

2018

# Computational optimization and prediction strategies for increasing communication rate in phoneme-based augmentative and alternative communication (AAC)

---

<https://hdl.handle.net/2144/32952>

*"Downloaded from OpenBU. Boston University's institutional repository."*

BOSTON UNIVERSITY  
SCHOOL OF MEDICINE

Dissertation

**COMPUTATIONAL OPTIMIZATION AND PREDICTION STRATEGIES FOR  
INCREASING COMMUNICATION RATE IN PHONEME-BASED  
AUGMENTATIVE AND ALTERNATIVE COMMUNICATION (AAC)**

by

**GABRIEL J. CLER**

B.S., Bradley University, 2007

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

2018

© 2018  
GABRIEL J. CLER  
All rights reserved except for Chapter 2,  
which is © 2017 by the authors (Cler &  
Stepp); publication rights licensed to  
ACM. doi: 10.1145/3132525.3132537

Approved by

First Reader

---

Cara E. Stepp, Ph.D.  
Associate Professor of Speech, Language & Hearing Sciences  
Sargent College of Health and Rehabilitation Sciences  
Associate Professor of Biomedical Engineering  
Associate Professor of Otolaryngology–Head and Neck Surgery

Second Reader

---

Frank H. Guenther, Ph.D.  
Professor of Speech, Language & Hearing Sciences  
Sargent College of Health and Rehabilitation Science  
Professor of Biomedical Engineering

Third Reader

---

Christopher A. Moore, Ph.D.  
Dean of Sargent College of Health and Rehabilitation Sciences  
Professor of Speech, Language & Hearing Sciences

## **DEDICATION**

I would like to dedicate this work to my foxhole buddies MKJ, LHM, JMV, VSM, & DA, other friends SD, SF, AL, DB, and KFN, my family and particularly my nephew Henry, the illustrious and unparalleled RLS and TM, my choir, the McElroys, and Pavement. None of this would have happened without you – feel free to take blame or credit as you see fit.

## ACKNOWLEDGMENTS

I would like to firstly thank my advisor, Cara Stepp, without whom none of this would be possible. Your mentorship and guidance has been invaluable, as have the huge wealth of opportunities that you have allowed me to take advantage of during my time with you. Further thank yous (thanks you?) are needed to my other mentors and committee members, Jay Bohland, Susan Fager, Frank Guenther, and Chris Moore. Thank each of you for the many conversations and lunches and Cornwallising throughout my time here. Thank you to the Graduate Program for Neuroscience for its support. Thanks to all aforementioned folks for their support and help with the design of this work, and thank you to Jenny Vojtech, Jake Noordzij, and Kat Kolin for their help in executing it.

**COMPUTATIONAL OPTIMIZATION AND PREDICTION STRATEGIES FOR  
INCREASING COMMUNICATION RATE IN PHONEME-BASED  
AUGMENTATIVE AND ALTERNATIVE COMMUNICATION (AAC)**

**GABRIEL J. CLER**

Boston University School of Medicine, 2018

Major Professor: Cara E. Stepp, Ph.D., Associate Professor of Speech,  
Language, & Hearing Sciences; Biomedical Engineering, and  
Otolaryngology-Head & Neck Surgery

**ABSTRACT**

Up to 1.2% of the population is unable to meet daily communication needs using typical speech and may use augmentative and alternative communication (AAC) strategies to communicate, including manual sign language, facial gestures, and aided strategies such as selecting targets on an onscreen keyboard. However, for individuals whose impairments affect both speech and non-speech motor systems (e.g., spinal cord injury, amyotrophic lateral sclerosis, multiple sclerosis), their ability to use manual sign and access computer systems are impacted. AAC access methods in this population remain inherently slow and effortful (e.g., eye-tracking, head-tracking, mechanical switches). Thus, optimizing communication interfaces for alternate access methods may provide significant improvements in communication rates and quality of life.

In this series of studies, we developed and evaluated methods for improving communication rates through optimization and prediction in communication interfaces. These interfaces enabled participants to select

sounds (phonemes) instead of letters and were computationally optimized offline via a model of human movement in order for targets likely to be selected together to be in close proximity. Online prediction was implemented such that likely targets were dynamically enlarged. Computational simulations suggested that optimized phonemic interfaces could increase communication rates by up to 30.9% compared to random phonemic interfaces. Communication rates were empirically evaluated in 36 participants without motor impairment using an alternate computer access method to produce messages with phonemic interfaces over 12 sessions. Results suggested that optimization increased communication rates by 10.5–23.0% compared to a random phonemic interface. Prediction increased communication rates during training sessions, but was not a significant factor in communication rates during the final session. Empirical evaluations in individuals with motor impairment revealed that all participants strongly agreed that they would improve with practice, and four out of six participants strongly preferred the interface with prediction.

Results of these studies suggest that optimized and predictive phonemic interfaces may provide increased communication rates for individuals with motor impairments affecting both oral communication and computer access. Methods for dynamically enlarging targets may also be applicable to other (non-phonemic) interfaces to increase communication rates. Further research is needed to fully translate these results into clinical practice.

## TABLE OF CONTENTS

<b>DEDICATION</b> .....	iv
<b>ACKNOWLEDGMENTS</b> .....	v
<b>LIST OF TABLES</b> .....	xiii
<b>LIST OF FIGURES</b> .....	xiv
<b>LIST OF ABBREVIATIONS</b> .....	xix
Chapter 1. Introduction .....	1
1.1 Augmentative and Alternative Communication.....	1
1.2 Access methods.....	3
1.2.1 Direct selection (continuous access) versus switch selection (binary input) 4	
1.2.2 Modeling AAC access .....	8
1.3 AAC Symbol Sets.....	9
1.3.1 Phonemic AAC Interfaces .....	10
1.3.2 Existing Phonemic Interfaces .....	11
1.3.3 Phonemic Targets and Labels.....	15
1.4 AAC Target Layout.....	16
1.4.1 Ten-finger typing .....	16
1.4.2 Serial input .....	17
1.4.3 Existing optimized (orthographic) interfaces .....	19

1.5	Prediction .....	20
1.5.1	Word Prediction in AAC .....	21
1.5.2	Character Prediction in AAC (Via Dynamic Keyboards).....	22
1.5.3	Reduced / Disambiguating Interfaces .....	23
1.5.4	Phonemic Prediction .....	24
1.6	Purpose of this work.....	27
Chapter 2. Development and Theoretical Evaluation of Optimized Phonemic		
	Interfaces.....	29
2.1	Abstract.....	29
2.2	Introduction .....	30
2.2.1	Phonemic Interfaces .....	30
2.2.2	Efficiency of Orthographic Interfaces .....	32
2.3	Methods .....	36
2.3.1	Phoneme Set .....	36
2.3.2	Corpora .....	38
2.3.3	Calculating Interface Efficiency .....	39
2.3.4	Metropolis Optimization Algorithm.....	43
2.3.5	Evaluation .....	48
2.4	Results .....	51
2.4.1	Corpora similarity .....	51
2.4.2	Interfaces .....	52
2.4.3	Interface efficiency .....	53

2.5	Discussion.....	53
2.5.1	Benefits of Alphabetic and Articulatory Interfaces.....	54
2.5.2	Other Efficiency Calculations.....	55
2.5.3	Clinically Meaningful Speed Improvements.....	58
2.5.4	Future Directions.....	59
2.6	Conclusions.....	60
2.7	Acknowledgments.....	61
Chapter 3. Empirical Evaluation of Optimized and Predictive Interfaces.....		62
3.1	Abstract.....	62
3.2	Introduction.....	63
3.2.1	Phonemic Interfaces.....	64
3.2.2	Quantitatively Optimized Interfaces.....	65
3.2.3	Prediction.....	66
3.2.4	Empirical Evaluation by Users with and without Motor Impairment.....	70
3.3	Methods.....	72
3.3.1	Interface Development.....	72
3.3.2	Optimization.....	73
3.3.3	Prediction.....	74
3.3.4	Surface Electromyographic (sEMG) Cursor.....	76
3.3.5	Participants.....	78
3.3.6	Experimental Designs.....	78
3.3.7	Statistical Analyses.....	85

3.4	Results .....	86
3.5	Discussion .....	90
3.5.1	Estimates of Accuracy.....	91
3.5.2	Sources of Mismatch between Prompt and Produced .....	92
3.5.3	Effects of Prediction .....	94
3.5.4	Effects of Interface Optimization .....	99
3.5.5	Effects of Training and Probes.....	101
3.5.6	Comparison to Other Communication Rates.....	102
3.5.7	Applications to Non-Phonemic Interfaces .....	103
3.5.8	Limitations and Future Directions.....	104
3.6	Conclusions.....	105
3.7	Acknowledgments .....	106
Chapter 4. Clinical Translation and Implications.....		107
4.1	Introduction .....	107
4.2	Methods .....	108
4.2.1	Participants .....	108
4.2.2	Study design .....	109
4.2.3	Analyses .....	110
4.3	Results .....	111
4.4	Discussion.....	118
4.4.1	Comparison to participants without motor impairment .....	118
4.5	Future Design Considerations for Clinical Translation .....	119

4.5.1	Speech synthesis .....	119
4.5.2	Voronoi diagram as predictive marker.....	120
4.5.3	Colors.....	121
4.5.4	Phoneme Labels .....	122
4.5.5	Effectiveness of Optimization .....	124
4.5.6	Flexibility .....	125
4.5.7	Other Elements .....	126
4.5.8	Training system.....	127
4.6	Conclusion .....	128
4.7	Acknowledgments .....	129
Chapter 5. Conclusions and Future Directions .....		130
5.1	Summary.....	130
5.2	Future directions .....	132
APPENDIX .....		135
New AAC interface follow-up .....		135
<b>BIBLIOGRAPHY</b> .....		137
<b>CURRICULUM VITAE</b> .....		152

## LIST OF TABLES

Table 1-1. Existing phonemic interfaces.....	14
Table 2-1. Reduced set of phonemes.....	37
Table 2-2. Corpora .....	39
Table 2-3. Time estimates to produce Suggested AAC corpus .....	57
Table 3-1. Sessions 1–9 mixed effects model - remaining factors after backwards stepwise regression.....	89
Table 3-2. Session 12 linear model; remaining factors after backwards stepwise regression.....	89
Table 3-3. Common mismatches between prompt and phonemes selected .....	93
Table 4-1. Participant characteristics.....	111

## LIST OF FIGURES

- Figure 2-1. Width ( $W_i$ ) and distance ( $D_{ij}$ ) calculations for the interfaces developed and evaluated in this study. Starting phoneme  $i$  outlined in green, with target phoneme  $j$  outlined in red. Width is calculated as the distance between the two intersection points of the ideal path from the center of the starting phoneme through the center of the target phoneme (blue dots).and after the optimization (max efficiency noted: 39.608 WPM via Suggested AAC corpus). ..... 41
- Figure 2-2. Different configurations of 39 targets; (A) shows a 10x10 interface before the Metropolis algorithm has run, in which 39 hexes have been randomly assigned a phoneme (blue) and the rest are unassigned (grey) (B) shows an interface with high efficiency after running the Metropolis algorithm, which has tightly clustered the targets (max efficiency noted: 39.655 WPM via Suggested AAC corpus). (C) has the 39 targets arranged in a consistent layout both before and after the optimization (max efficiency noted: 39.608 WPM via Suggested AAC corpus). ..... 45
- Figure 2-3. Each panel shows the random walk through the interface space via the Metropolis algorithm with a different value for scalar  $k$  with  $T$  (arbitrarily) at 10. Each data point represents the WPM for an interface arrangement with a higher efficiency than the “current” arrangement. .... 46
- Figure 2-4. Metropolis algorithm. Top panel shows one iteration of the algorithm, consisting of 8 million random swaps and annealing system temperature.

Panel B shows the typical process at the beginning of the algorithm, in which the efficiency quickly rises to the neighborhood of the final “optimized” version. Panel C shows how the annealing process (system temperature in green raising and lowering over time) allows the system to come out of local maxima in order to then approach the optimal solution (red dots). . . . . 47

Figure 2-5. Visual comparison of results of Metropolis algorithm. Left shows a random organization of phonemes; right shows an optimized interface. Width of lines between two targets represents the likelihood of transitioning between them in the AAC conversational phrases corpus. . . . . 49

Figure 2-6. Interfaces developed and evaluated in this study. Target colors show rough groupings of phonemes: simple vowels in green, complex vowels (diphthongs, r-colored vowels) in purple, fricatives and affricates in orange, stops in red, and liquids, nasals, and semivowels in blue. . . . . 50

Figure 2-7. Similarity between phoneme-to-phoneme transition probabilities from different corpora (color represents r-value of Pearson’s correlation between all diphone probabilities) . . . . . 51

Figure 2-8. Efficiency in words per minute (WPM) for each interface tested with each corpus. Highlights on the diagonal show when the interface is being tested against the same corpus used to optimize its layout. . . . . 54

Figure 2-9. Percent difference in efficiency compared to a random arrangement of phonemes. . . . . 57

Figure 3-1. sEMG mini-sensor locations (and associated grounds on chest and mastoids), placed to capture muscle activity during a particular facial gesture and subsequent cursor action: Left (half smile); right (half smile); up (eyebrow raise); down (chin contraction); click (wink). Combining gestures allows the cursor to move in any 360° direction, and magnitude of activity controls cursor speed (Cler & Stepp, 2015)..... 77

Figure 3-2. Four interfaces used in by different groups of participants. Top left: random/static interface. Top right: random/predictive interface. Bottom left: optimized/static interface. Bottom right: optimized/predictive interface. Phoneme labels are a standard set (Shoup, 1980). Colors are consistent across groups and were isoluminant. Colors denote rough phoneme category: simple vowels in green, complex vowels (diphthongs, r-colored vowels) in purple, fricatives and affricates in yellow, stops in red, and liquids, nasals, and semivowels in blue. .... 79

Figure 3-3. Experimental design. (A) Processes required to recreate a given prompt with the phonemic interface: translate the stimulus to the phoneme set, find phonemes on given interface, and use access method to move to and select the targets. (B): Main task, with outcome measure communication rate (phonemes/min). (C-E): Probes designed to assess participant acuity on each task: (C) Aural stimulus and phonemic representation with one phoneme missing are presented, and accuracy (% correct) and reaction time (responses/sec) were collected. (D) Participants indicated when they visually

located the given label (outcome measure: reaction time in responses/sec).

(E) Participants used facial sEMG cursor to select circular targets and were assessed on speed (selections/sec)..... 81

Figure 3-4. Communication rates per session averaged by group. Error bars are standard error. .... 88

Figure 3-5. Results of probes. Top left: Motor task speed. Top right: Visual search speed. Bottom left: phoneme identification speed. Bottom right: phoneme identification accuracy (% correct). Error bars are standard error. .... 88

Figure 3-6. Session in which each participant reached criterion of 80% accuracy of selecting [AY] on /aɪ/-initial trials over other vowel labels. Red (dark) bars: predictive groups. Note that these participants largely reached criterion in the first two sessions. Grey striped bars: static groups. Note that these participants took longer to reach criterion, and one participant never reached criterion (grey checked box). .... 99

Figure 4-1. Interfaces used in alternating blocks by participants. Left: optimized/static interface. Right: optimized/predictive interface. Phoneme labels are a standard set (Shoup, 1980). Colors are consistent across interfaces, isoluminant, and denote rough phoneme categories: simple vowels in green, complex vowels (diphthongs, r-colored vowels) in purple, fricatives and affricates in yellow, stops in red, and nasals and semivowels in blue. .... 112

Figure 4-2. Communication rates for the six participants with motor impairment (see Table 4-1 for participant characteristics). Participants all used optimized interfaces. Interfaces were either static (empty shapes) or predictive (filled shapes) in alternating blocks. When possible, participants completed blocks over two days; black dotted vertical lines indicate separation from day 1 to day 2. Error bars are standard deviation. .... 115

Figure 4-3. Survey results for “Did using the interface feel fast?” from “Very slow” to “Very fast” as a function of communication rates..... 116

Figure 4-4. Survey results for “Could you find the targets easily?” from “Not easily” to “Very easily”, as a function of communication rates. .... 116

Figure 4-5. Survey results for “Performance: how successful were you in accomplishing what you were asked to do?” from “Perfect” to “Failure”, as a function of communication rates..... 117

Figure 4-6. Survey results for “Was using the interface frustrating?” from “Not at all frustrating” to “Very frustrating”, as a function of communication rates. 117

## LIST OF ABBREVIATIONS

AAC	.....	Augmentative and alternative communication
ALS	.....	Amyotrophic lateral sclerosis
CI	.....	Confidence interval
CMUDict	.....	Carnegie Mellon University Pronouncing Dictionary
CP	.....	Cerebral palsy
GBS	.....	Guillain-Barré syndrome
IPA	.....	International phonetic alphabet
iScan	.....	Interactive Sound-based Communication Aid for Non-speakers
M	.....	Mean
MS	.....	Multiple sclerosis
NASA-TLX	.....	National Aeronautics and Space Administration Task Load Index
NIH	.....	National Institutes of Health
NSF	.....	National Science Foundation
SCI	.....	Spinal cord injury
sEMG	.....	Surface electromyography
SPEEC	.....	Sequences of Phonemes for Efficient English Communication
VAS	.....	Visual analog scale
wpm	.....	Words per minute

## **Chapter 1. Introduction**

### **1.1 Augmentative and Alternative Communication**

Up to 1.2% of the population is unable to meet daily communication needs using typical speech (Beukelman & Ansel, 1995), and over 500,000 individuals in the United States report that their oral speech cannot be understood at all (Brault, 2012). The inability to communicate disrupts every aspect of life.

Augmentative and alternative communication (AAC) strategies are thus needed by these individuals to communicate. AAC strategies range widely and include manual sign, facial gestures, writing, and choosing letters on a computer using eye-gaze. Individuals with additional motor impairments often cannot access manual sign, gestures, or typical computer access methods (mouse, keyboard) and thus can struggle to notify caregivers of their basic needs, maintain meaningful employment, or connect with family and friends.

Most AAC solutions involve a communication interface and a way to indicate which target the user wishes to select (herein: “access method”). Access methods vary based on an individual’s motor abilities and preferences and can include both direct selection methods like a finger or a head-tracker pointing to a target and indirect methods wherein the only signal is a binary yes/no signal (e.g., as generated by a sip and puff switch). Communication interfaces generally consist of a grid of targets on paper or on a computer screen. Interfaces can vary along a variety of dimensions: symbol set (e.g., letters, phrases, pictures representing given concepts), target layout (“standard”, such as QWERTY;

optimized manually/heuristically; or optimized methodically via modeling), prediction (predictive or non-predictive; word-prediction or icon/letter/character prediction); whether it is static or dynamic (shifting in shape or layout based on prediction).

Communication in individuals with motor impairments that impact both speech and computer access remains slow: 2–15 words per min (wpm), compared to 30–40 wpm by a skilled typist and 150–200 wpm in typical speech production (Beukelman & Mirenda, 2013; Copestake, 1997; Higginbotham, Shane, Russell, & Caves, 2007; Koester & Arthanat, 2017; Leshner, Moulton, & Higginbotham, 1998b; Newell, Langer, & Hickey, 1998). These rates are slow partially due to the motor impairments these individuals exhibit, which require the use of alternative access (e.g., head-tracking, mouthstick, eye-tracking). These access methods remain slow and inherently noisy<sup>1</sup>. Another barrier to achieving faster communication rates is in the design of the communication interface. Methods of optimizing AAC systems (both access and interfaces) may provide significant improvements in quality of life.

AAC strategies are employed by many different individuals. These include individuals with hearing impairment, developmental or congenital disorders, and acquired disorders (Beukelman & Mirenda, 2013). The work in this dissertation

---

<sup>1</sup> Here and throughout, we use “noisy” as a shorthand for the various difficulties inherent in using alternate access methods. In particular, we note that these methods are often dependent on calibrations that vary in stability and quality and are susceptible to both internal and external variability (e.g., changing lighting conditions interfere with eye-tracking; participant head movements can be unsteady based on fatigue).

focuses on a heterogeneous and underserved population of individuals: those who have motor impairment that affects both oral speech and precludes common access methods (manual sign; gesture; touchscreen use with a finger). This group of people includes individuals with developmental and acquired disorders of varying etiology, progression, and functional impairment. Disorders that often require AAC and alternate access methods include: high spinal cord injury, chronic Guillain-Barré syndrome, brain stem stroke, and cerebral palsy. However, this group of individuals is defined less by their specific injuries and more by their behavioral profile: limited motor output of any type, leading to slow and effortful communication.

## **1.2 Access methods**

AAC access often involves finger movements (on a touchscreen or keyboard) or hand movements (on a typical or adapted mouse). For individuals with motor impairments that preclude these access methods, alternative access methods range from devices that track head movements (Williams & Kirsch, 2008, 2016), eye movements (Frey, White, & Hutchinson, 1990; Higginbotham et al., 2007; Leshner, Moulton, & Higginbotham, 1998a), tongue movements (Huo, Wang, & Ghovanloo, 2008), sip and puff actions (e.g., Higginbotham et al., 2007), muscle signals (Cler, Nieto-Castañón, Guenther, Fager, & Stepp, 2016; Cler & Stepp, 2015; Williams & Kirsch, 2015), or brain signals (Brumberg, Nieto-Castanon, Kennedy, & Guenther, 2010; Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). However, all of these access methods remain

noisy and effortful for the user. Despite the technological advances achieved in these areas, target selection can remain slow and effortful, particularly in people with minimal movement capabilities (Beukelman, Fager, Ball, & Dietz, 2007; Fager, Beukelman, Fried-Oken, Jakobs, & Baker, 2012; Higginbotham et al., 2007; Koester & Arthanat, 2017). To improve communication speed and flexibility, further efforts are necessary to improve both AAC access methods and AAC communication interfaces.

### ***1.2.1 Direct selection (continuous access) versus switch selection (binary input)***

If an individual's motor output is limited, they may choose to use a binary switch to indicate yes/no. Switches are designed to capture a variety of inputs, including limited head or limb movements (sufficient to press a mechanical button), sip and puff respiratory actions, and even isolated muscle activity directly (Frick et al., 2017). Arrays of switches are often integrated with power wheelchairs and send messages to the system for such functions as "scroll down on the interface" and "select the highlighted option". When used to access communication, switches are typically paired with a scanning interface. A scanning system often consists of a grid of characters (or other targets) displayed on a computer screen or dedicated AAC device. The system then scans through the targets in one of a few configurations (e.g., linearly, row-column, row-column-group). Scan and selection protocols can also vary (e.g., automatic scanning by computer with one switch to select; two-switch scanning

with one to move between targets and one to select; press versus release to select).

In contrast to the binary options offered by switch and switch-enabled interfaces, continuous methods of computer control (also called direct selection) require more refined motor control, but are typically preferred due to speed and flexibility. Options include finger input (on a touchscreen or a piece of paper), typical or adapted mouse input, stylus input (e.g., mouthstick), and head- or eye-tracking. Direct selection access methods are generally considered to be less cognitively taxing than scanning methods (Sevcik & Ronski, 2000), and thus are typically chosen if AAC users have the physical capability for direct selection.

#### *1.2.1.1 Continuous direct control example – surface electromyographic cursor*

A promising continuous access method for individuals with minimal movement capabilities is that of a surface electromyography (sEMG)-based cursor (Choi, Rim, & Kim, 2011; Cler, Michener, & Stepp, 2014; Cler et al., 2016; Cler, Nieto-Castanon, Guenther, & Stepp, 2014; Cler & Stepp, 2015; Vernon & Joshi, 2011; Vojtech, Cler, Fager, & Stepp, 2018; Williams & Kirsch, 2008, 2016). In this access method, the electrical activity generated by spared muscles is detected from electrodes placed on the surface of the skin and translated to cursor movements. Much of our work has focused on using face- and shoulder-placed electrodes in a model of spared musculature following spinal cord injury (Cler, Michener, et al., 2014; Cler et al., 2016; Cler, Nieto-Castanon, et al., 2014;

Cler & Stepp, 2015), but electrodes can in fact be located over any muscle that can be activated consistently and independently, based on the user's abilities and preferences (Vojtech et al., 2018).

This type of access method offers some possible advantages over other available options for this population, including head-tracking and eye-tracking. For example, head-tracking requires full control of head muscles, which may be impaired in high spinal cord injury or in degenerative conditions such as MS or ALS. Further, head-tracking often utilizes cameras to assess head position, which is degraded if the lighting in the room changes or if the participant or the device move (Beukelman et al., 2007).

Eye-tracking requires high illumination, stable head positions, and complete control over eye movements (Beukelman et al., 2007); in addition, some users report fatigue in eye muscles, and the communication process can be further degraded during conversation due to the loss of directed gaze (Higginbotham et al., 2007). Although some people find success using eye-tracking (either immediately or with training), certain users are unable to effectively use eye-tracking (Bates & Istance, 2003). Further, eye-tracking requires that a user's eyes are used for both input (i.e., reading) and output (i.e., selection), which leads to unproductive and distracting cursor movements and selections while reading (Velichkovsky, Sprenger, & Unema, 1997).

sEMG-based systems do not require a particular positioning or lighting, have not yet been reported to cause fatigue, and do not require the use of the

eyes as both input and output. Further, sEMG can capture activity in hemiparetic muscles that are innervated but do not have adequate innervation and/or strength to support movement (Saxena, Nikolić, & Popović, 1995). Individuals can learn to control activity of even single motor neurons (Basmajian, 1972); this suggests that an sEMG cursor may be available to individuals with very little residual muscle control, which may not be detectable by mechanical interfaces or camera-based devices.

In our laboratory, sEMG cursors have produced information transfer rates (ITRs; represent speed and accuracy in one measure) at a mean of 69.6 bits/minute on the first day of use to 120.7 bits/min on the fourth session of use (Cler & Stepp, 2015). When using a phonemic interface but prompted with the sound labels to choose, participants used an sEMG cursor to produce ITRs of 53 bits/min during the first session to 111 bits/min during the eighth session (Cler et al., 2016). This is in the range of other relevant access methods, including eye-tracking (60–222 bits/min with predictive methods; Frey et al., 1990; Higginbotham et al., 2007; Liu et al., 2012; Majaranta, MacKenzie, Aula, & Rähä, 2006), head tracking (78 bits/min Williams & Kirsch, 2008), and non-invasive brain-computer-interfaces (1.8–24 bits/min; Blankertz, Dornhege, Krauledat, Müller, & Curio, 2007; Nijboer et al., 2008; Sellers, Krusienski, McFarland, Vaughan, & Wolpaw, 2006; Wolpaw et al., 2002).

### 1.2.2 Modeling AAC access

Computer access methods can be modeled via Fitts' law, a fundamental model of directed movements that applies to most human movements (Fitts, 1954). Fitts' law suggests that the time it takes to select a target is based on the target's distance and size – a nearer target is faster to select because it requires less movement, and a larger target is faster to select because it requires less precise movements.

Fitts' law (Equation 1-1) suggests that the movement time (MT) necessary to travel between targets  $i$  and  $j$  is related to the distance between the centers of target  $i$  and target  $j$  ( $D_{ij}$ ), and the width of the second target ( $W_j$ )<sup>2</sup>. Exact movement times are determined experimentally based on the pointing device employed; these are then used to derive the constants  $a$  and  $b$ .

$$MT = a + b \left[ \log_2 \left( \frac{D_{ij}}{W_j} + 1 \right) \right] \quad \text{Eq. 1-1}$$


Fitts' law has been used to define the characteristics of many human movements and computer access methods (Plamondon & Alimi, 1997), including: finger and wrist movement (Langolf, Chaffin, & Foulke, 1976), hand-held stylus (Fitts, 1954), joystick (Card, English, & Burr, 1978; Epps, 1986), typical computer mouse (Card et al., 1978; Epps, 1986), head-controlled computer input devices (Radwin, Vanderheiden, & Lin, 1990; Williams & Kirsch,

---

<sup>2</sup>Fitts originally used the equation  $MT = a + b \left[ \log_2 \left( \frac{2D_{ij}}{W_j} \right) \right]$ ; but the human-computer interaction community generally uses the "Shannon" version of the equation as shown in Eq. 1-1 (MacKenzie, 1992). This modification changes the fit of Fitts' law to better approximate very small target widths.

2008), and sEMG-based input devices (Choi & Kim, 2007; Vojtech et al., 2018; Williams & Kirsch, 2015). Interestingly, some reports have indicated that eye-tracking input follows Fitts' law (Miniotas, 2000; Vertegaal, 2008), and some reports suggest it does not (Sibert, Templeman, & Jacob, 2001; Ware & Mikaelian, 1986; Zhai, Morimoto, & Ihde, 1999).

### 1.3 AAC Symbol Sets

Targets on AAC interfaces couple a symbol with an underlying referent. The tightness of this coupling is the *iconicity* of the symbol (Bloomberg, Karlan, & Lloyd, 1990; Fuller & Lloyd, 1991; Fuller, Lloyd, & Schlosser, 1992), which can vary from translucent (such as a physical toothbrush representing the concept “toothbrush”) to opaque (such as the symbol  for the concept “machine”). Letters are opaque, as letters are grouped into graphemes that are somewhat consistent in mapping to a phoneme or set of phonemes. The physical form of the letter is not related to the phoneme; phonemes are not directly related to underlying meaning<sup>3</sup>.

AAC interfaces typically provide targets consisting of letters, whole words, or symbols (typically also representing whole words or phrases). Each presents benefits and drawbacks, typically compromising between speed, flexibility, and

---

<sup>3</sup> This particular concept has been debated since Plato and is not as straightforward as presented here. For example, words representing small things often contain /i/ (made with a small, constricted vocal tract), whereas those representing large objects often contain /a/, which is made with a large, open vocal tract (Miall, 2001; Schmidtke, Conrad, & Jacobs, 2014). This is consistent across a variety of languages and cultures, including French, Spanish, and Chinese (Shrum, Lowrey, Luna, Lerman, & Liu, 2012).

cognitive load. For example, one set of symbols created for AAC is called Blissymbols or Blissymbolics, in which a drawn symbol represents a given concept separate from English (Bristow & Fristoe, 1984; Hurlbut, Iwata, & Green, 1982; Mizuko, 1987). Blissymbols can be combined to create new concepts. However, they do not necessarily translate easily into spoken language and require much training. Alternately, an interface of letters (e.g., QWERTY on-screen keyboard) has the same symbol and referent and enables participants to produce flexible communication, but requires literacy, has a high working memory load if access is very slow, and is comparatively slow. Interfaces made up of pictures or line drawings each representing a word or phrase offer comparatively fast speeds to produce those words, but are less flexible (e.g., only those phrases are easily produced) and the opaqueness of the symbol may vary. Offering sounds as targets may give some speed advantage, while preserving flexibility. These will require the user to either have or develop phonological awareness.

### **1.3.1 Phonemic AAC Interfaces**

Some AAC interfaces have been developed that use phonemes (which represent a particular sound in a spoken language) as targets (Black, Waller, Pullin, & Abel, 2008; Cler et al., 2016; Cler, Nieto-Castanon, et al., 2014; Schroeder, 2005; Trinh, Waller, Vertanen, Kristensson, & Hanson, 2012; Vertanen, Trinh, Waller, Hanson, & Kristensson, 2012).

