JSLHR

Research Article

Optimized and Predictive Phonemic Interfaces for Augmentative and Alternative Communication

Gabriel J. Cler,[a,b] Katharine R. Kolin,[b] Jacob P. Noordzij Jr.,[b,c] Jennifer M. Vojtech,[b,c] Susan K. Fager,[d] and Cara E. Stepp[a,b,c,e]

Purpose: We empirically assessed the results of computational optimization and prediction in communication interfaces that were designed to allow individuals with severe motor speech disorders to select phonemes and 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 proximity. Target sizes were either static or predictive, such that likely targets were dynamically enlarged following each selection. Communication interfaces were evaluated by 36 users without motor impairments using an alternate access method. Each user was assigned to 1 of 4 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. Six participants with severe motor impairments used both the optimized/static and optimized/predictive interfaces in 1–2 sessions.
Results: In individuals without motor impairments, 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. Participants with motor impairments could use the interfaces with alternate access methods and generally rated the interface with prediction as preferred.
Conclusions: Optimization and prediction led to increases in communication rates in users without motor impairments. Predictive interfaces were preferred by users with motor impairments. Future research is needed to translate these results into clinical practice.
Supplemental Material: https://doi.org/10.23641/asha.8636948

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: two to 15 words per minute (wpm), compared to 30–40 wpm by a skilled typist and 150–200 wpm in typical speech (Beukelman & Mirenda, 2013; Copestake, 1997; Higginbotham, Shane, Russell, & Caves, 2007; Lesher, Moulton, & Higginbotham, 1998). These rates are slow partially due to the motor impairments these individuals exhibit requiring the use of alternative access. The design of the communication interface presents another barrier to achieving faster communication rates. Opportunities exist to research and

[a]Graduate Program for Neuroscience–Computational Neuroscience, Boston University, MA
[b]Department of Speech, Language, and Hearing Sciences, Boston University, MA
[c]Department of Biomedical Engineering, Boston University, MA
[d]Institute for Rehabilitation Science and Engineering, Madonna Rehabilitation Hospital, Lincoln, NE
[e]Department of Otolaryngology—Head and Neck Surgery, Boston University School of Medicine, MA
Correspondence to Gabriel J. Cler: gcler@bu.edu

Editor-in-Chief: Michael Hammer
Editor: Torrey Loucks
Received May 17, 2018
Revision received December 6, 2018
Accepted March 15, 2019
https://doi.org/10.1044/2019_JSLHR-S-MSC18-18-0187
Publisher Note: This article is part of the Forum: Selected Papers From the 2018 Conference on Motor Speech—Basic Science and Clinical Innovation.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

develop new interface options that demonstrate potential to increase rate and efficiency of message construction for individuals with severe motor impairments. This article describes the preliminary investigation of a new AAC interface that integrates phonemic targets, optimization, and prediction.

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, Fager, Ball, & Dietz, 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, Waller, Pullin, & Abel, 2008; Cler, Nieto-Castañón, Guenther, Fager, & Stepp, 2016; Cler, Nieto-Castañón, Guenther, & Stepp, 2014; Schroeder, 2005; Trinh, Waller, Vertanen, Kristensson, & Hanson, 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 resulting advantages in speed and flexibility may represent a significant improvement over other options for individuals with severe motor impairments. Improvements to the efficiency of phonemic interfaces may make this option even more appealing.

Quantitatively Optimized Interfaces

The standard orthographic keyboard layout, QWERTY (the Sholes keyboard, designed in 1873), is highly inefficient for 10-finger typing; it was specifically designed to be inefficient so as to minimize jamming typewriter keys (Noyes, 1983; Rumelhart & Norman, 1982). Alternate 10-finger typing layouts, such as Dvorak, show 4% improvements in typing speed (West, 1998). However, 10-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). Furthermore, individuals with sufficient 10-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 & Hu, 1994).

QWERTY keyboards are particularly inefficient for serial input, such as when individuals are entering text on a touch screen 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 required 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 proximity are faster to select (smaller distance to be traveled), 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, Hunter, & Smith, 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 randomly arranged phonemic interface (Cler & Stepp, 2017). However, these expected improvements in communication rate have not yet been empirically validated.

