = 3.2$) for the Leap Pinch-Air Keyboard, 7.6 *cm/s* ($SD = 1.7$) for the Predictive-Air Keyboard, and 15.2 *cm/s* ($SD = 5.1$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 121.9826$, $p\text{-value} = 2.0877e-41$, ($SD_{pooled} = 4.1$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.16, revealed significant differences in Hand Velocity between the Touch Screen Keyboard and all other keyboards with $p\text{-values} < 0.0001$. There were significant differences between the Leap Surface Keyboard and the Static-Air, Pinch-Air, and Predictive-Air keyboards with $p\text{-values} < 0.05$. There were also significant differences found between the Leap Bimodal-Air and all other mid-air keyboards with $p\text{-values} < 0.05$.

The hand velocities are all similar to the word-gesture velocities in Section 5.4.2, except for the Touch Screen Keyboard. This result is simply because of the difference in size between the interaction planes that participants used and the physical size of the Touch Screen used, as described in Section ?? TODO: REFER TO SECTION THAT SAYS SIZE OF TOUCH SCREEN AND OTHERS.

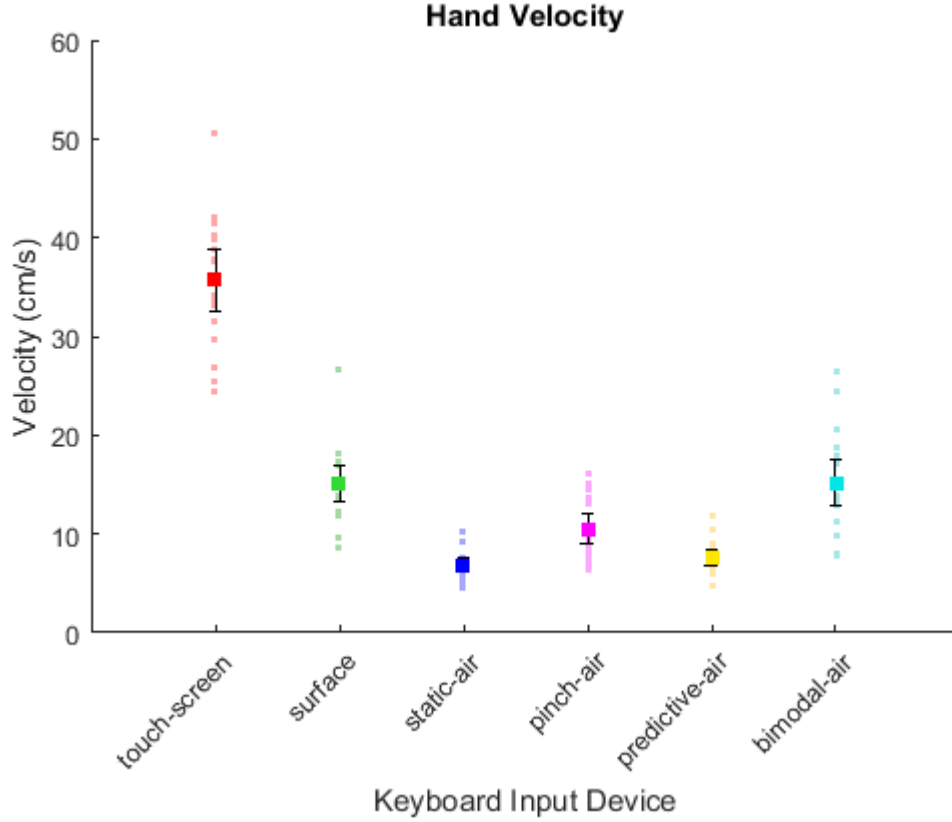


Figure 5.16. Mean Hand Velocity for each keyboard with error bars showing 95% confidence intervals.

5.5 Timing Measures

5.5.1 Word-Gesture Duration

Figure 5.17 shows the mean Word-Gesture Duration for each keyboard input device. Participants reached a mean Word-Gesture Duration of 1.65 s ($SD = 0.46$) for the Touch Screen Keyboard, 2.03 s ($SD = 1.17$) for the Leap Surface Keyboard, 4.01 s ($SD = 1.17$) for the Leap Static-Air Keyboard, 3.36 s ($SD = 0.76$) for the Leap Pinch-Air Keyboard, 3.87 s ($SD = 1.18$) for the Predictive-Air Keyboard, and 2.18 s ($SD = 0.36$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means,

$F(5, 102) = 28.1205$, $p\text{-value} = 8.2440e-18$, ($SD_{pooled} = 0.81$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

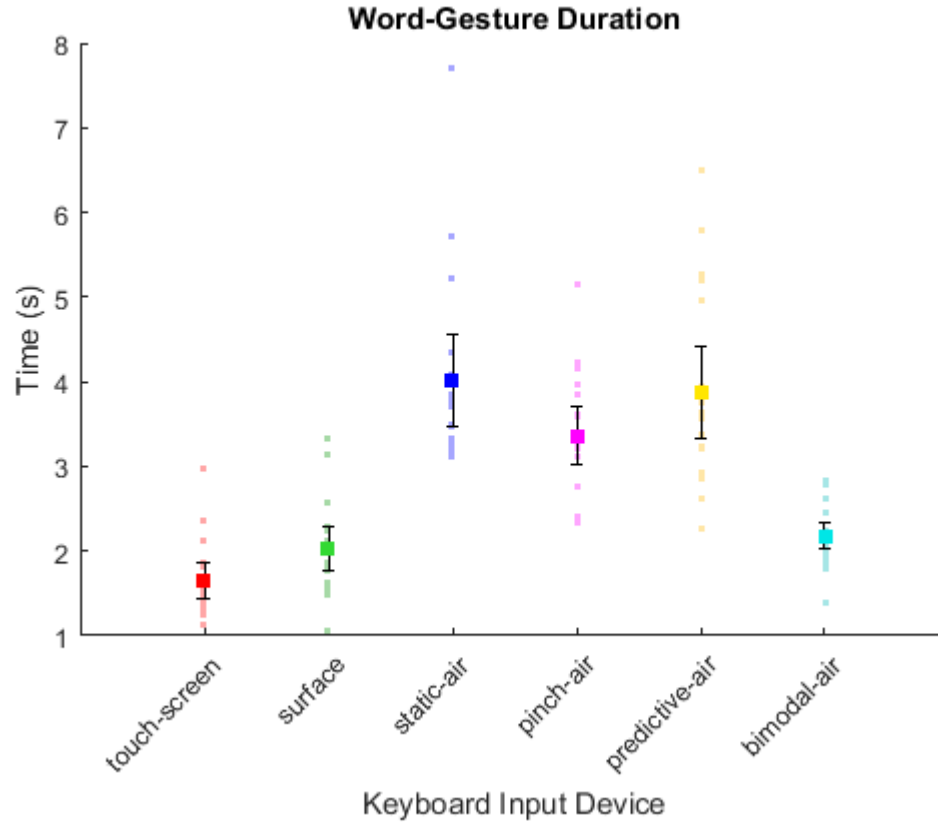


Figure 5.17. Mean Word-Gesture Duration for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.17, revealed significant differences in Word-Gesture Duration between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards, all with $p\text{-values} < 0.0001$. There were significant differences between the Leap Surface Keyboard and the Static-Air, Pinch-Air, and Predictive-Air keyboards, all with $p\text{-values} < 0.001$. There were also significant differences found between the Leap Bimodal-Air and all of the other mid-air keyboards with $p\text{-values} < 0.001$.

As expected, the Static-Air, Pinch-Air, and Predictive-Air keyboards had significantly longer word-gesture durations than the other keyboards. For the Static-Air and Predictive-Air, lower velocities and higher coupling mean that word-gestures take longer to perform, whereas for the Pinch-Air Keyboard, the longer distances traveled were the influence for the longer durations. The Bimodal-Air keyboard performed as well as the Touch Screen and Leap Surface keyboards as expected from its prior trends.

5.5.1.1 Word-Gesture Duration Modified-Shortest. Figure 5.18 shows the mean Word-Gesture Duration with the shortest-transcribed modification for each keyboard input device. Participants reached a mean Word-Gesture Duration of 1.60 *s* (*SD* = 0.40) for the Touch Screen Keyboard, 2.00 *s* (*SD* = 0.55) for the Leap Surface Keyboard, 3.75 *s* (*SD* = 0.93) for the Leap Static-Air Keyboard, 3.17 *s* (*SD* = 0.73) for the Leap Pinch-Air Keyboard, 3.56 *s* (*SD* = 0.86) for the Predictive-Air Keyboard, and 2.16 *s* (*SD* = 0.36) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 32.0355$, $p\text{-value} = 1.7009e-19$, ($SD_{pooled} = 0.67$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.18, revealed significant differences in Word-Gesture Duration with the shortest-transcribed modification between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards with $p\text{-values} < 0.0001$. There were significant differences between the Leap Surface Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards all with $p\text{-values} < 0.0001$. There were also significant differences found between the Leap Bimodal-Air and all of the other mid-air keyboards with $p\text{-values} < 0.001$.

As seen before, the shortest-transcribed modification for word-gesture duration had little effect on the total length of word-gesture durations. This again supports the

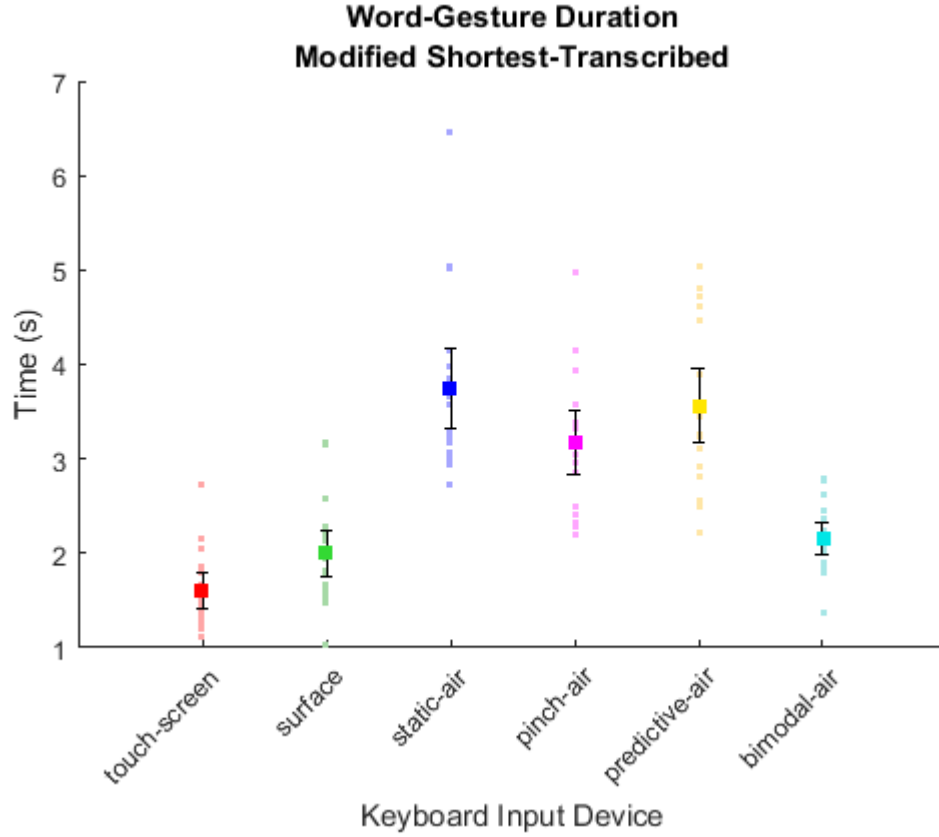


Figure 5.18. Mean Word-Gesture Dur using the shortest-transcribed modification for each keyboard with error bars showing 95% confidence intervals.

idea that participants mostly corrected errors as they made them while transcribing a word.

5.5.2 Reaction Time to Errors

Figure 5.19 shows the mean Reaction Time to Errors for each keyboard input device. Participants reached a mean Reaction Time to Errors of 0.22 s ($SD = 0.23$) for the Touch Screen Keyboard, 0.40 s ($SD = 0.29$) for the Leap Surface Keyboard, 1.02 s ($SD = 0.43$) for the Leap Static-Air Keyboard, 0.59 s ($SD = 0.30$) for the Leap Pinch-Air Keyboard, 0.78 s ($SD = 0.40$) for the Predictive-Air Keyboard, and 0.29 s ($SD = 0.24$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means,

$F(5, 102) = 16.3385$, $p\text{-value} = 8.1487e-12$, ($SD_{pooled} = 0.32$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

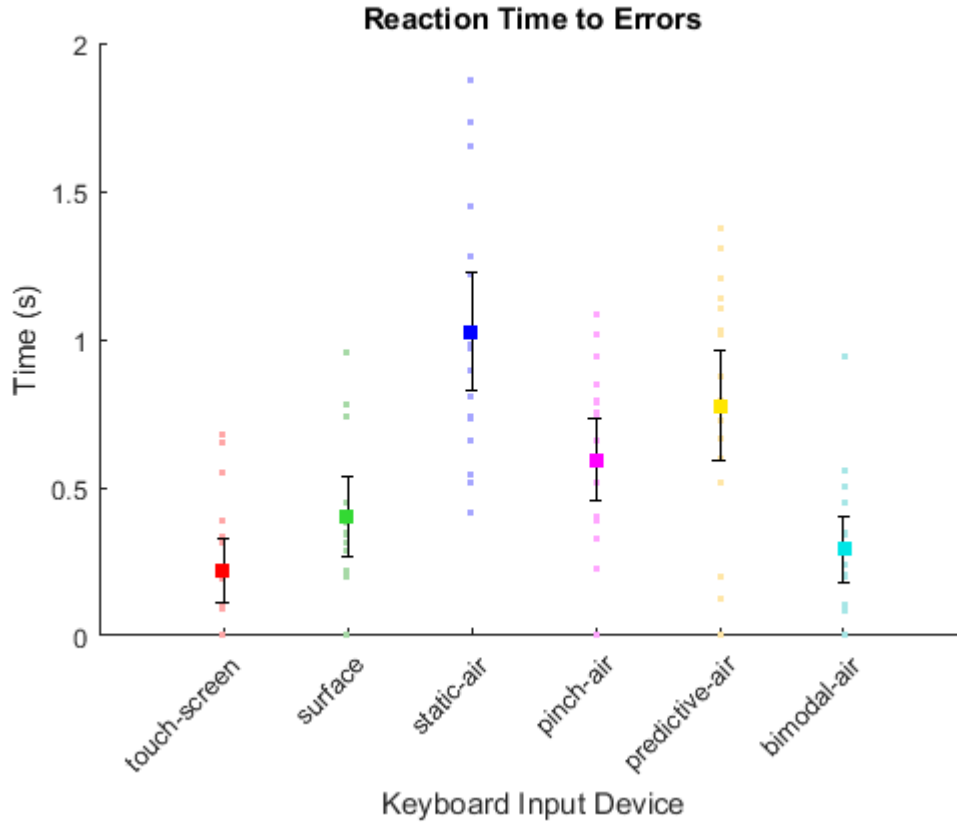


Figure 5.19. Mean Reaction Time to Errors for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.19, revealed significant differences in Reaction Time to Errors between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards, all with $p\text{-values} < 0.01$. There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards both with $p\text{-values} < 0.01$. There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, both with $p\text{-values} < 0.001$. Finally, there was a significant difference found between the Pinch-Air and Static-Air keyboards with a $p\text{-value} = 0.0017$.

As expected, keyboards involving the 3rd-Dimension, increasing complexity, had significantly slower response times to errors than keyboards that did not. This implies that these keyboards required a higher level of focus as well as increased coupling between the motor space and display.

5.5.3 Reaction Time to Simulate Touch

Figure 5.20 shows the mean Reaction Time to Simulate Touch for each keyboard input device. Participants reached a mean Reaction Time to Simulate Touch of 1.24 s ($SD = 0.21$) for the Touch Screen Keyboard, 1.22 s ($SD = 0.40$) for the Leap Surface Keyboard, 2.60 s ($SD = 0.82$) for the Leap Static-Air Keyboard, 1.71 s ($SD = 0.37$) for the Leap Pinch-Air Keyboard, 2.23 s ($SD = 0.83$) for the Predictive-Air Keyboard, and 1.41 s ($SD = 0.23$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 19.6476$, $p\text{-value} = 1.1541e-13$, ($SD_{pooled} = 0.54$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.20, revealed significant differences in Reaction Time to Simulate Touch between the Touch Screen Keyboard and the Leap Static-Air and Predictive-Air keyboards, both with $p\text{-values} < 0.0001$. There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards both with $p\text{-values} < 0.0001$. There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, both with $p\text{-values} < 0.001$. Finally, a significant difference was detected between the Pinch-Air and Static-Air keyboards with a $p\text{-value} = 0.0001$.

Again, the reaction time to simulate a touch was significantly slower for the Static-Air and Predictive-Air keyboards due to the introduction of the 3rd-Dimension and increased coupling between the word-gesture motor space and the display.

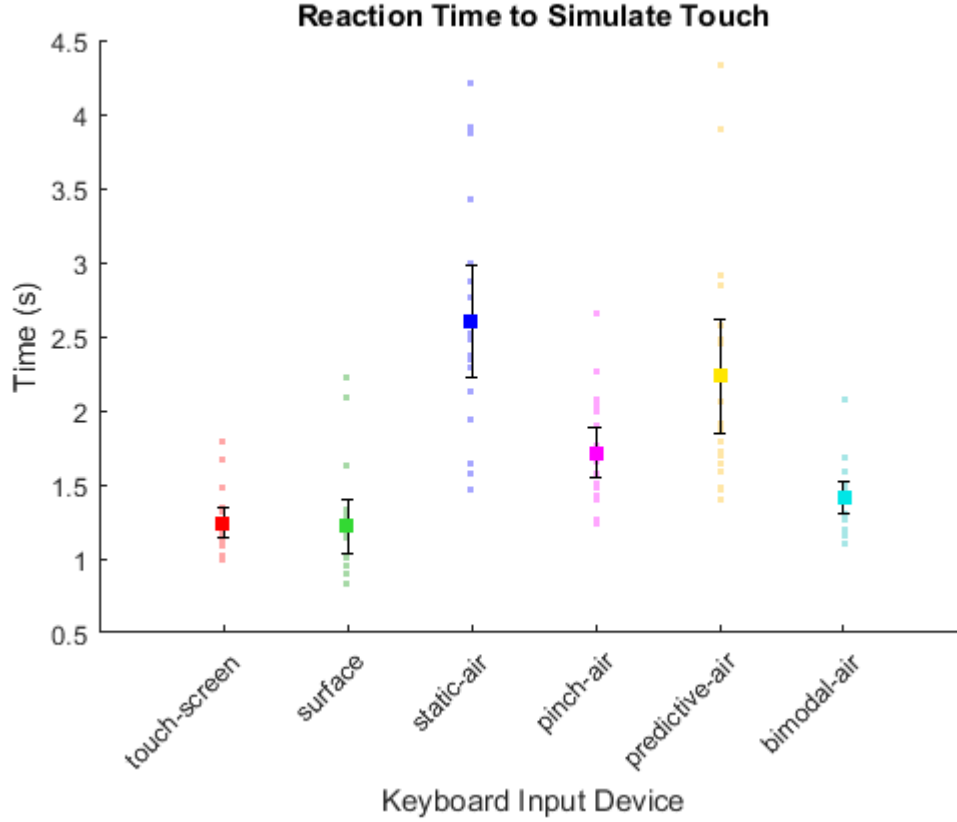


Figure 5.20. Mean Reaction Time to Simulate Touch for each keyboard with error bars showing 95% confidence intervals.