Phoneme selection allows individuals to create any set of sounds in their language, rather than relying on text-to-speech methods. Of particular interest to speakers with motor impairments, common AAC messages have 14–20% fewer phonemes than letters, depending on the set of vocabulary or messages evaluated (Cler et al., 2016). This may reduce the time needed to produce messages while retaining full flexibility. The drawback to using phonemes as interface targets is that one must learn to translate intended thoughts into a phoneme set rather than into letters. Typically we spend many years as children learning to translate thoughts into letters (i.e., writing). Sequencing phonemes (or syllables), although more similar to typical oral communication, is also likely to require training in order to produce intended messages. However, the speed and flexibility advantages to phonemic input suggest that it may be appropriate for some users, and thus effort should be expended to develop the most efficient phonemic interface possible. Further, there is some evidence that alternate communication forms could be used even in the event of brain damage leading to an inability to read standard text (Regard, Landis, & Hess, 1985).

### ***1.3.2 Existing Phonemic Interfaces***

Phonemic interfaces have previously been proposed for use by a variety of user populations, including children and adults with learning disabilities and/or motor impairments (Black et al., 2008; Schroeder, 2005; Trinh et al., 2012; Vertanen et al., 2012). Phonemic interfaces have primarily been suggested as a way to provide quick and intuitive access to oral language, and as such have

been shown to increase phonological awareness in adults with disabilities (Trinh, 2011). Previous phonemic interfaces and their attributes are shown in Table 1-1, including, where available, reported improvements in communication rate over orthographic interfaces. Based on the target population and available technology, these systems vary across a variety of dimensions including intended access method, phoneme set, target arrangement, and the availability of prediction (see Table 1-1 for summary).

The first phonemic interface was the Phonic Ear HandiVoice, developed in 1978 as an early voice-output communication aid (Glennen & DeCoste, 1997; Vanderheiden, 2002). Users memorized and typed three digit codes, which each represented one phoneme and provided early speech synthesis. Another early phonemic interface was the SPEEC (Sequences of Phonemes for Efficient English Communication) system, in which users were presented with 256 or 400 items consisting of phonemes and frequent phoneme sequences (Goodenough-Trepagnier & Prather, 1981). This partner-assisted interface involved users pointing to a given target, which was then pronounced by their communication partner. Empirical evaluations suggested a 30% increase over alphabetic input (Goodenough-Trepagnier, Tarry, & Prather, 1982), but suggested that some participants needed 4–8 months of training for proficiency (Goodenough-Trepagnier & Prather, 1981).

The purpose and thus design of more recent phonemic interfaces has varied. Some have been developed to improve literacy in children with minimal

spoken output (Black, 2011) or help children who are poor spellers produce written text (Schroeder, 2005). Some systems contain only a small set of phonemes (Black et al., 2008), or display a reduced set of phonemes on the screen at one time (Trinh et al., 2012), such that users must make several motor actions to select one phoneme (e.g., selecting one target to indicate that you wish to select a fricative, and then selecting a target on a second screen that appears to select /f/). Other systems use a reduced set of phonemes and then must disambiguate the intended selections based on prior selections (Vertanen et al., 2012).

Table 1-1. Existing phonemic interfaces

Interface	Access method	Phoneme set	Arrangement	Prediction	Improvement
Phonic Ear HandiVoice (described in Glennen & DeCoste, 1997)	3 digit codes (input by hand or stylus)	Full (48)	--	N	--
SPEEC system (Goodenough-Trepagnier & Prather, 1981; Goodenough-Trepagnier et al., 1982)	Partner-assisted; point to letter or a letter combination	256 or 400 (depending on version) phonemes and syllables	Ordered by frequency or alphabetical	N	30% increase
REACH Sound-It-Out (Schroeder, 2005)	Direct selection	Full (44)	Clustered by category of sound	Disables unlikely targets	Accuracy improvements
PhonicStick (Black, 2011; Black et al., 2008)	Integrated with joystick access	Limited (6)	In a circular arrangement	N	--
iScan (Trinh et al., 2012)	Integrated with joystick access	Full (42)	Multiple layers of 8 phonemes	Reorders phonemes; word completion	70% improvement by user with disabilities
Articulatory feature-based (Cler et al., 2016)	Touchscreen; typical or adapted or sEMG mouse	Full (39)	Ordered by articulatory features (place, manner, voicing)	N	--

### 1.3.3 Phonemic Targets and Labels

A variety of phoneme sets and labels have been suggested for American English. Speech language pathologists often learn to transcribe speech using the International Phonetic Alphabet (IPA; International Phonetic Association, 1999). This system provides a consistent set of 107 phonemes and was designed to be usable across all languages. It also provides methods for narrow transcriptions, in which transcribers can denote various specifics of the production via diacritics (e.g., voicing or nasalization on sounds that are typically voiceless or not nasalized). IPA contains symbols that are not typically contained in a particular typeface or understood by computers. Thus a machine readable set of 47 English phonemes called ARPABET was developed by the Advanced Research Projects Agency (Shoup, 1980). A modified version of ARPABET is used in the Carnegie Mellon University Pronouncing Dictionary (CMUDict; Weide, 2005); 39 phonemes are represented<sup>4</sup>. These 39 phonemes are used in our paper as a minimally sufficient set of English phonemes. Some dialects could further combine some sounds (e.g., [AO]-/ɔ/-“ought” and [AA]-/ɑ/-“father” are collapsed in Cler et al., 2016).

Phonemic interfaces have used different phoneme sets and labels as well.

The articulatory feature-based interface used researcher-defined phoneme sets

---

<sup>4</sup> The phonemes included in standard ARPABET but not the CMUDict are [AX]-/ə/-“about” (collapsed with [AH]-/ʌ/); [IX]-/i/-“debit”; [WH]-/m/-“which” (voiceless labial-velar fricative); [EL]-/l/-“bottle” (syllabic l); [EM]-/m/-“rhythm” (syllabic m); [EN]-/n/-“button” (syllabic n); [DX]-/r/-“batter” (flap); [Q]-/ʔ/- (glottal stop). In addition, CMUDict uses [NG] as a label instead of [NX]; we use [NG].

and labels (Cler et al., 2016; Cler, Nieto-Castanon, et al., 2014), as did the REACH Sound-It-Out keyboard (Schroeder, 2005). The iScan and the associated PhonicStick use a set of 42 phonemes used in a phonics literacy program (Lloyd, 1992); the labels used are pictures and optionally letters or digraphs. Here we used labels consistent with the CMUDict, which are one to two letter English transliterations of phonemes (Shoup, 1980; Weide, 2005; also see *Chapter 2 > 2.3.1 Phoneme Set > Table 2-1* for phoneme set).

## **1.4 AAC Target Layout**

Individuals with and without motor impairments use communication interfaces, particularly on computers and cellular phones. The most ubiquitous communication interface is the standard QWERTY keyboard (the Sholes keyboard, designed in 1873). This keyboard is highly inefficient for ten-finger typing and for serial input, as with a stylus; in fact, it was designed to be inefficient so as to minimize jamming typewriter keys (Noyes, 1983; Rumelhart & Norman, 1982). Alternate keyboards have been developed to increase communication rates for ten-finger typing and for serial input.

### **1.4.1 Ten-finger typing**

The most common alternate layout for English is the Dvorak keyboard (Dvorak & Dealey, 1932). This keyboard shows ~4% improvements in typing speed relative to a QWERTY keyboard (West, 1998). However, ten-finger typing is a parallel process, in which 90% of finger movements are initiated before the previous key is pressed (Gentner, Grudin, & Conway, 1980; Rumelhart &

Norman, 1982), and is thus difficult to model and optimize (Rumelhart & Norman, 1982).

Further, individuals with sufficient ten-finger motor control to type will likely not see large enough differences in typing rates to justify the cognitive and practical downsides to alternate keyboards. Professional typists, such as stenographers, do use alternate keyboard layouts. Interestingly, many systems of shorthand (methods designed for fast transcription) do in fact use phonetic transcriptions, rather than a symbolic representation per word or a shortened orthographic representation. Transactions of the “First International Shorthand Congress” (Axon, 1888) reveal that most attendees discussed a shorthand method called phonography (or Pitman shorthand), which indeed uses phonemic transcription. Stenographers use phonetic input as well (e.g., the word *cat* starts with the representation for *k*).

#### **1.4.2 Serial input**

QWERTY keyboards are particularly inefficient for serial input, such as when individuals are entering text on a touchscreen with a stylus or a finger, or when using a scanning interface. This process (serial input) is more easily modeled and has been optimized by a variety of research groups. Optimization for scanning typically includes re-ordering targets (often letters) by frequency (Leshner et al., 1998b) or changing the scanning pattern based on prediction (Baljko & Tam, 2006).

Direct selection can be modelled via Fitts' law (see 1.2.2 *Modeling AAC access* for more information). Briefly, Fitts' law states that the time required to select a target is a function of its size and the distance to be traveled to reach it (Fitts, 1954); near targets are faster to select (smaller distance to be travelled), as are large targets (less precision needed, thus faster movements are possible).

This can be expanded to calculate the efficiency of any particular arrangement of targets. The efficiency of a given interface can be calculated by multiplying the movement time required to travel between each targets by the likelihood those two targets will be selected in series (MacKenzie & Zhang, 1999; Zhai, Hunter, & Smith, 2002). Equation 1-2 shows the average movement time for an interface ( $\overline{MT}$ ), which is calculated as the sum of the probability of transitioning between each pair of targets (i and j) multiplied by the Fitts' law calculation of the time it would take to get from target i to target j.

$$\overline{MT} = a + \sum_{i=1}^N \sum_{j=1}^N Pr_{ij} * b \left[ \log_2 \left( \frac{D_{ij}}{W_j} + 1 \right) \right] \quad \text{Eq. 1-2}$$

Importantly, methods of calculating efficiency for direct selection rely on the frequency of letter-to-letter transitions, or *digraph statistics*, in which “digraph” means letter pairs. For example, if the word “the” appears in a corpus many times, the digraphs T→H and H→E (and space→T and E→space) will have high probabilities. For orthographic text entry, many researchers use digraph statistics from Mayzner and Tresselt (1965). However, some have noted that these traditional digraph likelihoods do not represent AAC usage (Wandmacher & Antoine, 2006). Various text and conversational corpora have been used to

calculate digraph likelihoods and to train language models for prediction. Results show that while testing and training models on the same corpus leads to the best keystroke savings, some savings can still be found even when training and testing on different text corpora (Wandmacher & Antoine, 2006).

### **1.4.3 Existing optimized (orthographic) interfaces**

A variety of optimized orthographic target arrangements have been developed for both physical and onscreen keyboards. Physical keyboard optimizations include the Dvorak typewriter keyboard (Dvorak & Dealey, 1932) which has empirically been shown to increase typing rates by 4% (West, 1998). An optimized keyboard for one-at-a-time direct selection (e.g., stylus held in mouth) was algorithmically optimized as early as 1986 (Getschow, Rosen, & Goodenough-Trepagnier, 1986). A variety of other optimizations followed (Chubon & Hester, 1988; Lewis, Kennedy, & LaLomia, 1999; Lewis, LaLomia, & Kennedy, 1999; MacKenzie & Zhang, 1999; Smith & Zhai, 2001; Textware Solutions, 1998; Zhai et al., 2002), ranging from 18.9% to 53.9% improvements in (varying theoretically- and empirically-derived) communication rates. Most of these were optimized using some combination of trial-and-error and incorporating letter frequency-of-use and letter-to-letter transition likelihoods. The keyboard with the highest theoretical improvement over QWERTY is the Metropolis interface, developed by Zhai and colleagues (2002). Methods in Chapter 2 for optimizing interfaces followed the same process, which used Fitts' law and its

extension for interfaces (Eq. 1-1 and Eq. 1-2) and was optimized using the Metropolis algorithm.

## 1.5 Prediction

A common way of increasing communication rates is to incorporate prediction. Previous studies have shown that adding prediction to orthographic (letter-based) interfaces improves communication rates by 58.6% (Trnka, Mccaw, Yarrington, Mccoy, & Pennington, 2009) and can improve communication rates in phonemic interfaces by 100% (Trinh et al., 2012; Vertanen et al., 2012). Prediction can be divided into word prediction and character prediction, but both generally function the same way: statistics are derived from large corpora of text and then used to generate predictions. For example, the letter “p” is often followed by “h”, “r”, or a vowel; thus if a user selects “p”, a character predictive system should suggest those letters. This example uses 1-character back to offer suggestions. It could use any number of characters back, or ‘n’. This type of prediction is called an n-gram. Word prediction typically uses this same method (in which each “gram” is a word instead of a character), but can also use language rules and labels to produce more sophisticated prediction. Regardless, prediction requires some bank of text or messages from which to derive predictions. Often these banks are based on generic corpora but then updated with user-specific vocabulary and probabilities.

### **1.5.1 Word Prediction in AAC**

Word prediction is typically implemented via n-grams and displayed to the user as a list of suggestions. Although word prediction is designed to increase communication rates, some empirical studies have found that it is not beneficial (Venkatagiri, 1993). For word prediction to be effective, the user must look at an alternate part of the computer interface (the word list, rather than the keyboard), search through several predicted options, decide if the correct word is there, and select the correct word if it appears (Horstmann & Levine, 1991). Each of these steps requires both time and cognitive effort, which may slow down communication rate while improving keystroke efficiency (Magnuson & Hunnicutt, 2002). Users prone to typing or spelling errors may prefer to use prediction to ensure properly spelled messages, but also may spend more time and cognitive effort searching a word-list for a word that will not appear due to an early spelling error. Some of these drawbacks can be remedied by more accurate or sophisticated predictive systems (e.g., Trnka et al., 2009), but some are inherent to any word prediction system. That is not to say that word prediction is never beneficial, however. Word prediction often reduces keystrokes at the cost of increased cognitive and perceptual load. Individuals who type quickly may find that word prediction methods are more distracting than useful. Individuals whose motor impairments make each selection effortful may benefit from word prediction in ways that are not reflected by communication rate.

### **1.5.2 Character Prediction in AAC (Via Dynamic Keyboards)**

Another approach to increasing communication rates is to dynamically update the position of targets on the screen, such that likely targets are in a highly-visible location. Dynamic keyboards are primarily designed for people who use scanning input (Heckathorne, Voda, & Leibowitz, 1987), although dynamic keyboards designed for direct selection have also been proposed. One example of such an approach is the Custom Virtual Keyboard designed by Pouplin and colleagues (2014), in which the arrangement of the targets changed after each selection. Empirical evaluations in primarily direct-selecting participants actually showed *reduced* communication rates by a mean of 37%, likely due to the increased visual search time and cognitive load involved. However, the one participant who used scanning access with this interface preferred the dynamic keyboard and saw some increase in communication rate.

A dynamic keyboard that did increase communication rate is the SpreadKey system (Merlin & Raynal, 2010), in which unlikely letters were dynamically replaced with likely letters. Thus likely letters were represented in multiple places on the interface. Participants gained approximately 20% in communication rates compared to a QWERTY keyboard. Even when simulations suggest that ideal use would increase communication rates, users have stated a preference for static keyboards over dynamic keyboards due to the cognitive load required for the dynamic keyboard (Merlin & Raynal, 2010; Pouplin et al., 2014).

While these dynamic keyboards keep a static overall layout while changing the target labels, an alternative text entry system called Dasher dynamically changes the size and layout of the targets. The targets are linearly displayed on the right side of the screen and move up and down on the screen based on the relative likelihoods of the different targets. This increases target selection speed (Ward & MacKay, 2002). As a result, the system can be distracting or disorienting, and users have reported that it requires a large amount of concentration to use (Tuisku, Majaranta, Isokoski, & R  ih  , 2008). Further, this method does not take advantage of enlarged targets as a visual search aid during training; because the position of the targets changes, users must visually search for every target, regardless of the level of training.

### ***1.5.3 Reduced / Disambiguating Interfaces***

Some prediction methods disambiguate words from an ambiguous entry (Kreifeldt, Levine, & Iyengar, 1989; Kushler, 1998; Leshner et al., 1998a; Levine & Goodenough-Trepagnier, 1990), which may be familiar from the T9 texting system (Kushler, 1998) or Swype (Smith & Chaparro, 2015), which disambiguate text from a reduced keyboard (such as on cellular phones with physical numerical buttons) or from a continuous finger drag (across an onscreen keyboard), respectively. These methods constrain possible selections to only those contained in the dictionary, reducing flexibility (Arnott & Javed, 1992; Kreifeldt et al., 1989; Leshner et al., 1998a). If misspellings occur, or the target word does not appear in the dictionary, users must spell letter-by-letter, which

takes 2-4 additional keystrokes per letter as compared to a typical (non-reduced) keyboard, depending on the number of letters per key and the order of the target letter.

#### **1.5.4 Phonemic Prediction**

Phonemic prediction is a subset of character or word prediction and has previously been implemented and empirically evaluated in two interfaces: the REACH Sound-It-Out Phonetic Keyboard (Schroeder, 2005) and iScan (Trinh et al., 2012), a touchscreen interface designed to be compatible with the PhonicStick (phoneme access via joystick; Black et al., 2008). A 12-key reduced/disambiguating phonemic keyboard with prediction was also developed and computationally evaluated (Vertanen et al., 2012).

##### *1.5.4.2 REACH Sound-It-Out Phonetic Keyboard*

The REACH Sound-it-out keyboard was developed to enable users with learning disabilities to produce orthographic text by entering phonemic sequences. It offers 44 targets (40 phonemes and 4 phoneme combinations). Users select phonemic targets with both a letter and picture exemplar (e.g., /f/ shows a fish and the letter “F”). After a user selects one target, all targets that do not follow that target in the loaded dictionary are disabled (letter prediction), and orthographic words that contain those sounds are offered with extra disambiguating text (word prediction). For example, if users select /nu/, the interface disables some subset of phonemic targets and offers “new (not old)”

and “knew (I knew that)” (Schroeder, 2005). The dictionaries used for prediction were not specified and likely proprietary.

Empirical evaluations in participants identified as typically-developing and poor spellers (categorized by a teacher, parent, or themselves as a poor speller) revealed an increase in accuracy for all groups to differing extents: 196% improvement in 9 children who were poor spellers; 54% in 11 typically-developing children; 122% improvement in 10 adults who were poor spellers; and 16% in typical adults. Communication rate was not reported as it was not the main objective of the interface and experiment.

#### *1.5.4.3 iScan*

The iScan (Interactive Sound-based Communication Aid for Non-speakers) was designed to be compatible with the PhonicStick and thus offers 9 targets at a time in a circular arrangement (Trinh et al., 2012). All 42 phonemes are available by first selecting a group of sounds on the first “layer” (e.g., plosives), selecting the appropriate specific phoneme on the next layer, and selecting a center target to confirm the selection. Once one target is selected, two things happen in order to help the user to quickly find the next target: first, the exemplars used to represent each category on the first layer are changed to be the likeliest choice in each category; second, the order of the offered phonemes on each lower layer is updated so that the likelier phonemes are near the exemplar and thus, ideally, faster to select. The groups are dictated by manner of articulation and color-coded by group (Trinh et al., 2012). In a

touchpad implementation, word prediction was also offered (one predicted word per entered phoneme).

Prediction was implemented with a 6-gram phoneme model and 3-gram word model. The 6-gram phoneme model predicted the likeliest next phoneme based on up to five of the preceding phonemes. The 3-gram word model predicted the next word using up to two preceding words. The models would ideally be trained on AAC corpora, but these corpora do not exist. Instead, these were trained on a set of AAC-like sentences from Twitter, blog, and Usenet datasets, selected to be similar to AAC-like sentences generated by crowdsourcing (Vertanen & Kristensson, 2011). The corpus was converted to phonemes automatically.

Empirical evaluations were completed on a touchpad in which predictive and non-predictive versions of the interface were presented to 16 university students without disabilities in three sessions. Average communication rates were 3.0 wpm on the non-predictive interface and 6.29 wpm on the predictive interface (improvement of 109%). One user with cerebral palsy and significant spelling difficulties (30% accuracy on a real-word spelling task) used the interfaces over two sessions with a speed of 0.74 wpm in non-predictive mode<sup>5</sup> and a mean of 1.72 wpm in predictive mode (improvement of 132%).

---

<sup>5</sup> The participant declined to continue using the non-predictive version and did not complete the trial; he transcribed one phrase only with the non-predictive version.

#### *1.5.4.4 Ambiguous Phonemic Keyboard*

A final predictive phonemic interface was produced by the same group as those who created iScan. It consists of a 12-key phoneme keyboard, set up like a telephone keypad. Phonemes are split into eight categories (front vowels, open vowels, rounded back vowels, voiceless plosives, voiced plosives, nasals and approximants, voiceless fricatives, and voiced fricatives). On a non-predictive setting, users would click the group repeatedly until the correct phoneme appeared and was selected. In the predictive setting, users would click the group and then the likeliest phoneme would appear in order. After each selection, five predicted words would also be presented.

The interface used the same phonemic 6-gram phoneme model and 3-gram word model (Vertanen et al., 2012). It further combined the two: given entered phonemes (“phoneme prefix”), the interface searched for matching words in the dictionary. Matching words were input to the word model to calculate their probabilities based on the two previously entered words. Although no empirical evaluation was reported, theoretical evaluations suggested a keystroke savings of 56.3% (Vertanen et al., 2012).

## **1.6 Purpose of this work**

This dissertation encompasses the development, theoretical evaluation, and empirical evaluations of optimized and predictive phonemic interfaces for the purpose of augmentative and alternative communication. Phonemic interfaces have not previously been optimized for speed. Although prediction has been

previously implemented in some phonemic interfaces, the method of alerting users to likely targets implemented here (dynamically enlarging them via Voronoi diagrams) has not previously been used in communication interfaces. The first chapter contains an introduction to relevant topics in this area; the second chapter consists of the development and a theoretical evaluation of optimized phonemic interfaces; the third chapter describes a systematic empirical evaluation of optimized and predictive phonemic interfaces in participants without motor impairments, the fourth chapter consists of a case-based evaluation of predictive phonemic interfaces in participants with motor impairments, and the final chapter consists of a summary as well as future directions.

## Chapter 2. Development and Theoretical Evaluation of Optimized Phonemic Interfaces

### 2.1 Abstract

In this paper, optimized communication interfaces in which users select phonemes (sounds) instead of letters or whole words are presented and evaluated. Optimization was based on phoneme transition likelihoods (i.e., the probability of transitioning from one phoneme to another in a particular communication corpus), similar to letter-to-letter transition likelihoods used to optimize orthographic interfaces. However, it is unknown to what extent phoneme transition likelihoods vary by corpus, nor how optimizing based on different corpora affects the final interface efficiency. Here we used computational evaluations to compare phoneme transition likelihoods between various phonemic corpora and optimize phonemic interfaces with each corpus. Each interface's efficiency was evaluated against all the corpora. Phoneme-to-phoneme transitions were highly correlated across corpora ( $r = .7-.86$ ). Optimization based on phoneme-to-phoneme transition likelihoods improved efficiency by around 20–30% compared to random phonemic layouts, regardless of the corpus used to optimize the interface. Optimizations using different corpora were similar, varying only by 3–5%. We conclude that, if possible, future phonemic interfaces should be optimized via a corpus from the intended user's communication. If this is not possible, however, optimization still improved

efficiency using all testing corpora, suggesting that optimizing via any relevant corpus is indicated over other layouts.

## **2.2 Introduction**

Some individuals use augmentative and alternative communication (AAC) methods to communicate, including those who have concomitant motor impairments. For these individuals, AAC use requires both an interface from which to select targets and a method by which to select those targets. Alternative access methods for people with motor impairments range from devices that track head movements (Williams & Kirsch, 2008, 2016), eye movements (Frey et al., 1990; Higginbotham et al., 2007; Leshner et al., 1998a), tongue movements (Huo et al., 2008), sip and puff actions (e.g., Higginbotham et al., 2007), or brain signals (Brumberg et al., 2010; Wolpaw et al., 2002). However, all of these access methods remain noisy and effortful for the user. Despite the technological advances achieved in these areas, target selection can remain slow and effortful, particularly in people with minimal movement capabilities (Beukelman et al., 2007; Higginbotham et al., 2007). To improve communication speed and flexibility, further efforts are necessary to improve both AAC access methods and AAC communication interfaces.

### **2.2.1 Phonemic Interfaces**

AAC interfaces typically provide targets consisting of letters, whole words, or symbols (typically also representing whole words or phrases). Each presents benefits and drawbacks, typically compromising between speed and flexibility.

Some AAC interfaces have been developed that use phonemes (which represent a particular sound in a spoken language) as targets instead. Phoneme selection allows individuals to create any set of sounds in their language, rather than relying on text-to-speech methods. Of particular interest to speakers with motor impairments, common AAC messages have 14-20% fewer phonemes than letters, depending on the set of vocabulary or messages evaluated (Cler et al., 2016). This may reduce the time needed to produce messages while retaining full flexibility. The drawback to using phonemes as interface targets is that one must learn to translate intended thoughts into a phoneme set rather than into letters. Typically we spend many years as children learning to translate thoughts into letters (i.e., writing). Sequencing phonemes (or syllables), although more similar to typical oral communication, is also likely to require training in order to produce intended messages. However, the speed and flexibility advantages to phonemic input suggest that it may be appropriate for some users, and thus effort should be expended to develop the most efficient phonemic interface possible.

Phonemic interfaces have previously been proposed for use by a variety of user populations, including children and adults with learning disabilities and/or motor impairments (Black et al., 2008; Schroeder, 2005; Trinh et al., 2012; Vertanen et al., 2012). Some systems contain only a small set of phonemes (Black et al., 2008), or display a reduced set of phonemes on the screen at one time (Trinh et al., 2012), such that users must make several motor actions to

select one phoneme (e.g., selecting one target to indicate that you wish to select a fricative, and then selecting a target on a second screen that appears to select /f/). Other systems use a reduced set of phonemes and then must disambiguate the intended selections based on prior selections (Vertanen et al., 2012), somewhat similar to the T9 texting system (e.g. Kushler, 1998). Finally, some phonemic interfaces display all possible phonemes, but disable phonemes that are unlikely to be selected next (Schroeder, 2005). Unfortunately, these final two methods for increasing efficiency restrict users to selecting only those words contained in the system's dictionary, without allowing for non-words, proper names, or novel utterances.

It has previously been shown that participants without motor impairments could use a noisy AAC access method (Cler & Stepp, 2015) to produce speech using a phonemic interface in which all phonemes are available to select at all times (Cler et al., 2016). This interface had phonemic targets arranged *a priori* based on articulatory features. However, other ways to improve efficiency of phonemic interfaces in which all phonemes are available to select at all times have not yet been explored.

### **2.2.2 Efficiency of Orthographic Interfaces**

A variety of methods for optimizing orthographic arrangements have been employed, for both physical keyboards (e.g., the Dvorak typewriter keyboard; Dvorak & Dealey, 1932) and onscreen keyboards (e.g., OPTI II, MacKenzie & Zhang, 1999; FITALY, Textware Solutions, 1998; or ATOMIK, Zhai et al., 2002).

Many of these were optimized using some combination of trial-and-error and manual incorporation of letter frequency-of-use and letter-to-letter transition likelihoods. Due to the ubiquity of QWERTY keyboards, most users (with and without motor impairments) do not choose a more efficient orthographic keyboard layout. AAC users do sometimes use an alphabetic arrangement or a frequency-based arrangement, particularly if using a very slow scanning method of communication access. However, if users choose to utilize a phoneme-based interface due to its flexibility and the reduced number of selections required, they will not have a previously-learned arrangement of targets (such as QWERTY in an orthographic interface) to produce interference, so learning an optimal target arrangement is likely not appreciably different from learning any other target arrangement.

#### *2.2.2.5 Optimizing Interface Efficiency*

Direct selection access methods (e.g., finger pointing, head-tracking, eye-tracking) are generally considered to be less cognitively taxing than scanning methods (Sevcik & Ronski, 2000), and thus are typically chosen if AAC users have the physical capability to directly select. Optimizing the layout of an interface used in switch scanning typically involves reordering the targets by frequency of use, such that those targets that are likeliest appear in the beginning of the scanning process (Leshner et al., 1998b). There are also methods of optimizing the arrangements of targets on an interface to be used with direct selection; while this efficiency optimization has been implemented for

orthographic keyboards (e.g. (Zhai et al., 2002)), it has not been applied to phonemic interfaces.

One way to calculate and then maximize the efficiency of an interface is based on Fitts' law, a fundamental model of directed movements that suggests that the time it takes to select a target is based on the target's distance and size – a nearer target is faster to select because it requires less movement, and a larger target is faster to select because it requires less precise movements. To optimize the efficiency of an interface, one can arrange the targets such that the distance between targets that are often selected sequentially is minimized.

Importantly, methods of calculating efficiency for direct selection rely on the frequency of letter-to-letter transitions, or *digraph statistics*, in which “digraph” means letter pairs. For example, if the word “the” appears in a corpus many times, the digraphs T→H and H→E (and space→T and E→space) will have high probabilities. For orthographic text entry, many researchers use digraph statistics from Mayzner and Tresselt (Mayzner & Tresselt, 1965). However, some have noted that these traditional digraph likelihoods do not represent AAC usage (Wandmacher & Antoine, 2006). Various text and conversational corpora have been used to calculate digraph likelihoods and to train language models for prediction. Results show that while testing and training models on the same corpus leads to the best keystroke savings, some savings can still be found even when training and testing on different text corpora (Wandmacher & Antoine,

2006). It is not clear whether this holds true with *diphone* likelihoods (phoneme-to-phoneme transition likelihoods) and resulting phonemic interfaces.

In this paper we present the results of optimizing and then testing the efficiency of phonemic interfaces using a variety of corpora to determine if phonemic interfaces optimized for AAC must be tailored to each user or if one generic keyboard (e.g., QWERTY, ATOMIK orthographic keyboards) is sufficiently efficient for all users.

#### *2.2.2.6 Research Questions and Motivation*

It is currently unknown to what extent phoneme transition likelihoods vary by corpus, nor how optimizing based on different corpora affects the final interface efficiency. In this study, we evaluated phoneme transition likelihoods between various phonemic corpora and optimize phonemic interfaces with each corpus. Then we evaluated each interface's efficiency by testing against all the corpora. If interface efficiency is highly impacted by the testing corpus, communication interfaces should be optimized per user. If efficiency is stable across testing corpora, we would expect AAC users to show similar performance using arrangements of phonemes based on any number of corpora. For reference, we additionally evaluated the efficiencies of two potential phonemic interfaces that were not explicitly optimized for efficiency, but potentially offer more immediate ease of use: a phonemic interface in which the phonemes are arranged alphabetically by their label ("Alphabetic"; developed for this study) and a phonemic interface in which phonemes are arranged by articulatory features

such as manner and place (“Articulatory”; developed previously and described in in Cler, et al. (2016)).

Here we present a theoretical evaluation of seven different phonemic interfaces against five different AAC/speech corpora. Thoroughly testing this many interface and testing set combinations in AAC users is infeasible, particularly as performance typically improves over time (thus necessitating many testing sessions per interface per user; Cler et al., 2016), and because access to these individuals is limited. This paper thus focuses on thoroughly detailing the quantitative processes involved in evaluating various corpora, optimizing interfaces, and performing theoretical evaluations as a means to reduce the set of interfaces upon which to perform the necessary empirical evaluations by AAC users.

## **2.3 Methods**

### **2.3.1 Phoneme Set**

The full set of phonemes used in American English is subject to some debate. For simplicity, the set of phonemes used in this study was the reduced set of phonemes used in the Carnegie Mellon Pronouncing Dictionary (Weide, 2005); this machine-readable dictionary was used to convert text corpora to phonemes, and its set of phonemes is similar to those used in the Buckeye Corpus (Pitt, Johnson, Hume, Kiesling, & Raymond, 2005; see Methods > Corpora). See Table 2-1 for the set of phonemes used for most of the interfaces.