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, Mccaw, Yarrington, Mccoy, & Pennington, 2009) and can improve communication rates in phonemic interfaces by 100% (Trinh et al., 2012; Vertanen, Trinh, Waller, Hanson, & Kristensson, 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 T9 texting (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 selection. 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 five-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 (Lesher et al., 1998).

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 (a) visually highlighting predicted targets to draw the user's attention (Magnien, Bouraoui, & Vigouroux, 2004; Sears, Jacko, Chu, & Moro, 2001) and (b) 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 impairments 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 impairments 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, Majaranta, Isokoski, & Räihä, 2008). Furthermore, 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, Boots, Sugihara, & Chiu, 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, Mioc, & Gold, 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.

Empirical Evaluation by Users With and Without Motor Impairments

Communication rates in individuals with motor impairments 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 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 105.6% 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-groups design, and therefore, each group must consist of relatively homogenous participants. Furthermore, 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 impairments are highly heterogeneous as a group and are difficult to recruit over many sessions. Although participants without motor impairments fit these requirements, their typical access methods (e.g., finger on a touch screen or a typical mouse) are overtrained and not representative of the noisy access methods generally available to participants with motor impairments. 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, Cler, Fager, & Stepp, 2018).[1]

Here, we present two empirical evaluations of these optimization and prediction strategies. First, four groups of individuals (36 in total) without motor impairments interacted with one of four phonemic interfaces in a 2 × 2 between-groups 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 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 using a within-participant design, six individuals with motor impairments used the optimized/static and optimized/predictive interfaces in alternating blocks and answered survey questions about their experience and preferences after each block.

Method

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, 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.

[1]It is frequently necessary to use non-AAC users as participants due to the difficultly 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 touch screen (Trnka, Yarrington, McCaw, McCoy, & Pennington, 2007). Although this does accurately model the speed of AAC use and (as suggested) prompt users to incorporate prediction more than a user with motor impairments 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 difficulty of alternate access in this population.

Optimization

Full description 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 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 may 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. This corpus is a set of 1,004 messages suggested by AAC specialists for individuals 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 impairments.

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. Although 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, e-mail) 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: 1,004 AAC messages suggested by AAC experts (Beukelman & Gutmann, 1999) translated to phonemes using the Carnegie Mellon University Pronunciation Dictionary (Weide, 2005). *N*-grams

(n = 1–3) 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.

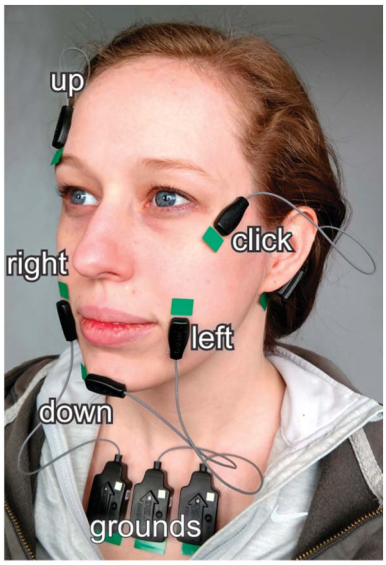
Likely targets were indicated to the user via weighted Voronoi diagrams. Seeds for each target were located at each target's center in a static grid, allowing users to retain knowledge of the phoneme arrangement and thus reducing the time required to visually search for the targets. The target weights (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 Material S1.

Surface Electromyographic Cursor

In the first experiment, participants without motor impairments used an alternate computer access method available to individuals with severe paralysis, a surface electromyographic (sEMG) facial cursor (Cler et al., 2016; Cler & Stepp, 2015; Vojtech et al., 2018). Full details of implementation are given in the study of Cler and 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 1). Muscle activity was captured with the Trigno sEMG system from Delsys, Inc. Each electrode consists 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).

Figure 1. Surface electromyographic minisensor 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), and click (wink). Combining gestures allows the cursor to move in any 360° direction, and magnitude of activity controls cursor speed (Cler & Stepp, 2015).



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.