5.5.4 Reaction Time for First Correct Letter

Figure 5.21 shows the mean Reaction Time for First Correct Letter for each keyboard input device. Participants reached a mean Reaction Time for First Correct Letter of 1.44 s ($SD = 0.33$) for the Touch Screen Keyboard, 1.92 s ($SD = 0.70$) for the Leap Surface Keyboard, 3.68 s ($SD = 1.52$) for the Leap Static-Air Keyboard, 2.34 s ($SD = 0.68$) for the Leap Pinch-Air Keyboard, 3.40 s ($SD = 1.40$) for the Predictive-Air Keyboard, and 1.49 s ($SD = 0.31$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 18.3416$, $p\text{-value} = 5.9470e-13$, ($SD_{pooled} = 0.95$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

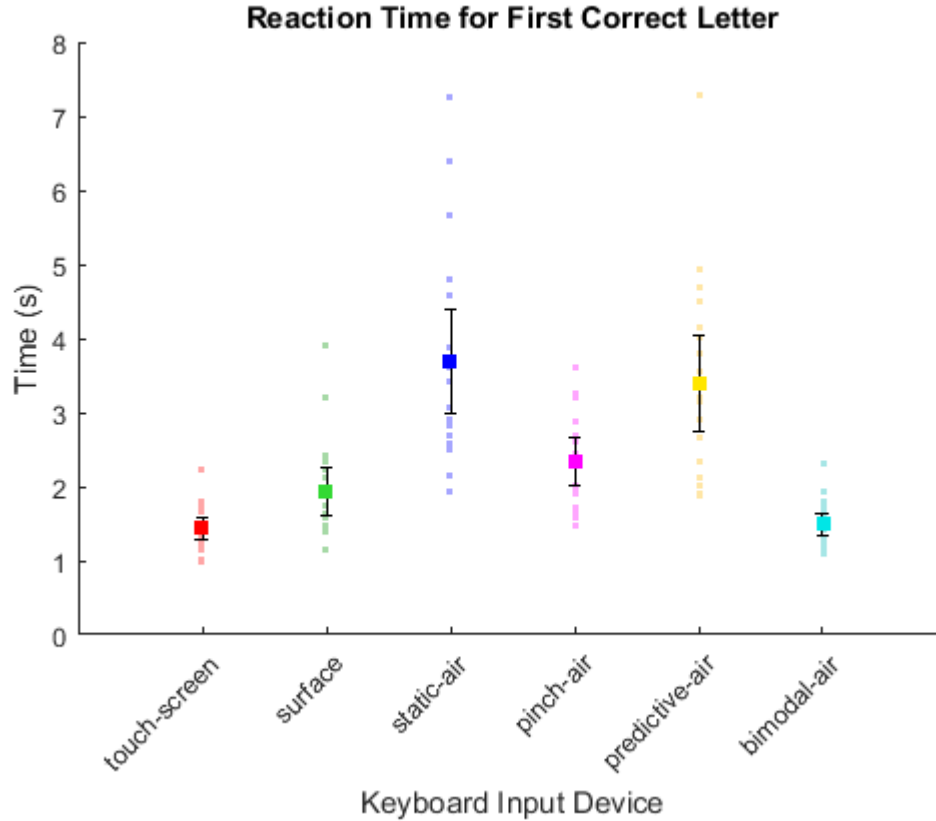


Figure 5.21. Mean Reaction Time for First Correct Letter for each keyboard with error bars showing 95% confidence intervals.

The multiple comparisons, seen in Table B.21, revealed significant differences in Reaction Time for First Correct Letter between the Touch Screen Keyboard and the Leap Static-Air and Predictive-Air keyboards, both with p -values < 0.0001 . There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards both with p -values < 0.001 . There were also significant differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, both with p -values < 0.0001 . Finally, there were significant differences found between the Leap Pinch-Air and the Static-Air and Predictive-Air keyboards both with p -values < 0.05 .

The Leap Bimodal-Air Keyboard performed at almost exactly the same rate as the Touch Screen Keyboard for hitting the first correct character of each word.

Although not a significant difference, it should also be noted that the Leap Surface Keyboard performed slightly slower than both the Bimodal-Air Touch Screen keyboards because participants had a tendency to look down at the printed keyboard. If participants had looked at the on-screen keyboard instead, there would have been a major increase in coupling and would have slowed the use of the Leap Surface Keyboard.

Again, the reaction time to simulate a touch was significantly slower for the Static-Air and Predictive-Air keyboards due to the introduction of the 3rd-Dimension and increased coupling between the word-gesture motor space and the display.

5.6 Quantitative Measures

5.6.1 Number of Touches Simulated

Figure 5.22 shows the mean Number of Touches Simulated for each keyboard input device. Participants reached a mean Number of Touches Simulated of 1.58 simulations ($SD = 0.79$) for the Touch Screen Keyboard, 1.62 simulations ($SD = 0.55$) for the Leap Surface Keyboard, 3.77 simulations ($SD = 2.21$) for the Leap Static-Air Keyboard, 2.42 simulations ($SD = 1.05$) for the Leap Pinch-Air Keyboard, 3.33 simulations ($SD = 1.62$) for the Predictive-Air Keyboard, and 1.67 simulations ($SD = 0.53$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means, $F(5, 102) = 10.0290$, $p\text{-value} = 7.9270e-08$, ($SD_{pooled} = 1.28$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.22, revealed significant differences in Number of Touches Simulated between the Touch Screen Keyboard and the Leap Static-Air and Predictive-Air keyboards, both with $p\text{-values} < 0.01$. There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards both with $p\text{-values} < 0.01$. There were also significant

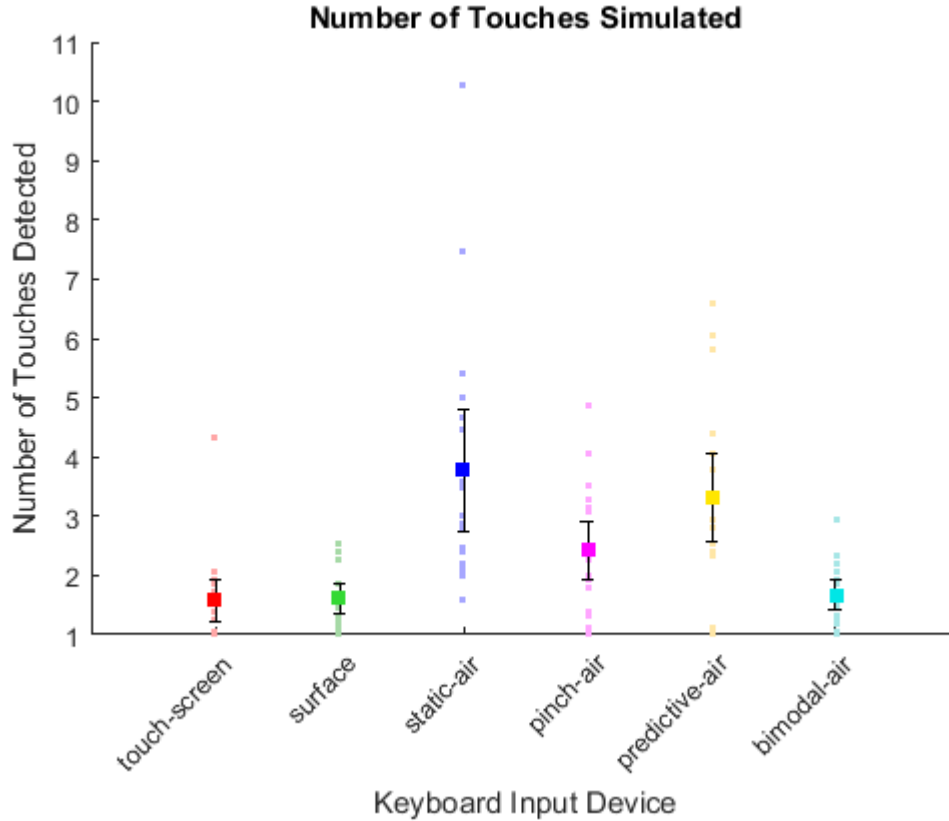


Figure 5.22. Mean Number of Touches Simulated for each keyboard with error bars showing 95% confidence intervals.

differences found between the Leap Bimodal-Air and the Static-Air and Predictive-Air keyboards, again both with p -values < 0.01 . Finally, there is a significant difference between the Pinch-Air keyboard and the Static-Air keyboard with a p -value = 0.0245.

Again, the same trends as before are seen. The Leap Surface and Bimodal-Air keyboards perform on par with the Touch Screen Keyboard, the Static-Air and Predictive-Air keyboards saw significantly more simulated touches, and the Pinch-Air keyboard was somewhere in the middle. It was observed by the researcher that the Static-Air keyboard suffered from a sort of "skimming" problem. The "skimming" problem is where participants would reach just far enough to barely simulate a touch on the interaction plane, and then when they would move their hand from one side of the keyboard to the other, due to natural arcing in the hand motion, participants'

hands would pull away from the interaction plane. This issue is expected to also be a culprit for some of the increased error rates.

5.6.2 Number of Practice Words

Figure 5.23 shows the mean Number of Practice Words for each keyboard input device. Participants reached a mean Number of Practice Words of 5.4 words ($SD = 4.3$) for the Touch Screen Keyboard, 7.7 words ($SD = 3.9$) for the Leap Surface Keyboard, 9.0 words ($SD = 6.7$) for the Leap Static-Air Keyboard, 8.1 words ($SD = 5.3$) for the Leap Pinch-Air Keyboard, 8.9 words ($SD = 6.8$) for the Predictive-Air Keyboard, and 10.2 words ($SD = 3.9$) for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, no significant differences were detected between keyboard means, $F(5, 102) = 1.6828$, $p\text{-value} = 0.1454$, ($SD_{pooled} = 5.3$), therefore no additional analysis was made. Typically, as observed by the researcher, there were two categories of participants when it came to performing practice words. The first category were participants who would perform a full set of 15 words for each keyboard. Rarely, participants felt the need to perform more than one practice set worth of words. The second category were participants who progressively performed less practice words for each consecutive keyboard, this is an example of why a Replicated Latin Squares design for counterbalancing was used.

5.7 Qualitative Measures

5.7.1 Level of Discomfort

Figure 5.24 shows the median Level of Discomfort for each keyboard input device rated by the participants. Keyboards reached a median Level of Discomfort of "Strongly Disagree" for the Touch Screen Keyboard, "Strongly Disagree" for the Leap Surface Keyboard, "Neutral" for the Leap Static-Air Keyboard, "Disagree" for the Leap Pinch-Air Keyboard, between "Disagree" and "Neutral" for the Predictive-Air Keyboard, and "Disagree" for the Leap Bimodal-Air Keyboard. Using a One-Way

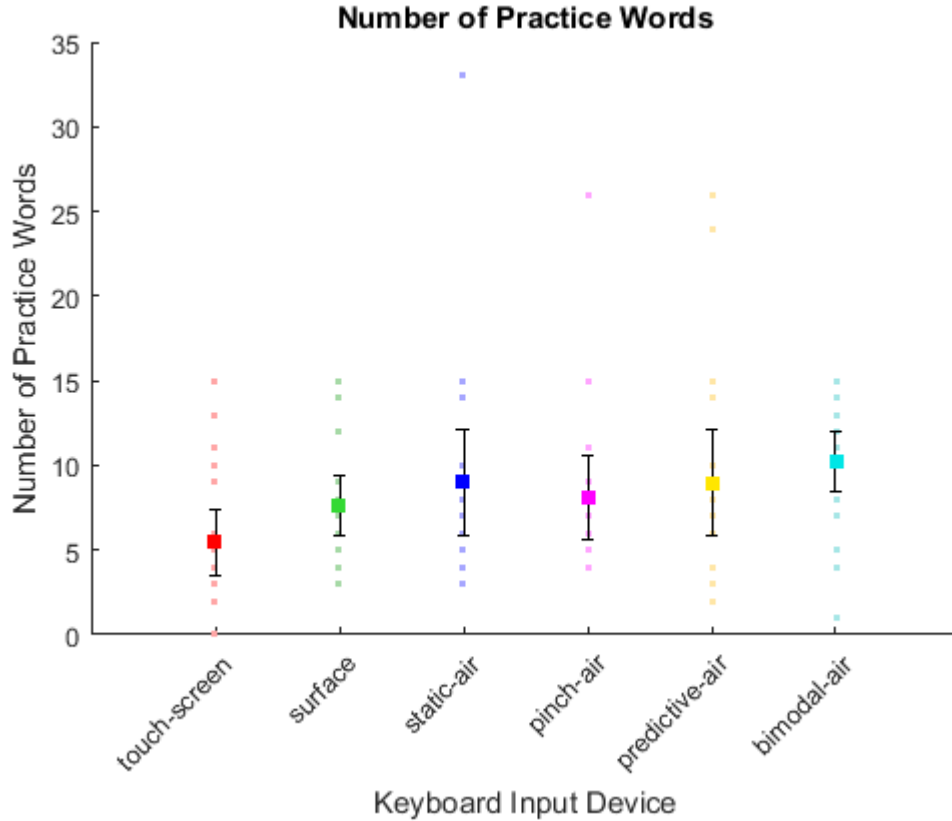


Figure 5.23. Mean Number of Practice Words for each keyboard with error bars showing 95% confidence intervals.

Analysis of Variance, a significant difference was detected between keyboard means for Level of Discomfort, $F(5, 102) = 7.5000$, $p\text{-value} = 4.9726e-06$, ($SD_{pooled} = 0.94$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.23, revealed significant differences in Level of Discomfort between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards, all with $p\text{-values} < 0.05$. There were significant differences between the Leap Surface Keyboard and the Static-Air and Predictive-Air keyboards both with $p\text{-values} < 0.01$. There was also a significant difference found between the Leap Bimodal-Air and the Static-Air keyboards with a $p\text{-value} = 0.0232$.

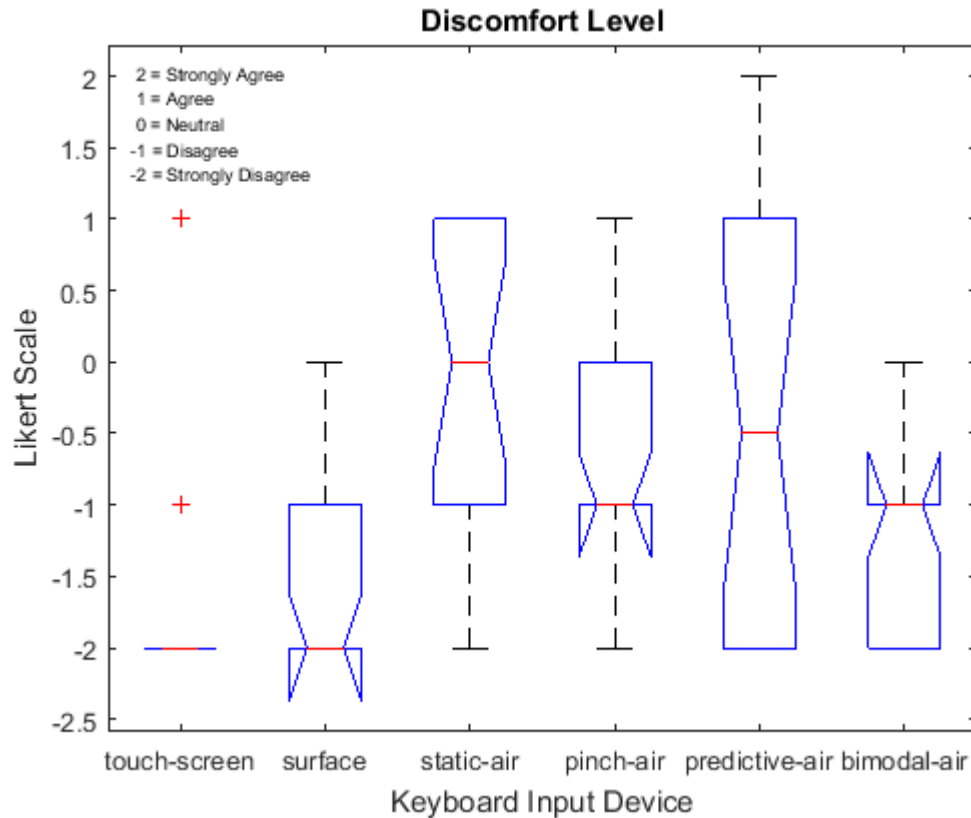


Figure 5.24. Median Level of Discomfort for each keyboard showing the 25th and 75th percentiles.

The level of discomfort experienced by participants, not surprisingly, was very similar to the trend seen in most other dependent measures. Discomfort was experienced more on the mid-air keyboards, and more extremely by the keyboards that utilized the 3rd-Dimension.

5.7.2 Level of Fatigue

Figure 5.25 shows the median Level of Fatigue for each keyboard input device rated by the participants. Keyboards reached a median Level of Fatigue of "Strongly Disagree" for the Touch Screen Keyboard, "Strongly Disagree" for the Leap Surface Keyboard, "Agree" for the Leap Static-Air Keyboard, between "Neutral" and "Agree" for the Leap Pinch-Air Keyboard, "Neutral" for the Predictive-Air Keyboard,

and "Disagree" for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means for Level of Discomfort, $F(5, 102) = 10.9910$, $p\text{-value} = 1.7659e-08$, ($SD_{pooled} = 1.12$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

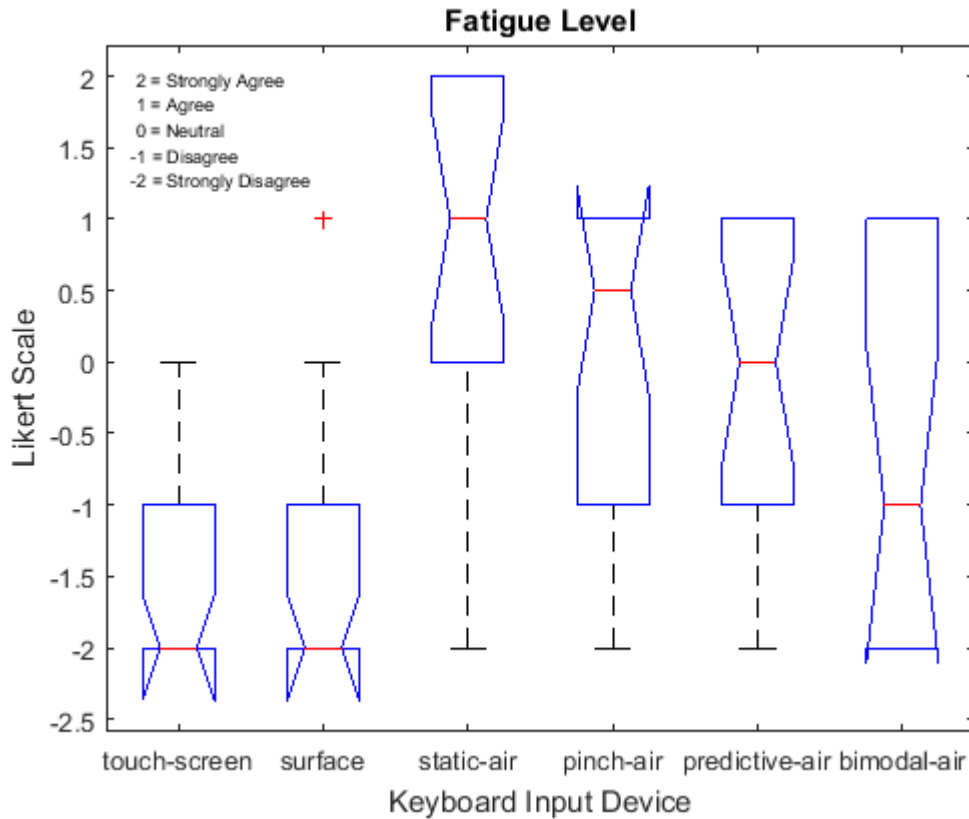


Figure 5.25. Median Level of Fatigue for each keyboard showing the 25th and 75th percentiles.

The multiple comparisons, seen in Table B.24, revealed significant differences in Level of Fatigue between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards, all with $p\text{-values} < 0.01$. There were significant differences between the Leap Surface Keyboard and the Static-Air, Pinch-Air, and Predictive-Air keyboards all with $p\text{-values} < 0.05$. There was also a significant

difference found between the Leap Bimodal-Air and the Static-Air keyboards with a p -value = 0.0043.

As expected, the fatigue levels experienced for mid-air keyboards were typically higher than those that were not mid-air.

5.7.3 Level of Difficulty

Figure 5.26 shows the median Level of Difficulty for each keyboard input device rated by the participants. Keyboards reached a median Level of Difficulty of "Strongly Disagree" for the Touch Screen Keyboard, "Strongly Disagree" for the Leap Surface Keyboard, between "Disagree" and "Neutral" for the Leap Static-Air Keyboard, "Disagree" for the Leap Pinch-Air Keyboard, "Disagree" for the Predictive-Air Keyboard, and "Disagree" for the Leap Bimodal-Air Keyboard. Using a One-Way Analysis of Variance, a significant difference was detected between keyboard means for Level of Discomfort, $F(5, 102) = 6.0351$, p -value = $6.2051e-05$, ($SD_{pooled} = 0.98$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.25, revealed significant differences in Level of Difficulty between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards, all with p -values < 0.05 . There was significant difference between the Leap Surface and the Static-Air keyboards with a p -values = 0.0042. There was also a significant difference found between the Leap Bimodal-Air and the Static-Air keyboards with a p -value = 0.0209.