**Table 2-1. Reduced set of phonemes**

Arpabet label	IPA label	Example word
AA*	ɑ	father
AE	æ	at
AH	ʌ, ə	hut
AO*	ɔ	ought
AW	aʊ	cow
AY	aɪ	hide
B	b	be
CH	tʃ	cheese
D	d	dee
DH	ð	that
EH	ɛ	red
ER	ɜ	hurt
EY	eɪ	ate
F	f	fee
G	g	green
HH	h	he
IH	ɪ	it
IY	i	eat
JH	dʒ	just
K	k	key
L	l	lay
M	m	man
N	n	no
NG	ŋ	sing
OW	oʊ	oat
OY	ɔɪ	toy
P	p	pay
R	r	read
S	s	sea
SH	ʃ	she
T	t	tier
TH	θ	think
UH	ʊ	hood
UW	u	two
V	v	veer
W	w	we
Y	j	yield
Z	z	zoo
ZH	ʒ	measure

\*These two phonemes are combined into one phoneme in the Articulatory interface

Note that the articulatory interface (see section Results > Interfaces > Articulatory Interface) collapsed the phonemes /AA/ and /AO/ into one target.

### 2.3.2 Corpora

The usage statistics used to optimize an interface impact its arrangement and thus its efficiency for the end-user. However, there is no one ideal corpus of AAC messages. Therefore, we have compared five corpora (see Table 2-2 for more details): an unabridged vocabulary list of one young adult AAC user (University of Nebraska-Lincoln, n.d.-b), a list of conversational phrases suggested by AAC specialists (University of Nebraska-Lincoln, n.d.-a), a bank of simulated AAC messages (Vertanen & Kristensson, 2011), and the Buckeye corpus of conversational speech (Pitt et al., 2005). Text corpora were converted to phoneme transition likelihoods by converting text to phonemes via the CMU Pronouncing Dictionary (Weide, 2005) with hand-corrections for words not contained in the dictionary (e.g., “aneurysm”). The Buckeye Corpus has two types of phonemic transcriptions: one that matches the dictionary entry for a given orthographic transcription (‘phonemic’ by their terminology, or ‘dictionary’ here for clarity) and one with actual phonemes produced by speaker (‘phonetic’ by their terminology, but ‘direct’ here). For example, one speaker said the phrase “tomorrow’s my dinner”; the dictionary transcription of “tomorrow’s” is [T-AH-M-AA-R-OW-Z], whereas the direct transcription is [T-M-AA-R-AH-Z]<sup>6</sup>. Two separate sets of transition likelihoods were calculated using the dictionary and

---

<sup>6</sup> Two phonemic transcription conventions are used throughout this paper. One is the International Phonetic Alphabet (IPA), which is likely familiar to readers and will be indicated with sounds between slashes ( / saundz / ). When relevant, we will also show transcriptions in ARPABET, which is a machine-friendly English transliteration and was used in this study as the target labels on the interfaces. ARPABET text will be indicated with sounds between square brackets ( [S-AW-N-D-Z] ).

direct transcriptions. For both, any transcriptions that included phonemes that were not in our set (e.g., 'AHN' for a nasalized 'AH'; syllabic 'EL') were converted

**Table 2-2. Corpora**

Corpus	Description	Number of words	Conversion process
AAC user (University of Nebraska-Lincoln, n.d.-b) ("Actual AAC")	Unabridged vocabulary list with use statistics from one young adult AAC user	49,718	Converted each word to phonemes via CMUDict (thus missing word-to-word transitions)
AAC conversational phrases (University of Nebraska-Lincoln, n.d.-a) ("Suggested AAC")	Context-specific message list compiled by AAC specialists	3,941	Converted each message to phonemes via CMUDict
Simulated AAC messages (Vertanen & Kristensson, 2011) ("Simulated AAC")	Mechanical Turk simulated AAC messages	25,182	Converted each message to phonemes via CMUDict
Buckeye Corpus - dictionary transcription (Pitt et al., 2005) ("Conversation-dictionary")	40 typical speakers conversing orally with interviewer	284,832	Converted to reduced phoneme set. Calculated transitions by message from dictionary phonemic entry of orthographic transcription
Buckeye Corpus - direct transcription (Pitt et al., 2005) ("Conversation - direct")	40 typical speakers conversing orally with interviewer	284,832	Converted to reduced phoneme set. Calculated transitions by message from direct phonetic transcription

to in-set phonemes (e.g., [AH]; [AH-L]).

### 2.3.3 Calculating Interface Efficiency

Fitts' law (Equation 2-1) suggests that the movement time (MT) necessary to travel between targets  $i$  and  $j$  is related to the distance between the centers of target  $i$  and target  $j$  ( $D_{ij}$ ), and the width of the second target ( $W_j$ ). Exact movement times are determined experimentally based on the pointing device employed; these are then used to derive the constants  $a$  and  $b$ .

$$MT = a + b \left[ \log_2 \left( \frac{D_{ij}}{W_j} + 1 \right) \right] \quad \text{Eq. 2-1}$$

The efficiency of a particular target layout is quantified via Equation 2-2 (Zhai et al., 2002), which is derived from Fitts' law, and suggests that the average movement time ( $\overline{MT}$ ) of an interface is characterized by the sum of the probability of transitioning between each pair of phonemes (i and j) multiplied by the time it would take to get from phoneme i to phoneme j. Average movement time is converted to words per minute (wpm) as shown in Equation 3 for human readability and comparison with other quantitative orthographic keyboards.

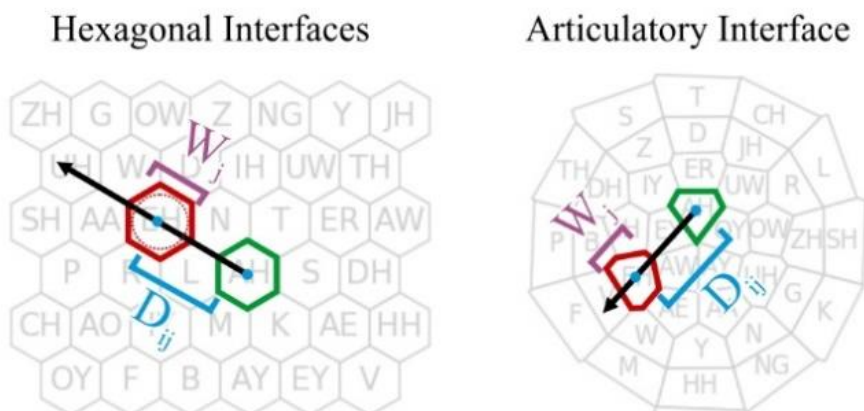
$$\overline{MT} = a + \sum_{i=1}^{39} \sum_{j=1}^{39} Pr_{ij} * b \left[ \log_2 \left( \frac{D_{ij}}{W_j} + 1 \right) \right] \quad \text{Eq. 2-2}$$

$$\text{Efficiency (words/min)} = \frac{1 \text{ word}}{5 \text{ phonemes}} \times \frac{1 \text{ phoneme}}{\overline{MT} \text{ (sec)}} \times \frac{60 \text{ sec}}{\text{minute}} \quad \text{Eq. 2-3}$$

### 2.3.3.7 Fitts' constants

The constants  $a$  and  $b$  in Equations 1 and 2 are Fitts' constants, which are experimentally derived aspects of the pointing device itself. Any change in these arising from choosing a different access method will affect the estimate of efficiency (i.e., some access methods are slower than others), but will not affect the *comparison* of two efficiencies using the optimizing algorithm. Therefore, we have used constants that apply to a stylus type pointing device, in which  $a$  is assumed to be 0,  $b = 1 / 4.9$ , and  $a = .127s$  when  $i = j$  in order to compare to other literature (MacKenzie & Zhang, 1999; Zhai et al., 2002). Thus, efficiencies

calculated will be considerably higher than those generated by AAC users with noisy access methods, across all interfaces.



**Figure 2-1. Width ( $W_j$ ) and distance ( $D_{ij}$ ) calculations for the interfaces developed and evaluated in this study. Starting phoneme  $i$  outlined in green, with target phoneme  $j$  outlined in red. Width is calculated as the distance between the two intersection points of the ideal path from the center of the starting phoneme through the center of the target phoneme (blue dots) and after the optimization (max efficiency noted: 39.608 WPM via Suggested AAC corpus).**

#### 2.3.3.8 Distance and widths

Distances between each pair of targets are calculated as the Euclidean distance between the centers of the targets  $i$  and  $j$ . Widths are calculated as the width of the second target  $j$  along the ideal path from the center of the starting phoneme through the center of the target phoneme. For all of the interfaces developed in this study, the target width is consistent, whereas for the articulatory interface (Cler et al., 2016), the target width varies (see Figure 2-1).

### 2.3.3.9 *Transition likelihoods*

Transition likelihoods were calculated from each corpus in Table 2-2 separately. Any text content was translated to phonemes via the CMU Pronouncing Dictionary (Weide, 2005) and hand-corrected. For the AAC user's vocabulary list, each word's transitions were counted and multiplied by the number of times it was used. For the remaining corpora utilizing messages, all phonemes from the message were concatenated, and each transition was counted. Then each set of counts was divided by the total number of transitions in the corpus, such that the sum of the probabilities was 1.

Note that the phoneme set in the articulatory interface was slightly different than all others (see section Results > Interfaces > Articulatory Interface). Therefore, to test this interface, separate transition likelihoods for each corpus were recalculated to collapse all  $X \rightarrow [AA]$ ,  $[AA] \rightarrow X$ ,  $X \rightarrow [AO]$ ,  $[AO] \rightarrow X$ ,  $[AA] \rightarrow [AO]$ , and  $[AO] \rightarrow [AA]$  likelihoods as appropriate.

### 2.3.3.10 *Words per minute*

Equation 3 includes a standard assumption used for orthographic keyboards, in which the average word is said to require five selections per word (four characters plus the space key). Theoretically, the average number of selections per word for a phonemic interface should be nearer 3, as no space key is necessary or provided, and as there are 14–20% fewer phonemes than letters per word depending on the corpus tested. Regardless, this is left at the orthographic standard of 5 selections/word; values in wpm are presented here

only for human readability and could be recalculated at need to represent a “truer” estimate of the phonemic wpm (see section *Discussion > Other Efficiency Calculations*).

### **2.3.4 Metropolis Optimization Algorithm**

Once an interface’s efficiency can be quantified, any number of optimization algorithms can be used. One such optimization algorithm is the Metropolis algorithm (see Beichl & Sullivan, 2000 for a review), which is a Markov chain Monte Carlo algorithm with a variety of applications; of particular interest here, it has previously been used to optimize orthographic keyboards.

In this case, the Metropolis algorithm is used as such: one interface of 39 phonemes is randomly generated, and its efficiency is calculated. The interface layout is then optimized via a random walk: two phonemes are randomly swapped and the keyboard efficiency is recalculated. If the new arrangement is more efficient than the current layout, then it is kept and the random walk continues. If the new arrangement is less efficient than the current layout, it may still be kept, according to Equation 2-4, in which the probability of keeping a less efficient arrangement is quantified by the difference in efficiency ( $\Delta E$ ) and the system temperature ( $T$ ) multiplied by a scalar ( $k$ ). The system temperature cycles over time in a process called annealing; this enables the system to break out of local  $\overline{MT}$  minima (Zhai et al., 2002).

$$\begin{aligned} \Pr(\text{keep}_{\text{new}}) &= 1, & \text{if } \overline{MT}_{\text{new}} < \overline{MT}_{\text{old}} & \\ &= e^{-\Delta E/(kT)}, & \text{if } \overline{MT}_{\text{new}} \geq \overline{MT}_{\text{old}} & \end{aligned} \quad \text{Eq. 2-4}$$

#### *2.3.4.11 Interface Shape*

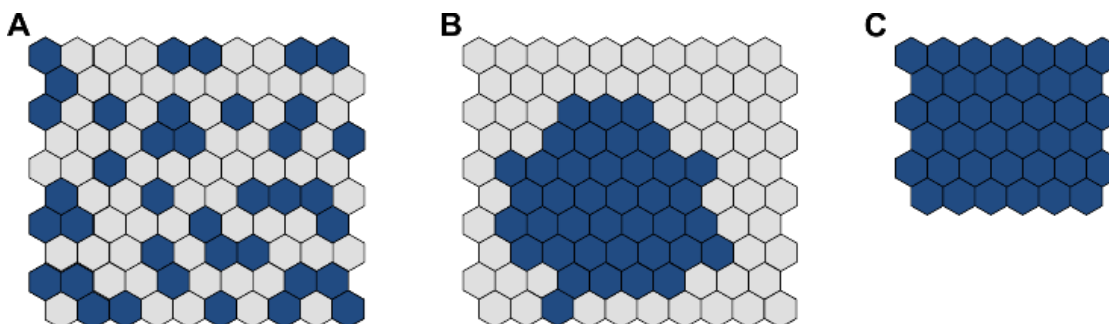
The most efficient interface is one in which the targets are tightly clustered, reducing the distances required to move the cursor. Thus target arrangements with hexagonal targets are more efficient than those with rectangular targets arranged in a grid.

The Metropolis algorithm can optimize any interface shape. Initial simulations were done with a large target space (e.g., 10 rows x 10 columns of target locations for the 39 phonemes; Figure 2-2A) to seek the most efficient layout. After running the algorithm, the most efficient arrangements had targets clustered together (Figure 2-2B). The tightly clustered targets were in a roughly circular arrangement, which maximizes efficiency (i.e., with phonemes assigned in a particular order, Figure 2-2B had an efficiency of 39.655 wpm, the maximum output of the algorithm through many iterations), but is less efficient in terms of screen space for end-users, who will need to have other programs on the screen. Therefore, the pre-set target arrangement in Figure 2-2C was chosen to maximize end-user usability and aesthetics while only slightly reducing efficiency (i.e., highest efficiency with this shape was 39.608 wpm).

#### *2.3.4.12 Determining Constants*

The width of each target was set at 10 to represent the circle circumscribed by the hexagon of the target. The distance between each button was calculated via the Euclidean distance between the centers of each target. As Equations 2-1 and 2-2 show, distance is divided by width, making the units

arbitrary as long as they are consistent.

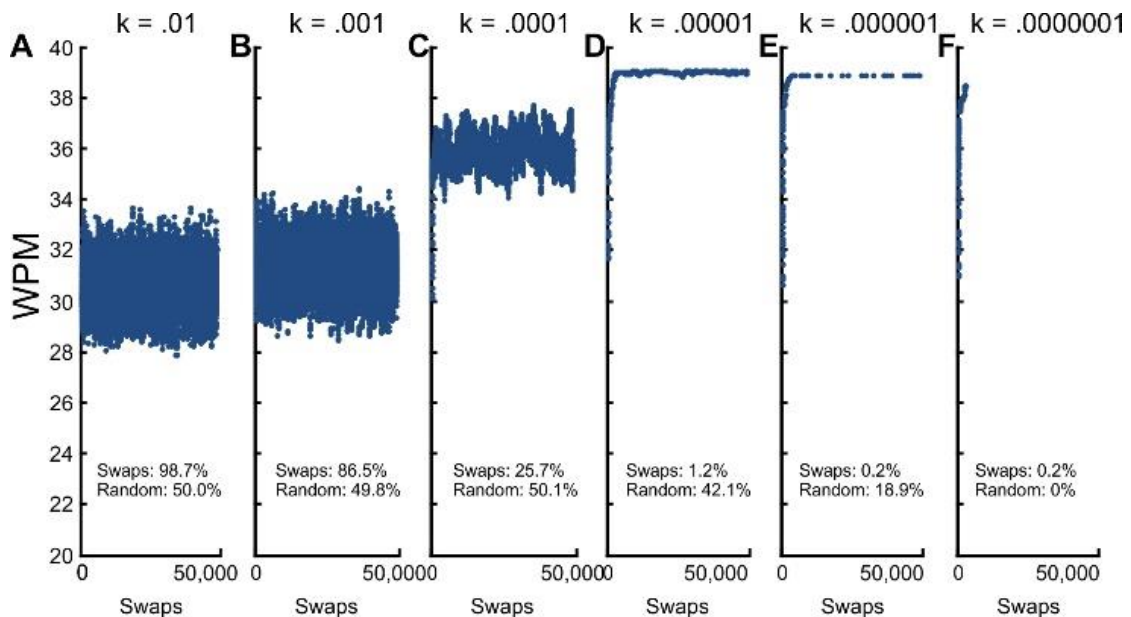


**Figure 2-2. Different configurations of 39 targets; (A) shows a 10x10 interface before the Metropolis algorithm has run, in which 39 hexes have been randomly assigned a phoneme (blue) and the rest are unassigned (grey) (B) shows an interface with high efficiency after running the Metropolis algorithm, which has tightly clustered the targets (max efficiency noted: 39.655 WPM via Suggested AAC corpus). (C) has the 39 targets arranged in a consistent layout both before and after the optimization (max efficiency noted: 39.608 WPM via Suggested AAC corpus).**

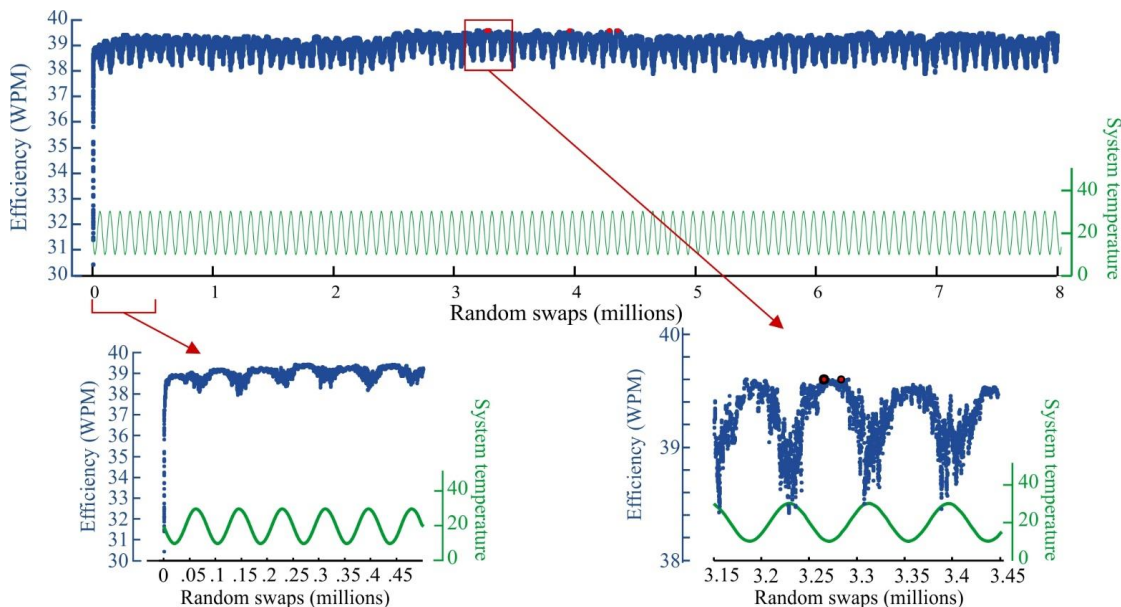
The probability of selecting an arrangement that is less efficient (Equation 4) is dependent on the difference in efficiency and two scaling variables,  $k$  (scalar) and  $T$  (system temperature, systematically varied over time). Zhai, Hunter, and Smith (2002) present this equation but do not suggest ranges for either  $k$  or  $T$ .

Figure 2-3 shows the results of example iterations of the Metropolis algorithm for different values of  $k$ , with a static  $T$  set at 10. Note that panels A and B stay near the mean random arrangement efficiency (~30 wpm); with this high of a  $k$ , many sub-optimal arrangements are kept, and thus the system never approaches an optimal value. Alternately, panels E and F have very few “kept” arrangements; these then are highly dependent on the starting arrangement and do not have the opportunity to get out of local minima. Panels C and D, however,

show behavior closer to the goal. Panel C shows a wpm that hovers above the mean. Panel D shows similar behavior to E and F, but with some minor dips in efficiency that eventually lead to higher wpm. Optimal behavior is likely in between these two numbers, as the algorithm optimizes by periodically choosing a less-optimal solution. Therefore,  $k$  was set at  $.00001$  (Panel D), and system temperature was set to vary between 10 and 35, such that the final behavior of the algorithm was between Panels C and D of Figure 2-3.



**Figure 2-3. Each panel shows the random walk through the interface space via the Metropolis algorithm with a different value for scalar  $k$  with  $T$  (arbitrarily) at 10. Each data point represents the WPM for an interface arrangement with a higher efficiency than the “current” arrangement.**



**Figure 2-4. Metropolis algorithm.** Top panel shows one iteration of the algorithm, consisting of 8 million random swaps and annealing system temperature. Panel B shows the typical process at the beginning of the algorithm, in which the efficiency quickly rises to the neighborhood of the final “optimized” version. Panel C shows how the annealing process (system temperature in green raising and lowering over time) allows the system to come out of local maxima in order to then approach the optimal solution (red dots).

For all interfaces, system temperature was varied sinusoidally between 10 and 35 at twelve cycles per million random swaps. The optimization ran for 8 million random swaps for each interface. An example optimization is shown in Figure 2-4.

#### 2.3.4.13 Verification of Optimization Outcome

Figure 2-4 shows the results of one iteration of the optimization algorithm, which consisted of 8 million random swaps. The efficiency first increases rapidly from an average random efficiency (~30 wpm for the suggested AAC messages corpus) to near the final optimum efficiency (~39 wpm; Figure 2-4B). Next the

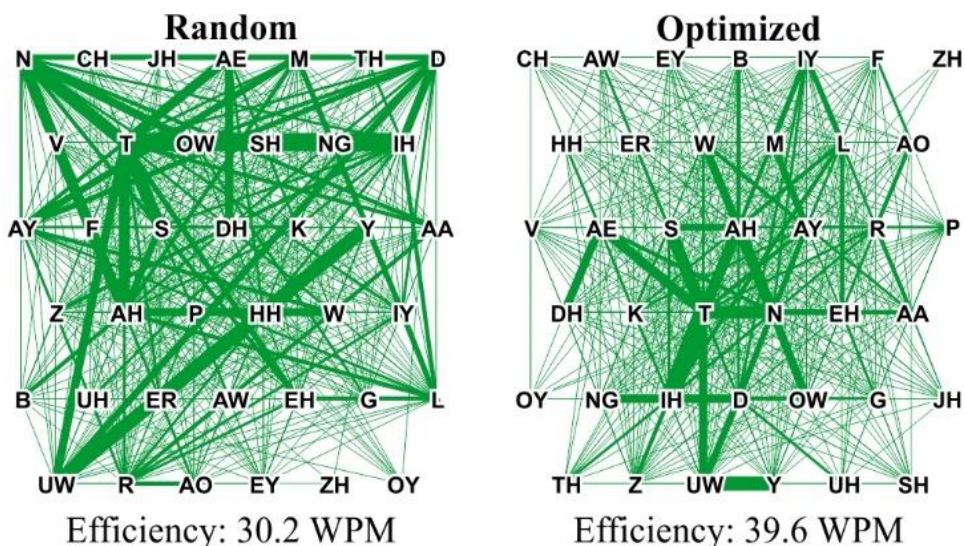
annealing process (in green; Figure 2-4C especially) systematically increases and decreases the system temperature, thus increasing and decreasing the likelihood that the algorithm will accept an arrangement with a worse efficiency, as in Equation 4. This allows the system to come out of local maxima in order to then approach the “optimal” solution.

In addition to the efficiency calculations performed at each step by the Metropolis algorithm, representations of the results of the optimization process were evaluated visually to verify the function of the algorithm. Figure 2-5 shows two different arrangements of 39 phonemes. The left is a random organization, whereas the right shows an optimized arrangement based on the Suggested AAC corpus. The width of the lines represent the transition likelihood between each pair of phonemes. Note that while thick lines occur throughout the random interface (left), the length of thick lines are minimized in the optimized interface; this suggests that when users try to produce the words and phrases common in the corpus (e.g., “you” or [Y-UW], and “don’t” or [D-OW-N-T]), they will not need to move as far to select the required targets. The efficiency of the random interface shown here (30.2 wpm) is also the mean efficiency of 100,000 random phoneme layouts using the Suggested AAC corpus to evaluate efficiency.

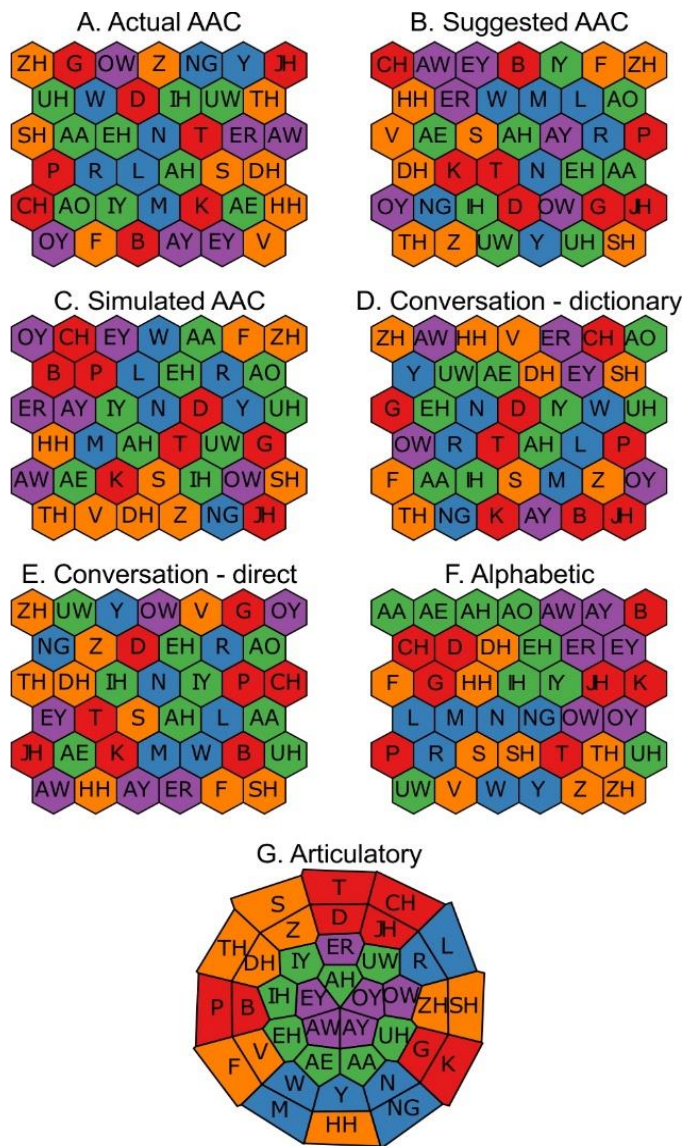
### **2.3.5 Evaluation**

Seven interfaces (Figure 2-6) were evaluated against five corpora (see Table 2-2). Interfaces A-E were optimized as stated in section 2.3.4 *Metropolis Optimization Algorithm* using diphone probability statistics from each of the five

respective corpora. Interface F (Alphabetic) was generated to be the same shape as Interfaces A-E, but with the phonemes arranged alphabetically according to their labels. The final interface (Figure 2-6G; Articulatory) was included to compare to previous studies (Cler et al., 2016; Cler, Nieto-Castanon, et al., 2014). Interfaces F and G are included primarily as non-optimized controls in order to evaluate the efficacy of the optimization. Further, while we hypothesized that the Alphabetic and Articulatory interfaces would not provide optimal efficiencies, these interfaces are organized by rules that people can learn, and thus may help users quickly learn where different targets are on the interface.



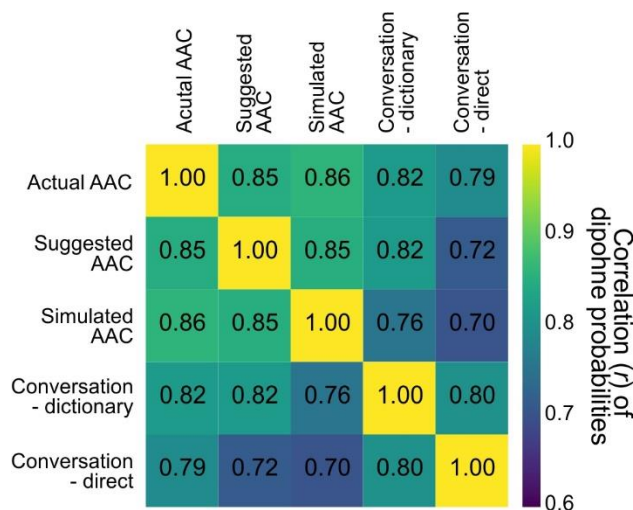
**Figure 2-5. Visual comparison of results of Metropolis algorithm. Left shows a random organization of phonemes; right shows an optimized interface. Width of lines between two targets represents the likelihood of transitioning between them in the AAC conversational phrases corpus.**



**Figure 2-6. Interfaces developed and evaluated in this study. Target colors show rough groupings of phonemes: simple vowels in green, complex vowels (diphthongs, r-colored vowels) in purple, fricatives and affricates in orange, stops in red, and liquids, nasals, and semivowels in blue.**

Although the interfaces were not evaluated empirically in this study, future studies may wish to do so. Additionally, if the differences in efficiency are small, users may choose to use the Articulatory or Alphabetic interfaces; whereas if the

differences in efficiency are large, they may choose to use an optimized interface instead.



**Figure 2-7. Similarity between phoneme-to-phoneme transition probabilities from different corpora (color represents r-value of Pearson's correlation between all diphone probabilities)**

## 2.4 Results

### 2.4.1 Corpora similarity

Figure 2-7 shows the correlation of diphone likelihoods between the different corpora tested here. Correlations were high, ranging from .70 to .86. Correlations between text resources (Actual AAC, suggested AAC, and simulated AAC) and conversational resources (Conversation-dictionary and Conversation-direct) were on the low-to-mid end of the range of correlations, ranging from .70 to .82, whereas those within text resources were the highest, from .85 to .86. Interestingly, the correlation between the Conversation-dictionary

and Conversation-direct probabilities was only .80, despite both being derived from the same conversational source.

## **2.4.2 Interfaces**

Figure 2-6 shows the different interfaces developed for this project (A-F) as well as the articulatory interface previously developed (G).

### *2.4.2.14 Optimized Interfaces*

The first five interfaces (Figure 2-6A-E) were generated with the Metropolis algorithm as in *Section 2.3.4 Metropolis Optimization Algorithm* using the diphone probabilities from each respective corpus.

### *2.4.2.15 Alphabetic Interface*

The alphabetic interface (Figure 2-6F) used the same layout and phonemes as the optimized interfaces, but arranged the phonemes in alphabetical order based on their label (see Figure 2-6).

### *2.4.2.16 Articulatory Interface*

The articulatory interface (Figure 2-6G) has previously been described in (Cler et al., 2016). Briefly, the targets on the interface were arranged manually in a circular layout, such that phonemes were organized based roughly on articulatory features (manner and place of articulation). Phonemes that are differentiated only by voicing (e.g., [TH] and [DH]) are located at the same angle but different radii. Only 38 phonemes were used for this interface instead of the set of 39 used in the other interfaces in this study; as noted in Table 2-1, the phonemes /AA/ (*father*) and /AO/ (*ought*) were collapsed into one phoneme.