Participants

Thirty-six adults without motor impairments participated in the first study.[2] 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, one nonbinary person; balanced across groups) had a mean age of 21.2 years (SD = 2.6). In the second study, six adults

[2]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 to have sufficient facial muscle control.

with motor impairments participated (participant characteristics in Table 1). Three participants were community dwelling, and three were inpatients at a rehabilitation hospital. Diagnoses were congenital (cerebral palsy) or acquired (multiple sclerosis, spinal cord injury [SCI], Guillain–Barré syndrome). Participants included those with stable (cerebral palsy; chronic SCI), degenerative (multiple sclerosis), and improving and/or stabilizing (Guillain–Barré syndrome, acute incomplete SCI) impairments. All participants provided consent in compliance with Boston University's Institutional Review Board; individuals with motor impairments provided either written consent or verbal consent witnessed by a communication partner as appropriate.

Experimental Designs

Study in Participants Without Motor Impairments

Participants without motor impairments completed 12 experimental sessions, each lasting 1–1.5 hr. Sessions occurred on separate days over 3 weeks with no more than 3 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 was assigned to one of the four different interfaces (see Figure 2).

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 2 min of free interaction with the interface using a typical mouse, followed by sEMG cursor application and calibration, 30 min of interaction with the interface via the sEMG cursor (the "main task"), and three short probe tasks.

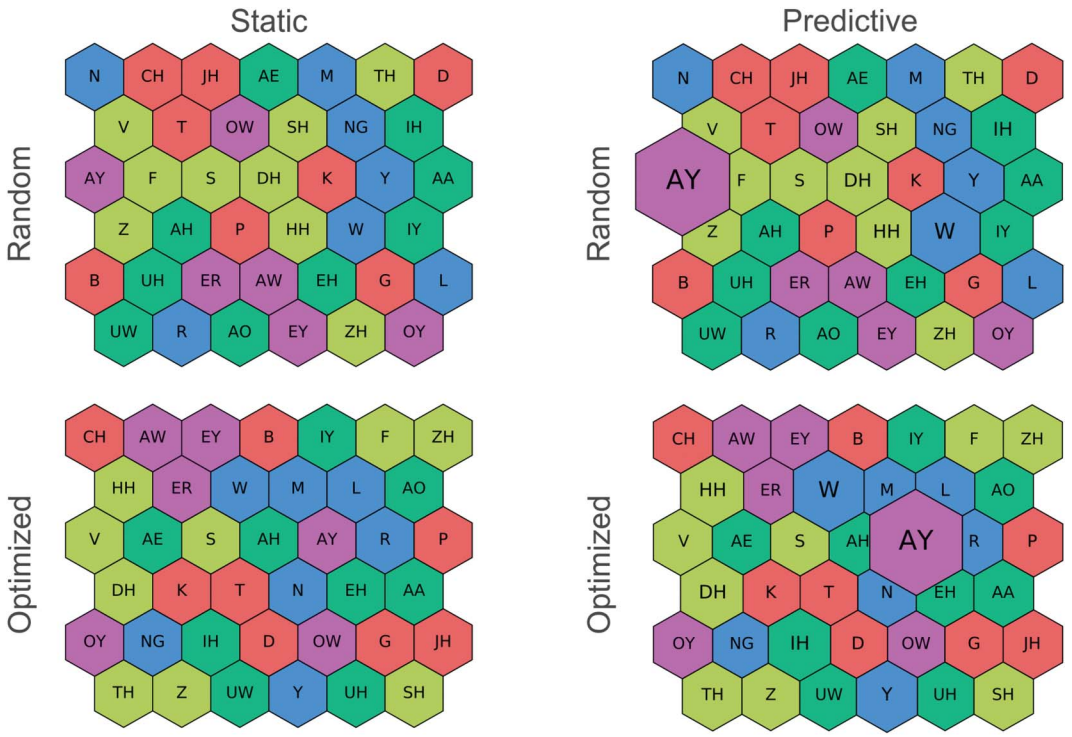
Most of the session was devoted to the main task, recreating aurally presented messages with the phonemic interface. This task required participants to translate the aural stimulus to our phoneme set, visually locate those phonemes on the interface, and move to and select the target (schematized in Figure 3a). The time it took to move and to select the target was governed by several factors: (a) the participant's proficiency with any particular access method, (b) the distance that must be traveled, and (c) 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. Participants were prompted with one message from a corpus of suggested AAC messages (1,004 messages; Beukelman & Gutmann, 1999) and then used the facial sEMG cursor to select the phonemes they wanted to use to recreate that message (see Figure 3b). Participants recreated different messages with the phonemic interface for at least 30 min each session (interactions were not automatically terminated after 30 min 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 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.