As anticipated, the mid-air keyboards, had a higher level of difficulty and were harder to use or understand than the non mid-air keyboards.

5.7.4 Preference Ranking

Figure 5.27 shows the mean Preference Ranking for each keyboard input device. Participants reached a mean Preference Ranking of 1.89 ($SD = 1.18$) for the Touch Screen Keyboard, 2.22 ($SD = 1.22$) for the Leap Surface Keyboard, 5.33

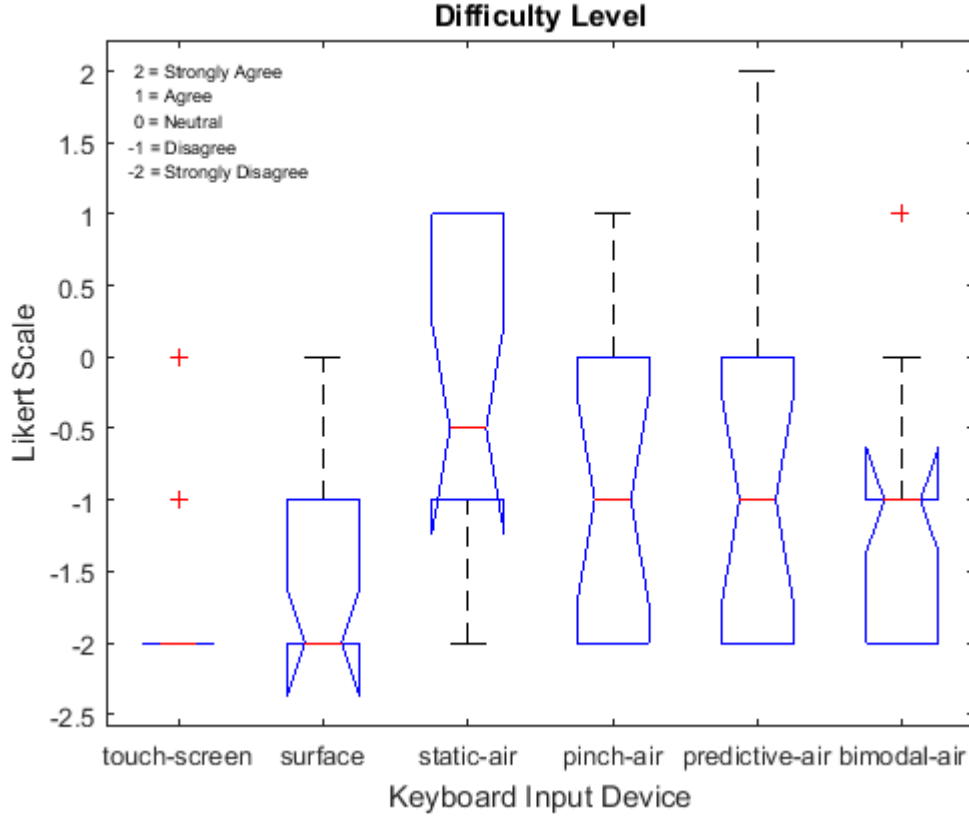


Figure 5.26. Median Level of Difficulty for each keyboard showing the 25th and 75th percentiles.

($SD = 0.84$) for the Leap Static-Air Keyboard, 4.22 ($SD = 1.06$) for the Leap Pinch-Air Keyboard, 4.50 ($SD = 1.42$) for the Predictive-Air Keyboard, and 2.83 ($SD = 1.29$) for the Leap Bimodal-Air Keyboard. Using Friedman's rank-test, Analysis of Variance, a significant difference was detected between keyboard means, $\chi^2(5, 85) = 49.1429$, $p\text{-value} = 2.0749e-09$, ($SD_{pooled} = 1.87$), prompting the use of Tukey's HSD with multiple-compare for further analysis.

The multiple comparisons, seen in Table B.26, revealed significant differences in Preference Ranking between the Touch Screen Keyboard and the Leap Static-Air, Pinch-Air, and Predictive-Air keyboards, all with $p\text{-values} < 0.01$. There were significant differences between the Leap Surface Keyboard and the Static-Air, Pinch-Air, and Predictive-Air keyboards all with $p\text{-values} < 0.05$. There were also significant

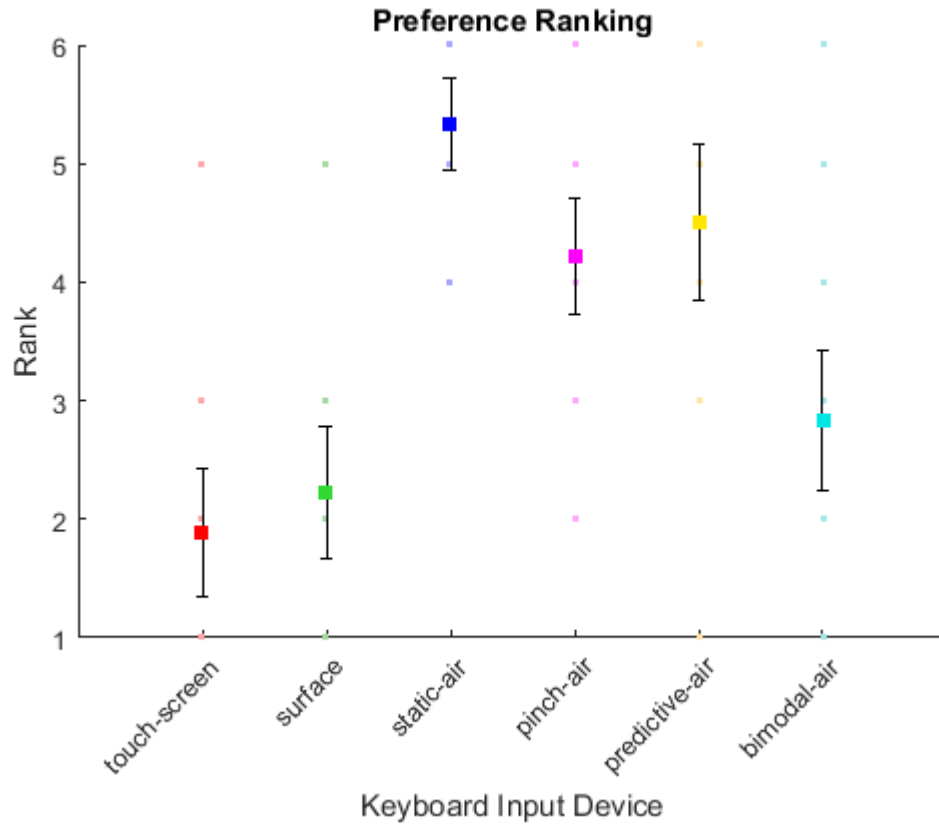


Figure 5.27. Mean Preference Rankings for each keyboard with error bars showing 95% confidence intervals.

differences found between the Leap Bimodal-Air and the Static-Air keyboards with a p -value = 0.0009.

The preference ranking of all of the keyboards very strongly reflects the trends seen in most of the other dependent measures. The Touch Screen Keyboard was the most preferred as expected, followed by the Leap Surface in a close second. The Leap Bimodal-Air was the most preferred mid-air keyboard whereas the Static-Air Keyboard was the least preferred.

CHAPTER SIX

Discussion, Future Work and Conclusion

6.1 Discussion

Vulture has shown that word-gesture keyboards are beneficial to mid-air text-entry by reaching a text-entry rate of 28 WPM, surpassing earlier work, however it utilizes pinching as a means of word separation TODO: VULTURE REFERENCE. This study aimed to find alternative solutions to word separation and interaction with a mid-air, word-gesture keyboard. Rather than pinching or inconvenient glove-use, this study utilized the 3rd-Dimension as well as a bimodal approach.

This study suffered from some of the same issues that were experienced in Vulture, text-entry rates were much slower for mid-air keyboards than the Touch Screen Keyboard. One reason for this is that participants are required to couple the gestures in motor space with those on the display TODO: REF VULTURE 24. This affect was described in detail for many of the results in Chapter 5. Another reason, as seen in Vulture and in other studies TODO: REF VULTURE AND VULTURE 32, participants heavily rely on the displayed feedback. This was seen mostly due to the small latency introduced when using the Leap Motion to track and display hand motions on the screen. Some participants commented having to slow down to ensure the display more accurately represented their hand movements. If movements were made that were too fast, the visual display would slightly lag behind and so some participants had a tendency to overshoot letters. A final reason, like in Vulture, is that many of the alternatives add to the complexity of text-entry by the forcing explicit delimiting of words.

6.1.1 *Pinching Interaction*

This study utilized a pinching method using the Leap Motion in order to have a metric for comparison against Vulture’s pinching since this study used a pseudo-implementation of word-gesturing, lacking word-recognition. The text-entry rate for pinching with a single session was 11.3 WPM which was consistent with the mean text-entry rate using Vulture ($M = 11.8$ WPM) for a single session TODO: REFERENCE VULTURE. This gave a solid measure for alternative solutions to compare against.

6.1.2 *3-Dimensional Interaction*

This study showed initial text-entry rates for utilizing the 3rd-Dimension as a means of word separation at 8.6 WPM for the Static-Air approach and 9.6 WPM for the Predictive-Air approach. These approaches underperformed because of the difficulty in coupling word-gestures in motor space with those displayed on the screen with the added factor of having to move in the 3rd-Dimension and interact with an invisible plane to delimit words.

The Predictive-Air Keyboard was not significantly worse than the Pinch-Air Keyboard and needs to be investigated further with a repeated-measures, multiple session study. Both the Static-Air and Predictive-Air suffered from an issue referred to as "skimming". The "skimming" issue is when a participant reaches towards the interaction plane and simulates a touch but as they move their hand during the gesturing motion, natural arcing of the participant’s hand causes them to lose the simulated touch TODO: GET PICTURE OF THIS. The natural arcing motion of the participants also seemed to cause a lot of errors after finishing writing a word because participants would pull their hand back in an arcing motion, especially if their elbow was rested. This would cause their finger to hit the key above the current key they were on and count it as a press TODO: GET PICTURE OF THIS TOO.

It’s very interesting to note that the Static-Air implementation was used to project a mid-air keyboard onto flat surfaces, simulating a touch screen. This was the

Leap Surface Keyboard. The Leap Surface was essentially indistinguishable from the Touch Screen Keyboard in terms of performance and text-entry rate, reaching 17.1 WPM, for a single session. The Leap Surface Keyboard handily proves how essential it is to avoid having a decoupled input plane and display plane. This hints that the Static-Air Keyboard may substantially benefit from an input plane displayed using augmented reality.

6.1.3 Bimodal Interaction

Bimodal mid-air interactions are shown to be very promising in this study and it should be noted that any other secondary input can be used as the source for touch simulation. The Bimodal-Air Keyboard reached a text-entry rate of 15.8 WPM for a single session which is a significant improvement over using pinching or utilizing the 3rd-Dimension. The Bimodal-Air Keyboard was also often times indistinguishable from the Touch Screen Keyboard for many other dependent measures, all detailed in Chapter 5. The Bimodal-Air keyboard needs to be further investigated in a repeated-measures, multiple session study on a word-gesture keyboard implemented with word-recognition.

The Bimodal-Air isn't without some of its own drawbacks though, as experienced with the other mid-air keyboards using the Leap Motion, there is an associated plane that the gesture is moving across as the participant moves their hand. This plane, if calibrated or oriented incorrectly, can lead to higher error rates and less precision overall.

6.1.4 Lacking Word-Recognition

There are some glaring limitations to using a pseudo-implementation of word-gesturing as opposed to using a full word-recognition implementation. A major difference was that the best-guess implementation analyzed the gesture as it was being created, showing participants real-time updates of what was happening to the path

they were creating. As the gesture was being created, the system attempted to guess the current character being pressed based on the deviations in the gesture and then compared those deviations with the current characters that the participant was trying to achieve and against the previously detected deviation. The limitation that this imposes, seeing real-time character-level updates, is that most participants would interrupt their current word-gesture in order to fix the errors being created mid-gesture which is unnatural for word-gesturing. As well, the fidelity for making errors was much higher, allowing these interruptions to sometimes occur frequently. The lack of word-recognition means that words that aren't real could be produced through gesturing which again is counter-intuitive and once one mistake was made, the likelihood to see more errors increased.

6.2 *Future Work*

There are many things to consider when looking at improves to be made or additional studies that can be conducted.

6.2.1 *Implement Word-Recognition*

The most important change to this study that needs to be explored is to re-conduct the study using a standardized word-gesture keyboard implementation with word-recognition. This single improvement will give better and more standardized results across the board. There would also be no need for any modified variables, and less error measures could have been observed. There should also be an increase in text-entry rates since participants won't be distracted with errors mid-gesture.

6.2.2 *Redesign Task*


The task for the study needs to be substantially redesigned to fit a word-recognition implementation of the word-gesture keyboards. First, the trials need to be phrases instead of single words to better track text-entry and error rates. Next, there shouldn't be any requirement to fix words, it should just feel natural for the

participant and they can fix errors where they feel it's necessary. The number of trials needs to be increased but a significant amount and as well there study needs to be designed for repeated sessions so that improvements in text-entry rate can be seen. With repeated sessions and actual word-recognition, the Bimodal-Air keyboard is expected to achieve text-entry rates superior to those found in Vulture.

6.2.3 *Augmented Reality*

In an effort to decrease the mental coupling required between the gesture motor-space and the keyboard display, augmented reality would be a huge boon for 3-Dimensional mid-air implementations. This is theorized to benefit the Static-Air Keyboard the most due to the success of the Leap Surface Keyboard which is just the Static-Air keyboard projected onto a surface. Participants being able to see the keyboard they are using would be expected to substantially increase text-entry rates.

6.2.4 *The "Skimming" Issue*

As a first thought to solve the "skimming" issue, there is a seemingly obvious solution to the problem. Simply, once a simulated touch has been made, increase the interaction plane's threshold in the participant's direction, enough so that the participant's hand doesn't leave the surface when moving from side to side or up and down.  TODO GET A PICTURE COMPARING SKIMMING PROBLEM TO THIS SOLUTION. This change should see improved results for the Static-Air Keyboard since there would be many less times where the user accidentally exits the interaction plane. A similar approach could also be used for the Predictive-Air because sometimes quick side to side movements were seen as a touch being released.

6.2.5 *Better Tracking*

A simple observation that was made by the researcher is that there were many times that the Leap Motion had issues detecting people's hands and palm. The main reasoning for this was the positioning of the Leap Motion itself as well as some

participants holding their hand in an upward position TODO: SHOW UPWARD POINTING HAND BLOCKING LEAP. A simple solution to this issue is to angle the leap so that so that it is facing the participant at a better angle TODO: SHOW LEAP AT AN ANGLE.

6.2.6 *Alternative Interaction Plane*

Something that was observed with all mid-air keyboards is that as the participants moved across the interaction plane, sometimes the side-to-side movements and up and down movements weren't represented as expected TODO: SHOW FIGURE THAT AN UP AND DOWN MOTION TRANSLATES TO AN ANGULAR MOTION. Better approaches to calibration as in and plane-creation Personal Space TODO: REFERENCE ALVIN need to be used such as those used and mentioned in TODO: REFERENCE ALVIN AND DARREN. A spherical or even curved interaction plane designed for the individual participant would greatly benefit both the all of the mid-air keyboards especially so that participants would be able to rest their arms. With a flat, quadrilateral plane, many participants suffered the same issues as mentioned in TODO: REFERENCE ALVIN, especially when moving to the extreme edges of the projected keyboard.

6.2.7 *Multi-functional Space*

Utilizing the 3rd-Dimension for mid-air word-gesturing keyboards, when not touching the interaction plane to simulate touch, hand gestures can be used to control a cursor or for other gestures. This would allow users to fully interact with screens of any kind, including projections, and not limit the user to a keyboard and mouse.

6.2.8 *Image Processing*

After using the Leap Surface Keyboard and projecting a virtual keyboard over just a printed sheet, an addition that could be added to make the feature much more

versatile is image processing. Being able to print and use any keyboard for word gesturing or otherwise could be interesting and highly beneficial.

6.2.9 Gaming Console Keyboards

An afterthought from using the Xbox Controller Keyboard in the pre-pilot and the pilot study, it would be very beneficial to apply what has been learned from this study and word-gesturing in general to gaming console keyboards. There are two options for implementing a word-gesture keyboard for gaming consoles. The first is to implement a mid-air word-gesture keyboard using the Xbox Kinect or using the Wii Remotes. The second option would be to transition from the standard console keyboards to a word-gesture keyboard just using standard console controllers. The word-gesture keyboard would most likely have to be bimodal. The user would hold the 'A' button while simulating touch and use the thumb stick to move a cursor around for word-gesturing. Either method would be a great improvement over single character text-entry currently seen on modern gaming consoles.

6.2.10 Accessibility

A major motivation for this research was the idea of applying it for amputees or those with disabilities that affect their performance using a standard keyboard. A proper study in accessibility should be performed to utilize the Bimodal-Air keyboard for users with disabilities.

6.3 Conclusion

Word-gesture keyboards show a promising means to efficient, mid-air text-entry by tracing word-gestures instead of slow, single-input text-entry. TODO: REFERENCE VULTURE. This study, demonstrated alternative ways to delimit the separation of words for mid-air, word-gesture keyboards. It was shown that utilizing the 3rd-Dimension as a means of word separation is too complex to be beneficial when paired with a decoupled gesture-space and keyboard display. As demonstrated by

projecting a Static-Air plane onto a surface with the Leap Surface Keyboard, better results may be achieved when using the 3rd-Dimension as a means of word-separation if they are complimented by an augmented reality design to re-couple the gesture-space and keyboard display. Another method to lessen the severity of a decoupled gesture-space and keyboard display, the interaction plane needs to use new techniques to be better calibrated for each individual user and to display more accurate results and control TODO: REFERENCE DARREN AND ALVIN. Finally, the empirical results from this study shows that using a bimodal technique as a means of word separation would greatly benefit mid-air, word-gesture keyboards. The Bimodal-Air keyboard was slower in text-entry rates than the Touch Screen Keyboard, however, the bimodal method was near indistinguishable from the touch screen for almost all other empirical results and was also better than pinching for nearly all results. This study needs to have a follow-up study with redesigned trials using a full word-gesture keyboard implementation with repeated measures and reoccurring sessions to further investigate bimodal techniques for mid-air text-entry. A bimodal approach just might be the future of mid-air text-entry for word-gesture keyboards.

APPENDICES

APPENDIX A
Word-Gesturing Implementation

APPENDIX B

Results Tables and Boxplots

The statistical methods used consisted of One-Way ANOVAs for each set of the dependent measures, followed by using Tukey's Honest Significant Difference (HSD) for multiple-compare for the post-hoc analysis. The ranking system from the exit survey used the Friedman's test for analysis in conjunction with Tukey's HSD for a post-hoc analysis.