Interface targets were allowed to be directly adjacent (Figure 2-6G) rather than leaving gaps in between targets as in (Cler et al., 2016); this was so that widths were as large as possible in relation to distance between targets, as those in the other hexagonal interfaces are tightly packed.

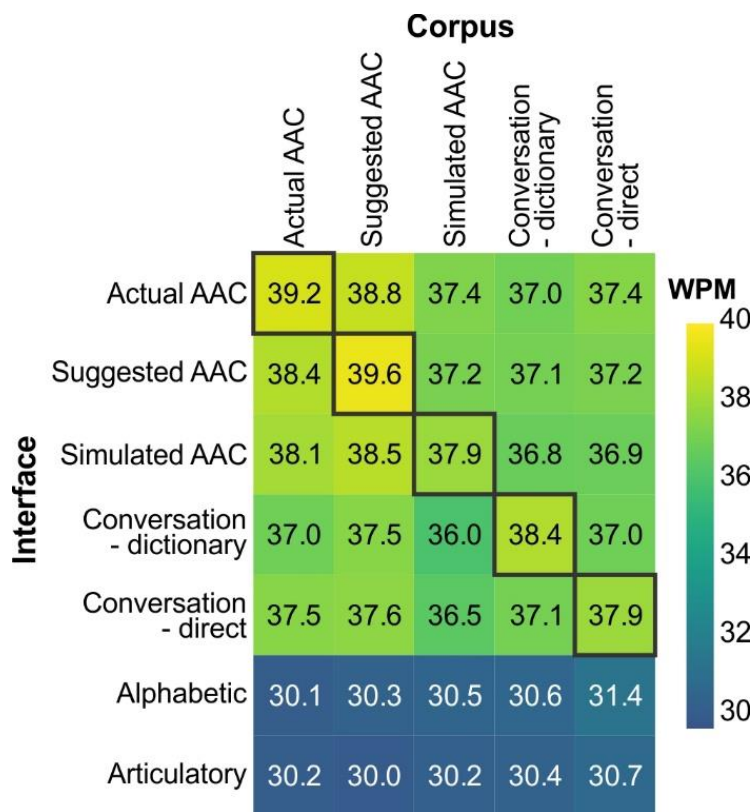
### **2.4.3 Interface efficiency**

Efficiency was calculated for each of the seven interfaces against each of the five testing corpora. Results are shown in Figure 2-8 in terms of wpm. The optimized interfaces had relatively high efficiencies across all testing sets, from 36.5 to 39.6 wpm. Generally efficiencies were highest when the testing corpus was the same as the corpus used to generate the diphone probabilities (shown in Figure 2-8 on the diagonal, highlighted in black). Variability in wpm for each interface across the different corpora was low, with ideal efficiency to lowest efficiency differing by only 1–3wpm. Efficiencies were lower with both interfaces not optimized by the Metropolis algorithm (30.0–31.4); these were even more consistent across testing corpora, with the Alphabetic only varying by 1.3 wpm across testing corpora, and the Articulatory only varying by 0.7 wpm depending on which corpus was used to evaluate its efficiency.

## **2.5 Discussion**

Interfaces not optimized for efficiency (Alphabetic; Articulatory) showed efficiencies around 30 wpm, which is near the mean random interface efficiency (30.2 wpm). Optimizing an interface using the Metropolis algorithm yielded an improvement of around 9 wpm over a random arrangement and the Alphabetic

and Articulatory interfaces. However, optimizing by one corpus and testing against another yielded differences around 1–3 wpm.



**Figure 2-8. Efficiency in words per minute (WPM) for each interface tested with each corpus. Highlights on the diagonal show when the interface is being tested against the same corpus used to optimize its layout.**

### **2.5.1 Benefits of Alphabetic and Articulatory Interfaces**

The Alphabetic and Articulatory interfaces may show advantages that are not captured in the efficiency calculation. The organization of the Alphabetic interface gives users some a priori information about where targets are located on the interface, as they are arranged via phonemic label. Figure 2-6F shows that targets arranged by label are also thus somewhat arranged by type of

phoneme (note that all nasals are together; vowels are largely grouped). This would likely improve early communication rates via reducing visual search time when users are first learning to use the interface. The Articulatory interface similarly shows organization (Figure 2-6G), such that all vowels are in the center of the interface, with consonants surrounding them. In addition, the Articulatory interface pairs phonemes that are similar by manner and place of articulation (e.g., /f/ and /v/ are neighbors), which may have initial and ongoing benefits. First, this organization may allow faster learning of the target locations, similar to the Alphabetic interface. In addition, however, this leads to some error tolerance that is not seen in orthographic interfaces or the other phonemic interfaces. If a user overshoots the target and accidentally selects the neighboring pair, the output of may still retain intelligibility that it would not with other interfaces. For example, if a user intends to select [V-OI-S] (“voice”) and instead selects [F-OI-S] or [V-OI-Z], a listener would still likely understand that production in context.

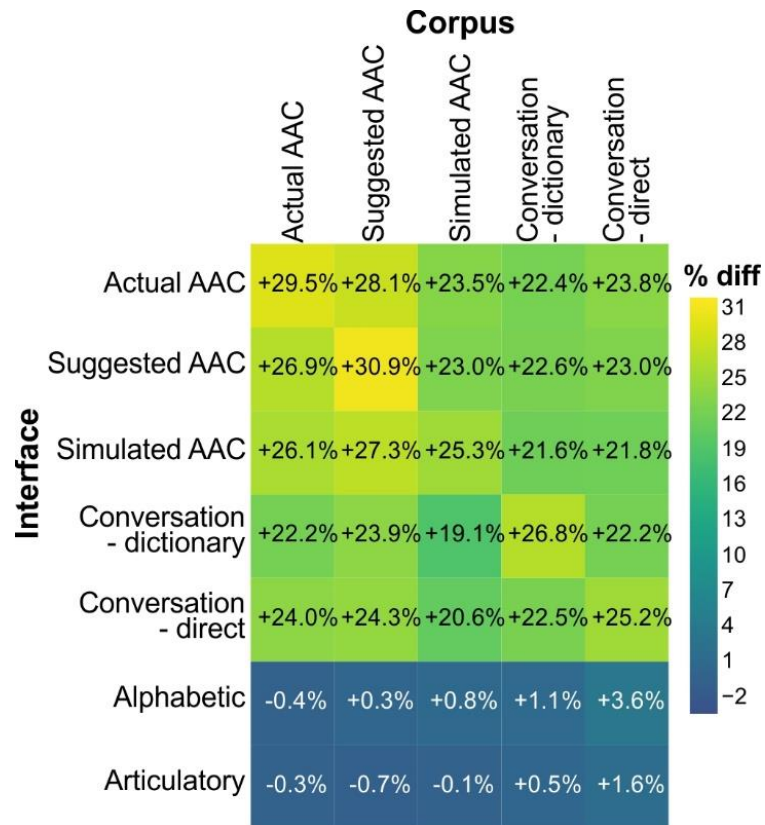
### **2.5.2 Other Efficiency Calculations**

Calculating efficiency in words per minute (wpm) here relied on two main assumptions that may not hold for phonemic interfaces used by individuals employing alternate access methods. First, the wpm calculation assumes five letter selections per word, which includes four characters and a space; phonemic interfaces do not require a space and have fewer characters per word; taking these both into account improves the efficiency calculation for phonemic interfaces.

In addition, alternate access methods are noisier than typical access methods, such as a stylus or physical keyboard. The calculations here used Fitts' constants from a typical stylus to compare to previous work (see section 2.3.3.7 *Fitts' constants*). If we used Fitts' constants derived from an individual with a spinal cord injury using electromyography (EMG) to control a cursor (Williams & Kirsch, 2016), the absolute difference between the interfaces changes. When recalculating Figure 6 using Fitts' constant  $b=1/2.6$ , (derived from line of best fit from Figure 5b in Williams & Kirsch, 2016), the number of estimated words per minute reduces from 30-39 to 8-9 wpm. However, these scale linearly (when using the common assumption that Fitts'  $a=0$ ). Figure 2-9 thus shows the improvement in efficiency in percent difference rather than in wpm, as wpm varies by input method, number of characters per word, and by individual skill.

A final way to consider the improvement of the optimization is not in words per minute, but rather as the time it would take to accomplish a task. Table 2-3 shows time estimates for producing a list of the 1004 suggested AAC messages (the Suggested AAC corpus) with different phonemic and orthographic on-screen interfaces, and using either a typical stylus movement time estimate (in which  $b=1/4.9$ ) or the EMG cursor movement time estimate (in which  $b=1/2.6$ ). For the phonemic interfaces, no spaces or punctuation were included. For the orthographic interfaces, no punctuation was included. Fitts' constant remained  $a=0$  for all calculations, as did the special case of movement time when tapping

the same target twice (.127s). In these calculations, then, no assumptions were made as to the number of phonemes or letters per word.



**Figure 2-9. Percent difference in efficiency compared to a random arrangement of phonemes.**

**Table 2-3. Time estimates to produce Suggested AAC corpus**

Interface	Access Method	
	Stylus ( $b = 1/4.9$ )	EMG cursor by person with spinal cord injury ( $b = 1/2.6$ )
Suggested AAC (phonemic)	54 hrs	102 hrs
Alphabetic (phonemic)	70 hrs	133 hrs
METROPOLIS (orthographic) (Zhai et al., 2002)	78 hrs	147 hrs
QWERTY (orthographic; square targets)	111 hrs	210 hrs

The times in Table 2-3 represent long-term interface usage (i.e., combined time to produce 1004 utterances) using different phonemic (Suggested AAC and Alphabetic, from this study) and orthographic interfaces (orthographic Metropolis interface from (Zhai et al., 2002); QWERTY keyboard in common use). The optimized phonemic interface shows substantial improvements over orthographic input methods. These improvements increase when the input method is noisy, such as the EMG cursor (Williams & Kirsch, 2016). Thus, while users with access to typical stylus use may not choose to switch to a phonemic interface, those for whom access is more time-consuming and difficult may find the initial costs of learning a phonemic interface worthwhile (e.g., decreasing time to produce messages from 210 hrs to 102 hrs – roughly 50%). Note, however, that these do not include any prediction, which improves both phonemic and orthographic input rates substantially (Trnka & McCoy, 2007; Vertanen et al., 2012).

### ***2.5.3 Clinically Meaningful Speed Improvements***

It is not yet clear what degree of improvement in efficiency is clinically meaningful, particularly as AAC users have many different access methods and preferences. Therefore it is also unclear whether the 5-8% improvements due to optimizing per corpus are worthwhile. While producing a new optimized interface once a corpus is obtained is not particularly difficult, it can be somewhat time consuming (converting a given corpus to phonemes often involves some level of hand-correcting for out-of-dictionary terms; running the actual optimization process as described in *2.3.4 Metropolis Optimization Algorithm* takes

approximately twelve hours of computing on a shared computing cluster).

Further, it can be difficult to obtain an appropriate corpus. If a user already has an AAC device, they may be willing to allow their AAC specialist to record their usage for some length of time. Although this would be an ideal situation for optimizing an individual AAC interface, only some AAC devices have this recording capability, and recording a person's communication output has privacy concerns.

#### **2.5.4 Future Directions**

##### *2.5.4.17 Empirical evaluations*

An empirical evaluation of these phonemic interfaces is a necessary next step. Although it was not possible to thoroughly evaluate all of the interfaces presented here in AAC users, an empirical evaluation of a small number of interfaces using only one testing set can now be completed to validate the theoretical results. This evaluation should be carried out by users with a variety of motor impairments and access methods. Empirical evaluations should be done to compare the interfaces developed in this study to other existing phonemic interfaces (Black et al., 2008; Schroeder, 2005; Trinh et al., 2012; Vertanen et al., 2012).

##### *2.5.4.18 Individualized optimizations*

This paper includes the technical details of the varied quantitative processes that were involved in generating and evaluating the interfaces. These are included specifically so that others can recreate them and produce interfaces

optimized based on any given corpus or with weighting factors other than just Fitts'-based efficiency. For example, if a particular user finds left-and-right movements less fatiguing than up-and-down movements, an additional weighting factor could be added for reduced efficiency when the next target is above or below the current target rather than on the same row. Alternate efficiency formulae and weighting could also allow these same algorithms to optimize interfaces intended for use with scanning rather than direct selection methods. In that case, efficiency would be calculated based on time from the onset of scanning to the target's selection, rather than the Euclidean distance between two targets.

#### *2.5.4.19 Prediction*

Finally, another vital way to optimize communication rates is to incorporate online prediction. Studies of existing phonemic interfaces have focused on prediction as the primary way of increasing communication rates (Schroeder, 2005; Trinh et al., 2012; Vertanen et al., 2012), without the offline optimizations shown here, and predictive language models can increase phonemic interface communication rates by as much as 100% (Trinh et al., 2012). Adding prediction to the interfaces presented here is needed in order to compare to other phonemic or orthographic interfaces and to maximize communication rates.

## **2.6 Conclusions**

Phoneme-to-phoneme transition likelihoods were highly correlated across corpora, particularly corpora generated from text or AAC use instead of from oral

conversation. Optimization based on any corpus increased efficiency from random layouts or from the Alphabetic and Articulatory interfaces by 19-31%. Optimizing and testing on the same corpus led to efficiency improvements of 5-8%, compared to testing on other corpora. Therefore, if possible, future phonemic interfaces should be optimized via a corpus from the intended user's speech. If this is not possible, however, optimization still improved efficiency using all testing corpora, suggesting that choosing an optimized corpus is indicated over other layouts. Future directions include empirical testing and adding prediction to further increase communication rates.

## **2.7 Acknowledgments**

This research was supported by NIH grant F31 DC014872 and NSF grant 1452169. Our thanks to Alfonso Nieto Castañón and Frank Guenther for access to their phonemic interface, which was developed under NIH grant R01DC002852.

## Chapter 3. Empirical Evaluation of Optimized and Predictive Interfaces

### 3.1 Abstract

**Purpose:** In this study, we empirically assessed the results of computational optimization and prediction in communication interfaces. These interfaces were designed to allow individuals with severe motor speech disorders to select phonemes to generate speech output.

**Method:** Interface layouts were either random or optimized, in which phoneme targets that were likely to be selected together were located in close proximity. Target sizes were either static or predictive, in which likely targets were dynamically enlarged following each selection. Communication interfaces were evaluated by 36 users without motor impairment using an alternate access method. Each user was assigned to one of four interfaces varying in layout and whether prediction was implemented (random/static; random/predictive; optimized/static; optimized/predictive) and participated in 12 sessions over a 3-week period.

**Results:** In individuals without motor impairment, prediction provided significantly faster communication rates during training (sessions 1–9), as users were learning the interface target locations and the novel access method. After training, optimization acted to significantly increase communication rates. The optimization likely became relevant only after training when participants knew the target locations and moved directly to the targets.

**Conclusions:** Optimization and prediction led to increases in communication rates in users without motor impairments. Future research is needed to translate these results into clinical practice.

### 3.2 Introduction

When motor speech disorders render speakers unable to communicate orally, individuals may use augmentative and alternative communication (AAC) strategies to communicate. Individuals with concomitant motor impairments (e.g., amyotrophic lateral sclerosis, spinal cord injury) may use an alternate access method (e.g., head-tracker, eye-tracker, switch-activated scanning) to choose letters or words on an onscreen interface to produce a synthesized speech output. Despite advances in access technologies, communication rates in this population remain slow: 2–15 words per min (wpm), compared to 30–40 wpm by a skilled typist and 150–200 wpm in typical speech (Beukelman & Mirenda, 2013; Copestake, 1997; Higginbotham et al., 2007; Koester & Arthanat, 2017; Leshner et al., 1998b). These rates are slow partially due to the motor impairments these individuals exhibit requiring the use of alternative access. Another barrier to achieving faster communication rates is in the design of the communication interface. Opportunities exist to research and develop new interface options that demonstrate potential to increase rate and efficiency of message construction for individuals with severe motor impairment. This paper describes the preliminary investigation of a new AAC interface that integrates phonemic targets, optimization, and prediction.

### **3.2.1 Phonemic Interfaces**

Most AAC interfaces provide targets consisting of letters, whole words, or symbols (typically representing whole words or phrases). The choice of targets is an important one, as each option offers a compromise between speed, flexibility, and cognitive load (Beukelman et al., 2007). Interfaces with symbols representing whole phrases, for example, provide very high speed to produce the given phrase, minimal flexibility (i.e., only certain phrases can be selected quickly or at all), and high cognitive load (Thistle & Wilkinson, 2013). Some interfaces use phonemes (which represent a particular sound in a spoken language) as targets (Black et al., 2008; Cler et al., 2016; Cler, Nieto-Castanon, et al., 2014; Schroeder, 2005; Trinh et al., 2012), which may provide a good balance of speed, flexibility, and cognitive load.

Phonemic targets enable full flexibility to produce any sequence of sounds and allow users to bypass complex text-to-speech methods employed by orthographic (alphabetic) interfaces. Of particular interest to individuals who use slow or effortful access methods, common AAC messages have 14–20% fewer phonemes than letters (Cler et al., 2016). The primary disadvantage of phonemic targets is that users must learn to translate their intended messages into phonemic components and then must find those targets on an interface. Typically, children spend many years learning to translate thoughts into orthographic targets (i.e., writing in English; typing on a QWERTY keyboard). While selecting a sequence of phonemes to create a message may be more

similar to the production of oral communication, individuals wishing to use phonemic interfaces are likely to require training. However, the speed and flexibility advantages may represent a significant improvement over other options for individuals with severe motor impairment. Improvements to the efficiency of phonemic interfaces may make this option even more appealing.

### ***3.2.2 Quantitatively Optimized Interfaces***

The standard orthographic keyboard layout, QWERTY (the Sholes keyboard, designed in 1873), is highly inefficient for ten-finger typing; in fact, it was specifically designed to be inefficient so as to minimize jamming typewriter keys (Noyes, 1983; Rumelhart & Norman, 1982). Alternate ten-finger typing layouts like Dvorak show 4% improvements in typing speed (West, 1998). However, ten-finger typing is a parallel process, in which 90% of finger movements are initiated before the previous key is pressed (Gentner et al., 1980; Rumelhart & Norman, 1982), and is thus difficult to model and optimize (Rumelhart & Norman, 1982). Further, individuals with sufficient ten-finger motor control to type largely will not see large enough differences in typing rates to justify the cognitive and practical downsides to alternate keyboards. Professional typists, such as stenographers, do use alternate keyboard layouts; interestingly, many shorthand systems (including stenography) use phonemic input (Beddoes & Zhongzhi Hu, 1994).

QWERTY keyboards are particularly inefficient for serial input, such as when individuals are entering text on a touchscreen with a stylus or a finger. This

process (serial input) is more easily modeled and optimized. Fitts' law, a fundamental model of human movement, can characterize the amount of time it takes to select a target using any pointing device (e.g., finger, typical mouse, stylus, head-tracker). Fitts' law states that the time required to select a target is a function of its size and the distance to be traveled to reach it (Fitts, 1954); targets within close proximity are faster to select (smaller distance to be travelled), as are large targets (less precision needed, thus faster movements are possible). The efficiency of a particular arrangement of targets can be calculated by multiplying the movement time required to travel between each pair of targets by the likelihood that those two targets will be selected in series (MacKenzie & Zhang, 1999; Zhai et al., 2002). In a previous study, we used computational simulations to optimize the layout of phonemic interfaces (Cler & Stepp, 2017). Simulations revealed an improvement of 30.9% in expected communications rates generated with an optimized phonemic interface compared to a random randomly arranged phonemic interface (Cler & Stepp, 2017). However, these expected improvements in communication rate have not yet been empirically validated.

### **3.2.3 Prediction**

Prediction is ubiquitous in cellular phone keyboards and in most high-tech AAC interfaces. Previous studies have shown that adding prediction to orthographic interfaces improves communication rates by 58.6% (Trnka et al., 2009) and can improve communication rates in phonemic interfaces by 100%

(Trinh et al., 2012; Vertanen et al., 2012). Two separate aspects must be considered when applying predictive methods to an AAC interface: how to determine likely targets, and how to indicate these likely targets to the user.

Prediction typically involves word or language use statistics (based on corpora of text plus the user's past selections) to predict the next character, the rest of the word, or the next word. This is often seen in cellular phones, which typically offer each of these options and can be implemented in a variety of ways in different systems (Garay-Vitoria & Abascal, 2006). Many of these methods increase selection speed at the cost of flexibility. For example, some prediction methods disambiguate words from an ambiguous entry, such as the T9 system (Kushler, 1998) or Swype (Smith & Chaparro, 2015) which disambiguate text from a reduced keyboard or from a continuous finger drag, respectively. These methods constrain possible selections to only those contained in the dictionary, reducing flexibility.

*Phoneme prediction.* Phonemic interfaces do not require spaces between words for intelligible production, as oral speech does not typically have pauses between words. This represents additional selection savings for individuals with motor impairments, but also removes word-level structure for word-completion type prediction or any language-based prediction. Thus, phonemic prediction is a form of character prediction in which the previous phonemes are used to predict the next phoneme. Character prediction can be generated from any corpus of messages, which typically consist of text gathered from written sources.

Character prediction is typically achieved through n-grams (blocks of characters). A table of frequencies of all 5-character strings (5-grams) enable the system to evaluate the likelihood of all characters after 4 (n-1) selections have been made. N-grams are also used in some AAC applications for scanning systems, in which dynamic scanning matrixes show the most probable characters (Leshner et al., 1998b).

*Alerting users to predicted targets.* Systems that do not automatically select highly-likely or disambiguated characters/words must display predicted options for the user to view and select. If the predicted words are too intrusive or inaccurate, they may be distracting. If they are located in a separate part of the screen than the keyboard, the user must remember to redirect their attention to a different location. If the user has made a misspelling early in the word or if the prediction is inaccurate, they may waste time checking the predicted list for a word that will not appear.

In this study, we have developed a novel system for alerting users to likely phonemes. After each selection, all targets are dynamically resized to enlarge likely targets. We hypothesize that this will improve communication rates by: (1) visually highlighting predicted targets to draw the user's attention (Magnien, Bouraoui, & Vigouroux, 2004; Sears, Jacko, Chu, & Moro, 2001) and (2) providing larger targets, which decreases the movement time required to select the second target based on Fitts' law (Fitts, 1954; McGuffin & Balakrishnan, 2005).

Expanding targets have been shown to increase selection rates in standard Fitts' law experiments in which users without motor impairment select one of a few targets on a screen (Zhai, Conversy, Beaudouin-Lafon, & Guiard, 2003) and in human-computer interface studies in which users without motor impairment select one target among a row of tightly-packed targets (e.g., the Mac OSX dock, in which icons are dynamically enlarged on hover; McGuffin & Balakrishnan, 2005). These have not been implemented in many AAC interfaces. An alternative text entry system called Dasher does incorporate dynamic weighting of targets based on target likelihoods, and thus can increase target selection speed (Ward & MacKay, 2002). This system uses orthographic entry rather than phonemes, and each target does not have a static location. Rather, the targets are linearly displayed on the right side of the screen and move up and down on the screen based on the relative likelihoods of the different targets. As a result, the system can be distracting or disorienting, and users have reported that it requires a large amount of concentration to use (Tuisku et al., 2008). Further, this method does not take advantage of enlarged targets as a visual search aid during training; because the position of the targets changes, users must visually search for every target, regardless of the level of training.

An alternate option for expanding targets in a grid that has not yet been applied to AAC interfaces is that of an algorithm common in computational geometry: Voronoi diagrams. A Voronoi diagram is built from a set of seeds scattered in a plane and segmented such that every point in the plane is

assigned to its nearest seed based on Euclidean distance (Okabe et al., 2009). A weighted Voronoi diagram is modified such that each seed has a weight, and points are assigned to a seed based on a function of both the weight and the distance (Anton et al., 1998). If seeds are defined in a grid and weights are assigned based on prediction, a new Voronoi diagram can be generated after each selection, and the likeliest targets will be dynamically enlarged. Because seed locations are static, the general layout does not change, thus mitigating the possible disorientation and increased visual search time associated with other methods.

#### ***3.2.4 Empirical Evaluation by Users with and without Motor Impairment***

Communication rates in individuals with motor impairment may be improved by reducing the motor actions required to complete a message. Offering phonemes as targets can theoretically reduce selection rates by 14–20% (Cler et al., 2016). In addition, organizing targets such that those that are often selected sequentially are placed in close proximity has been shown to reduce selection time (e.g., MacKenzie & Zhang, 1999; Zhai et al., 2002). Our computer simulations combining these strategies reveal an ideal communication rate improvement of 30.9% when using an optimized phonemic interface compared to a randomly arranged phonemic interface, and 51% compared to a QWERTY orthographic interface (Cler & Stepp, 2017). Additionally, adding prediction to a phonemic interface may improve communication rates by up to

100% (Trinh et al., 2012). However, these potential rate improvements are thus far only theoretical.

Assessing the differential effects of optimization and prediction empirically requires a between-group design, and therefore, each group must consist of relatively homogenous participants. Further, as ideal usage of the interfaces will only emerge with usage over time, participants must be available to use the interfaces over many days. Individuals with motor impairment are highly heterogeneous as a group and are difficult to recruit over many sessions. While participants without motor impairments fit these requirements, their typical access methods (e.g., finger on a touchscreen or a typical mouse) are over-trained and not representative of the noisy access methods generally available to participants with motor impairment. Thus, we recruited individuals with typical motor control but required them to interact with the interfaces using a noisy access method available to individuals with motor impairments: a computer cursor controlled via facial musculature (Cler et al., 2016; Cler & Stepp, 2015; Vojtech et al., 2018)<sup>7</sup>.

---

<sup>7</sup>It is frequently necessary to use non-AAC users as participants due to the difficulty in recruiting and evaluating people with different abilities and needs. For example, researchers have generated AAC-like conversational corpora by having users without impairments imagine that they have a disorder limiting their speech and type what messages they may wish to produce (Vertanen & Kristensson, 2011). One study evaluating prediction in AAC used people without motor impairments and modeled actual AAC users by implementing a 1.5 s pause after each selection on a touchscreen (Trnka, Yarrington, McCaw, McCoy, & Pennington, 2007). While this does accurately model the speed of AAC use and (as suggested) prompt users to incorporate prediction more than a user with motor impairment might, cognitive processing can continue during this pause (e.g., planning and locating the next selection on the interface) in a way that may not exactly model someone with a motor impairment; these individuals are concentrating on the motor action during the 1.5 s it takes to complete a selection. As such, we chose to have participants use an alternate access method that models both the speed and perhaps the

Here we present two empirical evaluations of these optimization and prediction strategies. First, four groups of individuals (36 total) without motor impairments interacted with one of four phonemic interfaces in a 2×2 between-group design permuting optimization and prediction. The layout of the targets was either random or optimized, such that phoneme targets that were likely to be selected together were located in close proximity. The interfaces were either static or predictive, meaning that highly likely targets were enlarged. Each user was assigned to one of four interfaces (optimized/static; optimized/predictive; random/static; random/predictive) and participated in 12 sessions over a 3-week period. Participants used an alternate input modality to act as a model of a motor-impaired AAC user. In a follow-up study, six individuals with motor impairment used the optimized/static and optimized/predictive interfaces in alternating blocks and answered survey questions about their experience and preferences after each block.

### **3.3 Methods**

#### **3.3.1 *Interface Development***

Interfaces and experimental architecture were developed in Python. Speech synthesis was accomplished via the MBROLA system (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996). Phoneme labels were from ARPABET, a machine-readable transliteration of English phonemes (Shoup,

---

difficulty of alternate access in this population.

1980). Colors were consistent across experimental groups, were isoluminant, and denoted rough phoneme category: simple vowels in green, complex vowels (diphthongs, r-colored vowels) in purple, fricatives and affricates in yellow, stops in red, and liquids, nasals, and semivowels in blue.

### **3.3.2 Optimization**

Full descriptions of the development and optimization of the interfaces are presented in Cler and Stepp (2017). Briefly, however: an interface's efficiency can be estimated via Fitts' law. The efficiency of any arrangement of targets can be calculated with the Fitts' law estimation of movement time between each pair of targets multiplied by the likelihood that the pair of targets will be selected in sequence (MacKenzie & Zhang, 1999; Zhai et al., 2002). Any optimization process could be used to maximize the efficiency of an interface by randomly producing target layouts and finding the most efficient arrangement.

An optimally efficient arrangement will have targets arranged such that the distance between targets that are often selected sequentially is minimized. This method has been implemented for orthographic keyboards (e.g., Zhai et al., 2002), but has not previously been applied to phonemic interfaces. A variety of optimized interfaces are developed and discussed in Cler and Stepp (2017). Results of computational simulations suggested that optimization should produce communication rate improvements around 20–30%, based on which corpora are used to optimize and then evaluate the interfaces. The interfaces used in the present studies were the random and optimized interfaces based on the

“Suggested AAC corpus” from Cler and Stepp (2017). This corpus is a set of 1004 messages suggested by AAC specialists for people with amyotrophic lateral sclerosis (Beukelman & Gutmann, 1999), which was converted into phonemes automatically using the Carnegie Mellon University Pronouncing Dictionary (CMUDict; Weide, 2005). This corpus also comprised the stimuli set in this experiment, as it consists of functional messages that are relevant to individuals with motor impairment.

### **3.3.3 Prediction**

Two separate aspects must be considered when applying predictive methods to an AAC interface: how to determine likely targets and how to indicate these likely targets to the user. Determining likely targets typically involves large corpora of text. While character prediction of text is relatively straightforward, standard textual corpora are not directly usable for phonemic AAC prediction. First, AAC messages are different in content from oral communication and written text (e.g., books, articles, email), due to their purpose and constraints (Trnka & McCoy, 2007). In addition, large corpora of AAC messages are not available, leading to many studies in this area combining text and spoken corpora, or using AAC messages generated by non-AAC users (Cler & Stepp, 2017; Trnka & McCoy, 2007; Vertanen & Kristensson, 2011). Our objective here was to evaluate a novel method of displaying likely targets to the user, so we did not attempt to overcome these issues. Instead, we used standard methods of prediction on a small corpus consisting of our stimuli set: 1004 AAC messages

suggested by AAC experts (Beukelman & Gutmann, 1999) translated to phonemes using the CMU Pronunciation Dictionary (Weide, 2005). N-grams (n=1 to 3) were generated automatically using the Natural Language Toolkit in Python (nltk; Bird, Klein, & Loper, 2009). These methods are easily replicable with other corpora as they become available or relevant, including large corpora of AAC messages and conversation or a corpus of an individual user's messages.

Likely targets were indicated to the user via weighted Voronoi diagrams. Seeds for each target were located at each target's center in a static grid, allowing users to retain knowledge of the phoneme arrangement and thus reducing the time required to visually search for the targets. The target weights (and thus size) were dynamically modified after each selection based on the likelihood that each phoneme will be selected next. Prediction weights were rescaled after each selection relative to currently predicted likelihoods rather than absolutely scaled across all prediction (i.e., at every time point, the most likely target had a prediction level of 1 and the least likely target had a prediction level of 0, with the other targets scaled in between;).