Table 1. Participant characteristics; study in individuals with motor impairments.

Participant	Age/sex	Diagnosis	Access method for this experiment	Community-dwelling or inpatient
P1	49/female	Multiple sclerosis (severe; 20 years postdiagnosis)	Mouthstick (stylus adapted to be held in mouth) on touch screen	Community dwelling
P2	45/male	Spinal cord injury (acute; 3 months postinjury)	Eye tracker with sip-and-puff switch for click (new to participant)	Inpatient
P3	63/male	Spinal cord injury (chronic; > 26 years post)	Stylus attached to stabilizing wrist guard on touch screen (Day 1: nondominant hand; Day 2: dominant hand)	Community dwelling
P4	21/female	Cerebral palsy	Nose on touch screen	Community dwelling
P5	59/male	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
P6	63/female	Guillain–Barré syndrome (acute; 3 months postonset)	Stylus in hand on touch screen	Inpatient

Figure 2. Four interfaces used 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.



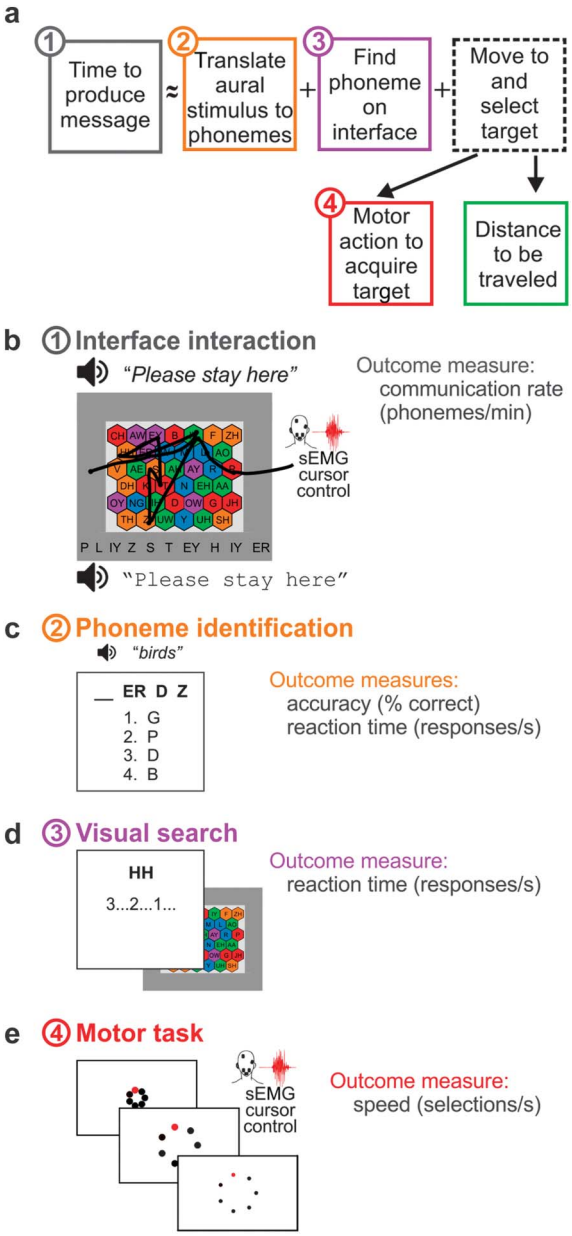
Accuracy was estimated using the minimum string distance between the phonemes selected and the phonemes expected based on automated dictionary transcription of prompts (Soukoreff & MacKenzie, 2001). Speed was calculated as the number of actual selections divided by the time it took to complete the trial. 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]-/stoʊr/[3] and [S-T-AO-R]-/stɔr/). Furthermore, these interfaces do not use spaces 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 intelligibility estimates infeasible. Thus, we focus here only on speed (selections per minute). Accuracy estimates are explored further in the discussion. Following the 30 min 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. To assess their ability to translate an aural stimulus to the phoneme set, participants completed 15 fill-in-the-blank questions during each session (see Figure 3c). 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"; see Figure 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 number keys on a standard QWERTY keyboard. Participants were instructed to complete this task as quickly and accurately as they could. Two outcome measures were obtained: accuracy and reaction time (responses per second).