B.1 Text-Entry Rate

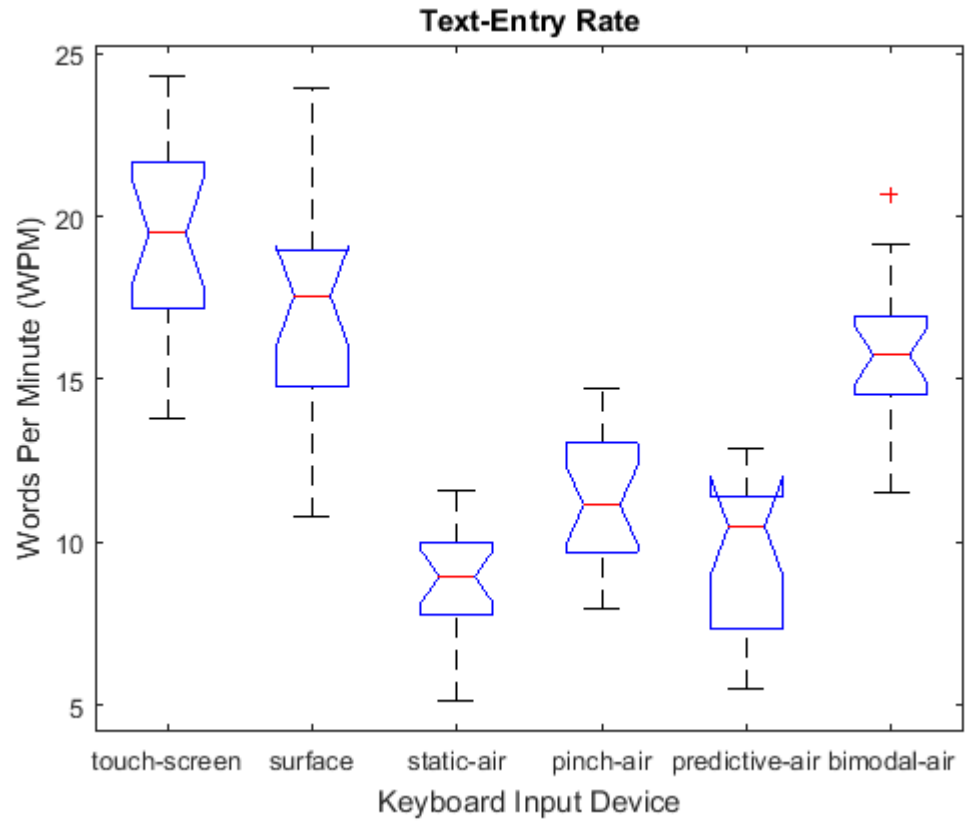


Figure B.1. Median Text-Entry Rates for each keyboard showing the 25th and 75th percentiles.

Table B.1. Tukey's Honest Significant Difference with multiple comparison for Text-Entry Rate.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-0.0837	2.3522	4.7882	0.0648
touchscreen	static-air	8.4479	10.8839	13.3198	0.0000
touchscreen	pinch-air	5.6920	8.1280	10.5639	0.0000
touchscreen	predictive-air	7.4227	9.8587	12.2946	0.0000
touchscreen	bimodal-air	1.2832	3.7192	6.1551	0.0003
surface	static-air	6.0957	8.5316	10.9676	0.0000
surface	pinch-air	3.3398	5.7757	8.2117	0.0000
surface	predictive-air	5.0705	7.5064	9.9424	0.0000
surface	bimodal-air	-1.0690	1.3669	3.8029	0.5809
static-air	pinch-air	-5.1919	-2.7559	-0.3199	0.0170
static-air	predictive-air	-3.4612	-1.0252	1.4108	0.8249
static-air	bimodal-air	-9.6007	-7.1647	-4.7287	0.0000
pinch-air	predictive-air	-0.7053	1.7307	4.1667	0.3146
pinch-air	bimodal-air	-6.8448	-4.4088	-1.9728	0.0000
predictive-air	bimodal-air	-8.5755	-6.1395	-3.7035	0.0000

B.1.1 Text-Entry Rate Modified-Shortest

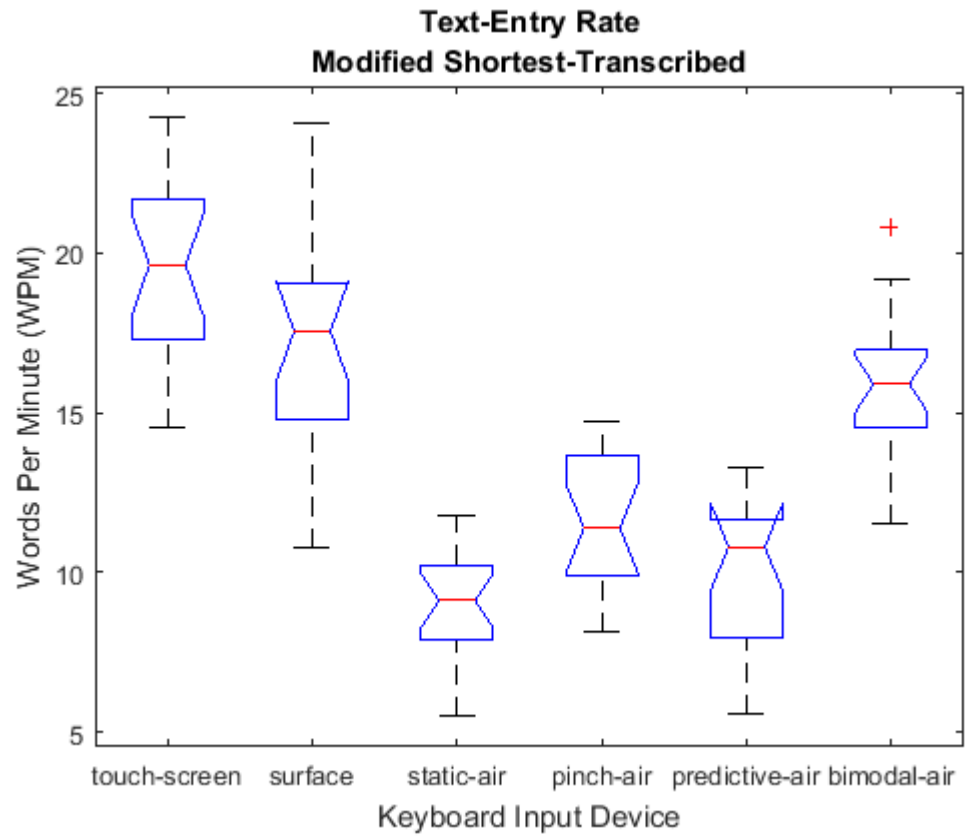


Figure B.2. Median Text-Entry Rates using the shortest-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.2. Tukey's Honest Significant Difference with multiple comparison for Text-Entry Rate modified with shortest-transcribed.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	p -value
		Lower End		Upper End	
touchscreen	surface	-0.0162	2.4063	4.8288	0.0526
touchscreen	static-air	8.4069	10.8295	13.2520	0.0000
touchscreen	pinch-air	5.6271	8.0496	10.4722	0.0000
touchscreen	predictive-air	7.3096	9.7321	12.1546	0.0000
touchscreen	bimodal-air	1.3652	3.7877	6.2103	0.0002
surface	static-air	6.0006	8.4232	10.8457	0.0000
surface	pinch-air	3.2208	5.6433	8.0659	0.0000
surface	predictive-air	4.9033	7.3258	9.7483	0.0000
surface	bimodal-air	-1.0411	1.3814	3.8040	0.5636
static-air	pinch-air	-5.2024	-2.7798	-0.3573	0.0148
static-air	predictive-air	-3.5199	-1.0974	1.3251	0.7757
static-air	bimodal-air	-9.4643	-7.0417	-4.6192	0.0000
pinch-air	predictive-air	-0.7401	1.6825	4.1050	0.3399
pinch-air	bimodal-air	-6.6844	-4.2619	-1.8394	0.0000
predictive-air	bimodal-air	-8.3669	-5.9443	-3.5218	0.0000

B.2 Error Rates

B.2.1 Minimum Word Distance

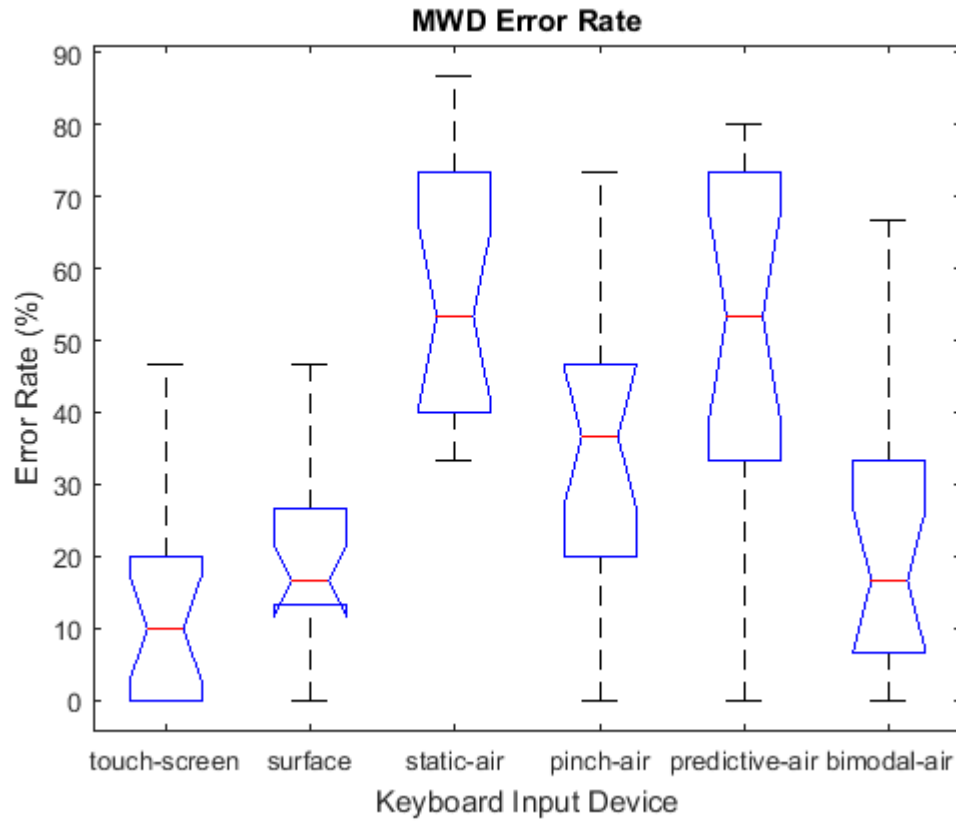


Figure B.3. Median Minimum Word Distance for each keyboard showing the 25th and 75th percentiles.

Table B.3. Tukey's Honest Significant Difference with multiple comparison for Minimum Word Distance.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	p -value
		Lower End		Upper End	
touchscreen	surface	-23.8140	-5.9259	11.9622	0.9287
touchscreen	static-air	-61.5918	-43.7037	-25.8156	0.0000
touchscreen	pinch-air	-39.3696	-21.4815	-3.5934	0.0091
touchscreen	predictive-air	-53.8140	-35.9259	-18.0378	0.0000
touchscreen	bimodal-air	-25.2955	-7.4074	10.4807	0.8346
surface	static-air	-55.6659	-37.7778	-19.8897	0.0000
surface	pinch-air	-33.4437	-15.5556	2.3326	0.1263
surface	predictive-air	-47.8881	-30.0000	-12.1119	0.0001
surface	bimodal-air	-19.3696	-1.4815	16.4066	0.9999
static-air	pinch-air	4.3341	22.2222	40.1103	0.0062
static-air	predictive-air	-10.1103	7.7778	25.6659	0.8043
static-air	bimodal-air	18.4082	36.2963	54.1844	0.0000
pinch-air	predictive-air	-32.3326	-14.4444	3.4437	0.1859
pinch-air	bimodal-air	-3.8140	14.0741	31.9622	0.2096
predictive-air	bimodal-air	10.6304	28.5185	46.4066	0.0002

B.2.1.1 MWD Modified-Shortest.

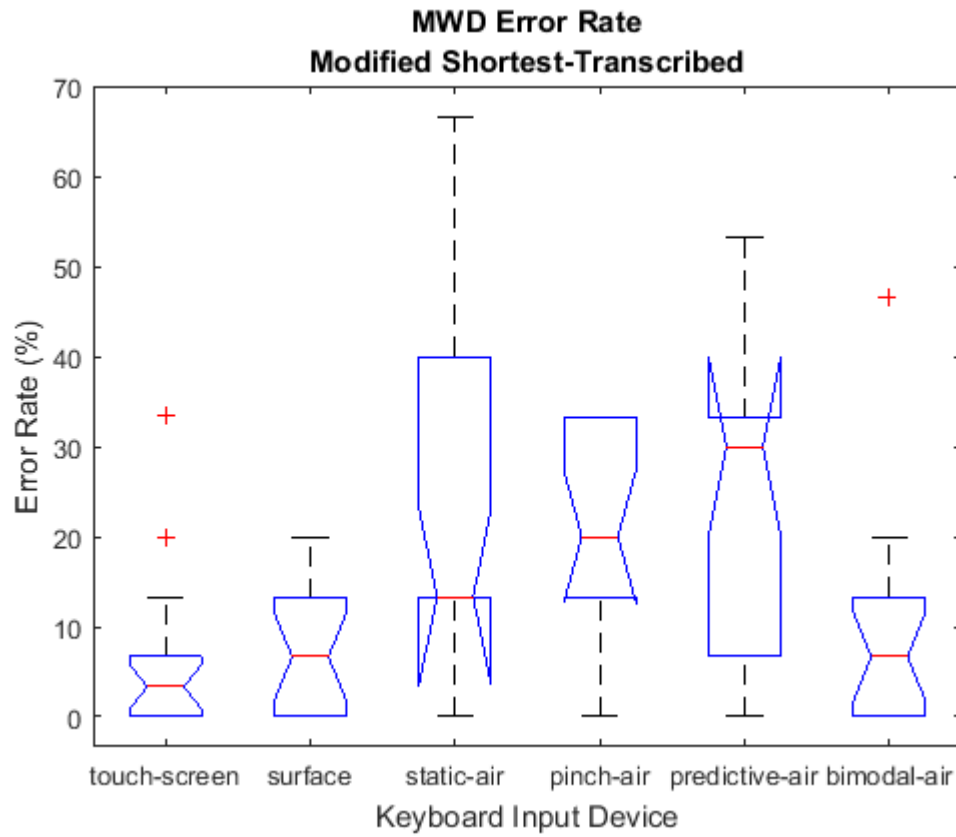


Figure B.4. Median Minimum Word Distance using the shortest-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.4. Tukey's Honest Significant Difference with multiple comparison for Minimum Word Distance using shortest-transcribed modification.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-13.0458	0.0000	13.0458	1.0000
touchscreen	static-air	-32.3051	-19.2593	-6.2135	0.0006
touchscreen	pinch-air	-26.7495	-13.7037	-0.6579	0.0336
touchscreen	predictive-air	-32.6754	-19.6296	-6.5838	0.0004
touchscreen	bimodal-air	-16.7495	-3.7037	9.3421	0.9623
surface	static-air	-32.3051	-19.2593	-6.2135	0.0006
surface	pinch-air	-26.7495	-13.7037	-0.6579	0.0336
surface	predictive-air	-32.6754	-19.6296	-6.5838	0.0004
surface	bimodal-air	-16.7495	-3.7037	9.3421	0.9623
static-air	pinch-air	-7.4902	5.5556	18.6013	0.8177
static-air	predictive-air	-13.4162	-0.3704	12.6754	1.0000
static-air	bimodal-air	2.5098	15.5556	28.6013	0.0099
pinch-air	predictive-air	-18.9717	-5.9259	7.1199	0.7737
pinch-air	bimodal-air	-3.0458	10.0000	23.0458	0.2349
predictive-air	bimodal-air	2.8801	15.9259	28.9717	0.0076

B.2.2 Keystrokes Per Character

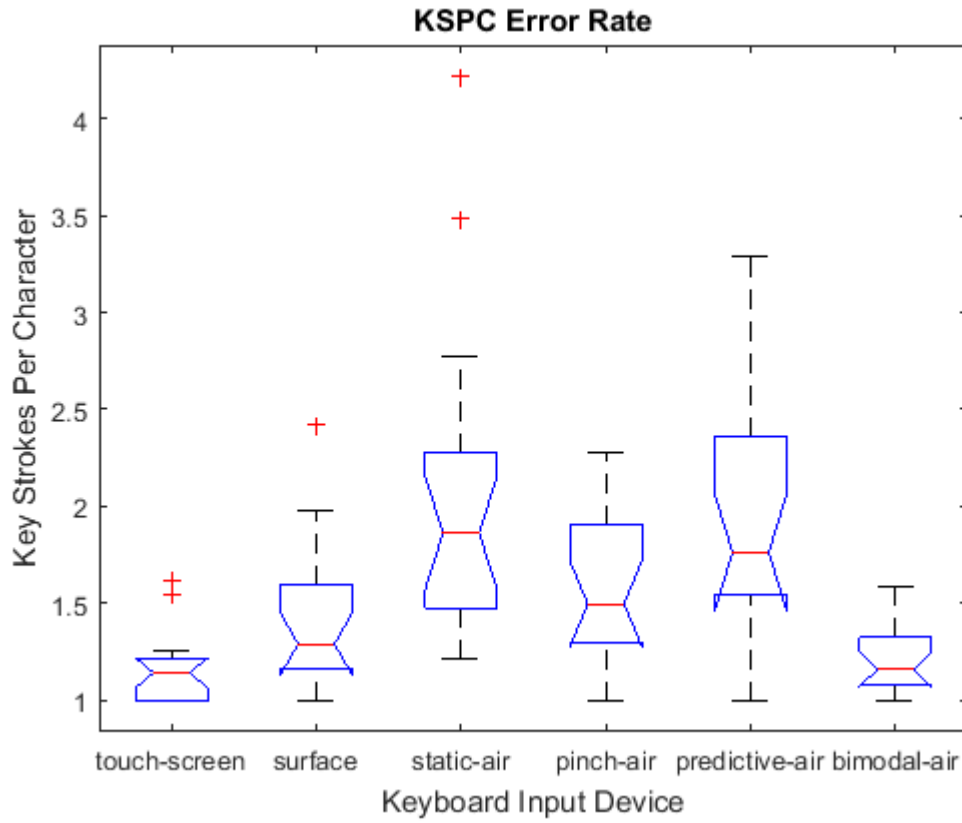


Figure B.5. Median Keystrokes Per Character for each keyboard showing the 25th and 75th percentiles.

Table B.5. Tukey's Honest Significant Difference with multiple comparison for Keystrokes Per Character.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-0.7298	-0.2597	0.2103	0.5972
touchscreen	static-air	-1.3559	-0.8859	-0.4158	0.0000
touchscreen	pinch-air	-0.8998	-0.4298	0.0403	0.0935
touchscreen	predictive-air	-1.2113	-0.7412	-0.2712	0.0002
touchscreen	bimodal-air	-0.5324	-0.0623	0.4077	0.9989
surface	static-air	-1.0962	-0.6262	-0.1561	0.0026
surface	pinch-air	-0.6401	-0.1701	0.3000	0.8993
surface	predictive-air	-0.9515	-0.4815	-0.0114	0.0414
surface	bimodal-air	-0.2726	0.1974	0.6675	0.8262
static-air	pinch-air	-0.0140	0.4561	0.9262	0.0626
static-air	predictive-air	-0.3254	0.1447	0.6147	0.9471
static-air	bimodal-air	0.3535	0.8236	1.2936	0.0000
pinch-air	predictive-air	-0.7815	-0.3114	0.1586	0.3935
pinch-air	bimodal-air	-0.1026	0.3675	0.8375	0.2157
predictive-air	bimodal-air	0.2089	0.6789	1.1490	0.0008

B.2.2.1 KSPC Modified-Shortest.