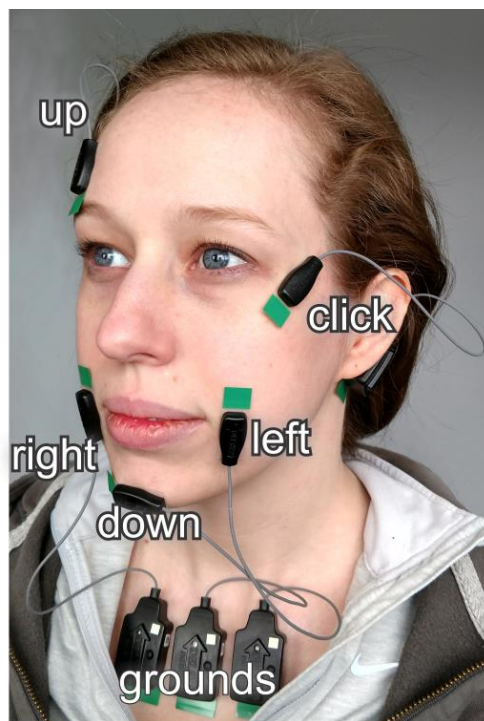
N-grams were calculated offline and stored. When a user selected a target, the appropriate set of probabilities were selected and used to generate a new weighted Voronoi diagram via the Python module pyvoro (Python bindings for Voro++; Rycroft, 2009), with the probabilities scaled from 2 to 8 and set as the weight parameter. Phoneme labels were also dynamically enlarged with this

same scaling. A video example of the online prediction is available in Supplemental Materials.

### **3.3.4 *Surface Electromyographic (sEMG) Cursor***

Participants without motor impairments used an alternate computer access method available to individuals with severe paralysis, an sEMG-based facial cursor (Cler et al., 2016; Cler & Stepp, 2015; Vojtech et al., 2018). Full details of implementation are given in (Cler & Stepp, 2015), but briefly: sEMG captures muscle activity from the surface of the skin and is presented as an alternative to eye-tracking or head-tracking for individuals who have spared muscle control. Electrodes are attached to the surface of the skin with double-sided tape to capture muscle activity from targeted (and surrounding/overlapping) muscles (see Figure 3-1). Muscle activity was captured with the Trigno™ sEMG system from Delsys, Inc (Natick, MA). Electrodes each consist of small sensors placed over the targeted muscle, short (200 mm) wires, and one larger ground per electrode. Electrodes are single-differential active electrodes with 4 mm bars. Grounds were placed on the chest and mastoids and communicated wirelessly to the sensor base, which acted as a data acquisition device. Five simultaneous sEMG signals were captured at 1000 Hz with custom Python code and evaluated every 100 ms to move the cursor (Cler et al., 2016). Maximum muscle activations

from each targeted facial muscle during a brief calibration (<5 min) were used to set thresholds per subject, session, and electrode. During the task, any muscle activation above the threshold moved the cursor in the direction of the associated facial gesture (e.g., eyebrow raise → cursor moves up). Combining facial gestures allowed users to move the cursor in any 360° direction, and the magnitude of the activation changed the speed of the cursor movement.



**Figure 3-1. sEMG mini-sensor locations (and associated grounds on chest and mastoids), placed to capture muscle activity during a particular facial gesture and subsequent cursor action: Left (half smile); right (half smile); up (eyebrow raise); down (chin contraction); click (wink). Combining gestures allows the cursor to move in any 360° direction, and magnitude of activity controls cursor speed (Cler & Stepp, 2015).**

### **3.3.5 Participants**

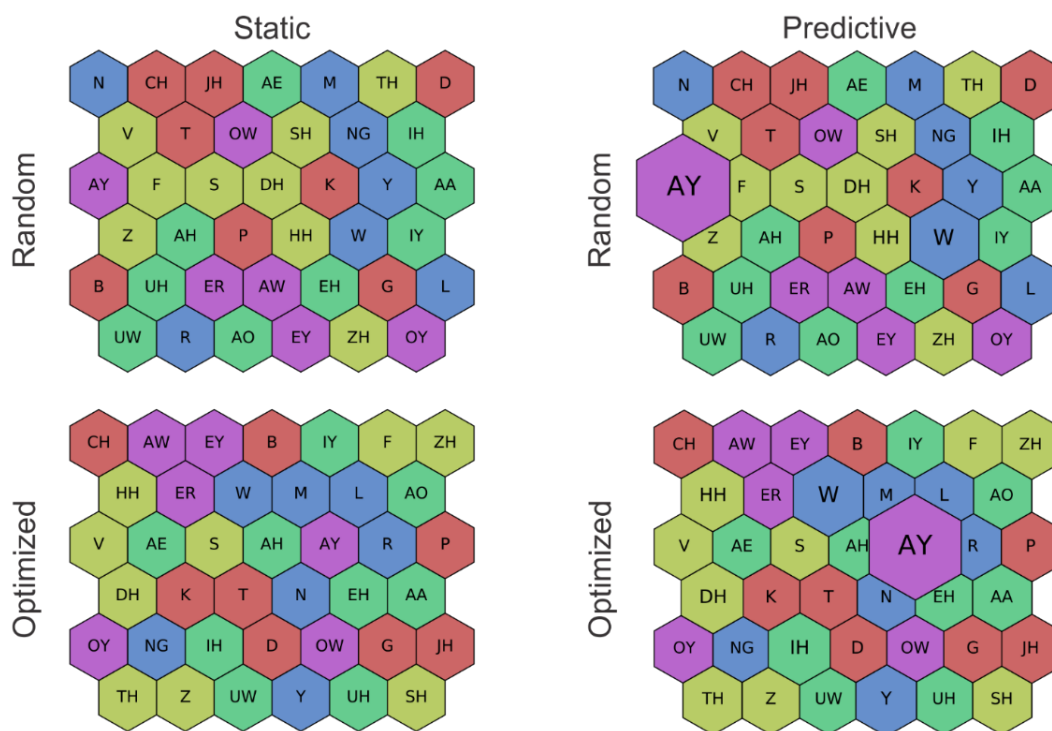
Thirty-six adults without motor impairments participated in the first study<sup>8</sup>. All were native speakers of American English and reported no history of speech, language, or hearing disorders. Participants were largely university students and were excluded if they had previous experience with sEMG research, phonemic keyboards, or transcription (e.g., speech language pathology students; singers). The participants (16 men, 19 women, 1 non-binary person; balanced across groups) had a mean age of 21.2 years (SD = 2.6).

### **3.3.6 Experimental Designs**

Participants without motor impairments completed twelve experimental sessions, each lasting 1–1.5 hours. Sessions occurred on separate days over three weeks with no more than three days between sessions. Participants used a facial sEMG-cursor to access the phonemic interfaces (Cler et al., 2016; Cler & Stepp, 2015; Vojtech et al., 2018). Participants were pseudorandomly assigned into one of four groups, balanced for age and reported gender. Each of the four groups were assigned to one of the four different interfaces (Figure 3-2).

---

<sup>8</sup> Three additional individuals were recruited but were unable to complete their participation. One completed seven sessions, but data were lost due to experimenter error, and thus the remaining sessions were cancelled. One had reported no neurological disorders but presented with a severe facial tic, so we chose to discontinue his participation. The final participant struggled to mimic the facial gestures used for the cursor control system (could not smile or move cheek on command) and chose to discontinue his participation at that point. We did not apply sEMG sensors or attempt to record sEMG data, so it is unclear whether the underlying musculature was activating or whether he could eventually have learned to use the cursor control system. AAC users have used the cursor with a variety of alternate placements (Cler et al., 2016; Vojtech et al., 2018). For homogeneity in this study, we did not offer alternate gestures or placements as we anticipated all participants would have sufficient facial muscle control.



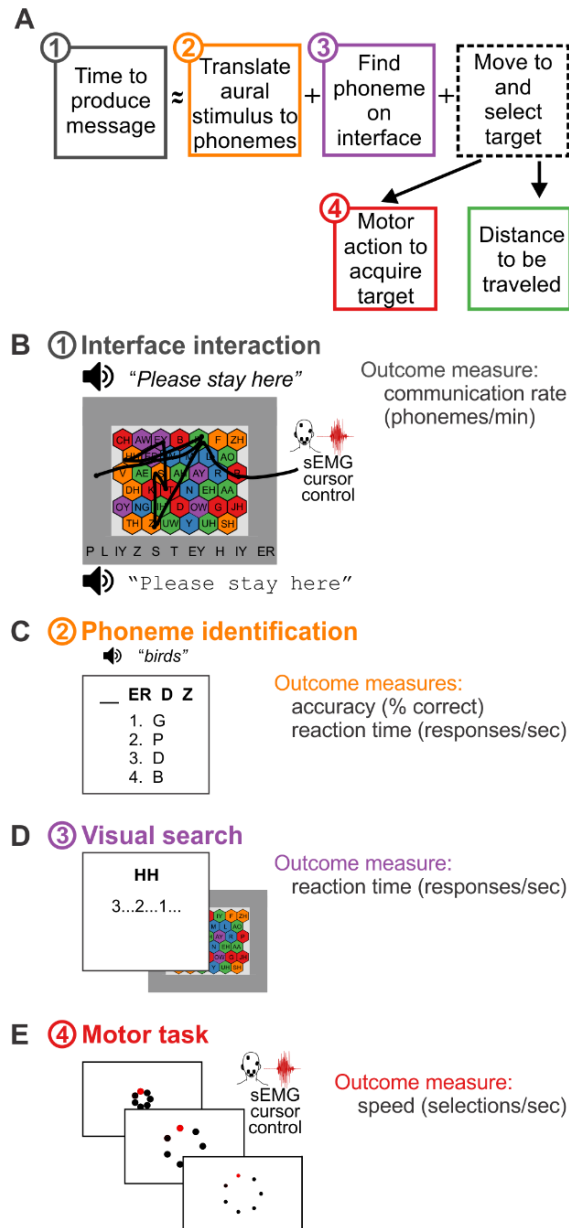
**Figure 3-2. Four interfaces used in by different groups of participants. Top left: random/static interface. Top right: random/predictive interface. Bottom left: optimized/static interface. Bottom right: optimized/predictive interface. Phoneme labels are a standard set (Shoup, 1980). Colors are consistent across groups and were isoluminant. Colors denote rough phoneme category: simple vowels in green, complex vowels (diphthongs, r-colored vowels) in purple, fricatives and affricates in yellow, stops in red, and liquids, nasals, and semivowels in blue.**

The first session began with a video showing each phoneme on the individual's assigned interface followed by its sound and an exemplar (e.g., “[CH], cheese; [ZH], measure”). Each session then had two minutes of free interaction with the interface using a typical mouse, followed by sEMG cursor application and calibration, 30 minutes of interaction with the interface via the sEMG cursor (the “main task”), and three short probe tasks.

Most of the session was devoted to the main task, recreating aurally-presented messages with the phonemic interface. This task required participants

to translate the aural stimulus to our phoneme set, visually locate those phonemes on the interface, and move to and select the target (schematized in Figure 3-3A). The time it took to move to and select the target was governed by several factors: (1) the participant's proficiency with any particular access method, (2) the distance that must be traveled, (3) the precision needed to select the target (determined by the target's size). The distance to be traveled was the only component that was modulated in the layout optimization process. Prediction modulated the precision needed and perhaps the speed of visually locating targets on the screen. As the other components of this task may vary across participants and should vary across sessions (as the participant's performance on the tasks improves), a series of probes were developed to assess each component.

*Main task (Figure 3-3B).* Participants were prompted with one message from a corpus of suggested AAC messages (1004 messages; Beukelman & Gutmann, 1999) and then used the facial sEMG cursor to select the phonemes they wanted to use to recreate that message. Participants recreated different messages with the phonemic interface for at least 30 minutes each session (interactions were not automatically terminated after 30 minutes if the participant was in the midst of a trial, but instead terminated after that trial was completed). The corpus of stimuli was also used to generate the phonemic transition properties used both in the optimization and prediction methods. Participants were not able to delete any accidental selections and were instructed to do their



**Figure 3-3. Experimental design. (A) Processes required to recreate a given prompt with the phonemic interface: translate the stimulus to the phoneme set, find phonemes on given interface, and use access method to move to and select the targets. (B): Main task, with outcome measure communication rate (phonemes/min). (C-E): Probes designed to assess participant acuity on each task: (C) Aural stimulus and phonemic representation with one phoneme missing are presented, and accuracy (% correct) and reaction time (responses/sec) were collected. (D) Participants indicated when they visually located the given label (outcome measure: reaction time in responses/sec). (E) Participants used facial sEMG cursor to select circular targets and were assessed on speed (selections/sec).**

best if they were not sure which sounds to select. Participants were instructed to complete each trial (message) as quickly and accurately as they could. The top left corner of the interface displayed the selections made during the current trial, and after the participant concluded the trial (by clicking the area surrounding the interface), the selected targets were synthesized as auditory feedback. After each trial, a popup box appeared with a number in it. Participants were instructed that that number represented an estimate of how quickly and accurately they had completed the message. This number was calculated online using information transfer rate (Wolpaw et al., 2000), which encapsulates both speed and accuracy in one number. Accuracy was estimated using the minimum string distance between the phonemes selected and the phonemes expected based on automated dictionary transcription of prompts (Soukoreff & MacKenzie, 2001). Speed was calculated as the number of actual selections divided by the time it took to complete the trial. Participants were instructed that the accuracy calculations were not always correct, but to just try to make the message sound as close to the prompt as possible, as quickly as possible. While an estimation of speed and accuracy were shown to the participants during the experiment, the main outcome measure used for the remaining analysis was speed. This is because messages can be created with a variety of phoneme choices and still be intelligible to the listener (e.g., consider the difference between [S-T-OW-R]-/stour/<sup>9</sup> and [S-T-AO-R]-/stɔr/). Further, these interfaces do not use spaces

---

<sup>9</sup> Two phonemic transcription conventions are used throughout this paper. One is the

between the words. While this does assist in speed, it makes error detection more difficult, as spaces serve as important orthographic markers of word boundaries. These factors make accurate automated error estimation impossible, and the total quantity of messages (>20,000) made perceptual intelligibility estimates infeasible. Thus, we focus here only on speed (selections per minute). Accuracy estimates are explored further in the discussion. Following the 30 minutes of interaction (henceforth, “main task”), participants completed three brief probes designed to capture skill learning of different aspects of the main task.

*Phoneme identification task* (Figure 3-3C). To assess their ability to translate an aural stimulus to the phoneme set, participants completed 15 fill-in-the-blank style questions during each session. Participants were aurally prompted with one of the messages from the message bank and one word was aurally repeated (e.g., “The birds are chirping... birds”; Figure 3-3C). Then participants were presented with a fill-in-the-blank question with the phonemic representation of that repeated word with one phoneme missing, using the experimental phoneme set and labels. Participants were instructed to determine which sound was missing and select the correct answer by hitting the 1-4 keys on the number row of the computer. Participants were instructed to complete this

---

International Phonetic Alphabet (IPA), which is likely familiar to readers and will be indicated with sounds between slashes ( / saundz / ). When relevant, we will also show transcriptions in ARPABET, which is a machine-friendly English transliteration and was used in this study as the target labels on the interfaces. ARPABET text will be indicated with sounds between square brackets ( [S-AW-N-D-Z] ). Auditory stimuli will be presented either with IPA or via orthographic text in quotes.

task as quickly and accurately as they could. Two outcome measures were obtained: accuracy and reaction time (responses/sec).

*Visual search task* (Figure 3-3D). To assess their ability to find phonemes on the interface, participants visually located 10 randomly-generated phoneme labels during each session. Participants were presented with a white screen with a particular phoneme label (e.g., “HH”; Figure 3-3D) and then the experimental interface presented a 3-2-1 countdown and disappeared. Participants were instructed to visually locate the prompted phoneme label and then hit the 0 key on a keyboard to indicate that they had found it. Phoneme labels were randomly selected on a trial-by-trial basis; this meant that occasionally the same label was presented twice in one session. These were removed in post-processing such that only the first presentation of any one label was used to calculate the outcome measure of visual search time (responses/sec).

*Motor task* (Figure 3-3E). To assess their ability to use the sEMG cursor, participants completed a task in which they selected dots on the screen using the cursor during each session. Participants were presented with a circle of black dots of three possible distance and sizes, selected to represent three different difficulties (Fitts' law indices of difficulty [ID] of 2, 3, and 4). One dot would turn red; participants were instructed to select this dot as quickly as possible. Once selected, a dot across the circle in a standard order would turn red and the participant would select that dot, and so forth, until all dots in one difficulty level were selected. All three difficulty levels were presented in random order each

session. The outcome measure was speed (selections/sec).

### **3.3.7 Statistical Analyses**

All statistical analyses were completed in R (R Core team, 2015). The outcome measure in participants without impairments was communication rate, and factors included participant, session, interface, prediction, and the measures from probes: motor task performance (selections/sec), phoneme identification–accuracy (%), phoneme identification–reaction time (responses/sec), and visual search performance (responses/sec). Parameters were analyzed for normality via visual inspection of Quantile-Quantile plots. All factors were normalized ( $M=0$ ;  $SD=1$ ) and multicollinearity between factors was assessed and rejected. Separate statistical models were calculated to answer our two questions: (1) How do optimization and prediction affect learning? (2) How do optimization and prediction affect performance after learning?

To assess learning, a linear mixed effects model (Bates, Machler, Bolker, & Walker, 2015) was performed on data from sessions 1–9 with communication speed as the outcome measure, participant as a random factor, and session, interface, prediction, probe measures, and all relevant interactions as factors. Sessions 1–9 were chosen via visual inspection of communication rates across all groups and sessions (see Figure 3-4) to include approximately linear learning slopes. To assess communication rate after learning, a linear model was performed on the data from the final session (12) only, with communication speed as the outcome measure and interface, prediction, probe measures, and

all relevant interactions as factors. For both models, backwards stepwise-regressions were performed in order to determine which, if any, of the probe measures captured individual variation relevant to the task<sup>10</sup>. Unstandardized  $\beta$  coefficients are provided as a proxy for effect sizes. For the mixed-effect model, marginal and conditional  $R^2$  were calculated to represent the variability accounted for by the fixed effects alone and the fixed and random effects in the model respectively (Lefcheck, 2015; Nakagawa & Schielzeth, 2013).

### 3.4 Results

Participants without motor impairments completed 20,849 trials across a total of 432 sessions. Participants showed an increase in communication rate

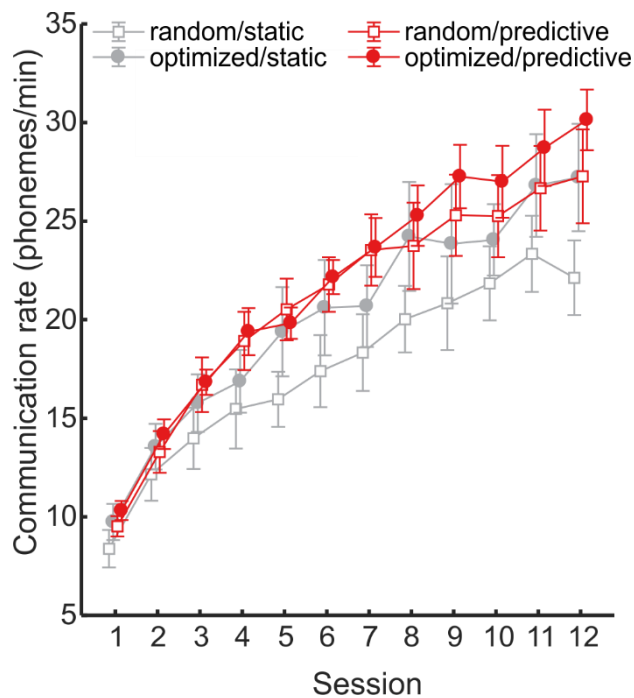
---

<sup>10</sup> Full model inserted into first backwards stepwise-regression (training sessions):  
 $\text{lmer}(\text{communication rate} \sim (1 | \text{participant}) + \text{interface} + \text{prediction} + \text{session} + \text{motor task speed} + \text{phoneme identification} - \text{accuracy} + \text{phoneme identification} - \text{reaction time} + \text{visual search time} + \text{interface} : \text{prediction} + \text{interface} : \text{session} + \text{prediction} : \text{session} + \text{interface} : \text{prediction} : \text{session} + \text{interface} : \text{motor task speed} + \text{prediction} : \text{motor task speed} + \text{session} : \text{motor task speed} + \text{interface} : \text{prediction} : \text{motor task speed} + \text{interface} : \text{session} : \text{motor task speed} + \text{prediction} : \text{session} : \text{motor task speed} + \text{interface} : \text{prediction} : \text{session} : \text{motor task speed} + \text{interface} : \text{phoneme identification} - \text{accuracy} + \text{prediction} : \text{phoneme identification} - \text{accuracy} + \text{session} : \text{phoneme identification} - \text{accuracy} + \text{interface} : \text{prediction} : \text{phoneme identification} - \text{accuracy} + \text{interface} : \text{session} : \text{phoneme identification} - \text{accuracy} + \text{prediction} : \text{session} : \text{phoneme identification} - \text{accuracy} + \text{interface} : \text{prediction} : \text{session} : \text{phoneme identification} - \text{accuracy} + \text{interface} : \text{phoneme identification} - \text{reaction time} + \text{prediction} : \text{phoneme identification} - \text{reaction time} + \text{session} : \text{phoneme identification} - \text{reaction time} + \text{interface} : \text{prediction} : \text{phoneme identification} - \text{reaction time} + \text{interface} : \text{session} : \text{phoneme identification} - \text{reaction time} + \text{prediction} : \text{session} : \text{phoneme identification} - \text{reaction time} + \text{interface} : \text{prediction} : \text{session} : \text{phoneme identification} - \text{reaction time} + \text{interface} : \text{visual search time} + \text{prediction} : \text{visual search time} + \text{session} : \text{visual search time} + \text{interface} : \text{prediction} : \text{visual search time} + \text{interface} : \text{session} : \text{visual search time} + \text{prediction} : \text{session} : \text{visual search time} + \text{interface} : \text{prediction} : \text{session} : \text{visual search time})$

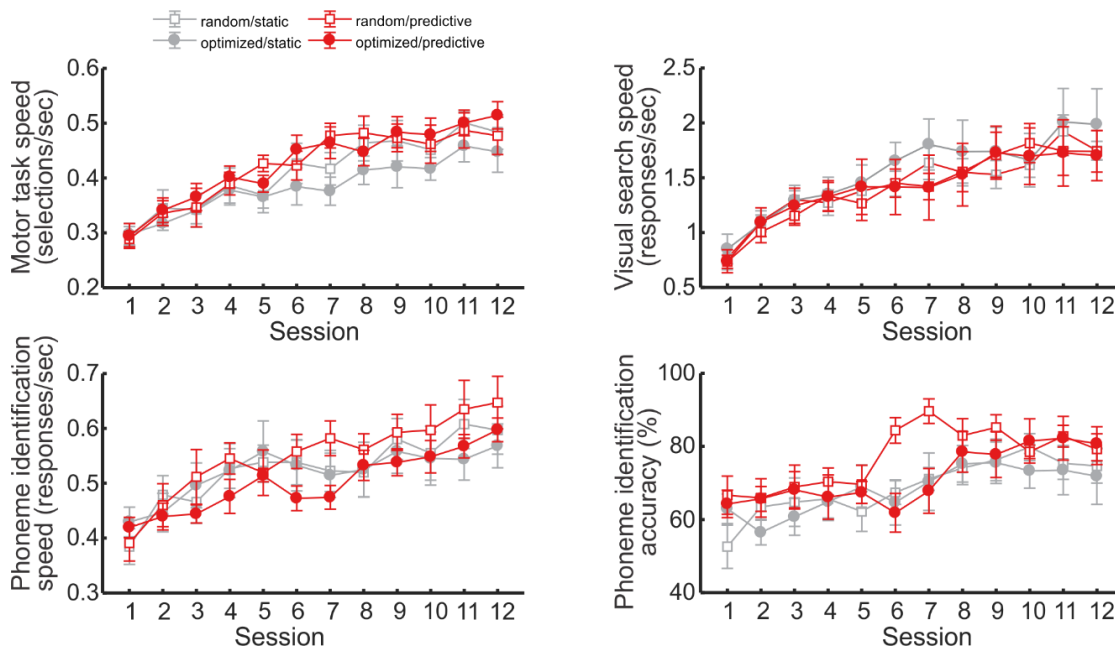
Full model inserted into second backwards stepwise-regression (final session only):  
 $\text{lm}(\text{communication rate} \sim \text{interface} + \text{prediction} + \text{motor task speed} + \text{phoneme identification} - \text{accuracy} + \text{phoneme identification} - \text{reaction time} + \text{visual search time} + \text{interface} + \text{prediction} + \text{interface} : \text{prediction} + \text{interface} : \text{motor task speed} + \text{prediction} : \text{motor task speed} + \text{interface} : \text{prediction} : \text{motor task speed} + \text{interface} : \text{phoneme identification} - \text{accuracy} + \text{prediction} : \text{phoneme identification} - \text{accuracy} + \text{interface} : \text{prediction} : \text{phoneme identification} - \text{accuracy} + \text{interface} : \text{phoneme identification} - \text{reaction time} + \text{prediction} : \text{phoneme identification} - \text{reaction time} + \text{interface} : \text{prediction} : \text{phoneme identification} - \text{reaction time} + \text{interface} : \text{visual search time} + \text{prediction} : \text{visual search time} + \text{interface} : \text{prediction} : \text{visual search time})$

across the 12 training sessions. Average communication rates across groups ranged from 9.4 phonemes/min (SD: 2.3) in session 1 to 26.7 phonemes/min (SD: 6.9) in session 12. Communication rates between groups are shown in Figure 3-4, which suggests that the optimized/predictive interface provides the highest communication rates, the random/static interface provides the lowest, and the optimized/static and random/predictive provide similar communication rates. Probe results are shown in Figure 3-5 and show that all measures increase with session and with communication rate (Figure 3-4), as expected with participant learning. Probe measures generally show overlapping error bars, suggesting similar performance across groups.

Results of the first linear model on data from sessions 1–9 are shown in Table 3-1; this model accounted for 86.5% of the variance in the data (conditional  $R^2$  including random factor: 86.5%; marginal  $R^2$ : 66.9%). Significant main effects were prediction, session, motor task speed, and phoneme identification–reaction time. Interface was not significant. Results of the linear model on session 12 data are shown in Table 3-2; this model accounted for 67.5% of the variance in the data ( $R^2$ ). Significant main effects were interface, motor task speed, and phoneme identification–reaction time. Prediction and all interactions were not significant.



**Figure 3-4. Communication rates per session averaged by group. Error bars are standard error.**



**Figure 3-5. Results of probes. Top left: Motor task speed. Top right: Visual search speed. Bottom left: phoneme identification speed. Bottom right: phoneme identification accuracy (% correct). Error bars are standard error.**

**Table 3-1. Sessions 1–9 mixed effects model - remaining factors after backwards stepwise regression**

	Communication Rate		
	$\beta$	<i>CI</i>	<i>p</i>
(Intercept)	12.05	9.61 – 14.48	<.001
interface	0.83	-2.57 – 4.23	.635
prediction	3.68	0.90 – 6.47	.015
session	0.88	0.63 – 1.12	<.001
motor task speed	1.94	0.86 – 3.02	<.001
phoneme identification–reaction time	0.95	0.44 – 1.46	<.001
visual search	0.32	-0.78 – 1.43	.567
interface x prediction	-3.00	-6.94 – 0.93	.145
interface x session	0.41	0.08 – 0.74	.017
interface x motor task speed	-1.14	-2.73 – 0.45	.16
session x motor task speed	0.01	-0.15 – 0.18	.86
session x phoneme identification–accuracy	-0.07	-0.13 – -0.00	.043
interface x visual search time	0.39	-1.02 – 1.80	.586
prediction x visual search time	1.77	0.51 – 3.04	.006
interface x session x motor task speed	0.37	0.13 – 0.61	.002
interface x prediction x visual search time	-2.04	-3.68 – -0.39	.016
Observations	324		
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	.669 / .865		

**Table 3-2. Session 12 linear model; remaining factors after backwards stepwise regression**

	Communication Rate		
	$\beta$	<i>CI</i>	<i>p</i>
(Intercept)	17.65	14.12 – 21.18	<.001
interface	7.27	2.74 – 11.80	.003
prediction	1.63	-3.40 – 6.67	.511
motor task speed	3.87	1.91 – 5.82	<.001
phoneme identification–accuracy	1.14	-3.11 – 5.39	.586
phoneme identification–reaction time	2.47	0.66 – 4.28	.009
interface x prediction	-1.69	-8.75 – 5.37	.627
interface x phoneme identification–accuracy	-0.45	-5.38 – 4.49	.853
prediction x phoneme identification–accuracy	5.11	-1.23 – 11.45	.11
interface x prediction x phoneme identification–accuracy	-6.11	-13.79 – 1.56	.114
Observations	36		
R <sup>2</sup> / adj. R <sup>2</sup>	.675 / .563		

### 3.5 Discussion

The regression models accounted for a moderate amount of the variance in individual performance:  $R^2=86.5\%$  for sessions 1–9 and  $R^2=67.5\%$  for the final session (12). Interestingly, the significant factors were different between these models, which is also reflected in the group differences evident in Figure 3-4. During early sessions, the two predictive groups (red) appear to have similar, high communication rates. By the final session, the groups appear to have stratified and somewhat stabilized. Visually (that is, without accounting for individual differences captured by the probes), differences are seen in the groups differing by interface (in Figure 3-4, note the large differences between the red circle and red square and between the grey circle and grey square) and by prediction (note the differences between the red and grey circles and between the red and grey squares). When accounting for individual performance differences captured by the probes, prediction is significant during the early sessions, whereas interface is significant during the final session. These effects are explored in detail in following sections. Briefly, however, prediction likely provided faster communication rates during training because it enabled users to learn the interface target locations and provided larger targets when precision was more difficult (this is, when participants were still learning the novel access method). Optimization acted to increase communication rates during the final session (and not during the earlier training sessions). As the optimization assumes that participants will move directly to the targets, it perhaps only

became beneficial in later sessions when participants knew the target locations and were skilled in the access method, and thus moved directly to the targets.

### **3.5.1 Estimates of Accuracy**

Many studies of communication effectiveness do not include measures of accuracy, instead focusing only on speed (words per minute). Although accuracy is somewhat easy to estimate in most studies using orthographic text entry (e.g., Soukoreff & MacKenzie, 2001), reliable accuracy measures are difficult to estimate in this study. Automated accuracy measures typically compare entered text to target text and may or may not accept misspellings and count deletions. However, the targets in this study were auditory, which could be reliably transcribed in a variety of ways (e.g., the word “tomorrow’s” could be transcribed in the dictionary spelling [T-AH-M-AA-R-OW-Z] or said aloud [T-M-AA-R-AH-Z] or [T-UW-M-AO-R-OW-Z] or a variety of other ways). This is a strength of phonemic interfaces: users can make messages that sound exactly as they wish, and previous reports have suggested that phonemic interfaces may also be robust to certain types of substitution errors (Cler et al., 2016). Further, these interfaces do not use spaces between the words, as they are not necessary to the speech synthesis and represent additional unnecessary motor activity. However, this also makes automated accuracy estimation more difficult, as without spaces to serve as boundary markers, it can be very difficult to determine what word or sound is a participant’s current goal. The appropriate accuracy measure for these messages is likely thus intelligibility, but the total quantity of messages (>20,000)

make perceptual intelligibility estimates infeasible. Thus, while the statistical analyses in this paper focus only on speed (selections per minute), other qualitative and quantitative analyses of errors may be of benefit to understand the effects of optimization and prediction.