Visual search task. To assess their ability to find phonemes on the interface, participants visually located

[3]Two phonemic transcription conventions are used throughout this article. One is the International Phonetic Alphabet, which is likely familiar to readers and will be indicated with sounds between slashes (/saʊndz/). 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 International Phonetic Alphabet or via orthographic text in quotes.

Figure 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 per minute). (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 per second) were collected. (d) Participants indicated when they visually located the given label (outcome measure: reaction time in responses per second). (e) Participants used facial surface electromyographic (sEMG) cursor to select circular targets and were assessed on speed (selections per second).



10 randomly generated phoneme labels during each session (see Figure 3d). Participants were presented with a white screen with a particular phoneme label (e.g., "HH"; see Figure 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 postprocessing such that only the first presentation of any one label was used to calculate the outcome measure of visual search time (responses per second).

Motor task. 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 (see Figure 3e). 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 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 per second).

Study in Participants With Motor Impairments

Participants with motor impairments 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). Each block consisted of 10 min of interaction with one of the two interfaces (same as "main task," above), followed by a survey to capture their experiences using the interfaces. Responses were solicited from these participants as they had personal experience using assistive technology and may have input on the design and usability of these AAC interfaces beyond those considered by the researchers and participants without motor impairments.

User preference survey. The 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 Material S2 (further details are also available in Cler, 2018). If participants were able, they completed the forms themselves. Otherwise, the experimenter read the questions aloud and dragged a pen across each VAS line until the participant indicated their preferred stopping place. Participant reactions and responses were also transcribed during

the experiment and during the survey assessment and are presented here qualitatively.

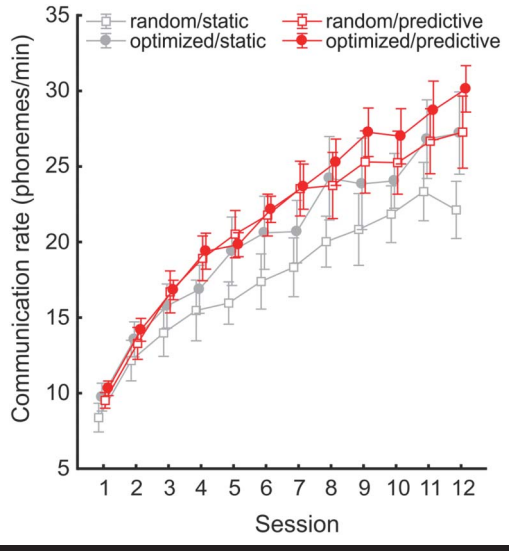
Statistical Analyses

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 per second), phoneme identification–accuracy (%), phoneme identification–reaction time (responses per second), and visual search performance (responses per second). 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: (a) How do optimization and prediction affect learning? (b) How do optimization and prediction affect performance after learning?

To assess learning, a linear mixed-effects model (Bates, Mächler, Bolker, & Walker, 2015) was performed on data from Sessions 1 to 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 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, backward stepwise regressions were performed in order to determine which, if any, of the probe measures captured individual variation relevant to the task. Unstandardized β 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, 2016; Nakagawa & Schielzeth, 2013).

In participants with motor impairments, communication rates and survey responses were tabulated to assess user effort and preferences. Quantitative survey responses were measured as distance from the left end of each line and as such are reported from 0 to 10 cm. No statistical tests were performed on the VAS, but responses and preferences are presented descriptively.

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



Results

Participants without motor impairments completed 20,849 trials across a total of 432 sessions. Participants showed an increase in communication rate 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 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 showed that all measures increase with session and with communication rate (see Supplemental Material S3), 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 to 9 are shown in Table 2; 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; 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.

Communication rates from participants with motor impairments are shown in Figure 5. Participants completed three to seven blocks of trials (10 min 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.

User Preference Survey and Comments

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 of six participants strongly preferred the interface with prediction over the static interface (P1, P3,

Table 2. Sessions 1–9 mixed-effects model; remaining factors after backward stepwise regression.