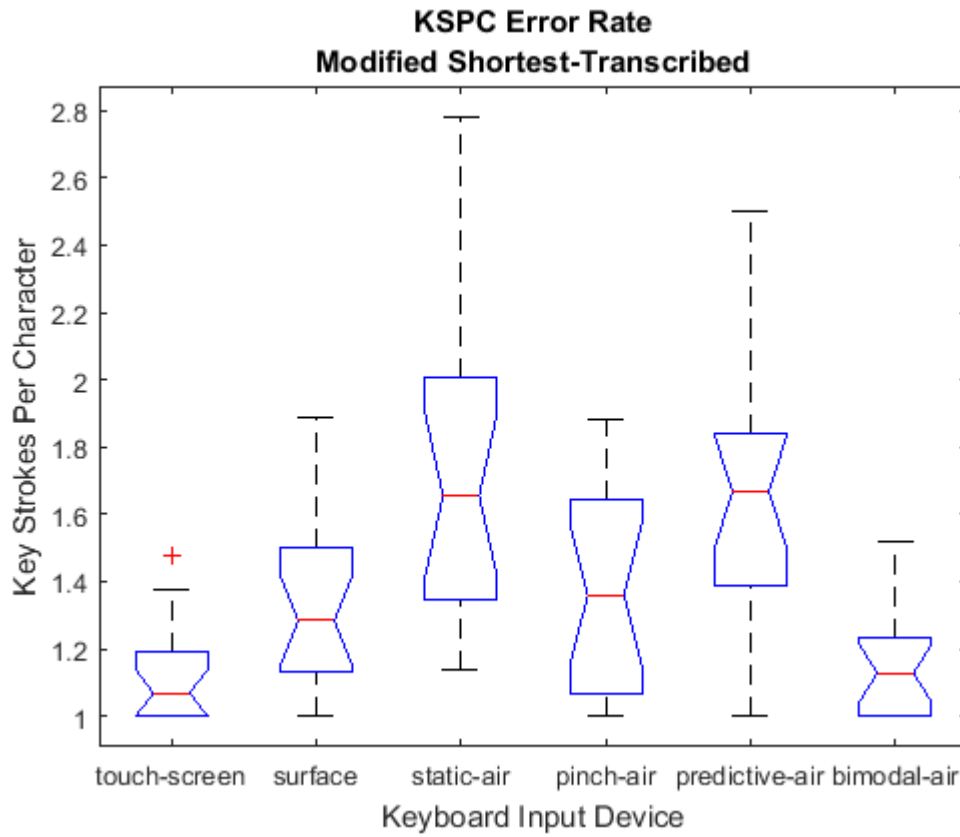


Figure B.6. Median Keystrokes Per Character using the shortest-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.6. Tukey’s Honest Significant Difference with multiple comparison for Keystrokes Per Character using shortest-transcribed modification.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	p -value
		Lower End		Upper End	
touchscreen	surface	-0.5181	-0.2055	0.1070	0.4020
touchscreen	static-air	-0.9366	-0.6240	-0.3115	0.0000
touchscreen	pinch-air	-0.5532	-0.2407	0.0719	0.2304
touchscreen	predictive-air	-0.8488	-0.5363	-0.2237	0.0000
touchscreen	bimodal-air	-0.3346	-0.0220	0.2906	0.9999
surface	static-air	-0.7310	-0.4185	-0.1059	0.0024
surface	pinch-air	-0.3477	-0.0351	0.2774	0.9995
surface	predictive-air	-0.6433	-0.3307	-0.0182	0.0316
surface	bimodal-air	-0.1290	0.1835	0.4961	0.5313
static-air	pinch-air	0.0708	0.3833	0.6959	0.0072
static-air	predictive-air	-0.2248	0.0877	0.4003	0.9641
static-air	bimodal-air	0.2895	0.6020	0.9146	0.0000
pinch-air	predictive-air	-0.6082	-0.2956	0.0169	0.0748
pinch-air	bimodal-air	-0.0939	0.2187	0.5312	0.3316
predictive-air	bimodal-air	0.2017	0.5143	0.8269	0.0001

B.2.3 Minimum String Distance

B.2.3.1 MSD Modified-Shortest.

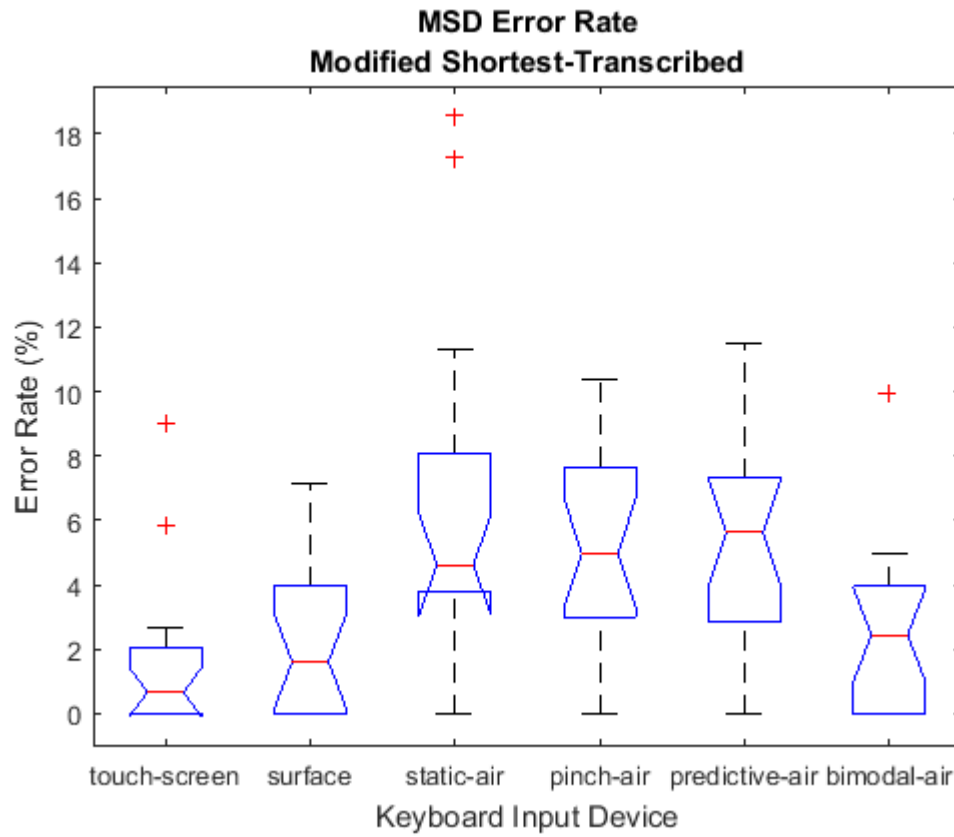


Figure B.7. Median Minimum String Distance using the shortest-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.7. Tukey’s Honest Significant Difference with multiple comparison for Minimum String Distance using shortest-transcribed modification.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-3.7747	-0.6145	2.5457	0.9931
touchscreen	static-air	-7.9973	-4.8372	-1.6770	0.0003
touchscreen	pinch-air	-6.7038	-3.5437	-0.3835	0.0186
touchscreen	predictive-air	-6.9953	-3.8351	-0.6749	0.0081
touchscreen	bimodal-air	-4.0980	-0.9378	2.2223	0.9546
surface	static-air	-7.3828	-4.2227	-1.0625	0.0025
surface	pinch-air	-6.0893	-2.9292	0.2310	0.0856
surface	predictive-air	-6.3808	-3.2206	-0.0604	0.0431
surface	bimodal-air	-3.4835	-0.3233	2.8368	0.9997
static-air	pinch-air	-1.8667	1.2935	4.4537	0.8412
static-air	predictive-air	-2.1581	1.0021	4.1622	0.9403
static-air	bimodal-air	0.7392	3.8993	7.0595	0.0067
pinch-air	predictive-air	-3.4516	-0.2914	2.8687	0.9998
pinch-air	bimodal-air	-0.5543	2.6058	5.7660	0.1677
predictive-air	bimodal-air	-0.2629	2.8973	6.0574	0.0919

B.2.4 Total Error Rate

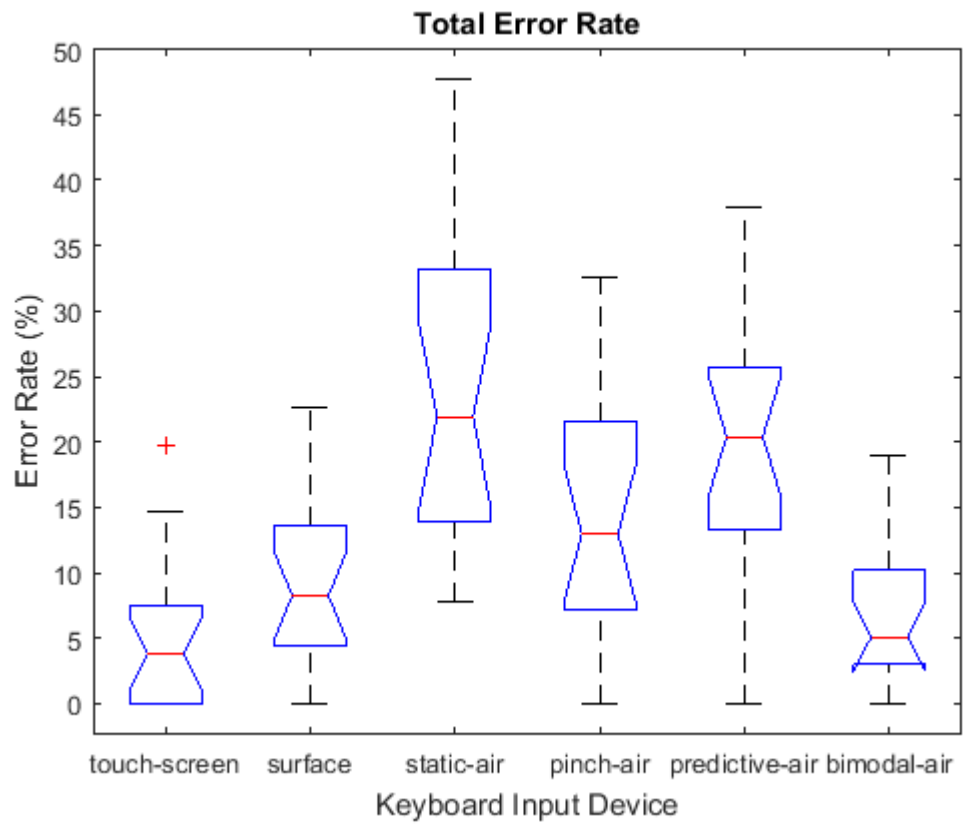


Figure B.8. Median Total Error Rate for each keyboard showing the 25th and 75th percentiles.

Table B.8. Tukey's Honest Significant Difference with multiple comparison for Total Error Rate.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-12.9197	-4.3716	4.1766	0.6743
touchscreen	static-air	-27.2196	-18.6714	-10.1233	0.0000
touchscreen	pinch-air	-17.6640	-9.1158	-0.5677	0.0295
touchscreen	predictive-air	-23.6696	-15.1214	-6.5733	0.0000
touchscreen	bimodal-air	-10.4357	-1.8875	6.6606	0.9875
surface	static-air	-22.8480	-14.2999	-5.7517	0.0001
surface	pinch-air	-13.2924	-4.7443	3.8039	0.5926
surface	predictive-air	-19.2980	-10.7498	-2.2017	0.0054
surface	bimodal-air	-6.0641	2.4840	11.0322	0.9584
static-air	pinch-air	1.0075	9.5556	18.1038	0.0191
static-air	predictive-air	-4.9981	3.5500	12.0982	0.8329
static-air	bimodal-air	8.2358	16.7839	25.3321	0.0000
pinch-air	predictive-air	-14.5537	-6.0056	2.5425	0.3270
pinch-air	bimodal-air	-1.3198	7.2283	15.7764	0.1473
predictive-air	bimodal-air	4.6858	13.2339	21.7820	0.0003

B.2.4.1 Total Error Rate Modified-Shortest.

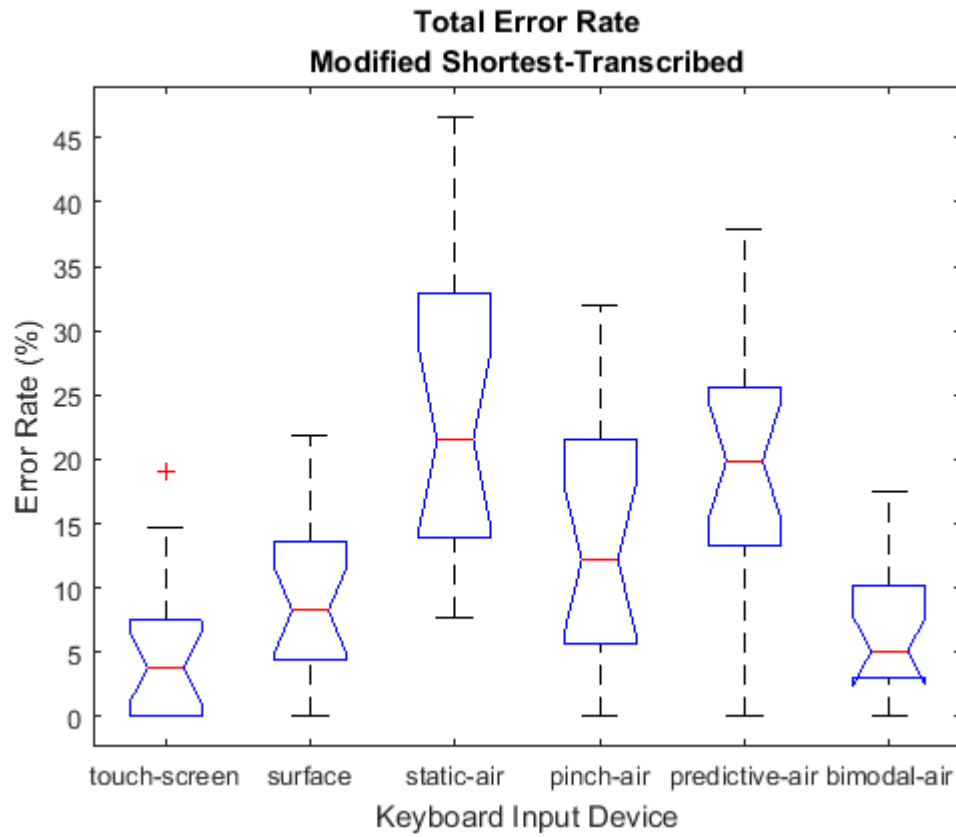


Figure B.9. Median Total Error Rate using the shortest-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.9. Tukey's Honest Significant Difference with multiple comparison for Total Error Rate using shortest-transcribed modification.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-12.6316	-4.3367	3.9581	0.6532
touchscreen	static-air	-26.7322	-18.4374	-10.1426	0.0000
touchscreen	pinch-air	-16.8635	-8.5686	-0.2738	0.0386
touchscreen	predictive-air	-22.7732	-14.4783	-6.1835	0.0000
touchscreen	bimodal-air	-10.1405	-1.8456	6.4492	0.9871
surface	static-air	-22.3955	-14.1007	-5.8058	0.0000
surface	pinch-air	-12.5268	-4.2319	4.0629	0.6765
surface	predictive-air	-18.4364	-10.1416	-1.8468	0.0075
surface	bimodal-air	-5.8037	2.4911	10.7859	0.9523
static-air	pinch-air	1.5739	9.8688	18.1636	0.0101
static-air	predictive-air	-4.3358	3.9591	12.2539	0.7351
static-air	bimodal-air	8.2969	16.5918	24.8866	0.0000
pinch-air	predictive-air	-14.2045	-5.9097	2.3851	0.3116
pinch-air	bimodal-air	-1.5718	6.7230	15.0178	0.1826
predictive-air	bimodal-air	4.3379	12.6327	20.9275	0.0003

B.3 Correctness

B.3.1 Fréchet Distance

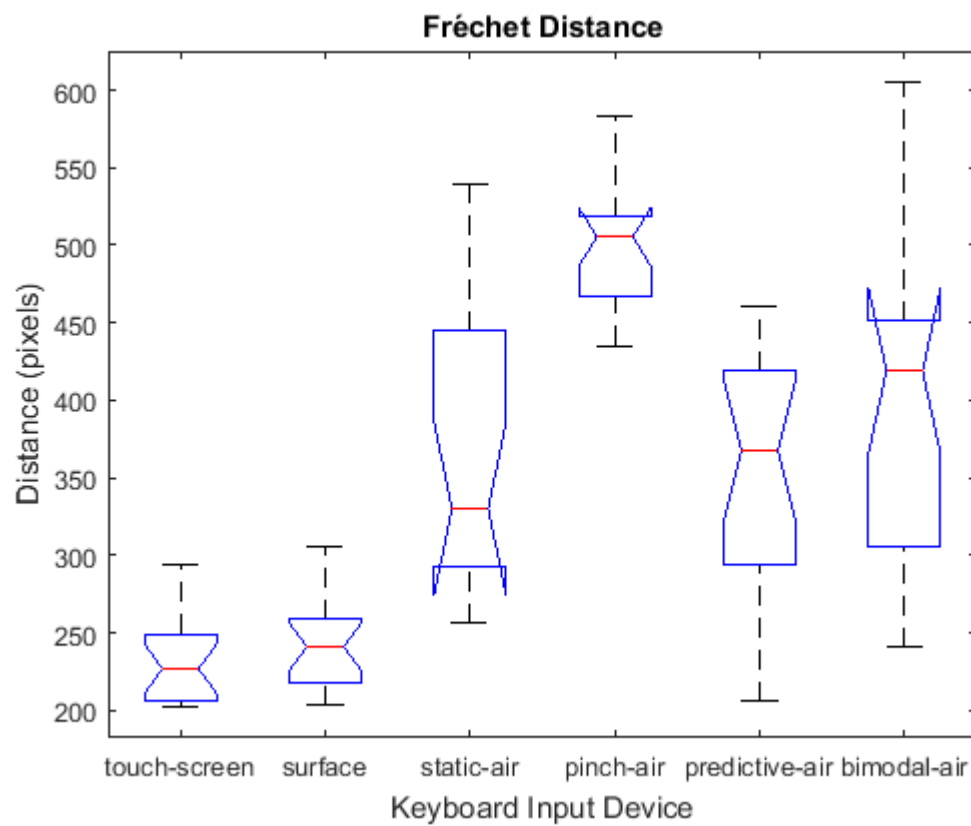


Figure B.10. Median Fréchet Distance for each keyboard showing the 25th and 75th percentiles.

Table B.10. Tukey's Honest Significant Difference with multiple comparison for Fréchet Distance.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	p -value
		Lower End		Upper End	
touchscreen	surface	-78.8540	-11.6328	55.5884	0.9960
touchscreen	static-air	-200.3012	-133.0800	-65.8588	0.0000
touchscreen	pinch-air	-337.2356	-270.0144	-202.7931	0.0000
touchscreen	predictive-air	-186.6273	-119.4061	-52.1849	0.0000
touchscreen	bimodal-air	-232.5464	-165.3252	-98.1040	0.0000
surface	static-air	-188.6684	-121.4472	-54.2260	0.0000
surface	pinch-air	-325.6027	-258.3815	-191.1603	0.0000
surface	predictive-air	-174.9945	-107.7733	-40.5521	0.0001
surface	bimodal-air	-220.9136	-153.6924	-86.4712	0.0000
static-air	pinch-air	-204.1555	-136.9343	-69.7131	0.0000
static-air	predictive-air	-53.5473	13.6739	80.8951	0.9914
static-air	bimodal-air	-99.4664	-32.2452	34.9760	0.7309
pinch-air	predictive-air	83.3870	150.6083	217.8295	0.0000
pinch-air	bimodal-air	37.4679	104.6891	171.9103	0.0002
predictive-air	bimodal-air	-113.1403	-45.9191	21.3021	0.3586

B.3.1.1 *Fréchet Distance Modified-Shortest.*

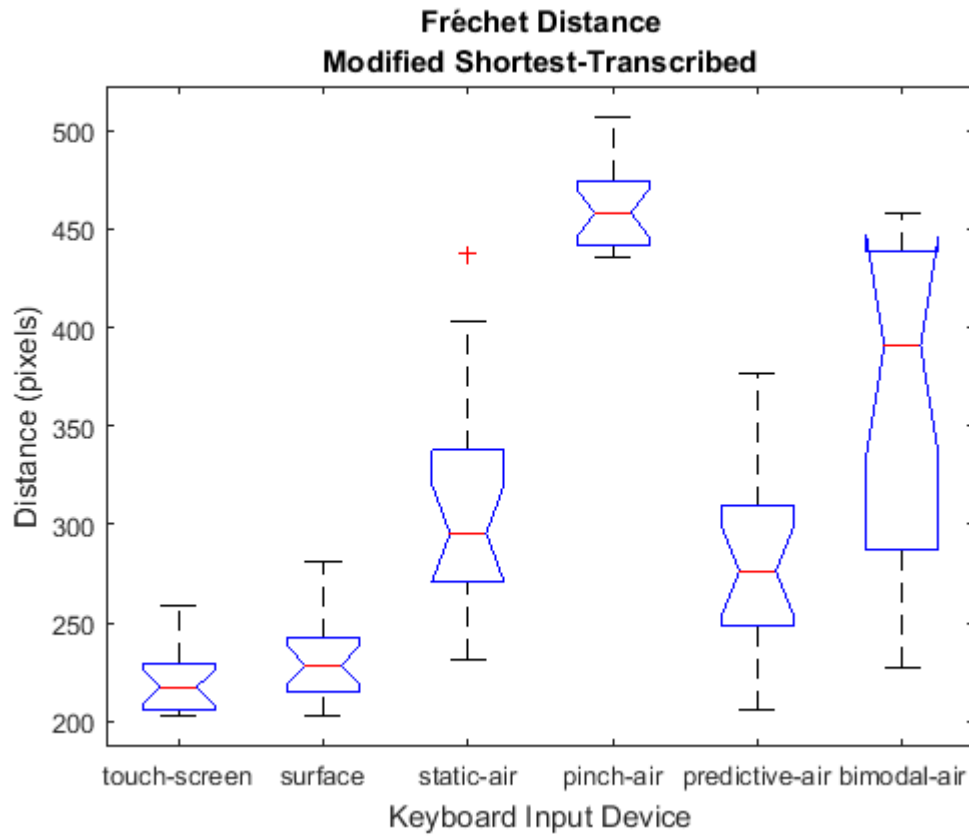


Figure B.11. Median Fréchet Distance using the shortest-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.11. Tukey's Honest Significant Difference with multiple comparison for Fréchet Distance using shortest-transcribed modification.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	p -value
		Lower End		Upper End	
touchscreen	surface	-55.6656	-9.9934	35.6788	0.9881
touchscreen	static-air	-131.1279	-85.4556	-39.7834	0.0000
touchscreen	pinch-air	-283.8592	-238.1870	-192.5147	0.0000
touchscreen	predictive-air	-106.2951	-60.6229	-14.9506	0.0027
touchscreen	bimodal-air	-193.0612	-147.3890	-101.7167	0.0000
surface	static-air	-121.1345	-75.4622	-29.7900	0.0001
surface	pinch-air	-273.8658	-228.1936	-182.5213	0.0000
surface	predictive-air	-96.3017	-50.6295	-4.9572	0.0207
surface	bimodal-air	-183.0678	-137.3956	-91.7233	0.0000
static-air	pinch-air	-198.4036	-152.7313	-107.0591	0.0000
static-air	predictive-air	-20.8395	24.8328	70.5050	0.6140
static-air	bimodal-air	-107.6056	-61.9333	-16.2611	0.0020
pinch-air	predictive-air	131.8919	177.5641	223.2363	0.0000
pinch-air	bimodal-air	45.1258	90.7980	136.4702	0.0000
predictive-air	bimodal-air	-132.4383	-86.7661	-41.0939	0.0000