### **3.5.2 Sources of Mismatch between Prompt and Produced**

There are several possible sources of mismatches between the dictionary transcription of the prompt and what the user produced. One category of mismatch occurred when participants misheard the prompt (e.g., “Let’s go” as “let go”, or dropping a quiet “the” at the start of a prompt). Another is motor-based selection errors, in which users accidentally selected phonemes other than the ones they meant to select because of unfamiliarity with or noise in the access method. The final category of errors are those in which the user accurately selected their intended targets, and those intended targets differed from the prompt. These differences could be due to: (1) the user’s intended pronunciation differing from the prompt, (2) the user not knowing the correct phoneme, or (3) the user knowing the “correct” phoneme, but not knowing which phoneme label corresponded to that sound. These final two categories of error may have varied between the groups and are thus of interest for further analysis.

There were several common phoneme mismatches seen across participants in different groups that may represent pronunciation, phoneme error, and/or phoneme label confusion. These common mismatches were identified by researchers during the experimental sessions, and examples of different

categories of mismatches are given in Table 3-3. Repeated vowel substitutions could be due to any one of these three causes (single vowel substitutions could also be due to a motor-based selection error). Phoneme confusion was likely responsible for voicing contrast errors. For example, plurals are typically indicated orthographically with an added “s” at the end of a word. Phonemically, however, those may be pronounced as /s/ or /z/. Participants in this study often used [S] for both cases, but likely would pronounce the word appropriately. In addition, participants often used [OW]-/ou/ instead of the correct selection [AW]-/aʊ/, likely because the label OW forms the English word “ow” (/aʊ/), rather than because participants did not know which sound belonged in the word of interest. Some of these mismatches differed between groups and are illustrated further in *Discussion > Effects of Prediction > Effects of prediction on mismatches between prompt and produced*.

**Table 3-3. Common mismatches between prompt and phonemes selected**

Type of mismatch	Examples	Possible source(s) of mismatch
Voicing error	Plurals using [S] instead of [Z]	Phonemic error
	“the” produced as [TH-AH]-/θə/ instead of correctly [DH-AH]-/ðə/	Phonemic error, possible label confusion
Vowel substitution	“I” produced with sounds other than [AY]	Label confusion
	“How” produced as [HH-OW]-/hou/ instead of the correct [HH-AW]-/haʊ/	Label confusion
	“Been” (dictionary: [B-IH-N]-/bɪn/; acceptable US variants: [B-IY-N]-/bɪn/, [B-EH-N]-/bɛn/)	Pronunciation/accent differences or label confusion

These likely represent different sources of mismatches that might be inconsequential or could be remedied. Listeners may not even notice a voicing error or may be able to understand the message even with the error. If these errors do impact communication efficiency, they may be remedied by further unstructured practice (e.g., the 2 min warm-up each session), semi-structured practice (e.g., the main task), or formal instruction, such that speech-language pathology students are taught. We gave no explicit instruction on phonemic transcription to our participants. End-users of the interfaces would likely have speech therapy, during which these types of instruction could take place.

### **3.5.3 Effects of Prediction**

Adding prediction had a significant effect on communication rate, which may have resulted from a variety of factors, both intended and inadvertent. Statistical analyses of the training sessions (Table 3-1) showed a significant main effect of prediction. As statistical models consider only communication rate as an outcome, there are likely additional effects that are not reflected in these statistical results; these include differences in the quantity and type of mismatches between the prompt message and the produced message.

*Effects of prediction on mismatches between prompt and produced.* One possible effect of prediction relates to the earlier discussion of mismatches between the sounds in the prompt and what participants produced (*Discussion > Sources of Mismatch between Prompt and Produced*). Predictive interfaces had the effect of drawing the users' attention to phoneme labels that they may not

have chosen otherwise. Mismatching selections could have many different sources: a motor error (i.e., clicked a target accidentally), phoneme identification error (i.e., could not identify that the word started with an /aɪ/ sound), or a target label identification error (i.e., the participant knew the sound was /aɪ/ but not which target represented that sound).

We illustrate the possible influence of prediction on these mismatches with three different examples. First, accuracy measures of the produced sentences will not necessarily map directly to intelligibility, as listeners can comprehend sentences with voicing errors or different phonemes; this is why we have used only communication rate in measures thus far. However, the prediction methods steer the user towards using the dictionary transcriptions, and users were also given automated feedback on their speed and accuracy. We can measure the extent to which the produced messages exactly match the dictionary prompt. Of all of the 20,849 trials, 3962 matched the prompt exactly (19.0%); this ranged per individual from 2.3% to 45.4%. Individuals in the static groups produced 13.4% completely “correct” messages, while individuals in the predictive groups produced 23.9% completely correct messages. Although we do not know if there is an intelligibility difference between the groups, it is likely that the prediction at least trained users to produce messages using the dictionary transcriptions.

Next, to explore voicing errors, we tallied trials with either a [TH] or [DH] in the prompt<sup>11</sup> and calculated the types of differences seen across all participants. In trials with a [TH] in the prompt, participants used [TH] correctly 93.3% of the time, with substitutions of DH (5.4%), T-HH (0.3%), or T (0.8%). Of trials with [DH], participants correctly used [DH] 37.6% of the time, with substitutions of TH (60.5%), T-HH (0.6%), or D (1.0%). These mismatches are likely the result of both phonemic errors (that is, users do not consciously realize that /θ/ and /ð/ are different or are different than /t h/) and label errors (users may know that /θ/ and /ð/ are different, but not that they are represented by [TH] and [DH] on the interface). These may also represent pronunciation differences, as the common words “with” and “thank” can be variably pronounced with either phoneme. The “correctness” of the [DH] trials varied by cohort, with the static groups producing correct [DH] trials only 23.3% of the time, whereas predictive groups produced 53.7% correct [DH] trials. This suggests that the prediction may have indicated pronunciation, phoneme, and phoneme label suggestions to the users.

---

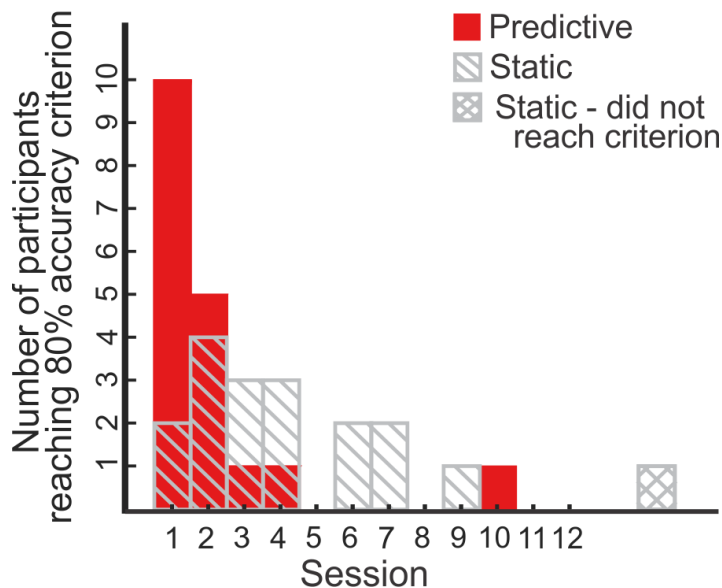
<sup>11</sup> We considered trials with only a [TH] or [DH] in the prompt and excluded those with both for simplicity. Of the remaining trials, 1079 were correct [TH] trials, 1429 were correct [DH] trials, 62 were [TH] prompts with [DH] selected, 2309 were [DH] prompts with [TH] selected, and 277 could not be automatically assessed and needed to be manually classified. Of those 277, 27 were correct [TH] trials, 33 were correct [DH] trials, three were [TH] prompts with [DH] selected, and 41 were [DH] prompts with [TH] selected. Of the remaining, 62 were trials where the participant ended the trial early, before getting to the [TH/DH]; 19 were those in which the participant appeared to miss the word with the [TH/DH] entirely (often a quiet “the” at the beginning of the prompt); 28 trials had [T-HH] instead of the [TH/DH]; 18 used just [T] for the [TH/DH]; 40 used just [D] for the [TH/DH]. The remaining six prompts were unclassifiable. These varied responses highlight the difficulty of assessing accuracy automatically. Even the [T-HH] or [T]- or [D]-only trials are generally intelligible to the listener.

Finally, to disambiguate phoneme from phoneme label mismatches, we have illustrated the effect of prediction on trials with the initial /aɪ/ sound (the English word “I”). Many of the sentences in the stimuli set begin with the word “I” or “I’m” and thus the sound /aɪ/. This is reflected by the size of the [AY] target in the starting configuration of the predictive interfaces, shown in in Figure 3-2. These are unlikely to be just phoneme errors, as the phonological mapping of the word “I” to the sound /aɪ/ is simple and consistent across dialects. If participants select only sounds with confusable labels, then we can infer that the likely cause of the errors is the label. If they select targets surrounding [AY], this would indicate a motor-based error, in which participants attempted to select the correct target but hit a nearby target instead. Of all trials starting with an /aɪ/ sound, participants used the correct label, [AY], 85.8% of the time. Most other trials (12.5%) began instead with sounds with easily-confused labels ([IY, IH, EY, AH, AE]; range from 4.6% to 0.7% each), with only 1.7% of trials starting with any other sound (<0.2% each). These errors varied across time and by cohort: Figure 3-6 indicates during which session a participant reached an (arbitrary) accuracy criterion of 80% correct in selecting [AY] in /aɪ/-initial trials. Note that participants using interfaces with no prediction (grey striped bars) took longer to reach 80% accuracy than those in predictive cohorts (red bars), with one participant in a static cohort who never reached criterion. All groups heard the phonemes that they selected synthesized together as auditory feedback, and all groups produced the same message bank (in a random order). Thus, it would appear

that prediction increased the accuracy of the messages produced by participants. These possible accuracy increases are not explicitly incorporated in the main results in Figure 3-4 or in the statistical results, as the main outcome measure (communication rate) does not consider accuracy. The results may implicitly reflect these differences if in fact prediction allowed participants to select sounds faster; that is, they perhaps hesitated less or better remembered the location of the intended targets on the interface. Although it is clear that prediction increased communication rates and affected how participants learned the target labels, further research could reveal the precise mechanisms behind these improvements.

*Effects of prediction during final session.* During the final session, prediction was no longer a significant factor in communication rate. This suggests that the effects expected via Fitts' law (that is, that larger targets are faster to select) were not consistent. This could be due to a variety of factors. In particular, the underlying prediction could have been inaccurate. Only 19% of the messages produced by participants completely matched those that were used to build the prediction that was based on dictionary-based automated transcription. This suggests that the effects of prediction were not maximized here. Future work could base prediction, at least in part, on the series of sounds these participants used, rather than a dictionary transcription. A final interface delivered to end users should certainly incorporate each users' selection history into the prediction algorithm. Because the participants often used non-dictionary transcriptions for

their messages, they perhaps learned to disregard the prediction entirely. In this way, the prediction could have actively made their performance worse if it made their preferred targets smaller and thus harder to select.



**Figure 3-6. Session in which each participant reached criterion of 80% accuracy of selecting [AY] on /aɪ/-initial trials over other vowel labels. Red (dark) bars: predictive groups. Note that these participants largely reached criterion in the first two sessions. Grey striped bars: static groups. Note that these participants took longer to reach criterion, and one participant never reached criterion (grey checked box).**

### **3.5.4 Effects of Interface Optimization**

Interface optimization had a significant main effect during the final session. During the training sessions, individuals in the optimized group saw extra gains based on session and motor task performance; this suggests that the later the session and the better able the participant was to use the access method, the

larger communication rate increases were seen from the (motor-based) optimization.

During the final session, the optimized interface had a large positive effect ( $\beta = 7.27$ ). Previous work suggested that an optimized interface should show 30.9% increase in communication rate, assuming ideal motor access and ideal phoneme selection. Final communication rates for the static groups were 22.1 versus 27.2 phonemes/min (random/static and optimized/static, respectively), whereas the predictive groups showed communication rates of 27.3 versus 30.1 phonemes/min (random/predictive and optimized/predictive, respectively). As a result, optimization improved communication rates by 23.0% and 10.2% in the final session. The reason for this discrepancy is likely due to the difference between the transition likelihoods used to create the optimizations (based on dictionary transcriptions of the stimuli set of messages) and the targets actually used by the participants. For example, the target combinations [AY] to [M] and [DH] to [AE] are near each other on the optimized interface, due to the high number of occurrences of the words “I’m” and “that” in the stimuli set. However, if participants routinely used [EY-M] and [TH-AE] instead (or other common errors shown in Table 3-3), their communication rates would not be increased over someone using the random layout.

Interface was not a significant contributor to communication rate during the training sessions. The optimization assumes that users go directly from one target to the next by Euclidean distance. However, individuals in this study were

contending with two additional issues that preclude this usage (aside from previous remarks about accuracy). First, they were learning the access method. Previous work suggests that during early training sessions using this access method, participants used separate facial gestures (e.g., first left and then up), but learned to coordinate gestures to go directly to the target diagonally by the fourth session (Cler & Stepp, 2015). Fitts' law optimizations used the Euclidean distance between targets to determine optimized layout, under the assumption that participants would move directly to the targets using coordinated facial gestures; it is likely that they were not doing this until later sessions. Second, participants had to learn which phonemes were in each message and where those targets were on the interface. This likely led to additional cognitive and visual search time between selections, masking possible effects of the optimization. As they got more experience with the interface and the task, these cognitive demands and search times decreased. Thus, in the final session, differences between the random and optimized interfaces were evident.

### ***3.5.5 Effects of Training and Probes***

As expected, session had a significant main effect in the training model (the model of the final day did not include session as a factor). In addition, during training, proficiency at the motor task had a large positive main effect on communication rate, suggesting that individuals who were faster with the access method saw increases in communication rate. During the final session, proficiency at the motor task had a large positive effect on communication rate.

Phoneme identification–reaction time was significantly related to communication rate during training and in the final session, but phoneme identification–accuracy was only a significant predictor as an interaction with session in the first model. This is likely due to the fact that the outcome measure in this study (communication rate) does not incorporate accuracy, but only speed. Finally, visual search time did not have a significant main effect in either model.

### ***3.5.6 Comparison to Other Communication Rates***

The maximum group communication rate was 30.1 phonemes/min in the final session using the optimized/predictive interface. This translates to approximately 7.5 words per minute (wpm), using a standard of four phonemes per word and no spaces needed. While this rate is comparable to or higher than other studies using this access method (Cler et al., 2016; Cler & Stepp, 2015), this is much slower than oral speech: 596 phonemes/min (counts diphthongs as separate phonemes; Osser & Peng, 1964) or 152 words/min (Maclay & Osgood, 1959).

A previous study using this access method with an alphabetical interface saw communication rates of 29 selections/min during the final (fourth) session. This is very similar to the rate of 30.1 selections/min seen in the optimized/predictive group in the final (twelfth) session. However, phonemes carry more information per selection than letters, and phonemic input does not require spaces. Thus the 30.1 phonemes/min represents 7.5 wpm, whereas the 29 selections/min on the alphabetic interface represents 4.8 wpm. While both are

much slower than oral speech, phonemic input does seem to lead to increased communication rates using alternate access methods, if we consider character-by-character input with no word completion.

### ***3.5.7 Applications to Non-Phonemic Interfaces***

Some of these advances may be applied to orthographic or symbol-based interfaces. Orthographic interfaces have already been optimized with these methods (e.g., MacKenzie & Zhang, 1999; Zhai et al., 2002). However, these interfaces have not generally been adopted, likely because users have a large amount of experience with QWERTY interfaces. The method used to indicate predicted targets, however, has not been explored in AAC or in other computer interface applications. Expanding targets have been shown to increase selection speed in center-out tasks (Zhai et al., 2003) or in a line of tightly-packed targets (e.g., the Mac OSX dock, in which icons are dynamically enlarged on hover; McGuffin & Balakrishnan, 2005). Visually highlighting targets on a keyboard via bolding or increasing the font size of labels on predicted targets (Magnien et al., 2004; Sears et al., 2001) has similarly been shown to increase selection rates, even if prediction is noisy. However, the use of expanding targets in a grid (which are also paired with increased font sizes) is novel and likely to be beneficial in a variety of uses. This prediction could be applied to orthographic interfaces or even interfaces with grids of symbolic targets. The underlying algorithm requires only a set of seeds (here, positioned at the center of each target) and a set of weights; further, this algorithm implementation (via pyvoro) is fast enough that it

is usable with even over-trained access methods (e.g., typical mouse and touchscreen input) without a noticeable delay.

### **3.5.8 *Limitations and Future Directions***

In order to assess the different aspects of these phonemic interfaces in a longitudinal design, we recruited 36 individuals without motor impairments. This enabled large cohorts over many time-points but may not represent how individuals who use AAC will use the interfaces. The participants did use an alternate access method to interact with the interfaces, and the access method was designed for individuals with motor impairments who use AAC (Cler et al., 2016; Cler & Stepp, 2015; Vojtech et al., 2018). However, this also meant that the participants were learning to use the access method at the same time that they were learning to use the phonemic interfaces. This may not always be the case in AAC users, as some may be long-term users of a particular access method who start to use a phonemic interface, or who may use a phonemic interface with a variety of access methods as their abilities and preferences change. Some AAC users may learn to use an access method and interface simultaneously (e.g., those with spinal cord injury). In addition, we were not able to complete a direct comparison to an orthographic interface. Further study will involve benchmarking these interfaces against orthographic interfaces with a variety of access methods.

There are a variety of different aspects of these interfaces that could be evaluated and refined. As previously mentioned, different phoneme labels would

likely expedite learning of sound/label mappings. In this study, we provided limited formal instruction: participants were shown a 1 min video on the first day that selected each sound and provided an exemplar. They were not permitted to watch the video again, and were provided no feedback (beyond motivation, e.g., “That one sounded good!”) or answers to specific questions (“Which one is /aɪ/?”, “What does ‘DH’ mean?”). Clinical implementation would likely involve a structured training program, which would include adjusting settings for phoneme labels and the degree of scaling for predicted targets as well as specific instruction in translating intended messages to phonemes.

Finally, while the prediction was beneficial in this study, there are additional improvements that would make it more effective. Our method of generating predictions based only on the stimuli set is limiting. Previous work suggests that text prediction for AAC is best when prediction is trained on a large set of text combined with a small set of AAC or AAC-like messages (Vertanen & Kristensson, 2011). Future evaluation should involve broader prediction strategies, including larger corpora and more sophisticated markers to assign prediction weights (e.g., language rules; a user’s past selection history or eye-gaze), as well as more refinement of the method of indicating prediction to the user.

### **3.6 Conclusions**

This study empirically assessed the effects of computational optimization and prediction on communication rates generated by participants without motor

impairments. Optimization was derived from corpus-based statistics and involved organizing phonemic targets so that targets likely to be selected in sequence were located in close proximity. Predicted targets were dynamically enlarged based on past selections and corpus statistics. Empirical evaluations revealed that dynamically enlarging targets based on prediction provided faster communication rates for participants without motor impairments during training (sessions 1–9), as users were learning the interface target locations and the novel access method. After training, optimization acted to increase communication rates. The optimization likely became relevant only after training when participants knew the target locations and moved directly to the targets. Future work is needed to validate these novel methods of optimization and prediction for AAC and to translate these results into clinical practice.

### **3.7 Acknowledgments**

The authors would like to thank Jaime Kim, Rebecca Glover, and Tiffany Peters for helping GJC, KRK, and JPN to run participants, Andreas Singer for recording the message bank, and Jay Bohland, Frank Guenther, and Chris Moore for providing input on the design of the study in participants without motor impairments. This work was supported by the National Institutes of Health - National Institute on Deafness and Other Communication Disorders under grant F31 DC014872 (GJC) and the National Science Foundation under grants 1452169 (CES) and 1247312 (JMV).

## Chapter 4. Clinical Translation and Implications

### 4.1 Introduction

Computational simulations suggest that phonemic interfaces may increase communication rates by up to 51% compared to orthographic interfaces (optimal use; no prediction), and that optimized phonemic interfaces should increase communication rates by up to 30.9% compared to phonemic interfaces with random layouts (Chapter 2; Cler & Stepp, 2017). However, these estimates do not include the additional time for cognitive processing and the effort required by users to learn to use a phonemic interface. Evaluations in participants without motor impairment suggested that optimization increased communication rates by 10.5–23.0% in the final session (percent difference between optimized and random interfaces in groups with and without prediction respectively; *Chapter 3 > 3.4 Results > Figure 3-4*).

We also evaluated the effect of prediction on communication rates in participants without motor impairment and found that prediction increased communication rates during training sessions (1–9), but was not a significant predictor of communication rate in the final session. However, the predictive groups did have higher communication rates than the static groups, with increases of 23.2% and 10.7% in the random and optimized groups respectively (*Chapter 3 > 3.4 Results > Figure 3-4*). These results suggest that further evaluation of prediction is warranted, particularly as the methods for indicating prediction are novel (see *Chapter 3 > 3.3 Methods > 3.3.3 Prediction* for

implementation). Thus we completed an additional evaluation of these predictive and static interfaces in individuals with motor impairment and solicited their feedback on the interfaces.

## **4.2 Methods**

### **4.2.1 Participants**

Six adults with motor impairments participated (participant characteristics in Table 4-1). Three participants were community dwelling and three were inpatients at a rehabilitation hospital. Diagnoses were congenital (cerebral palsy [CP]) or acquired (multiple sclerosis [MS], spinal cord injury [SCI], Guillain-Barré syndrome [GBS]). Participants included those with stable (CP; chronic SCI), degenerative (MS), and improving and/or stabilizing (GBS, acute incomplete SCI) impairments. Community-dwelling participants used a variety of computer access methods in their daily lives, including stylus access (held in mouth: P1; attached to wrist-guard: P3) and touchscreen access with their nose or eye-tracking (P4). Participants who were inpatients (P2, P5, P6) had used a variety of changing access methods as their conditions and needs progressed (e.g., head arrays of switches for wheelchair control; voice control for phones), but were not yet expert in any particular access methods. All participants provided consent in compliance with Boston University's Institutional Review Board; individuals with motor impairment provided either written consent or verbal consent witnessed by a communication partner as appropriate.

### **4.2.2 Study design**

Participants with motor impairment completed one or two sessions based on availability. For this within-subject design, participants used both the optimized/static and optimized/predictive interfaces in alternating blocks (counterbalanced; see interfaces in Figure 4-1). Each block consisted of 10 minutes of interaction with one of the two interfaces (same as “main task” in evaluation in participants without motor impairment; Chapter 3) followed by surveys to capture their experiences using the interfaces.

The first survey was the NASA Task Load Index (NASA-TLX), a brief survey that asks participants to rate the preceding task on six dimensions: mental demand, physical demand, temporal demand, performance, frustration, and effort (Hart & Staveland, 1988). Ratings are produced on a visual analog scale (VAS) with 20 bins; most are anchored with “very low” and “very high”, except performance, which is anchored at “perfect” and “failure”. The second survey was custom designed for this experiment and asked a variety of questions on a 10 cm visual analog scale (VAS). Questions included: “Do you think you could improve with practice?” (no improvement – lots of improvement); “I preferred the interface” (without prediction – with prediction); “I thought the enlarged targets:” (got too large – didn’t get large enough); and “I thought the enlarged targets:” (helped me learn the location of targets – made it harder to learn the location of targets). The full survey is included in Supplemental Materials. The experimenter read the questions aloud and dragged a pen across each VAS line until the

participant indicated their preferred stopping place (P1, P2, P3, P5, P6); one participant wished to fill out the forms herself and did so with a pen held in one hand (P4).

### **4.2.3 Analyses**

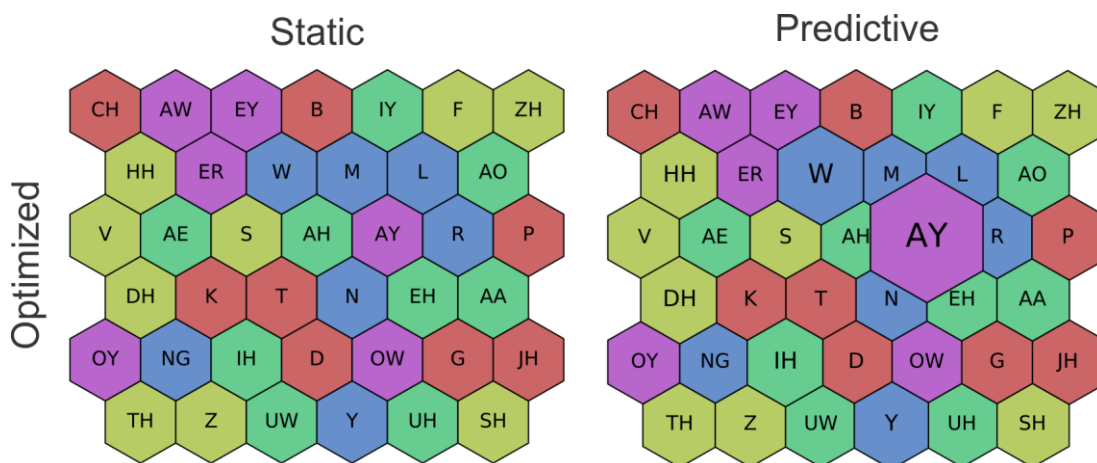
Communication rates and survey responses were tabulated to assess user effort and preferences. Communication rates were evaluated in phonemes per minute (as in the experiment in participants without motor impairment; detailed in Chapter 3). Survey responses were solicited from these participants in order to capture their experiences and insight into the design and usability of these AAC interfaces; we anticipated that their personal experience with noisy and effortful access methods and various AAC interfaces would provide valuable information beyond those considered by the researchers and participants without motor impairments. Survey responses were calculated as distance from the left line anchor divided by total length of the line (9.85 cm for NASA-TLX; 10 cm for all VAS questions). No statistical tests were performed; rather, responses and preferences are presented descriptively. Numbers presented in the text are the responses from the last set of surveys the participant filled out, measured in cm from the left anchor.

**Table 4-1. Participant characteristics**

	<b>Age/ Sex</b>	<b>Diagnosis</b>	<b>Access method for this experiment</b>	<b>Community- dwelling or inpatient</b>
<b>P1</b>	49/F	Multiple sclerosis (20 years post-diagnosis)	Mouthstick (stylus controlled with mouth) on touchscreen	Community-dwelling
<b>P2</b>	45/M	Spinal cord injury (acute; 3 months post injury)	Eye-tracker with sip-and-puff switch for click (new to participant)	Inpatient
<b>P3</b>	63/M	Spinal cord injury (chronic; >26 years post)	Stylus attached to stabilizing wrist guard on touchscreen (day 1: non-dominant hand; day 2: dominant hand)	Community-dwelling
<b>P4</b>	21/F	Cerebral palsy	Nose on touchscreen	Community-dwelling
<b>P5</b>	59/M	Spinal cord injury (acute; 6 weeks post)	Eye-tracker with physical switch for click, mounted on wheelchair for outside leg access (new to participant)	Inpatient
<b>P6</b>	63/F	Guillain-Barré syndrome (acute; 3 months post onset)	Stylus in hand on touchscreen	Inpatient

### 4.3 Results

Communication rates from participants with motor impairments are shown in Figure 4-2. Participants completed 3–7 blocks of trials (10 mins per block) over one or two sessions (sessions denoted by vertical dotted line). Community-dwelling participants (P1, P3, P4) completed more blocks than inpatients (P2, P5, P6) due to fatigue and availability.



**Figure 4-1. Interfaces used in alternating blocks by participants. Left: optimized/static interface. Right: optimized/predictive interface. Phoneme labels are a standard set (Shoup, 1980). Colors are consistent across interfaces, isoluminant, and denote rough phoneme categories: simple vowels in green, complex vowels (diphthongs, r-colored vowels) in purple, fricatives and affricates in yellow, stops in red, and nasals and semivowels in blue.**

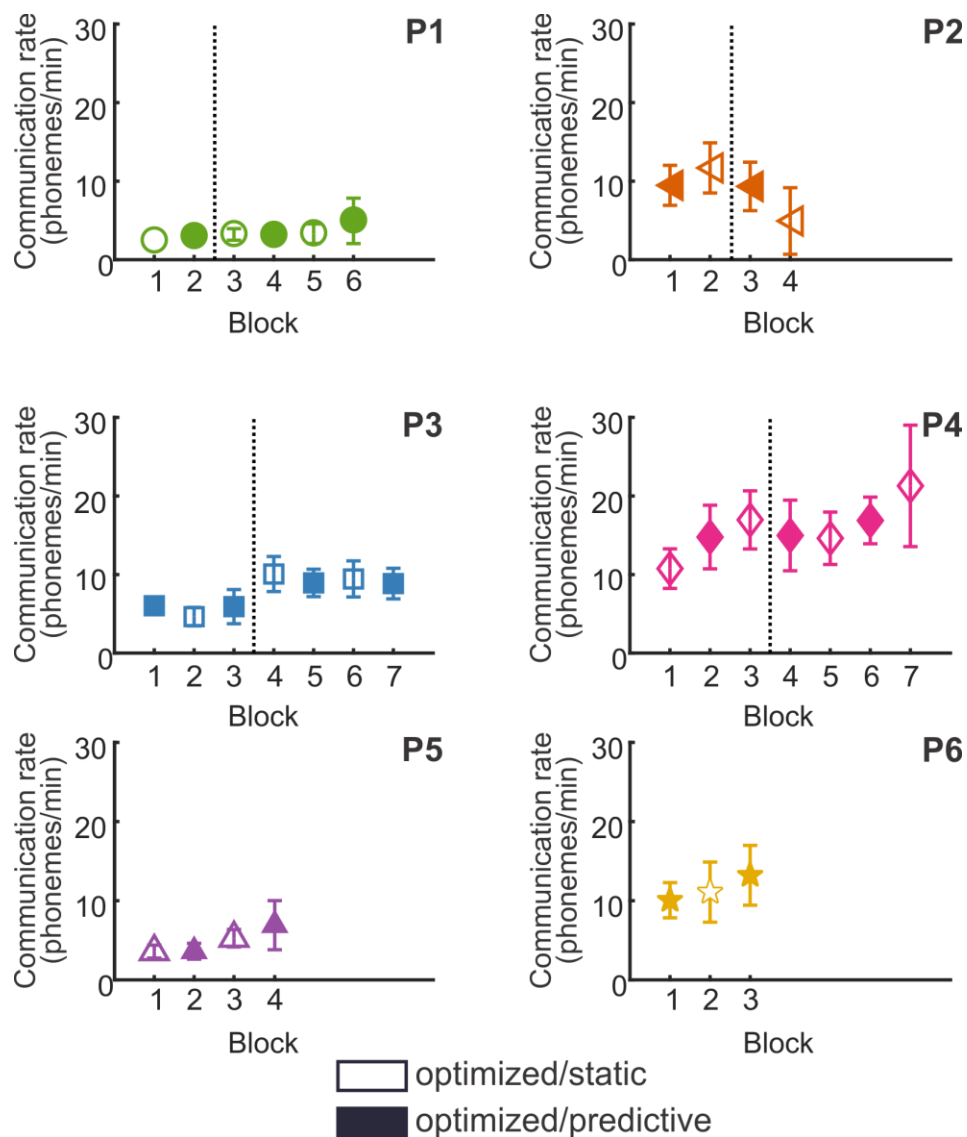
All participants strongly agreed that they would improve with practice ( $M=9.8$  cm,  $SD=0.6$ ; in which 0 cm indicated “no improvement” and 10 cm indicated “lots of improvement”). Four out of six participants strongly preferred the interface with prediction over the static interface (P1, P3, P4, P6; responses: 10 cm, 10 cm, 9.4 cm, 10 cm, in which 0 cm indicated a complete preference for static and 10 cm indicated a complete preference for prediction), whereas two participants moderately and strongly preferred the static interface (P2, P5; 2 cm and 0 cm). Participants generally agreed that the targets enlarged the right amount ( $M=4.5$  cm;  $SD=1.2$ , in which 0 cm was anchored at “got too large”, 5 was informally described as “about the right amount”, and 10 cm was anchored at “didn’t get large enough”) and that the prediction helped them to learn the location of targets ( $M=2.2$  cm;  $SD=2.6$ , in which 0 cm was “helped me learn the

location of targets” and 10 cm was “made it harder to learn the location of targets”).