B.3.1.2 *Fréchet Distance Modified-Backspace.*

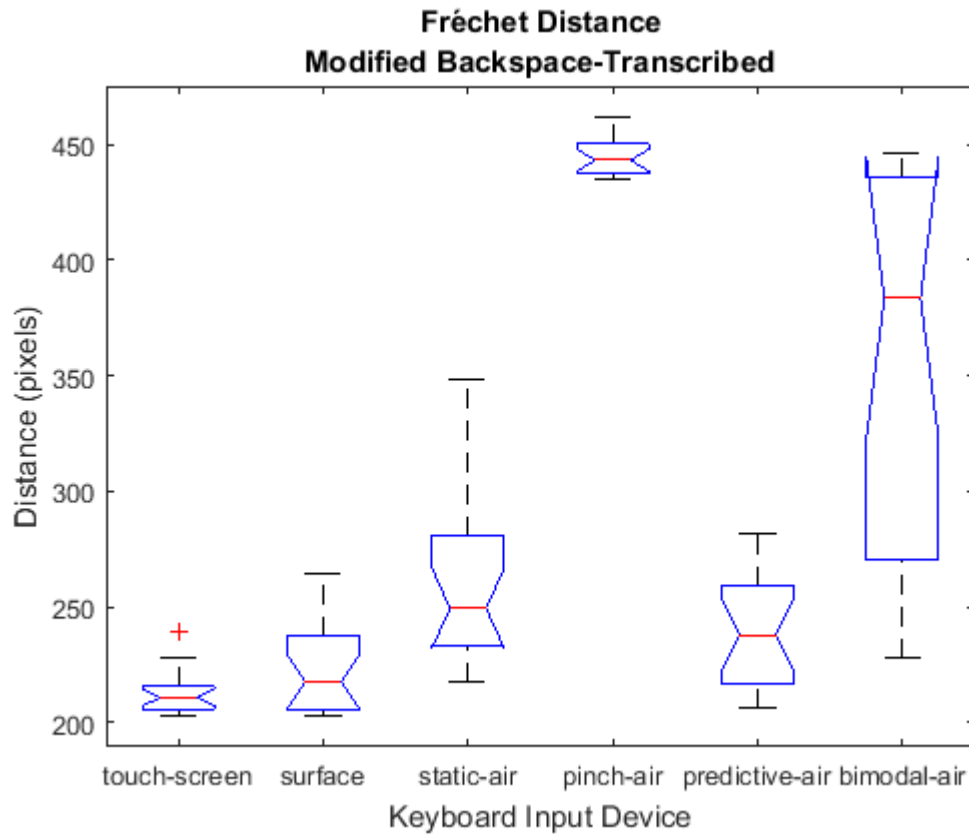


Figure B.12. Median Fréchet Distance using the backspace-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.12. Tukey's Honest Significant Difference with multiple comparison for Fréchet Distance using backspace-transcribed modification.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	p -value
		Lower End		Upper End	
touchscreen	surface	-49.6728	-11.2179	27.2370	0.9577
touchscreen	static-air	-85.0013	-46.5465	-8.0916	0.0084
touchscreen	pinch-air	-270.8703	-232.4155	-193.9606	0.0000
touchscreen	predictive-air	-63.8306	-25.3757	13.0791	0.3981
touchscreen	bimodal-air	-182.9460	-144.4911	-106.0363	0.0000
surface	static-air	-73.7834	-35.3285	3.1263	0.0907
surface	pinch-air	-259.6524	-221.1975	-182.7427	0.0000
surface	predictive-air	-52.6127	-14.1578	24.2970	0.8924
surface	bimodal-air	-171.7281	-133.2732	-94.8183	0.0000
static-air	pinch-air	-224.3239	-185.8690	-147.4141	0.0000
static-air	predictive-air	-17.2842	21.1707	59.6256	0.6011
static-air	bimodal-air	-136.3995	-97.9447	-59.4898	0.0000
pinch-air	predictive-air	168.5848	207.0397	245.4946	0.0000
pinch-air	bimodal-air	49.4694	87.9243	126.3792	0.0000
predictive-air	bimodal-air	-157.5703	-119.1154	-80.6605	0.0000

B.4 Distance Measures

B.4.1 Word-Gesture Distance

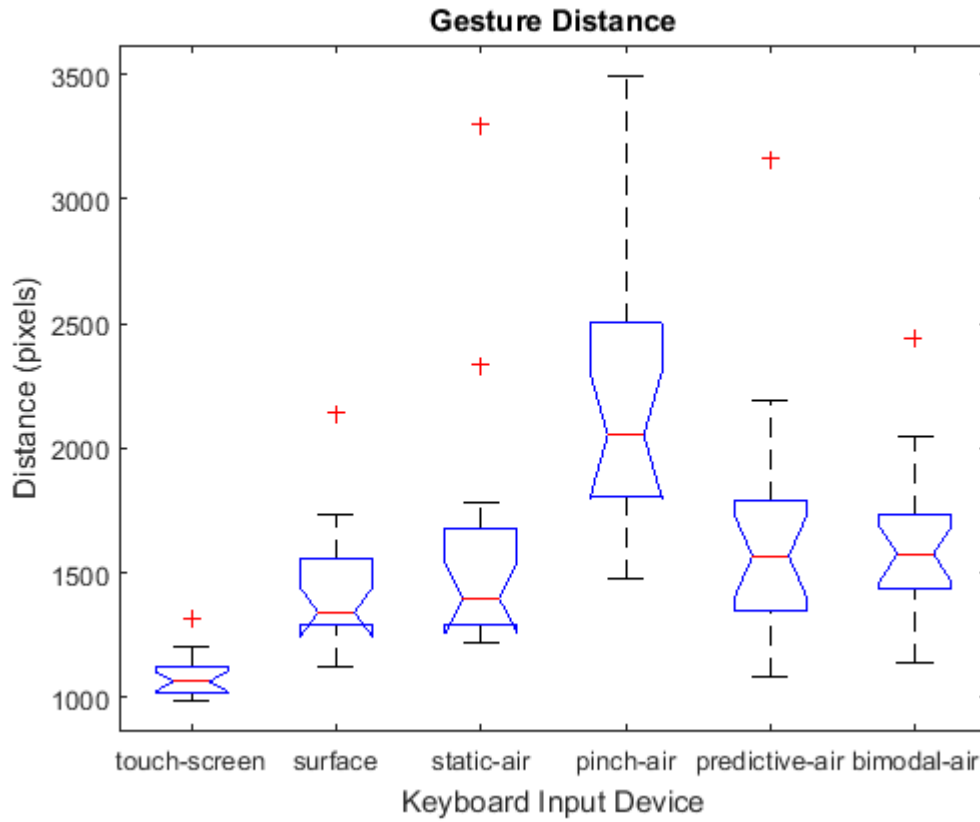


Figure B.13. Median Word-Gesture Distance for each keyboard showing the 25th and 75th percentiles.

Table B.13. Tukey's Honest Significant Difference with multiple comparison for Word-Gesture Distance.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	p -value
		Lower End		Upper End	
touchscreen	surface	-722.3100	-336.4100	49.4910	0.1245
touchscreen	static-air	-889.2300	-503.3300	-117.4300	0.0034
touchscreen	pinch-air	-1463.6000	-1077.7000	-691.8200	0.0000
touchscreen	predictive-air	-933.1100	-547.2100	-161.3100	0.0011
touchscreen	bimodal-air	-926.0700	-540.1700	-154.2700	0.0013
surface	static-air	-552.8200	-166.9200	218.9800	0.8077
surface	pinch-air	-1127.2000	-741.3100	-355.4100	0.0000
surface	predictive-air	-596.7000	-210.8000	175.1000	0.6092
surface	bimodal-air	-589.6600	-203.7600	182.1500	0.6435
static-air	pinch-air	-960.3000	-574.3900	-188.4900	0.0005
static-air	predictive-air	-429.7800	-43.8820	342.0200	0.9995
static-air	bimodal-air	-422.7400	-36.8390	349.0600	0.9998
pinch-air	predictive-air	144.6100	530.5100	916.4100	0.0017
pinch-air	bimodal-air	151.6500	537.5500	923.4600	0.0014
predictive-air	bimodal-air	-378.8600	7.0433	392.9400	1.0000

B.4.1.1 Word-Gesture Distance Modified-Shortest.

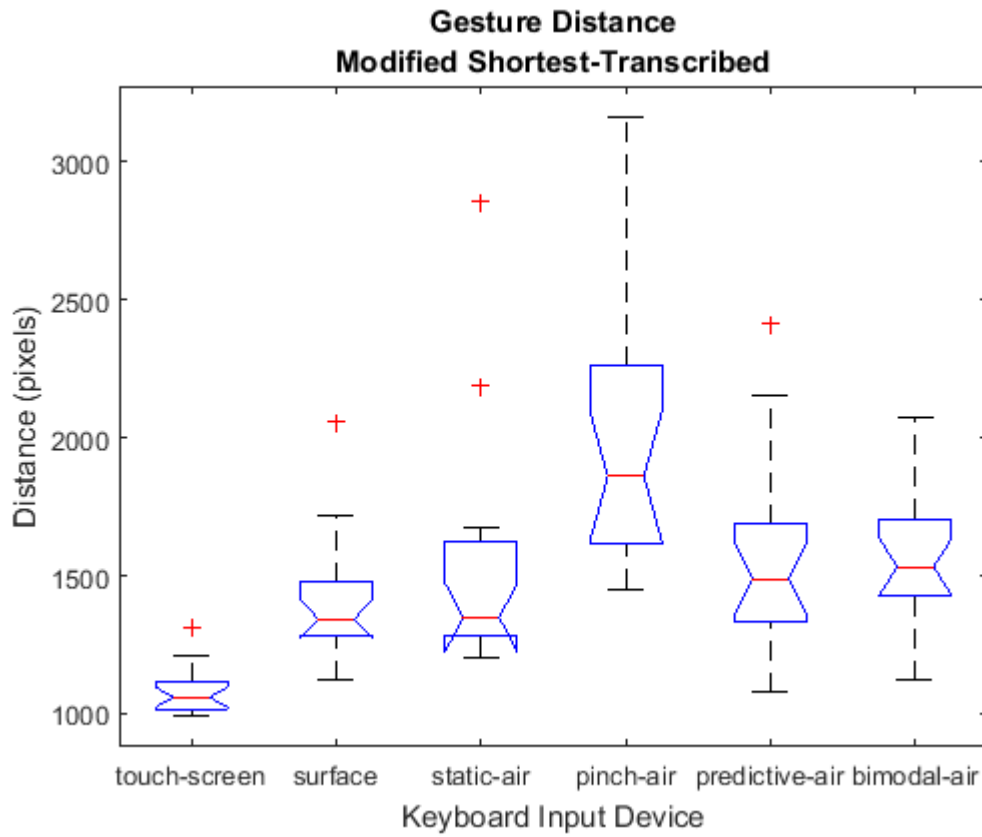


Figure B.14. Median Word-Gesture Distance using the shortest-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.14. Tukey's Honest Significant Difference with multiple comparison for Word-Gesture Distance using shortest-transcribed modification.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-630.0100	-315.2300	-0.4611	0.0494
touchscreen	static-air	-758.2800	-443.5100	-128.7300	0.0012
touchscreen	pinch-air	-1214.9000	-900.1000	-585.3300	0.0000
touchscreen	predictive-air	-770.2100	-455.4400	-140.6600	0.0008
touchscreen	bimodal-air	-799.9600	-485.1900	-170.4200	0.0003
surface	static-air	-443.0400	-128.2700	186.5000	0.8437
surface	pinch-air	-899.6400	-584.8700	-270.0900	0.0000
surface	predictive-air	-454.9800	-140.2000	174.5700	0.7878
surface	bimodal-air	-484.7300	-169.9600	144.8200	0.6211
static-air	pinch-air	-771.3700	-456.6000	-141.8200	0.0008
static-air	predictive-air	-326.7100	-11.9330	302.8400	1.0000
static-air	bimodal-air	-356.4600	-41.6840	273.0900	0.9989
pinch-air	predictive-air	129.8900	444.6600	759.4400	0.0011
pinch-air	bimodal-air	100.1400	414.9100	729.6900	0.0030
predictive-air	bimodal-air	-344.5200	-29.7520	285.0200	0.9998

B.4.2 Word-Gesture Velocity

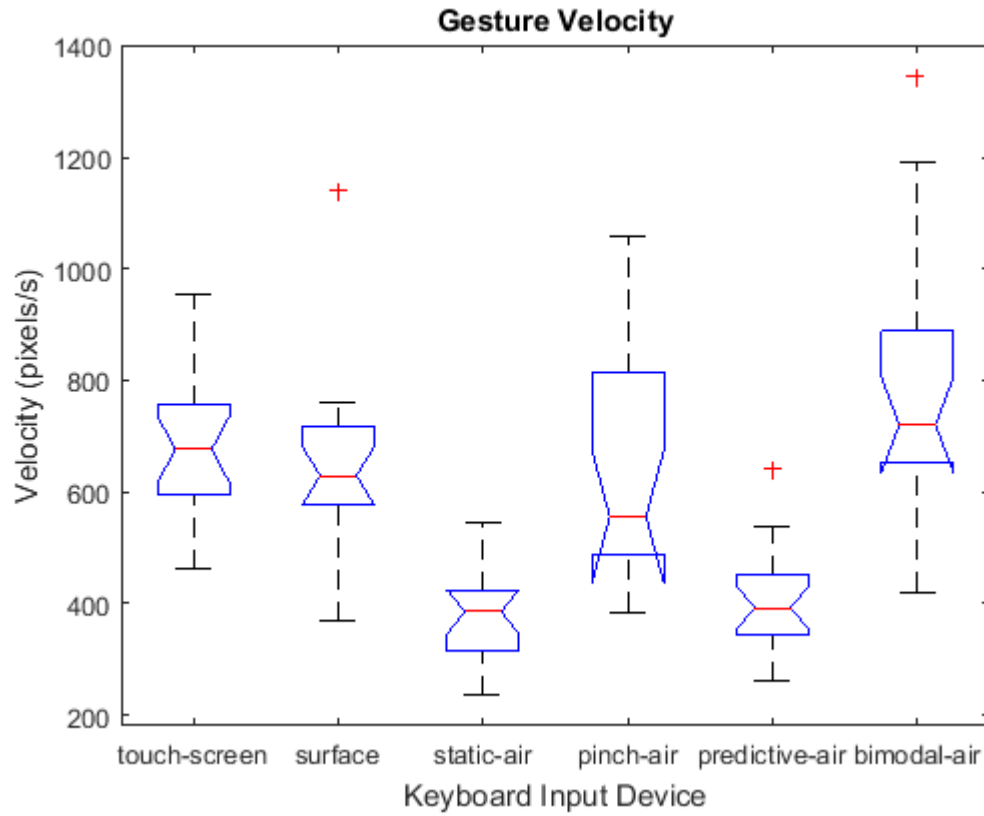


Figure B.15. Median Word-Gesture Velocity for each keyboard showing the 25th and 75th percentiles.

Table B.15. Tukey's Honest Significant Difference with multiple comparison for Word-Gesture Velocity.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-124.4400	32.2030	188.8500	0.9910
touchscreen	static-air	135.7300	292.3800	449.0200	0.0000
touchscreen	pinch-air	-127.1800	29.4680	186.1200	0.9940
touchscreen	predictive-air	111.0400	267.6900	424.3400	0.0000
touchscreen	bimodal-air	-260.9500	-104.3100	52.3400	0.3877
surface	static-air	103.5300	260.1700	416.8200	0.0001
surface	pinch-air	-159.3800	-2.7346	153.9100	1.0000
surface	predictive-air	78.8390	235.4900	392.1300	0.0004
surface	bimodal-air	-293.1600	-136.5100	20.1370	0.1248
static-air	pinch-air	-419.5500	-262.9100	-106.2600	0.0001
static-air	predictive-air	-181.3300	-24.6870	131.9600	0.9974
static-air	bimodal-air	-553.3300	-396.6800	-240.0400	0.0000
pinch-air	predictive-air	81.5740	238.2200	394.8700	0.0004
pinch-air	bimodal-air	-290.4200	-133.7800	22.8720	0.1397
predictive-air	bimodal-air	-528.6400	-372.0000	-215.3500	0.0000

B.4.3 Hand Velocity

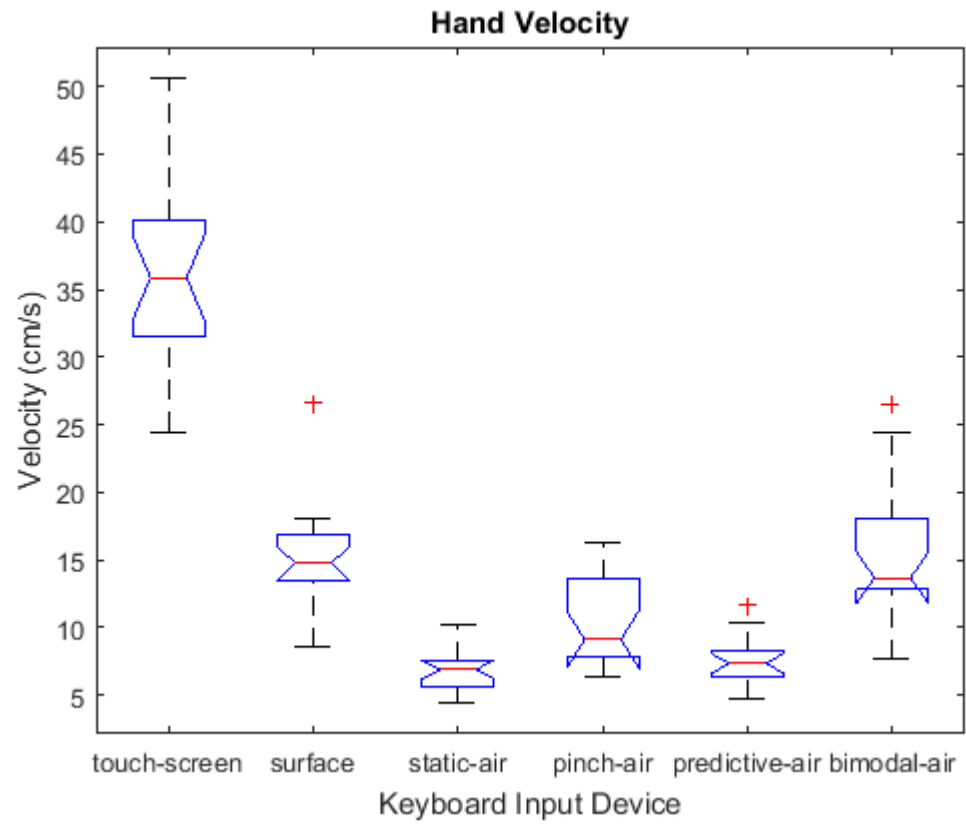


Figure B.16. Median Hand Velocity for each keyboard showing the 25th and 75th percentiles.