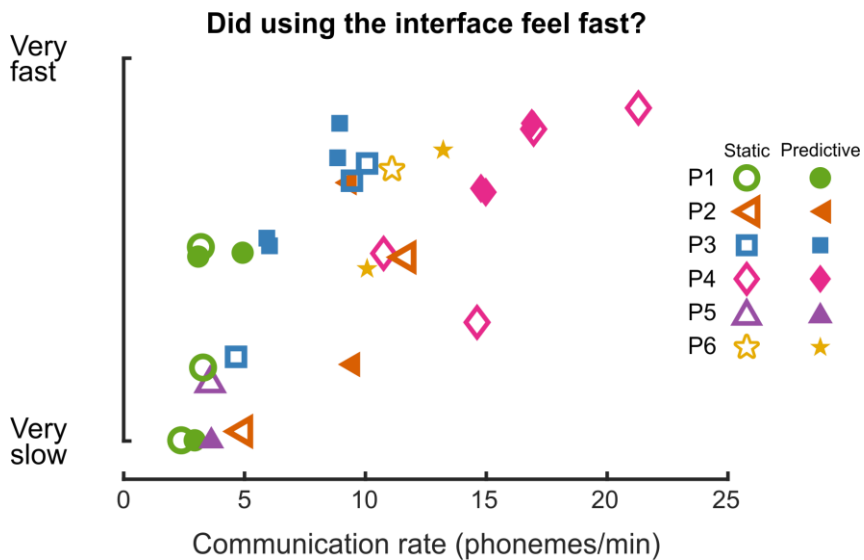
Participants remarked that they would improve with practice: “If I had this at home, I would go through every one of those sounds and I think you could get to where you could get pretty good speeds” and that the phonemic input was flexible: “As I use it more, I can see that you could get it to do the dictation just as you would want” (P3). One participant noted that the orthographic labels on the sounds interfered with her ability to select the right sounds: “It’s easier when you don’t know how it’s spelled,” but also noted that it got easier with practice: “I’m getting used to the sounds now” (P4). This was a common theme: “Boy did it go a lot easier. I felt more confident. I’m getting to know what /aɪ/ needs to be” (P3, on day two). One participant initially said he preferred the interface without prediction, but later highly preferred prediction as he got used to it (P5).

Although most participants preferred prediction, one participant who preferred the static interface remarked that he “liked to figure it out himself,” and that the prediction led him in a direction that he did not want (P2). The other participant who preferred static said that he did not use the prediction; “It didn’t matter, because it wasn’t the sound I was looking for, so I didn’t use it” (P5). However, this participant also remarked about a large target “I don’t remember what sound that makes... oh well, I’ll pick it anyway”, suggesting that he did in fact use cues from the prediction periodically (P5).

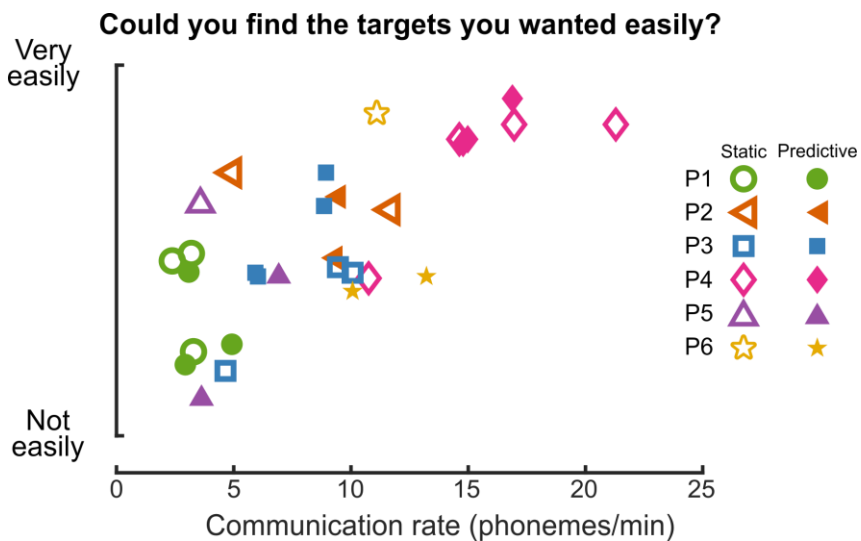
Some of the survey responses may have been modulated by communication rate, both between and within participant (across different blocks of trials and sessions). For example, Figure 4-3 shows survey responses for the question “Did using the interface feel fast?”. Participant responses show a positive correlation with communication rate, both within participant (that is, when they produced messages faster, they reported that the interface felt faster) and between participants. Similarly, Figure 4-4 shows responses from “Could you find the targets easily?”, which ranged from “Not easily” to “Very easily”. Participants who reported that they could easily find the targets had higher communication rates. Interestingly, the participant ratings for performance (from perfect to failure) shown in Figure 4-5 show a similar trajectory both within and between participants, despite each participant having no context for how their performance might relate to others’. A final comparison is shown in Figure 4-6, which depicts survey results for the question “Was using the interface frustrating”, from “not at all” to “very”. Participants with the lowest communication rates (P1 and P5) rated the interface as very frustrating, and their ratings of frustration were not obviously modulated by communication rates. Participants with moderate and high communication rates all rated the interfaces as minimally frustrating, with no apparent modulation from communication rate during a given block.



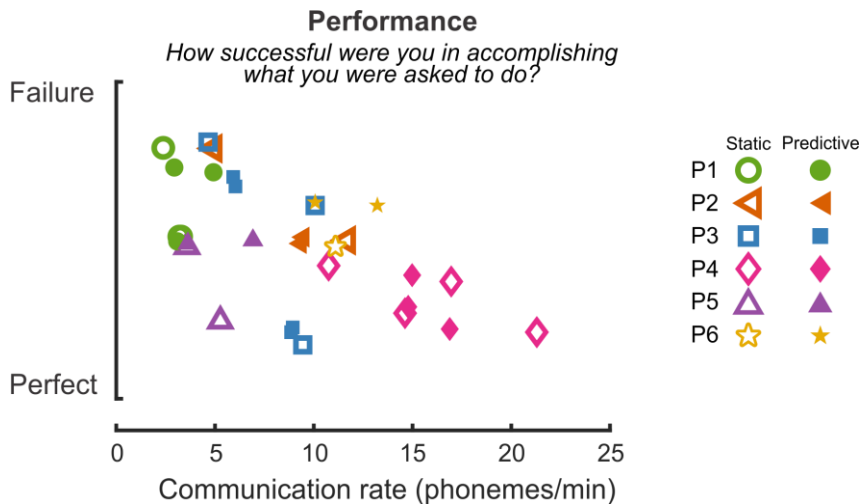
**Figure 4-2. Communication rates for the six participants with motor impairment (see Table 4-1 for participant characteristics). Participants all used optimized interfaces. Interfaces were either static (empty shapes) or predictive (filled shapes) in alternating blocks. When possible, participants completed blocks over two days; black dotted vertical lines indicate separation from day 1 to day 2. Error bars are standard deviation.**



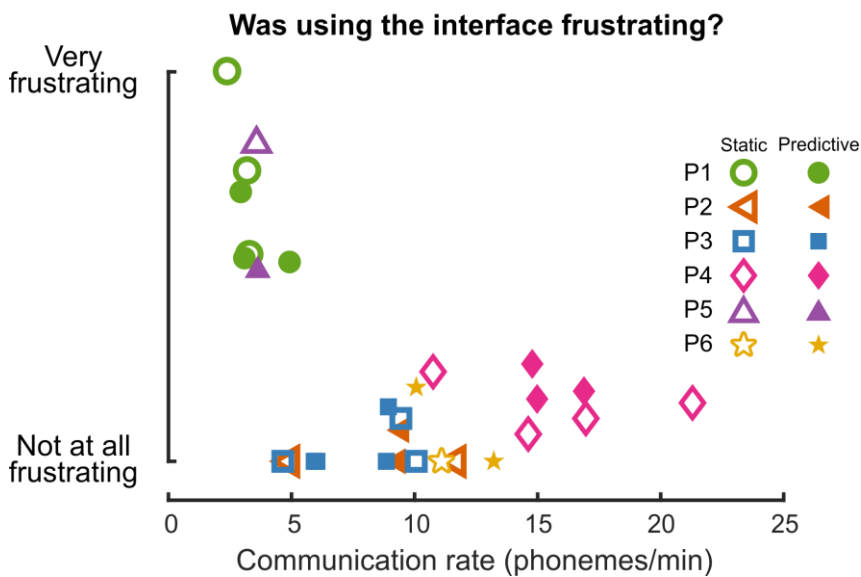
**Figure 4-3. Survey results for “Did using the interface feel fast?” from “Very slow” to “Very fast” as a function of communication rates**



**Figure 4-4. Survey results for “Could you find the targets easily?” from “Not easily” to “Very easily”, as a function of communication rates.**



**Figure 4-5. Survey results for “Performance: how successful were you in accomplishing what you were asked to do?” from “Perfect” to “Failure”, as a function of communication rates.**



**Figure 4-6. Survey results for “Was using the interface frustrating?” from “Not at all frustrating” to “Very frustrating”, as a function of communication rates.**

## 4.4 Discussion

Due to the heterogeneity of the participants in terms of capabilities, preferences, and access methods, a between-participant evaluation of optimization in participants with motor impairment was not possible. Within-participant assessments of prediction did enable us to gather participant preferences and reactions, as well as overall communication rates. Results of the experiment in participants without motor impairment in Chapter 3 suggested that prediction helped participants to learn the location and identity of various targets. However, the within-participant experimental design of the current study alternated blocks of predictive and static interfaces, meaning that any learning effects from prediction would likely carry over to the static blocks and thus be washed out. As a result, the trends in Figure 4-2 that show no block-to-block changes in communication rates with predictive interfaces versus static interfaces were expected.

### ***4.4.1 Comparison to participants without motor impairment***

Participants without motor impairment (Chapter 3) used the sEMG cursor to produce communication rates of 5.0 to 15.6 phonemes/min on the first day ( $M=9.5$ ;  $SD=2.2$ ). Participants with motor impairment used a variety of access methods to produce communication rates of 2.8 to 11.6 phonemes/min ( $M=8.4$ ;  $SD=4.2$ ) on their first day. Participants with motor impairments produced between 4 and 35 trials on their first day ( $M=22$ ;  $SD=10$ ), whereas participants without motor impairments produced between 10 and 35 trials on their first day. While

this suggests that participants without motor impairments were a reasonable model of participants with motor impairments on their first day, it is not clear whether their learning trajectories would be the same. For example, all participants without motor impairments were learning a new access method, whereas participants with motor impairments either used their daily access methods (P1, P3, P4, P6) or used new-to-them access methods due to the recency of their injuries (P2, P5).

#### **4.5 Future Design Considerations for Clinical Translation**

The results of this case series in individuals with motor impairment suggest that these interfaces are promising for clinical translation. Participants were able to use the interfaces with and without prediction to produce messages. Their subjective impressions were generally positive. However, based on their feedback, there are a variety of factors that should be investigated further and refined before full clinical trials or commercial transfer could occur.

##### **4.5.1 *Speech synthesis***

The speech synthesis used in these studies was implemented with a freely-available synthesizer, MBROLA (Dutoit et al., 1996), chosen for cost and the ease of integration with a phoneme-based input method (MBROLA is at base a phoneme synthesizer that can be fitted with extra text-to-phoneme layers for different languages). However, customized synthesizers could be built or modulated further for more intelligible and flexible synthesis. For example, in this implementation, we set each phoneme to have a target synthesis length of

100ms and a stable fundamental frequency (pitch). These could be modified to be phoneme-specific, as vowels and different categories of consonants have different habitual lengths. Further, the interface does not currently incorporate prosodic markers, but future implementations could offer a variety of pre-specified prosodic contours (e.g., rising pitch fitted over an utterance ended with a button press of a “?”; higher amplitude and shorter segments with a downward pitch inflection to indicate anger; Murray & Arnott, 1993). The interfaces could also be modified to indicate overall “stress” of particular phonemes, indicated by the user by selecting the same phoneme multiple times, and implemented by simultaneous increases in pitch, duration, and amplitude.

#### ***4.5.2 Voronoi diagram as predictive marker***

Displaying predicted targets via a Voronoi diagram was novel, and a variety of interface factors could be manipulated to optimize the display of the prediction. For example, the prediction weights were rescaled every time the interface was refreshed (i.e., after each selection) relative to the other likelihoods, rather than absolutely scaled. Thus at every time point, the most likely target had a prediction level of 1 and the least likely target had a prediction level of 0, with the other targets scaled in between. This means that the size of predicted targets was not consistent across trials. That is, a target that was actually  $\text{Pr}(.04)$  with all other targets at  $\text{Pr}(.025)$  was scaled the same as if a target were  $\text{Pr}(.8)$  with all other targets  $\text{Pr}(.005)$ . Within a trial, this suggested the likeliest target at any time. However, this method does not emphasize how certain the algorithm was

about the prediction. Future evaluation could compare these two different methods and determine any possible effect on learning and communication rates. In addition, the magnitude of the scaling (that is, how large predicted targets were allowed to get) was determined by trial and error, such that likely targets were obviously larger but not so large as to make smaller targets difficult to click with a noisy access method. Although participants with motor impairments largely agreed that the degree of scaling was appropriate, this could also be modulated and assessed to determine an ideal setting. An interface delivered to an end-user could easily have an adjustable setting to modulate the scaling. Finally, a variety of other markers could be used to denote prediction. Here, we simultaneously increased the font size as well as the target size, which conflates the effects of prediction based on visual search with those based on Fitts' law. Evaluations could be designed to tease apart these effects. Target colors used were chosen to be isoluminant, but an early version of the interface also modulated the brightness of the colors with prediction; this factor was dropped because we felt that it might be distracting, but this could also be tested empirically.

#### **4.5.3 Colors**

The targets in this interface were color-coded to denote rough phoneme category: simple vowels in green, complex vowels (diphthongs, r-colored vowels) in purple, fricatives and affricates in yellow, stops in red, and liquids, nasals, and semivowels in blue. This matches roughly how the iScan phonemic interface is

color-coded: groups are formed by manner of articulation and vowels in warm colors with consonants in cool colors (Trinh et al., 2012).

Color coding targets has a long history in AAC interfaces. The Fitzgerald key system was designed to give children with hearing impairment extra scaffolding on the parts of speech (e.g., diagramming sentences with different colors reflecting “who”, “what”, “where”, “when”, “verbs”, and “modifiers”). These have been widely extrapolated into AAC in order to provide additional visual cues as well as structured language training for users. However, recent work has suggested that color cues are not as beneficial as previously assumed, at least for children (Wilkinson & Snell, 2011). Thus further evaluation could assess the effectiveness of color coding the different phoneme groups.

#### **4.5.4 Phoneme Labels**

The phoneme labels utilized here were a standard set (Shoup, 1980; Weide, 2005). Previous phonemic interfaces have either provided their own set of letter/digraph labels (Cler et al., 2016; Cler, Nieto-Castanon, et al., 2014), or a set of pictures and/or custom digraphs (Black, 2011; Black et al., 2008; Schroeder, 2005; Trinh et al., 2012).

In this study, participants with and without motor impairments remarked on the labels and how they did not seem to represent the sounds they thought. For example, one participant with motor impairment remarked “It's easier when you don't know how it's spelled.” This likely had to do with the interference caused between knowing the orthographic representation of a given word and the

orthographic representation of the phonemes. The choice of labels may have influenced the results of the predictive versus static evaluations, particularly in participants with motor impairments. In order for prediction to be beneficial, users must trust its accuracy. This is one reason that using interfaces over time is necessary to determine the full benefits of prediction (Magnuson & Hunnicutt, 2002). In this case, users were explicitly instructed that the prediction was trying to help them by suggesting likely targets. Even so, some users were reluctant to trust it. For example, one participant remarked that the prediction let him in a direction he did not want; this could be due either to inaccurate prediction overall, inaccurate previous phonemic input leading to unhelpful predictions, or accurate prediction of targets that the user did not know were in fact the correct sounds. Another participant, when asked about what he thought about the prediction, remarked, "It didn't matter, because it wasn't the sound I was looking for, so I didn't use it." A further remark indicated both that the labels were confused and that the prediction did in fact prompt him to choose the enlarged sounds, however: "I don't remember what sound that makes....oh well, I'll pick it anyway."

Future implementations of this interface will not use the phoneme label set used here, but will offer the choice of several sets: IPA, researcher-defined, pictures as in Trinh et al. (2012) and Schroeder (2005), and the option for a particular user to input preferred labels for each sound. The labels could also include both a picture or digraph and an exemplar, as in Cler et al. (2016), in which vowel digraphs were presented with a relevant example: **bee**; **boot**; iy bye.

#### **4.5.5 Effectiveness of Optimization**

The computational results from Chapter 2 suggest that the optimized interface should show improvements of around 30% when optimizing and testing on the Suggested AAC corpus of messages as automatically translated into phonemes (as used here for optimizing and as stimuli for the empirical evaluation). The empirical evaluation in Chapter 3 showed an improvement of 10.5% in the individuals using the optimized/predictive interface over the random/predictive interface and 23.0% increase in communication rates in individuals using the optimized/static interface over the random/static interface. These suggest that the optimization is effective, but individuals are not reaching the optimal performance improvements suggested by computational estimates. This could be because the optimization process assumes that participants will use the dictionary transcriptions of the messages, and that is not the case. Even the calculations in Chapter 2 using different corpora still assume that users will select the correct phonemes. Once phonemic corpora are available, phonemic transitions could be recalculated and a better optimized interface could be produced. In addition, the percent improvement from the computational simulations is based on the assumption that communication rate is entirely determined by the motor action of moving from one phoneme to the next. It does not consider any additional time for cognitive processing or visual search. Both cognitive processing and visual search should be reduced based on further experience using the interfaces.

#### **4.5.6 Flexibility**

One of the possible benefits of a phonemic interface is that it allows users to produce any sequence of sounds that they like. In children without oral speech, phonemic interfaces have been suggested as a way to provide phonemic “babbling” of a sort, to increase phonological awareness and eventual literacy (Black, 2011; Black et al., 2008). In adults with motor impairment, we propose that the main benefits of phonemic interface are the possible communication rate increases. However, there is certainly a need in the AAC-using community for flexible, controllable speech synthesis. One of the participants with motor impairment remarked upon this strength: “I can really see how if someone didn't have any speech at all and sat down to learn these, you could really get it to say whatever you wanted.” One participant was a college student at the time of recording and had recently given a presentation in class. She noted that she had tried nearly a dozen different (free) speech synthesizers to get one that sounded reasonable to her, and she still needed to spend a lot of time editing her exact script so that the speech synthesis was smooth and accurate. This represents a remarkable amount of extra time required to complete a project, particularly for a person whose motor control makes text entry laborious. Speech synthesis can be awkward or inaccurate because it relies on convoluted text-to-speech rules. Perhaps if a phonemic interface were integrated with other speech synthesizers and interfaces, users could enter entire

speeches or at least out-of-dictionary words or phrases that would be produced exactly as they intended.

#### **4.5.7 Other Elements**

The interfaces used here were limited for experimental purposes. Future implementations would provide delete buttons, the ability to repeat a given selection, and user-specific settings as discussed previously (e.g., phoneme labels; the degree of scaling of predicted targets; preferred markers of prediction; toggle prediction on/off). A fully functional interface would also include the ability to input, store, and quickly retrieve commonly-used phrases. Finally, the prediction used here was simple character prediction implemented in a novel way via Voronoi diagrams and trained only on the set of 1004 suggested AAC messages that formed the stimuli set. Additional word-prediction could be implemented (Trinh et al., 2012), and both word- and character-prediction should be based on larger corpora. The data collected from participants without motor impairments (*Chapter 3*) comprises a large set of 20,000 messages that indicate how participants actually create phonemic messages, rather than how a dictionary transcription might suggest. These could be used to modify the dictionary-based transcriptions of large corpora used in phonemic prediction (Vertanen et al., 2012).

#### **4.5.8 Training system**

Participants were not given explicit phonemic training in these experimental paradigms, but instead were presented with a very brief introduction to the different sounds via a video and instructed to recreate messages as best they could. We chose this method in order to assess learning over time as if the user were getting no additional support, to give all users equal information (rather than, for example, answering questions that some participants thought to ask and some did not), as well as to assess whether the optimization or prediction also affected their ability to learn to use the phoneme set. However, this was somewhat frustrating for some participants. The participants without motor impairment were given two minutes per session to interact with the interfaces using their typical access modality (typical mouse) and were instructed to use this time to attempt to resolve any confusions they were having. However, participants still had difficulty differentiating some sounds. For example, [Y]/-j/ and [IY]/-i/ are very different sounds, but because the semivowel [Y]/-j/ is not intended to be selected without a vowel, the synthesizer produces it in isolation as something that is hard to perceptually code as a listener. Similarly, [AA]/-a/-father and [AO]/-ɔ/-ought are difficult to differentiate in some dialects of American English, and may be even harder to differentiate if distorted slightly through the synthesizer. Finally, [DH]/-ð/-“there” and [TH]/-θ/-“think” are easily perceptually differentiated by voicing, but participants likely do not know what the difference is when they hear them, nor do they know when to use which.

These confusions would easily be resolved with a slightly more regimented training program. This could involve existing phonics-based literacy training programs (e.g., Lloyd, 1992), or even simply instead of a video of all phonemes being clicked, an interactive interface where upon clicking, each sound played by itself and played an exemplar. Further, support by a speech-language pathologist could include when to choose voiced and voiceless cognates and provide the understanding that listeners will likely understand the output even if they make the incorrectly voiced selection.

It is promising that the participants with motor impairment noted that they were improving, even without explicit training beyond using the interface. For example, one participant said on the second day “I’m getting used to the sounds now.” Another participant noted, when asked to rate the interface’s usability, “The more you use it, the more usable it is.” He also noted the possibilities for training: “If I had this at home, I would go through every one of those sounds and I think you could get to where you could get pretty good speeds.”

#### **4.6 Conclusion**

The interfaces developed and evaluated in these studies show promise for clinical translation. Participant responses to the interfaces and to the prediction in particular were generally positive. Results of computational and empirical evaluation suggest that these interfaces may provide individuals with motor impairments a fast and flexible means of producing synthesized speech.

#### **4.7 Acknowledgments**

The authors would like to thank Tabatha Sorensen for helping to recruit participants. This work was supported by the National Institutes of Health - National Institute on Deafness and Other Communication Disorders under grant F31 DC014872 (GJC) and the National Science Foundation under grant 1452169 (CES).

## **Chapter 5. Conclusions and Future Directions**

### **5.1 Summary**

The ability to communicate is essential for unrestricted participation in all domains of human activity. Individuals with a variety of abilities are unable to rely on oral speech and thus use augmentative and alternative communication (AAC) strategies to communicate. However, communication remains slow for individuals who have restricted oral speech and limb movement for computer access, due in part to the access methods available to this population (e.g., head-tracking, eye-tracking, switch access). In this series of studies, we have taken the approach of optimizing communication interfaces in order to provide improvements in communication rates and thus quality of life.

First, we developed and optimized communication interfaces in which users select sounds (phonemes) instead of letters or whole words. The optimization was based on phoneme transition likelihoods (i.e., the probability of transitioning from one phoneme to another in a particular communication corpus), following previous research using letter-to-letter transition likelihoods to optimize orthographic interfaces. Using these likelihoods, computational simulations were used to calculate estimated interface efficiency based on the distance between targets, following Fitts' law. Regardless of the communication corpus used to optimize interfaces, optimization improved estimated efficiency by 20–30%.

Next, we added prediction to the optimized and random interfaces and tested the effects of these changes empirically. Prediction was implemented such that likely targets were dynamically enlarged following each selection by a user. The computational optimization and prediction were empirically assessed in 36 users without motor impairment using an alternate access method. Each user was assigned to one of four interfaces varying in layout and whether prediction was implemented (random/static; random/predictive; optimized/static; optimized/predictive) and participated in 12 sessions over a 3-week period. We found that prediction provided significantly faster communication rates during training (sessions 1–9), as users were learning the interface target locations and the novel access method. After training, optimization acted to significantly increase communication rates. The optimization likely became relevant only after training when participants knew the target locations and were able to move directly to the targets.

Finally, we completed a within-subject evaluation of the predictive and static interfaces in individuals with motor impairment and solicited their feedback on the interfaces. Both predictive and static interfaces had the optimized layout. Participants completed 3–7 blocks of trials (10 mins per block) over one or two sessions. All participants strongly agreed that they would improve with practice, and four out of six participants strongly preferred the interface with prediction over the static interface. Participants generally agreed that the targets enlarged

the right amount and that the prediction helped them to learn the location of targets.

Taken together, these studies suggest that optimized and predictive phonemic interfaces may provide increased communication rates for individuals with motor impairments affecting both oral communication and computer access. Methods for dynamically enlarging targets may also be applicable to other (non-phonemic) interfaces to increase communication rates. However, further research is needed to fully translate these results into clinical practice.

## **5.2 Future directions**

A variety of elements of the interfaces could be modulated and their effects could be empirically evaluated, including those discussed in Chapter 4. For example, the parameters of the Voronoi diagram could be modulated and empirically evaluated to determine the optimal setting for the extent of the enlargement. In addition, the singular and additive effects of other predictive markers could be evaluated, including color, brightness, and font size of the target labels. The target labels themselves were a source of confusion for the participants, and alternate sets should be evaluated empirically.

An additional important consideration is that of the applicability of the optimization and prediction methods for use with a variety of access methods. The Fitts' law optimization in Chapter 2 modulates the distance between targets and assumes that movements will be straight lines between sequential targets. This may not be the case when a user is first introduced to the interface and

must search for each target. Further, it is not clear that all access methods necessarily follow Fitts' law. Fitts' law has been used to characterize a variety of pointing devices: hand-held stylus (Fitts, 1954), joystick (Card et al., 1978; Epps, 1986), typical computer mouse (Card et al., 1978; Epps, 1986), head-controlled computer input devices (Radwin et al., 1990; Williams & Kirsch, 2008), and sEMG-based input devices (Choi & Kim, 2007; Vojtech et al., 2018; Williams & Kirsch, 2015). However, reports of eye-tracking and Fitts' law are mixed, with some groups reporting that eye-tracking does not follow Fitts' law (Sibert et al., 2001; Ware & Mikaelian, 1986; Zhai et al., 1999), and some groups reporting that it does (Miniotas, 2000; Vertegaal, 2008). We have also reported that the percent improvement in optimization should scale linearly with different access methods. This is only true if we make the common assumption that  $a=0$  in Fitts' law. This constant can represent any number of factors (Zhai, 2004), including aspects of the pointing device that do not vary with distance or target size (e.g., time to click a typical mouse; dwell time for the case of head- or eye-tracking). If dwell-time is used for selection, estimations of improvement using an optimized interface may be inaccurate and should be reconsidered before recommending an optimized phonemic interface as an option to increase communication rate. Alternate methods of assessing efficiency could be used with access methods that do not obey Fitts' law. For example, a phonemic interface could be optimized for switch access, in which the calculation depends not on the Euclidean distance but on the number of targets in the scan order between any two targets.

Future study is needed to determine how and when to introduce users to a phonemic interface. For example, we do not yet know which individuals have sufficient language capabilities to operate a phonemic interface. Some evidence suggests that individuals who grow up with no oral speech may have reduced phonological processing (Card & Dodd, 2006; Peeters, Verhoeven, de Moor, & van Balkom, 2009). Further, we would not be surprised if individuals with aphasia performed poorly with the interface due to impaired language affecting several aspects of the task (Blumstein, 1998; Caramazza, Berndt, & Basili, 1983). Thus the individuals likeliest to benefit from these optimized interfaces are those with acquired motor impairment that does not affect language, such as spinal cord injury. Other disorders with a primary motor impairment may or may not impact language (e.g., multiple sclerosis, amyotrophic lateral sclerosis, cerebral palsy). Other phonemic interfaces have been designed for various types of literacy or speech/language impairment, and these incorporate different numbers and types of targets and prediction based on the population. Empirical evaluation could determine which, if any, sets of targets and prediction are appropriate for different types of users.

Finally, the prediction we used in the evaluation in Chapter 3 was based purely on the dictionary translations of the stimuli set. Ideally, prediction would be based on an AAC-based corpus as well as an individual user's selections over time. Prediction could also be based on additional input modalities, such as real-time updates of a user's gaze.

APPENDIX

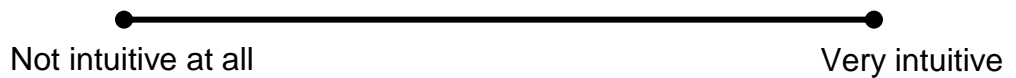
**New AAC interface follow-up**

*Asked to the person who uses AAC; answers indicated via typical communication modality or by a communication partner slowly moving from left to right with the user indicating where to make a mark*

How usable was this interface?



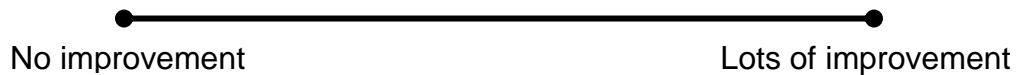
How intuitive was this interface?



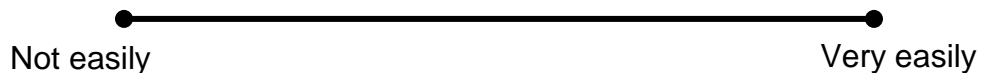
How easy was it to understand the layout?



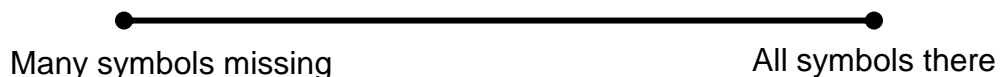
Do you think you could improve with practice?



Could you find the letters/phonemes you wanted easily?



Could you find all of the symbols you wanted to use?



Was using the interface frustrating?



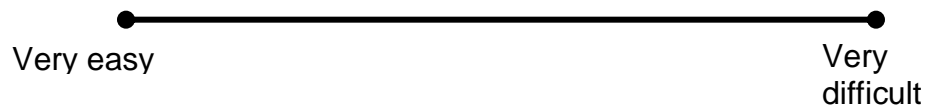
Did using the interface feel fast?



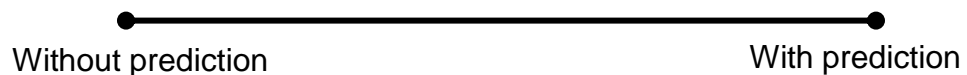
How difficult was it to learn to use your CURRENT method of communication?



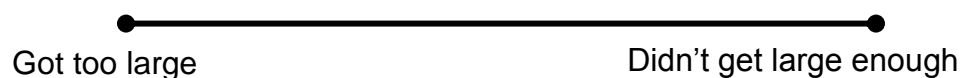
How difficult was it to learn to use this new method of communication?



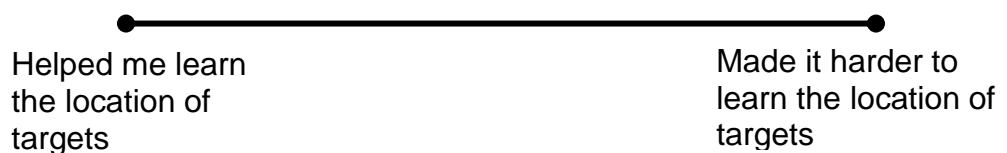
I preferred the interface:



I thought the enlarged targets:



I thought the enlarged targets:



## BIBLIOGRAPHY

- Arnott, J., & Javed, M. (1992). Probabilistic character disambiguation for reduced keyboards using small text samples. *Augmentative and Alternative Communication*, 8(3), 215–223.  
<https://doi.org/10.1080/07434619212331276203>
- Axon, W. E. A. (1888). Shorthand Literature. *The Academy*, 21(846), 1869–1902. Retrieved from  
<https://search.proquest.com/docview/8245452/fulltextPDF/E5175277FDA439BPQ/1?accountid=9676>
- Baljko, M., & Tam, A. (2006). Indirect text entry using one or two keys. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility - Assets '06* (p. 18). Portland, Oregon: ACM Press. <https://doi.org/10.1145/1168987.1168992>
- Basmajian, J. (1972). Electromyography comes of age. *Science*, 176(4035), 603–609. <https://doi.org/10.1126/science.176.4035.603>
- Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bates, R., & Istance, H. O. (2003). Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices. *Universal Access in the Information Society*, 2(3), 280–290. <https://doi.org/10.1007/s10209-003-0053-y>
- Beddoes, M. P., & Zhongzhi Hu. (1994). A chord stenograph keyboard: a possible solution to the learning problem in stenography. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(7), 953–960. <https://doi.org/10.1109/21.297785>
- Beichl, I., & Sullivan, F. (2000). The Metropolis algorithm. *Computing in Science & Engineering*, 2(1), 65–69.
- Beukelman, D., & Ansel, B. (1995). Research priorities in augmentative and alternative communication. *Augmentative and Alternative Communication*, 11(2), 131–134. <https://doi.org/10.1080/07434619512331277229>
- Beukelman, D., & Gutmann, M. (1999). Generic message list for AAC users with ALS. Retrieved from [http://aac.unl.edu/ALS\\_Message\\_List1.htm](http://aac.unl.edu/ALS_Message_List1.htm)
- Beukelman, D. R., Fager, S., Ball, L., & Dietz, A. (2007). AAC for adults with

- acquired neurological conditions: A review. *Augmentative and Alternative Communication*, 23(3), 230–242.  
<https://doi.org/10.1080/07434610701553668>
- Beukelman, D. R., & Mirenda, P. (2013). *Augmentative and Alternative Communication: Supporting Children and Adults with Complex Communication Needs (Fourth Edition)*. In *Baltimore, MD: Paul H. Brookes*.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly. <https://doi.org/10.1017/CBO9781107415324.004>
- Black, R. (2011). The PhonicStick: A joystick to generate novel words using phonics. In *13th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '11* (p. 325). Dundee, Scotland, UK: ACM Press. <https://doi.org/10.1145/2049536.2049630>
- Black, R., Waller, A., Pullin, G., & Abel, E. (2008). Introducing the PhonicStick: Preliminary evaluation with seven children. In *13th Biennial Conference of the International Society for Augmentative and Alternative Communication*. Montreal, Canada.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., & Curio, G. (2007). The non-invasive Berlin Brain–Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2), 539–550.  
<https://doi.org/10.1016/J.NEUROIMAGE.2007.01.051>
- Bloomberg, K., Karlan, G. R., & Lloyd, L. L. (1990). The Comparative Translucency of Initial Lexical Items Represented in Five Graphic Symbol Systems and Sets. *Journal of Speech Language and Hearing Research*, 33(4), 717. <https://doi.org/10.1044/jshr.3304.717>
- Blumstein, S. E. (1998). Phonological Aspects of Aphasia. In M. T. Sarno (Ed.), *Acquired Aphasia* (Third, pp. 157–185). <https://doi.org/10.1016/B978-012619322-0/50021-X>
- Brault, M. W. (2012). *Americans with disabilities: 2010. Current Population Reports*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau. Retrieved from <http://www.census.gov/prod/2012pubs/p70-131.pdf>
- Bristow, D., & Fristoe, M. (1984). Learning of Blissymbols and Manual Signs. *Journal of Speech and Hearing Disorders*, 49(2), 145.  
<https://doi.org/10.1044/jshd.4902.145>
- Brumberg, J. S., Nieto-Castanon, A., Kennedy, P. R., & Guenther, F. H. (2010).

- Brain-computer interfaces for speech communication. *Speech Communication*, 52(4), 367–379.  
<https://doi.org/10.1016/j.specom.2010.01.001>
- Caramazza, A., Berndt, R. S., & Basili, A. G. (1983). The selective impairment of phonological processing: A case study. *Brain and Language*, 18(1), 128–174. [https://doi.org/10.1016/0093-934X\(83\)90011-1](https://doi.org/10.1016/0093-934X(83)90011-1)
- Card, R., & Dodd, B. (2006). The phonological awareness abilities of children with cerebral palsy who do not speak. *Augmentative and Alternative Communication*, 22(3), 149–159.  
<https://doi.org/10.1080/07434610500431694>
- Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics*, 21(8), 601–613.  
<https://doi.org/10.1080/00140137808931762>
- Choi, C., & Kim, J. (2007). A real-time EMG-based assistive computer interface for the upper limb disabled. *2007 IEEE 10th International Conference on Rehabilitation Robotics, ICORR'07, 00(c)*, 459–462.  
<https://doi.org/10.1109/ICORR.2007.4428465>
- Choi, C., Rim, B. C., & Kim, J. (2011). Development and evaluation of a assistive computer interface by SEMG for individuals with spinal cord injuries. In *IEEE International Conference on Rehabilitation Robotics*.  
<https://doi.org/10.1109/ICORR.2011.5975386>
- Chubon, R. A., & Hester, M. R. (1988). An enhanced standard computer keyboard system for single-finger and typing-stick typing. *Journal of Rehabilitation Research and Development*, 25(4), 17–24. Retrieved from <https://pdfs.semanticscholar.org/7f70/c379330c94670518ddb907668b2d957933dc.pdf>
- Cler, G. J., & Stepp, C. E. (2017). Development and theoretical evaluation of optimized phonemic interfaces. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '17* (pp. 230–239). Baltimore, MD: ACM Press.  
<https://doi.org/10.1145/3132525.3132537>
- Cler, M. J., Michener, C., & Stepp, C. E. (2014). Discrete vs. continuous surface electromyographic interface control. In *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 26 – 30 August* (Vol. 2014).  
<https://doi.org/10.1109/EMBC.2014.6944593>

- Cler, M. J., Nieto-Castañón, A., Guenther, F. H., Fager, S. K., & Stepp, C. E. (2016). Surface electromyographic control of a novel phonemic interface for speech synthesis. *Augmentative and Alternative Communication*, 32(2), 120–130. <https://doi.org/10.3109/07434618.2016.1170205>
- Cler, M. J., Nieto-Castanon, A., Guenther, F. H., & Stepp, C. E. (2014). Surface electromyographic control of speech synthesis. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014* (pp. 5848–5851). Chicago, IL. <https://doi.org/10.1109/EMBC.2014.6944958>
- Cler, M. J., & Stepp, C. E. (2015). Discrete versus continuous mapping of facial electromyography for human-machine-interface control: Performance and training effects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(4), 572–580. <https://doi.org/10.1109/TNSRE.2015.2391054>
- Copestake, A. (1997). Augmented and alternative NLP techniques for augmentative and alternative communication. In *Proceedings of the ACL workshop on Natural Language Processing for Communication Aids* (pp. 37–42). Madrid, Spain. Retrieved from <http://www.aclweb.org/anthology/W97-0506>
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Fourth International Conference on Spoken Language Processing* (pp. 1393–1396). Philadelphia, PA. Retrieved from <http://ai2-s2-pdfs.s3.amazonaws.com/7b1f/dadf05b8f968a5b361f6f82852ade62c8010.pdf>
- Dvorak, A., & Dealey, W. L. (1932). *US2040248*. United States. Retrieved from <https://www.google.com/patents/US2040248?dq=2040248>
- Epps, B. W. (1986). Comparison of Six Cursor Control Devices Based on Fitts' Law Models. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 30, pp. 327–331). <https://doi.org/10.1177/154193128603000403>
- Fager, S., Beukelman, D. R., Fried-Oken, M., Jakobs, T., & Baker, J. (2012). Access Interface Strategies. *Assistive Technology*, 24(1), 25–33. <https://doi.org/10.1080/10400435.2011.648712>
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391.

- Frey, L. A., White, K. P., & Hutchinson, T. E. (1990). Eye-Gaze Word Processing. *IEEE Transactions on Systems, Man and Cybernetics*, 20(4), 944–950. <https://doi.org/10.1109/21.105094>
- Frick, V., Girouard, K., Shiakolas, E., Cler, G. J., Fager, S. K., & Stepp, C. E. (2017). Design of an electromyographic switch for communication systems. In *Boston Speech Motor Control Mini-Symposium*. Boston, MA.
- Fuller, D., & Lloyd, L. (1991). Toward a common usage of iconicity terminology. *Augmentative and Alternative Communication*, 7(3), 215–220. <https://doi.org/10.1080/07434619112331275913>
- Fuller, D., Lloyd, L., & Schlosser, R. (1992). Further development of an Augmentative and Alternative Communication symbol taxonomy. *Augmentative and Alternative Communication*, 8(1), 67–74. <https://doi.org/10.1080/07434619212331276053>
- Garay-Vitoria, N., & Abascal, J. (2006). Text prediction systems: a survey. *Universal Access in the Information Society*, 4, 188–203. <https://doi.org/10.1007/s10209-005-0005-9>
- Gentner, D. R., Grudin, J., & Conway, E. (1980). *Skilled finger movements in typing*. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a085985.pdf>
- Getschow, C. O., Rosen, M. J., & Goodenough-Trepagnier, C. (1986). A systematic approach to design a minimum distance alphabetical keyboard. In *RESNA (Rehabilitation Engineering Society of North America) 9th Annual Conference* (pp. 396–398). Minneapolis, Minnesota.
- Glennen, S., & DeCoste, D. (1997). *The Handbook of Augmentative and Alternative Communication*. San Diego: Singular. Retrieved from [http://www.nwoet.org/oatdlp2/augcom/documents/chap3/chap3\\_sec1.pdf](http://www.nwoet.org/oatdlp2/augcom/documents/chap3/chap3_sec1.pdf)
- Goodenough-Trepagnier, C., & Prather, P. (1981). Communication Systems for the Nonvocal Based on Frequent Phoneme Sequences. *Journal of Speech Language and Hearing Research*, 24(3), 322. <https://doi.org/10.1044/jshr.2403.322>
- Goodenough-Trepagnier, C., Tarry, E., & Prather, P. (1982). Derivation of an Efficient Nonvocal Communication System. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 24(2), 163–172. <https://doi.org/10.1177/001872088202400202>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in*

- Psychology*, 52, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Heckathorne, C., Voda, J., & Leibowitz, L. (1987). Design rationale and evaluation of the Portable Anticipatory Communication Aid—PACA. *Augmentative and Alternative Communication*, 3(4), 170–180. <https://doi.org/10.1080/07434618712331274489>
- Higginbotham, D. J., Shane, H., Russell, S., & Caves, K. (2007). Access to AAC: present, past, and future. *Augmentative and Alternative Communication (Baltimore, Md. : 1985)*, 23(3), 243–257. <https://doi.org/10.1080/07434610701571058>
- Horstmann, H. M., & Levine, S. P. (1991). The effectiveness of word prediction. In *Proceedings of the 14th Annual RESNA Conference* (pp. 100–102). Kansas City, MO.
- Huo, X., Wang, J., & Ghovanloo, M. (2008). Introduction and preliminary evaluation of the Tongue Drive System: Wireless tongue-operated assistive technology for people with little or no upper-limb function. *Journal of Rehabilitation Research & Development*, 45(6), 921–930. <https://doi.org/10.1682/JRRD.2007.06.0096>
- Hurlbut, B. I., Iwata, B. A., & Green, J. D. (1982). Nonvocal Language Acquisition In Adolescents With Severe Physical Disabilities: Blissymbol Versus Iconic Stimulus Formats. *Journal of Applied Behavior Analysis*, 15, 241–258. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1308268/pdf/jaba00040-0053.pdf>
- International Phonetic Association. (1999). *Handbook of the International Phonetic Association*. Cambridge Univ Press. <https://doi.org/10.1017/S0025100311000089>
- Koester, H. H., & Arthanat, S. (2017, April 3). Text entry rate of access interfaces used by people with physical disabilities: A systematic review. *Assistive Technology*, pp. 1–13. <https://doi.org/10.1080/10400435.2017.1291544>
- Kreifeldt, J. G., Levine, S. L., & Iyengar, C. (1989). Reduced Keyboard Designs Using Disambiguation. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 33, pp. 441–444). SAGE PublicationsSage CA: Los Angeles, CA. <https://doi.org/10.1518/107118189786759642>
- Kushler, C. (1998). AAC: Using a Reduced Keyboard. In *CSUN Conference on Technology for persons with disabilities*. California State University, Northridge CA.

- Langolf, G. D., Chaffin, D. B., & Foulke, J. A. (1976). An Investigation of Fitts' Law Using a Wide Range of Movement Amplitudes. *Journal of Motor Behavior*, 8(2), 113–128. <https://doi.org/10.1080/00222895.1976.10735061>
- Lefcheck, J. S. (2015). piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, 7, 573–579. <https://doi.org/10.1111/2041-210X.12512>
- Leshner, G. W., Moulton, B. J., & Higginbotham, D. J. (1998a). Optimal character arrangements for ambiguous keyboards. *IEEE Transactions on Rehabilitation Engineering*, 6(4), 415–423. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9865889>
- Leshner, G. W., Moulton, B. J., & Higginbotham, D. J. (1998b). Techniques for augmenting scanning communication. *AAC: Augmentative and Alternative Communication*, 14(2), 81–101. <https://doi.org/10.1080/07434619812331278236>
- Levine, S. H., & Goodenough-Trepagnier, C. (1990). Customised text entry devices for motor-impaired users. *Applied Ergonomics*, 21(1), 55–62. [https://doi.org/10.1016/0003-6870\(90\)90074-8](https://doi.org/10.1016/0003-6870(90)90074-8)
- Lewis, J. R., Kennedy, P. J., & LaLomia, M. J. (1999). Development of a Digram-Based Typing Key Layout for Single-Finger/Stylus Input. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(5), 415–419. <https://doi.org/10.1177/154193129904300505>
- Lewis, J. R., LaLomia, M. J., & Kennedy, P. J. (1999). Evaluation of typing key layouts for stylus input. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(5), 420–424. <https://doi.org/10.1177/154193129904300506>
- Liu, S. S., Rawicz, A., Rezaei, S., Ma, T., Zhang, C., Lin, K., & Wu, E. (2012). An eye-gaze tracking and human computer interface system for people with ALS and other locked-in diseases. *Journal of Medical and Biological Engineering*, 32(2), 37–42.
- Lloyd, S. (1992). *The Phonics Handbook* (Third). Essex, United Kingdom: Jolly Learning Ltd.
- MacKenzie, I. S. (1992). Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction*, 7(1), 91–139. [https://doi.org/10.1207/s15327051hci0701\\_3](https://doi.org/10.1207/s15327051hci0701_3)
- MacKenzie, I. S., & Zhang, S. X. (1999). The design and evaluation of a high-

- performance soft keyboard. *CHI 99 Conference on Human Factors in Computing Systems*.
- Maclay, H., & Osgood, C. E. (1959). Hesitation Phenomena in Spontaneous English Speech. *WORD*, *15*(1), 19–44.  
<https://doi.org/10.1080/00437956.1959.11659682>
- Magnien, L., Bouraoui, J. L., & Vigouroux, N. (2004). Mobile text input with soft keyboards: optimization by means of visual clues. In *Mobile Human-Computer Interaction-MobileHCI 2004* (pp. 337–341). Springer.
- Magnuson, T., & Hunnicutt, S. (2002). Measuring the effectiveness of word prediction: The advantage of long-term use. *TMH-QPSR*, *43*(1), 57–67.
- Majaranta, P., MacKenzie, I. S., Aula, A., & Rähkä, K.-J. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Universal Access in the Information Society*, *5*(2), 199–208.
- Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, *Vol 1*(2), 13–32.
- McGuffin, M. J., & Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(4), 388–422.
- Merlin, B., & Raynal, M. (2010). Evaluation of SpreadKey system with motor impaired users. In *ICCHP 2010: 12th International Conference on Computers Helping People with Special Needs* (Vol. 6180 LNCS, pp. 112–119). Vienna, Austria, Austria: Springer, Berlin, Heidelberg.  
[https://doi.org/10.1007/978-3-642-14100-3\\_18](https://doi.org/10.1007/978-3-642-14100-3_18)
- Miall, D. S. (2001). Sounds of contrast: An empirical approach to phonemic iconicity. *Poetics*, *29*, 55–70. Retrieved from [www.elsevier.nl/locate/poetic](http://www.elsevier.nl/locate/poetic)
- Miniotas, D. (2000). Application of Fitts' law to eye gaze interaction. In *CHI '00 extended abstracts on Human factors in computing systems - CHI '00* (p. 339). The Hague, The Netherlands: ACM Press.  
<https://doi.org/10.1145/633292.633496>
- Mizuko, M. (1987). Transparency and Ease of Learning of Symbols Represented by Blissymbols, PCS, and Picsyms.  
<https://doi.org/10.1080/07434618712331274409>
- Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic

- speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2), 1097–1108.  
<https://doi.org/10.1121/1.405559>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Newell, A., Langer, S., & Hickey, M. (1998). The rôle of natural language processing in alternative and augmentative communication. *Natural Language Engineering*, 4(1), 1–16.
- Nijboer, F., Sellers, E. W., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., ... Kübler, A. (2008). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119(8), 1909–1916. <https://doi.org/10.1016/j.clinph.2008.03.034>
- Noyes, J. (1983). The QWERTY keyboard: a review. *International Journal of Man-Machine Studies*, 18(3), 265–281. [https://doi.org/10.1016/S0020-7373\(83\)80010-8](https://doi.org/10.1016/S0020-7373(83)80010-8)
- Osser, H., & Peng, F. (1964). A cross cultural study of speech rate. *Language and Speech*, 7(2), 120–125. <https://doi.org/10.1177/002383096400700208>
- Peeters, M., Verhoeven, L., de Moor, J., & van Balkom, H. (2009). Importance of speech production for phonological awareness and word decoding: The case of children with cerebral palsy. *Research in Developmental Disabilities*, 30(4), 712–726. <https://doi.org/10.1016/J.RIDD.2008.10.002>
- Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication*, 45, 89–95.
- Plamondon, R., & Alimi, A. M. (1997). Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*.  
<https://doi.org/10.1017/S0140525X97001441>
- Pouplin, S., Robertson, J., Antoine, J.-Y., Blanchet, A., Kahloun, J. L., Volle, P., ... Bensmail, D. (2014). Effect of dynamic keyboard and word-prediction systems on text input speed in persons with functional tetraplegia. *Journal of Rehabilitation Research and Development*, 51(3), 467–480.  
<https://doi.org/10.1682/JRRD.2012.05.0094>
- R Core team. (2015). R Core Team. *R: A Language and Environment for*

*Statistical Computing*. Vienna, Austria, Austria: R Foundation for Statistical Computing.

Radwin, R. G., Vanderheiden, G. C., & Lin, M. L. (1990). A method for evaluating head-controlled computer input devices using Fitts' law. *Human Factors*, 32(4), 423–438. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2150065>

Regard, M., Landis, T., & Hess, K. (1985). Preserved Stenography Reading in a Patient With Pure Alexia. *Archives of Neurology*, 42(4), 400. <https://doi.org/10.1001/archneur.1985.04060040114026>

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a Skilled Typist: A Study of Skilled Cognitive-Motor Performance. *Cognitive Science*, 6, 1–36. Retrieved from [https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0601\\_1](https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0601_1)

Rycroft, C. H. (2009). VORO++: A three-dimensional Voronoi cell library in C++. *Chaos*. <https://doi.org/10.1063/1.3215722>

Saxena, S., Nikolić, S., & Popović, D. (1995). An EMG-controlled grasping system for tetraplegics. *Journal of Rehabilitation Research and Development*, 32(1), 17–24. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7760263>

Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology*, 5, 80. <https://doi.org/10.3389/fpsyg.2014.00080>

Schroeder, J. E. (2005). Improved spelling for persons with learning disabilities. In *20th Annual International Conference on Technology and Persons with Disabilities*. Northridge, CA.

Sears, A., Jacko, J. A., Chu, J., & Moro, F. (2001). The role of visual search in the design of effective soft keyboards. *Behaviour & Information Technology*, 20(3), 159–166.

Sellers, E. W., Krusienski, D. J., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2006). A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. *Biological Psychology*, 73(3), 242–252. <https://doi.org/10.1016/j.biopsycho.2006.04.007>

Sevcik, R. A., & Ronski, M. B. T.-A. L. (2000). AAC: More Than Three Decades of Growth and Development, 5(19), 5. Retrieved from [http://libraries.state.ma.us/login?gwurl=http://go.galegroup.com/ps/i.do?p=GPS&sw=w&u=mlln\\_b\\_bumml&v=2.1&it=r&id=GALE%7CA66665379&asid=9](http://libraries.state.ma.us/login?gwurl=http://go.galegroup.com/ps/i.do?p=GPS&sw=w&u=mlln_b_bumml&v=2.1&it=r&id=GALE%7CA66665379&asid=9)

63c77cdf2607ef72e1b346d1b7771dd

- Shoup, J. (1980). Phonological Aspects of Speech Recognition. In W. A. Lea (Ed.), *Trends in speech recognition* (pp. 125–138). New York: Prentice-Hall.
- Shrum, L. J., Lowrey, T. M., Luna, D., Lerman, D. B., & Liu, M. (2012). Sound symbolism effects across languages: Implications for global brand names. *International Journal of Research in Marketing*, 29(3), 275–279. <https://doi.org/10.1016/J.IJRESMAR.2012.03.002>
- Sibert, L. E., Templeman, J. N., & Jacob, R. J. K. (2001). *Evaluation and Analysis of Eye Gaze Interaction*. Washington, DC. Retrieved from <http://www.dtic.mil/docs/citations/ADA393229>
- Smith, A. L., & Chaparro, B. S. (2015). Smartphone Text Input Method Performance, Usability, and Preference With Younger and Older Adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(6), 1015–1028. <https://doi.org/10.1177/0018720815575644>
- Smith, B. A., & Zhai, S. (2001). Optimised Virtual Keyboards with and without Alphabetical Ordering – A Novice User Study. In *INTERACT 2001 – IFIP TC13 International Conference on Human-Computer Interaction* (pp. 92–99). Tokyo, Japan.
- Soukoreff, R. W., & MacKenzie, I. S. (2001). Measuring errors in text entry tasks. In *ACM SIGCHI - human-computer interaction* (p. 319). New York, USA: ACM Press. <https://doi.org/10.1145/634067.634256>
- Textware Solutions. (1998). The Fitaly one-finger keyboard. Retrieved from <http://fitaly.com/fitaly/fitaly.htm>
- Thistle, J. J., & Wilkinson, K. M. (2013). Working Memory Demands of Aided Augmentative and Alternative Communication for Individuals with Developmental Disabilities. *Augmentative and Alternative Communication*, 29(3), 235–245. <https://doi.org/10.3109/07434618.2013.815800>
- Trinh, H. (2011). Using a computer intervention to support phonological awareness development of nonspeaking adults. In *13th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '11* (pp. 329–330). Dundee, Scotland, UK: ACM Press. <https://doi.org/10.1145/2049536.2049632>
- Trinh, H., Waller, A., Vertanen, K., Kristensson, P. O., & Hanson, V. L. (2012). iSCAN: A Phoneme-based Predictive Communication Aid for Nonspeaking Individuals. *ASSETS'12*. Boulder, Colorado.

- Trnka, K., Mccaw, J., Yarrington, D., Mccoy, K. F., & Pennington, C. (2009). User Interaction with Word Prediction : The Effects of Prediction Quality. *ACM Transactions on Accessible Computing*, 1(3), 1–34. <https://doi.org/10.1145/1497302.1497307>.http
- Trnka, K., & McCoy, K. F. (2007). Corpus studies in word prediction. *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility - Assets '07*, 195. <https://doi.org/10.1145/1296843.1296877>
- Trnka, K., Yarrington, D., McCaw, J., McCoy, K. F., & Pennington, C. (2007). The effects of word prediction on communication rate for AAC. In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers, (NAACL-HLT-2007)* (pp. 173–176). Rochester, NY. <https://doi.org/10.3115/1614108.1614152>
- Tuisku, O., Majaranta, P., Isokoski, P., & Rähkä, K.-J. (2008). Now Dasher! Dash away!: longitudinal study of fast text entry by Eye Gaze. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 19–26). ACM.
- University of Nebraska-Lincoln. (n.d.-a). Context Specific Messages (Suggested by AAC Specialists). Retrieved from <http://cehs.unl.edu/aac/aac-messaging-and-vocabulary/>
- University of Nebraska-Lincoln. (n.d.-b). Unabridged Vocabulary Lists with Use Statistics - AAC User. Retrieved from <http://cehs.unl.edu/aac/aac-messaging-and-vocabulary/>
- Vanderheiden, G. C. (2002). A journey through early augmentative communication and computer access. *Journal of Rehabilitation Research and Development*, 39(6), 39–54. Retrieved from <http://0-web.ebscohost.com/impulse.ucdenver.edu/ehost/pdfviewer/pdfviewer?sid=ab92ae0c-0618-45cf-a3a6-e83eceed713e%40sessionmgr112&vid=1&hid=108>
- Velichkovsky, B., Sprenger, A., & Unema, P. (1997). Towards gaze-mediated interaction: Collecting solutions of the “Midas touch problem.” In *Human-Computer Interaction INTERACT '97* (pp. 509–516). Boston, MA: Springer US. [https://doi.org/10.1007/978-0-387-35175-9\\_77](https://doi.org/10.1007/978-0-387-35175-9_77)
- Venkatagiri, H. (1993). Efficiency of lexical prediction as a communication acceleration technique. *Augmentative and Alternative Communication*, 9(3), 161–167. <https://doi.org/10.1080/07434619312331276561>

- Vernon, S., & Joshi, S. S. (2011). Brain-muscle-computer interface: Mobile-phone prototype development and testing. *IEEE Transactions on Information Technology in Biomedicine*, 15(4), 531–538. <https://doi.org/10.1109/TITB.2011.2153208>
- Vertanen, K., & Kristensson, P. O. (2011). The imagination of crowds: conversational AAC language modeling using crowdsourcing and large data sources. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 700–711). Edinburgh, United Kingdom: Association for Computational Linguistics.
- Vertanen, K., Trinh, H., Waller, A., Hanson, V. L., & Kristensson, P. O. (2012). Applying prediction techniques to phoneme-based AAC systems. *NAACL-HLT 2012 Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*. Montreal, Canada.
- Vertegaal, R. (2008). A Fitts Law comparison of eye tracking and manual input in the selection of visual targets. In *10th international conference on Multimodal interfaces - IMCI '08* (pp. 241–248). Chania, Crete, Greece: ACM Press. <https://doi.org/10.1145/1452392.1452443>
- Vojtech, J. M., Cler, G. J., Fager, S. K., & Stepp, C. E. (2018). Predicting optimal augmentative and alternative communication device control in individuals with motor speech disorders using surface electromyography. In *Conference on Motor Speech*. Savannah, GA.
- Wandmacher, T., & Antoine, J.-Y. (2006). Training language models without appropriate language resources: Experiments with an AAC system for disabled people. In *Proceedings of LREC*.
- Ward, D. J., & MacKay, D. J. C. (2002). Fast hands-free writing by gaze direction. *Nature*, 418(6900), 838. <https://doi.org/10.1038/418838a>
- Ware, C., & Mikaelian, H. H. (1986). An evaluation of an eye tracker as a device for computer input. *CHI '87 Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, 17(SI), 183–188. <https://doi.org/10.1145/30851.275627>
- Weide, R. (2005). The Carnegie Mellon Pronouncing Dictionary [cmudict. 0.6]. Retrieved from <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>
- West, L. J. (1998). *The Standard and Dvorak Keyboards Revisited: Direct Measures of Speed*. Sante Fe Institute Working Papers. Sante Fe, CA, CA. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.8.6886&rep=rep1>

&type=pdf

- Wilkinson, K. M., & Snell, J. (2011). Facilitating children's ability to distinguish symbols for emotions: the effects of background color cues and spatial arrangement of symbols on accuracy and speed of search. *American Journal of Speech-Language Pathology*, *20*(4), 288–301. [https://doi.org/10.1044/1058-0360\(2011/10-0065\)](https://doi.org/10.1044/1058-0360(2011/10-0065))
- Williams, M. R., & Kirsch, R. F. (2008). Evaluation of head orientation and neck muscle EMG signals as command inputs to a human-computer interface for individuals with high tetraplegia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *16*(5), 485–496. <https://doi.org/10.1109/TNSRE.2008.2006216>
- Williams, M. R., & Kirsch, R. F. (2015). Evaluation of head orientation and neck muscle EMG signals as three-dimensional command sources. *Journal of NeuroEngineering and Rehabilitation*, *12*(1). <https://doi.org/10.1186/s12984-015-0016-6>
- Williams, M. R., & Kirsch, R. F. (2016). Case study: Head orientation and neck electromyography for cursor control in persons with high cervical tetraplegia. *Journal of Rehabilitation Research and Development*, *53*(4), 519–530. <https://doi.org/10.1682/JRRD.2014.10.0244>
- Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., ... Vaughan, T. M. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 164–173. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10896178>
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, *113*(6), 767–791. [https://doi.org/10.1016/S1388-2457\(02\)00057-3](https://doi.org/10.1016/S1388-2457(02)00057-3)
- Zhai, S. (2004). Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing. *International Journal of Human-Computer Studies*, *61*(6), 791–809. <https://doi.org/10.1016/J.IJHCS.2004.09.006>
- Zhai, S., Conversy, S., Beaudouin-Lafon, M., & Guiard, Y. (2003). Human on-line response to target expansion. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 177–184). ACM.
- Zhai, S., Hunter, M., & Smith, B. A. (2002). Performance optimization of virtual

keyboards. *Human-Computer Interaction*, 17(2–3), 229–269.  
<https://doi.org/10.1080/07370024.2002.9667315>

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99* (pp. 246–253). Pittsburgh, Pennsylvania, USA: ACM Press.  
<https://doi.org/10.1145/302979.303053>

## CURRICULUM VITAE