Table B.16. Tukey's Honest Significant Difference with multiple comparison for Hand Velocity.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	16.7046	20.6775	24.6504	0.0000
touchscreen	static-air	24.8793	28.8522	32.8251	0.0000
touchscreen	pinch-air	21.2448	25.2177	29.1906	0.0000
touchscreen	predictive-air	24.1600	28.1329	32.1058	0.0000
touchscreen	bimodal-air	16.5779	20.5508	24.5237	0.0000
surface	static-air	4.2018	8.1747	12.1476	0.0000
surface	pinch-air	0.5673	4.5403	8.5132	0.0154
surface	predictive-air	3.4825	7.4554	11.4284	0.0000
surface	bimodal-air	-4.0996	-0.1267	3.8462	1.0000
static-air	pinch-air	-7.6074	-3.6345	0.3385	0.0932
static-air	predictive-air	-4.6922	-0.7193	3.2536	0.9950
static-air	bimodal-air	-12.2743	-8.3014	-4.3285	0.0000
pinch-air	predictive-air	-1.0577	2.9152	6.8881	0.2798
pinch-air	bimodal-air	-8.6399	-4.6670	-0.6940	0.0116
predictive-air	bimodal-air	-11.5551	-7.5821	-3.6092	0.0000

B.5 Timing Measures

B.5.1 Word-Gesture Duration

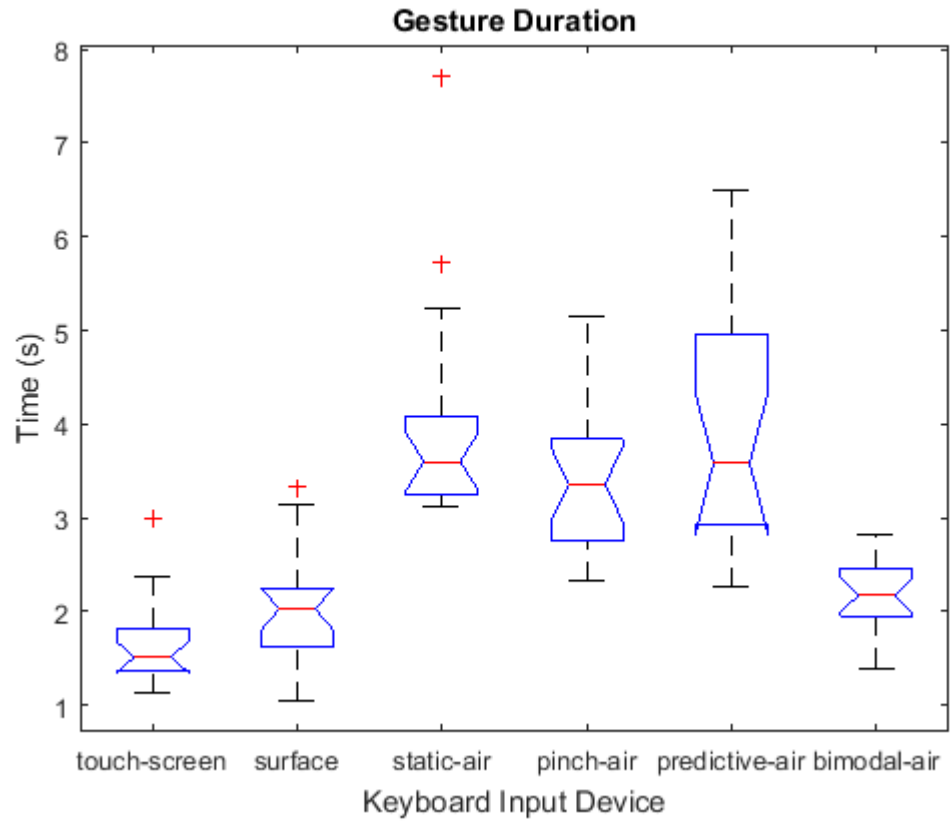


Figure B.17. Median Word-Gesture Duration for each keyboard showing the 25th and 75th percentiles.

Table B.17. Tukey's Honest Significant Difference with multiple comparison for Word-Gesture Duration.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-1.1732	-0.3842	0.4047	0.7181
touchscreen	static-air	-3.1457	-2.3568	-1.5679	0.0000
touchscreen	pinch-air	-2.5019	-1.7130	-0.9241	0.0000
touchscreen	predictive-air	-3.0051	-2.2162	-1.4272	0.0000
touchscreen	bimodal-air	-1.3204	-0.5315	0.2574	0.3743
surface	static-air	-2.7615	-1.9726	-1.1836	0.0000
surface	pinch-air	-2.1177	-1.3288	-0.5398	0.0001
surface	predictive-air	-2.6208	-1.8319	-1.0430	0.0000
surface	bimodal-air	-0.9362	-0.1473	0.6417	0.9943
static-air	pinch-air	-0.1451	0.6438	1.4327	0.1766
static-air	predictive-air	-0.6483	0.1406	0.9296	0.9954
static-air	bimodal-air	1.0364	1.8253	2.6142	0.0000
pinch-air	predictive-air	-1.2921	-0.5031	0.2858	0.4373
pinch-air	bimodal-air	0.3926	1.1815	1.9704	0.0005
predictive-air	bimodal-air	0.8957	1.6847	2.4736	0.0000

B.5.1.1 Word-Gesture Duration Modified-Shortest.

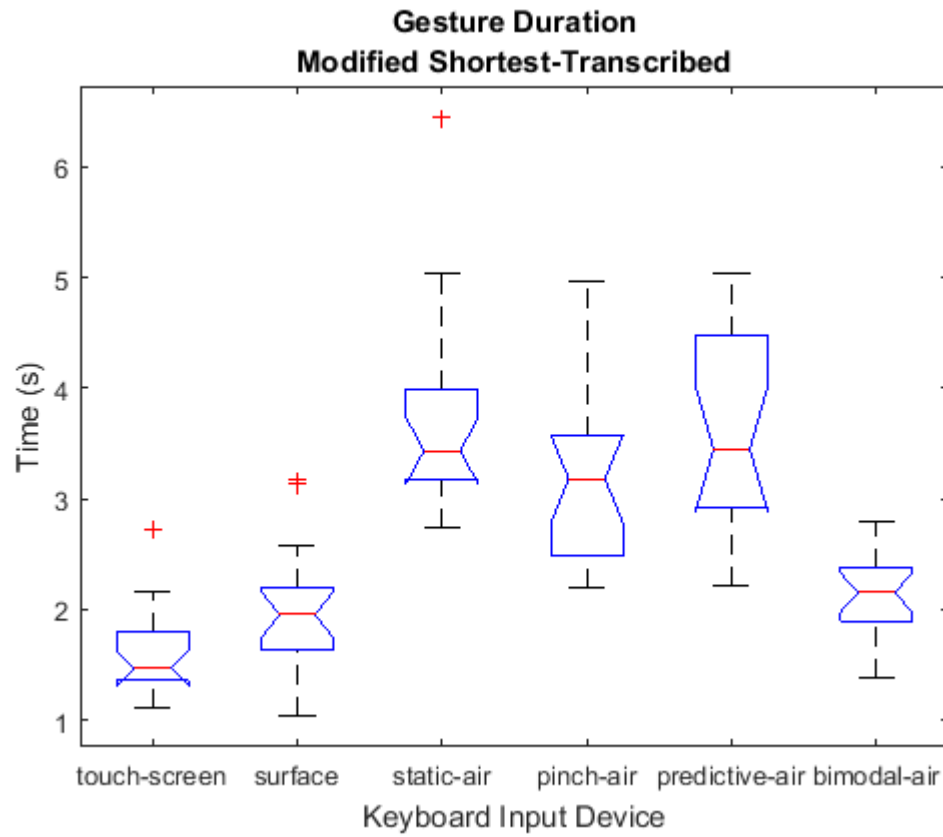


Figure B.18. Median Word-Gesture Duration using the shortest-transcribed modification for each keyboard showing the 25th and 75th percentiles.

Table B.18. Tukey's Honest Significant Difference with multiple comparison for Word-Gesture Duration using shortest-transcribed modification.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-1.0464	-0.3939	0.2587	0.5004
touchscreen	static-air	-2.7937	-2.1411	-1.4885	0.0000
touchscreen	pinch-air	-2.2167	-1.5641	-0.9115	0.0000
touchscreen	predictive-air	-2.6087	-1.9561	-1.3036	0.0000
touchscreen	bimodal-air	-1.2034	-0.5508	0.1018	0.1488
surface	static-air	-2.3998	-1.7472	-1.0947	0.0000
surface	pinch-air	-1.8228	-1.1702	-0.5177	0.0000
surface	predictive-air	-2.2148	-1.5623	-0.9097	0.0000
surface	bimodal-air	-0.8095	-0.1569	0.4957	0.9817
static-air	pinch-air	-0.0756	0.5770	1.2296	0.1147
static-air	predictive-air	-0.4676	0.1850	0.8375	0.9626
static-air	bimodal-air	0.9378	1.5903	2.2429	0.0000
pinch-air	predictive-air	-1.0446	-0.3920	0.2605	0.5058
pinch-air	bimodal-air	0.3608	1.0133	1.6659	0.0002
predictive-air	bimodal-air	0.7528	1.4054	2.0579	0.0000

B.5.2 Reaction Time to Errors

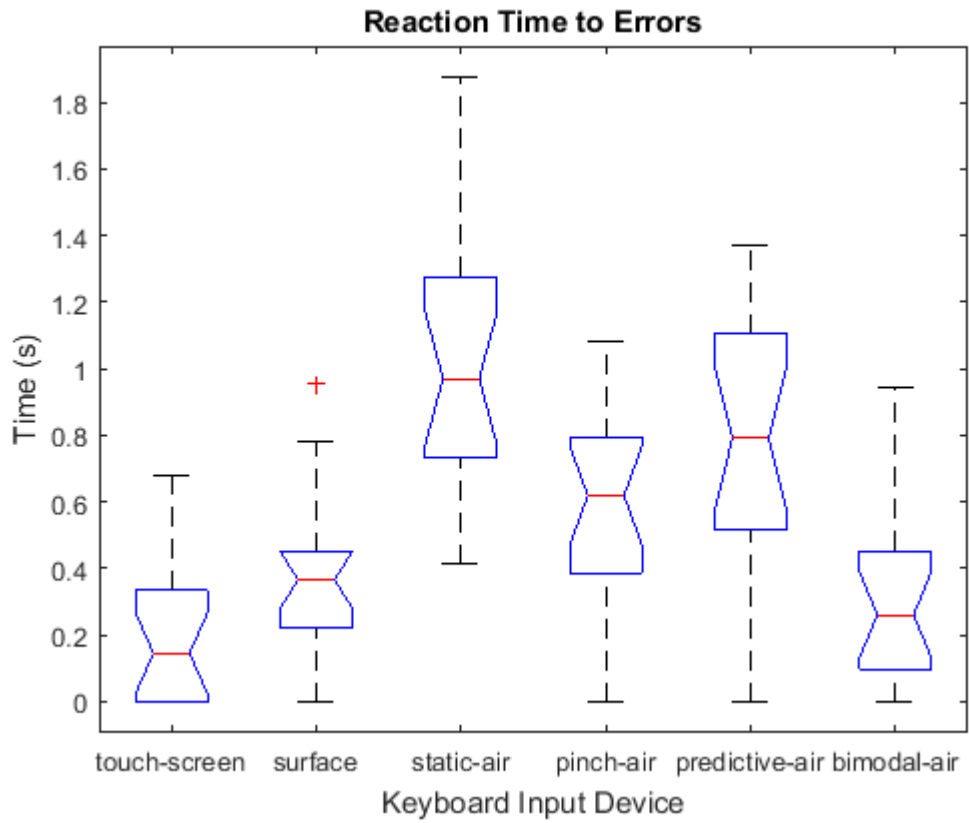


Figure B.19. Median Reaction Time to Errors for each keyboard showing the 25th and 75th percentiles.

Table B.19. Tukey's Honest Significant Difference with multiple comparison for Reaction Time to Errors.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-0.4967	-0.1824	0.1318	0.5441
touchscreen	static-air	-1.1210	-0.8068	-0.4925	0.0000
touchscreen	pinch-air	-0.6899	-0.3756	-0.0614	0.0096
touchscreen	predictive-air	-0.8725	-0.5582	-0.2440	0.0000
touchscreen	bimodal-air	-0.3874	-0.0731	0.2411	0.9842
surface	static-air	-0.9386	-0.6244	-0.3101	0.0000
surface	pinch-air	-0.5075	-0.1932	0.1210	0.4792
surface	predictive-air	-0.6901	-0.3758	-0.0616	0.0096
surface	bimodal-air	-0.2050	0.1093	0.4235	0.9136
static-air	pinch-air	0.1169	0.4311	0.7454	0.0017
static-air	predictive-air	-0.0657	0.2486	0.5628	0.2047
static-air	bimodal-air	0.4194	0.7336	1.0479	0.0000
pinch-air	predictive-air	-0.4968	-0.1826	0.1317	0.5431
pinch-air	bimodal-air	-0.0117	0.3025	0.6168	0.0662
predictive-air	bimodal-air	0.1708	0.4851	0.7993	0.0003

B.5.3 Reaction Time to Simulate Touch

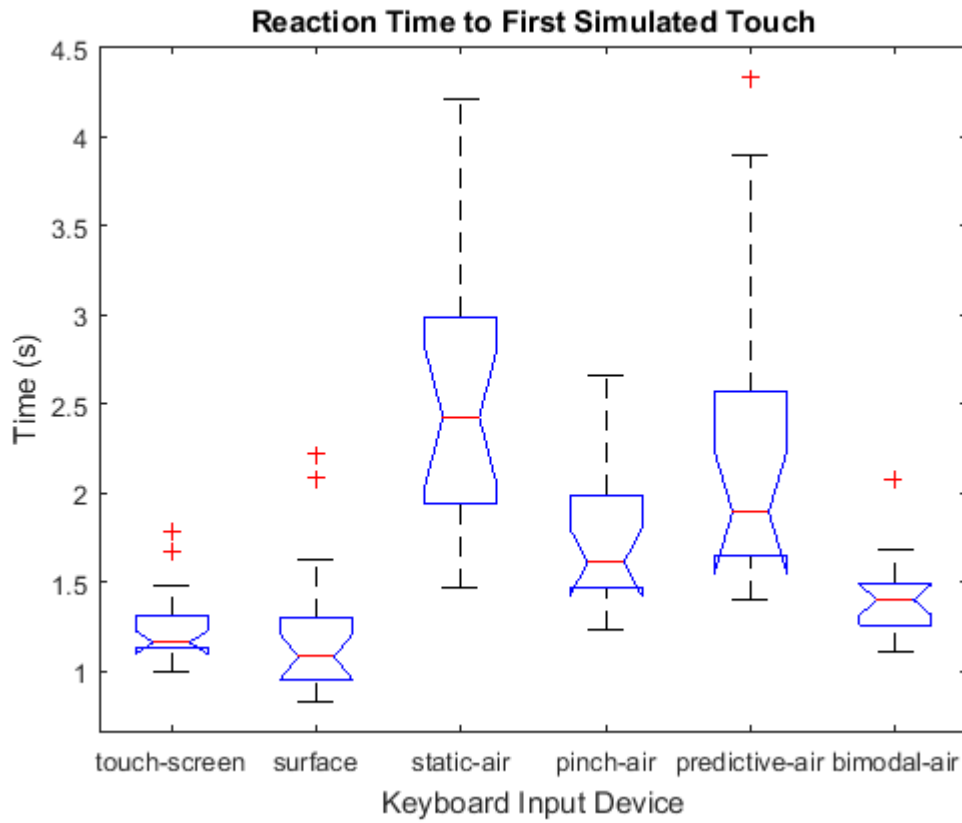


Figure B.20. Median Reaction Time to Simulate Touch for each keyboard showing the 25th and 75th percentiles.

Table B.20. Tukey's Honest Significant Difference with multiple comparison for Reaction Time to Simulate Touch.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-0.5034	0.0211	0.5456	1.0000
touchscreen	static-air	-1.8777	-1.3532	-0.8287	0.0000
touchscreen	pinch-air	-0.9956	-0.4711	0.0533	0.1044
touchscreen	predictive-air	-1.5133	-0.9888	-0.4643	0.0000
touchscreen	bimodal-air	-0.6888	-0.1644	0.3601	0.9431
surface	static-air	-1.8988	-1.3743	-0.8498	0.0000
surface	pinch-air	-1.0167	-0.4922	0.0323	0.0789
surface	predictive-air	-1.5344	-1.0099	-0.4854	0.0000
surface	bimodal-air	-0.7099	-0.1854	0.3391	0.9079
static-air	pinch-air	0.3576	0.8821	1.4066	0.0001
static-air	predictive-air	-0.1601	0.3644	0.8889	0.3395
static-air	bimodal-air	0.6644	1.1889	1.7134	0.0000
pinch-air	predictive-air	-1.0422	-0.5177	0.0068	0.0552
pinch-air	bimodal-air	-0.2177	0.3068	0.8313	0.5356
predictive-air	bimodal-air	0.3000	0.8245	1.3490	0.0002

B.5.4 Reaction Time for First Correct Letter

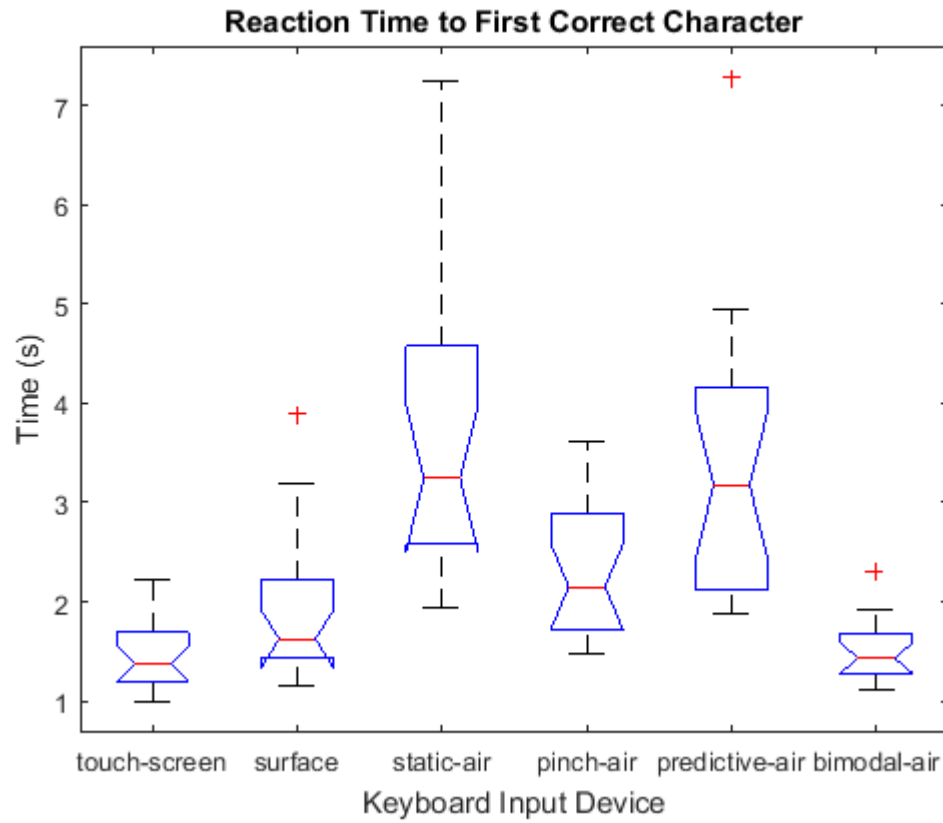


Figure B.21. Median Reaction Time for First Correct Letter for each keyboard showing the 25th and 75th percentiles.

Table B.21. Tukey's Honest Significant Difference with multiple comparison for Reaction Time for First Correct Letter.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-1.4101	-0.4882	0.4337	0.6406
touchscreen	static-air	-3.1678	-2.2460	-1.3241	0.0000
touchscreen	pinch-air	-1.8268	-0.9050	0.0169	0.0575
touchscreen	predictive-air	-2.8838	-1.9620	-1.0401	0.0000
touchscreen	bimodal-air	-0.9788	-0.0569	0.8649	1.0000
surface	static-air	-2.6796	-1.7578	-0.8359	0.0000
surface	pinch-air	-1.3386	-0.4168	0.5051	0.7772
surface	predictive-air	-2.3956	-1.4738	-0.5519	0.0001
surface	bimodal-air	-0.4906	0.4313	1.3531	0.7512
static-air	pinch-air	0.4192	1.3410	2.2629	0.0007
static-air	predictive-air	-0.6379	0.2840	1.2059	0.9469
static-air	bimodal-air	1.2672	2.1891	3.1109	0.0000
pinch-air	predictive-air	-1.9789	-1.0570	-0.1352	0.0149
pinch-air	bimodal-air	-0.0738	0.8480	1.7699	0.0899
predictive-air	bimodal-air	0.9832	1.9051	2.8269	0.0000

B.6 Quantitative Measures

B.6.1 Number of Touches Simulated

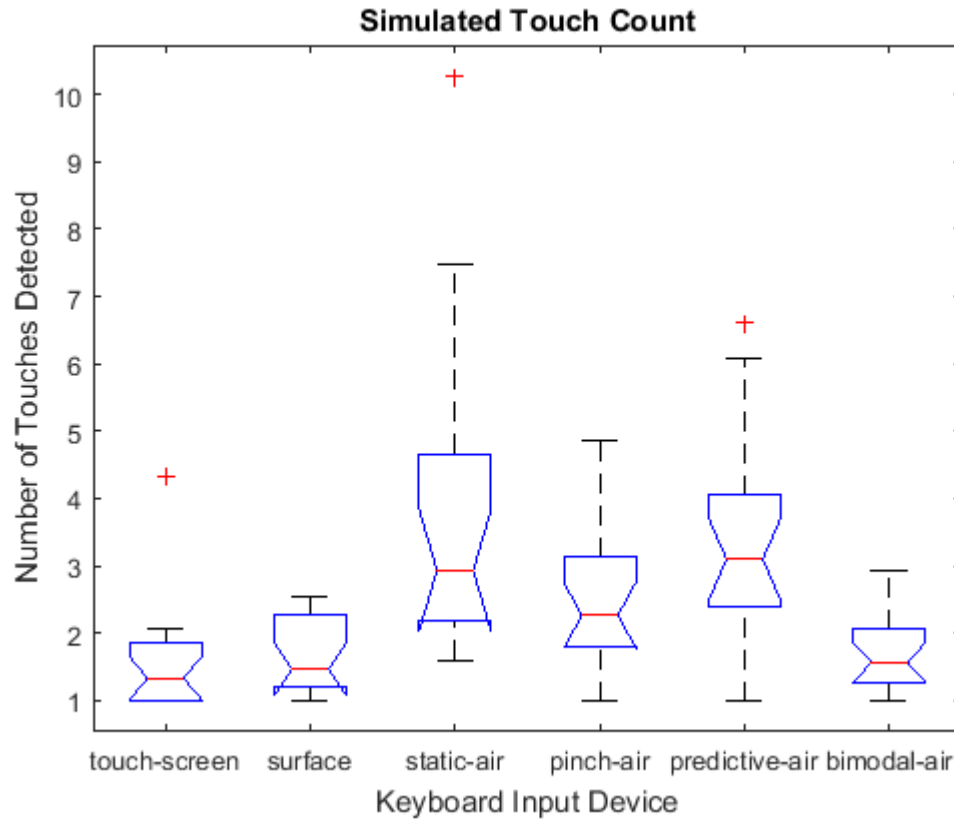


Figure B.22. Median Number of Touches Simulated for each keyboard showing the 25th and 75th percentiles.

Table B.22. Tukey's Honest Significant Difference with multiple comparison for
Number of Touches Simulated.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	<i>p</i> -value
		Lower End		Upper End	
touchscreen	surface	-1.2824	-0.0444	1.1935	1.0000
touchscreen	static-air	-3.4306	-2.1926	-0.9546	0.0000
touchscreen	pinch-air	-2.0824	-0.8444	0.3935	0.3602
touchscreen	predictive-air	-2.9861	-1.7481	-0.5102	0.0011
touchscreen	bimodal-air	-1.3269	-0.0889	1.1491	0.9999
surface	static-air	-3.3861	-2.1481	-0.9102	0.0000
surface	pinch-air	-2.0380	-0.8000	0.4380	0.4222
surface	predictive-air	-2.9417	-1.7037	-0.4657	0.0017
surface	bimodal-air	-1.2824	-0.0444	1.1935	1.0000
static-air	pinch-air	0.1102	1.3481	2.5861	0.0245
static-air	predictive-air	-0.7935	0.4444	1.6824	0.9022
static-air	bimodal-air	0.8657	2.1037	3.3417	0.0000
pinch-air	predictive-air	-2.1417	-0.9037	0.3343	0.2852
pinch-air	bimodal-air	-0.4824	0.7556	1.9935	0.4878
predictive-air	bimodal-air	0.4213	1.6593	2.8972	0.0024

B.6.2 Number of Practice Words

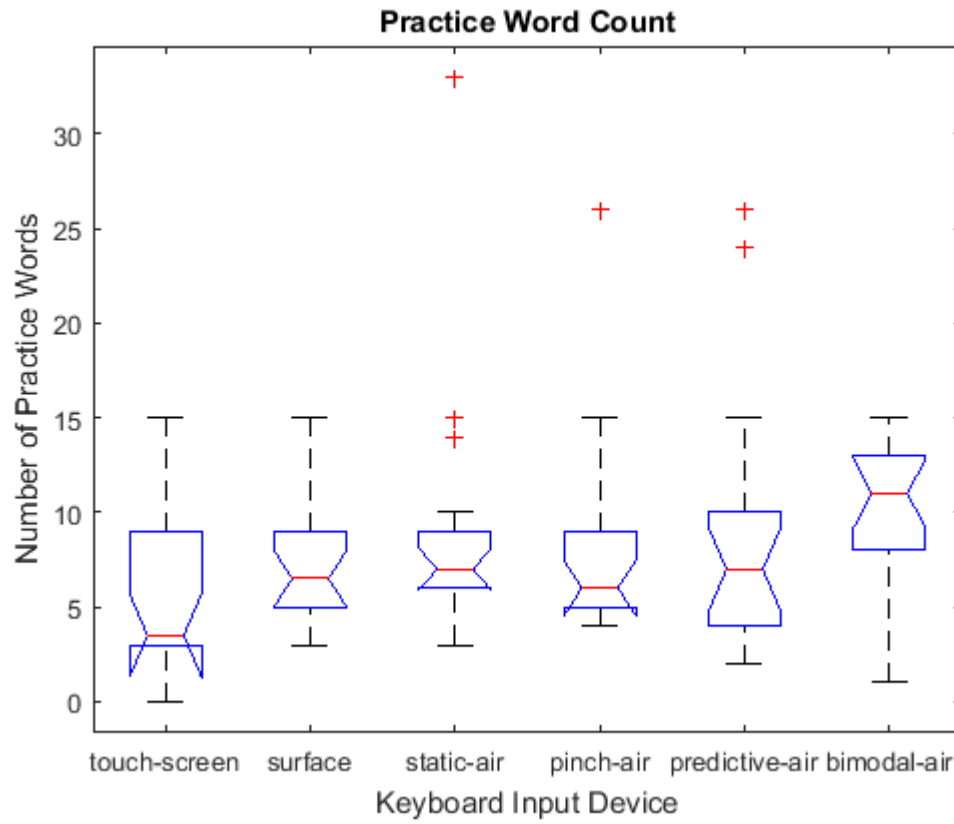


Figure B.23. Median Number of Practice Words for each keyboard showing the 25th and 75th percentiles.

B.7 Qualitative Measures

B.7.1 Level of Discomfort

Table B.23. Tukey's Honest Significant Difference with multiple comparison for Level of Discomfort.

Group 1	Group 2	Confidence Interval	Mean Difference	Confidence Interval	<i>p</i> -value
		Lower End	$\mu_1 - \mu_2$	Upper End	
touchscreen	surface	-1.0795	-0.1667	0.7462	0.9948
touchscreen	static-air	-2.3573	-1.4444	-0.5316	0.0002
touchscreen	pinch-air	-1.8573	-0.9444	-0.0316	0.0381
touchscreen	predictive-air	-2.2462	-1.3333	-0.4205	0.0007
touchscreen	bimodal-air	-1.3573	-0.4444	0.4684	0.7184
surface	static-air	-2.1906	-1.2778	-0.3649	0.0013
surface	pinch-air	-1.6906	-0.7778	0.1351	0.1414
surface	predictive-air	-2.0795	-1.1667	-0.2538	0.0044
surface	bimodal-air	-1.1906	-0.2778	0.6351	0.9496
static-air	pinch-air	-0.4128	0.5000	1.4128	0.6063
static-air	predictive-air	-0.8017	0.1111	1.0239	0.9993
static-air	bimodal-air	0.0872	1.0000	1.9128	0.0232
pinch-air	predictive-air	-1.3017	-0.3889	0.5239	0.8174
pinch-air	bimodal-air	-0.4128	0.5000	1.4128	0.6063
predictive-air	bimodal-air	-0.0239	0.8889	1.8017	0.0610

B.7.2 Level of Fatigue

Table B.24. Tukey's Honest Significant Difference with multiple comparison for Level of Fatigue.

Group 1	Group 2	Confidence Interval	Mean Difference	Confidence Interval	<i>p</i> -value
		Lower End	$\mu_1 - \mu_2$	Upper End	
touchscreen	surface	-1.4179	-0.3333	0.7513	0.9475
touchscreen	static-air	-3.4735	-2.3889	-1.3043	0.0000
touchscreen	pinch-air	-2.6402	-1.5556	-0.4709	0.0009
touchscreen	predictive-air	-2.5846	-1.5000	-0.4154	0.0015
touchscreen	bimodal-air	-2.0846	-1.0000	0.0846	0.0886
surface	static-air	-3.1402	-2.0556	-0.9709	0.0000
surface	pinch-air	-2.3068	-1.2222	-0.1376	0.0177
surface	predictive-air	-2.2513	-1.1667	-0.0821	0.0273
surface	bimodal-air	-1.7513	-0.6667	0.4179	0.4797
static-air	pinch-air	-0.2513	0.8333	1.9179	0.2326
static-air	predictive-air	-0.1957	0.8889	1.9735	0.1729
static-air	bimodal-air	0.3043	1.3889	2.4735	0.0043
pinch-air	predictive-air	-1.0291	0.0556	1.1402	1.0000
pinch-air	bimodal-air	-0.5291	0.5556	1.6402	0.6728
predictive-air	bimodal-air	-0.5846	0.5000	1.5846	0.7626

B.7.3 Level of Difficulty

Table B.25. Tukey's Honest Significant Difference with multiple comparison for Level of Difficulty.

Group 1	Group 2	Confidence Interval	Mean Difference	Confidence Interval	<i>p</i> -value
		Lower End	$\mu_1 - \mu_2$	Upper End	
touchscreen	surface	-1.2306	-0.2778	0.6750	0.9578
touchscreen	static-air	-2.4528	-1.5000	-0.5472	0.0002
touchscreen	pinch-air	-2.0083	-1.0556	-0.1028	0.0209
touchscreen	predictive-air	-2.0083	-1.0556	-0.1028	0.0209
touchscreen	bimodal-air	-1.3972	-0.4444	0.5083	0.7535
surface	static-air	-2.1750	-1.2222	-0.2694	0.0042
surface	pinch-air	-1.7306	-0.7778	0.1750	0.1763
surface	predictive-air	-1.7306	-0.7778	0.1750	0.1763
surface	bimodal-air	-1.1195	-0.1667	0.7861	0.9958
static-air	pinch-air	-0.5083	0.4444	1.3972	0.7535
static-air	predictive-air	-0.5083	0.4444	1.3972	0.7535
static-air	bimodal-air	0.1028	1.0556	2.0083	0.0209
pinch-air	predictive-air	-0.9528	0.0000	0.9528	1.0000
pinch-air	bimodal-air	-0.3417	0.6111	1.5639	0.4308
predictive-air	bimodal-air	-0.3417	0.6111	1.5639	0.4308

B.7.4 Preference Ranking

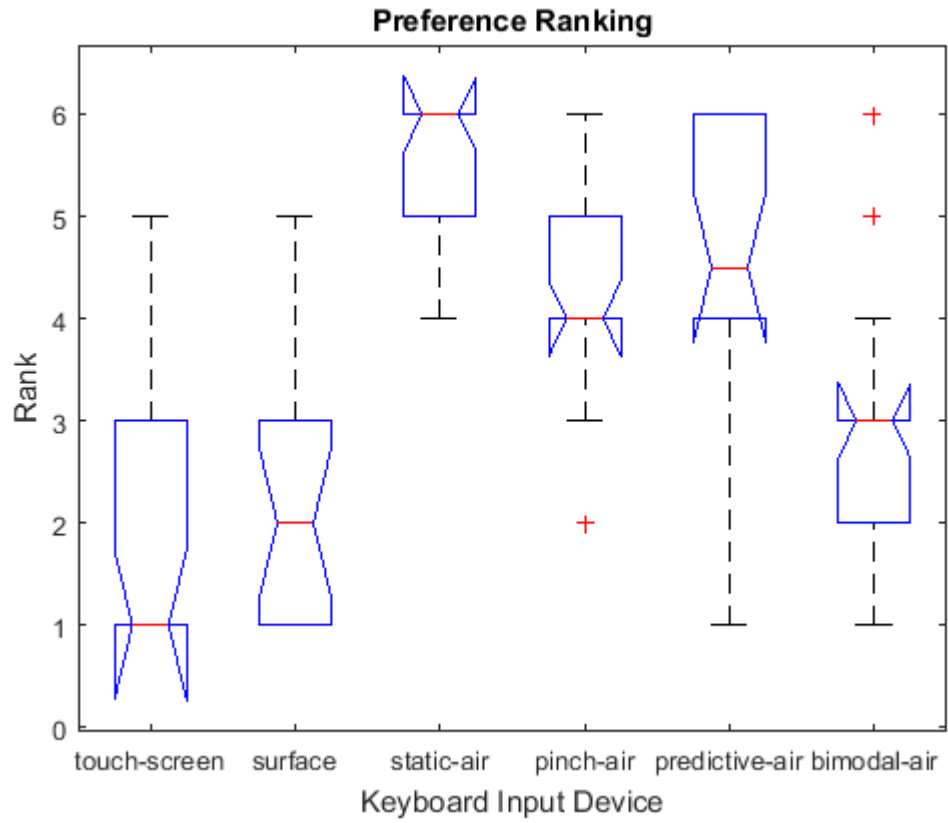


Figure B.24. Median Preference Ranking for each keyboard showing the 25th and 75th percentiles.

Table B.26. Tukey's Honest Significant Difference with multiple comparison for Preference Ranking.

Group 1	Group 2	Confidence Interval	Mean Difference $\mu_1 - \mu_2$	Confidence Interval	p -value
		Lower End		Upper End	
touchscreen	surface	-2.1104	-0.3333	1.4438	0.9948
touchscreen	static-air	-5.2215	-3.4444	-1.6673	0.0000
touchscreen	pinch-air	-4.1104	-2.3333	-0.5562	0.0025
touchscreen	predictive-air	-4.3882	-2.6111	-0.8340	0.0004
touchscreen	bimodal-air	-2.7215	-0.9444	0.8327	0.6549
surface	static-air	-4.8882	-3.1111	-1.3340	0.0000
surface	pinch-air	-3.7771	-2.0000	-0.2229	0.0169
surface	predictive-air	-4.0549	-2.2778	-0.5007	0.0035
surface	bimodal-air	-2.3882	-0.6111	1.1660	0.9244
static-air	pinch-air	-0.6660	1.1111	2.8882	0.4776
static-air	predictive-air	-0.9438	0.8333	2.6104	0.7648
static-air	bimodal-air	0.7229	2.5000	4.2771	0.0009
pinch-air	predictive-air	-2.0549	-0.2778	1.4993	0.9978
pinch-air	bimodal-air	-0.3882	1.3889	3.1660	0.2251
predictive-air	bimodal-air	-0.1104	1.6667	3.4438	0.0808

BIBLIOGRAPHY

- Banerjee, A., J. Burstyn, A. Girouard, and R. Vertegaal (2012). Multipoint: Comparing laser and manual pointing as remote input in large display interactions. In *International Journal of Human - Computer Studies*, Volume 70(10) of *CHI '11*, pp. 690–702. Elsevier Ltd.
- Cockburn, A., P. Quinn, C. Gutwin, G. Ramos, and J. Looser (2011). Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback. In *International Journal of Human-Computer Studies*, Volume 69(6), pp. 401–414. Elsevier Ltd.
- Devoe, D. (1967). Alternatives to handprinting in the manual entry of data. In *IEEE Transactions on Human Factors in Electronics*, Volume HFE-8(1), pp. 21–32. IEEE.
- Har-Peled, S. and B. Raichel (2014, January). The frÉchet distance revisited and extended. *ACM Trans. Algorithms* 10(1), 3:1–3:22.
- Jones, B., R. Sodhi, D. Forsyth, B. Bailey, and G. Maciocci (2012). Around device interaction for multiscale navigation. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '12, pp. 83–92. ACM.
- Jota, R., M. A. Nacenta, J. A. Jorge, S. Carpendale, and S. Greenberg (2010). A comparison of ray pointing techniques for very large displays. In *Proceedings - Graphics Interface*, pp. 269–276. Canadian Information Processing Society.
- Kristensson, P. O. (2007). *Discrete and Continuous Shape Writing for Text Entry and Control*. Doctoral dissertation, Linköping University, Sweden.
- Malik, S. and J. Laszlo (2004). Visual touchpad: A two-handed gestural input device. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, ICMI '04, pp. 289–296. ACM.
- Markussen, A., M. R. Jakobsen, and K. Hornbæk (2013). Selection-based mid-air text entry on large displays. In *Human-Computer Interaction INTERACT 2013*, Volume 8117(1) of *Lecture Notes in Computer Science*, pp. 401–418. Springer Berlin Heidelberg.
- Markussen, A., M. R. Jakobsen, and K. Hornbæk (2014). Vulture: A mid-air word-gesture keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 1073–1082. ACM.
- Nancel, M., J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay (2011). Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 177–186. ACM.

- Ni, T., D. Bowman, and C. North (2011). Airstroke: Bringing unistroke text entry to freehand gesture interfaces. In *Proceedings of the SIGCHI Conference on human factors in computing systems*, CHI '11, pp. 2473–2476. ACM.
- Piumsomboon, T., A. Clark, M. Billinghamurst, and A. Cockburn (2013). User-defined gestures for augmented reality. In *Human-Computer Interaction INTERACT 2013*, Volume 8118(2) of *Lecture Notes in Computer Science*, pp. 282–299. Springer Berlin Heidelberg.
- Shoemaker, G., L. Findlater, J. Q. Dawson, and K. S. Booth (2009). Mid-air text input techniques for very large wall displays. In *Proceedings - Graphics Interface*, pp. 231–238. Canadian Information Processing Society.
- Shoemaker, G., A. Tang, and K. S. Booth (2007). Shadow reaching: A new perspective on interaction for large displays. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, pp. 53–56. ACM.
- Sutton, J. (2013). Air painting with corel painter freestyle and the leap motion controller: A revolutionary new way to paint! In *ACM SIGGRAPH 2013 Studio Talks*, SIGGRAPH '13, pp. 21:1–21:1. ACM.
- Toimil, M., N. Shin, and A. Eisner (2013). Gesture technology: An innovative leap for workflow. In *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government*, EEE '13, pp. 291. WorldComp.
- Vikram, S., L. Li, and S. Russell (2013). Writing and sketching in the air, recognizing and controlling on the fly. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pp. 1179–1184. ACM.
- Vogel, D. and R. Balakrishnan (2005). Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, UIST '05, pp. 33–42. ACM.
- Zhai, S. and P. O. Kristensson (2012). The word-gesture keyboard: Reimagining keyboard interaction. *Communications of the ACM* 55(9), 91–101.
- Zhai, S., P. O. Kristensson, P. Gong, M. Greiner, S. A. Peng, L. M. Liu, and A. Dunnigan (2009). Shapewriter on the iphone: From the laboratory to the real world. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pp. 2667–2670. ACM.