Factor	Communication rate		
	β	CI	*p*
(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 × Prediction	–3.00	[–6.94, 0.93]	.145
Interface × Session	0.41	[0.08, 0.74]	.017
Interface × Motor Task Speed	–1.14	[–2.73, 0.45]	.16
Session × Motor Task Speed	0.01	[–0.15, 0.18]	.86
Session × Phoneme Identification–Accuracy	–0.07	[–0.13, 0.00]	.043
Interface × Visual Search Time	0.39	[–1.02, 1.80]	.586
Prediction × Visual Search Time	1.77	[0.51, 3.04]	.006
Interface × Session × Motor Task Speed	0.37	[0.13, 0.61]	.002
Interface × Prediction × Visual Search Time	–2.04	[–3.68, –0.39]	.016
Observations		324	
Marginal R^2/conditional R^2		.669/.865	

Note. CI = confidence interval.

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 and 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 cm 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"). 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"; on Day 2). 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" (on Day 2). One participant initially said he preferred the interface without prediction, but later highly preferred prediction as he got used to it.

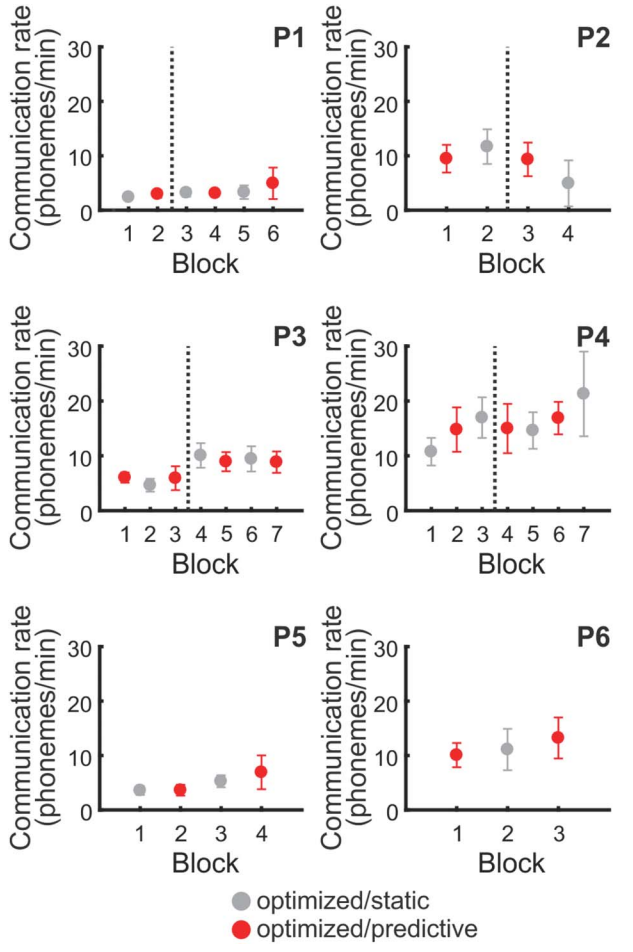
Although most participants preferred prediction, one participant who preferred the static interface remarked that he "liked to figure it out himself" and that the prediction

Table 3. Session 12 linear model; remaining factors after backward stepwise regression.

Factor	Communication rate		
	β	CI	*p*
(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 × Prediction	–1.69	[–8.75, 5.37]	.627
Interface × Phoneme Identification–Accuracy	–0.45	[–5.38, 4.49]	.853
Prediction × Phoneme Identification–Accuracy	5.11	[–1.23, 11.45]	.11
Interface × Prediction × Phoneme Identification–Accuracy	–6.11	[–13.79, 1.56]	.114
Observations		36	
R^2/adjusted R^2		.675/.563	

Note. CI = confidence interval.

Figure 5. Communication rates for the six participants with motor impairments (see Table 1 for participant characteristics). Participants all used optimized interfaces. Interfaces were either static (gray) or predictive (red, dark) in alternating blocks. When possible, participants completed blocks over 2 days; black dotted vertical lines indicate separation from Day 1 to Day 2. Error bars are standard deviation.



led him in a direction that he did not want. 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." 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 the prediction.

Discussion

Integrating the results of these two studies reveals the promise and potential pitfalls of phonemic targets in communication interfaces for AAC. The results in users without motor impairments reveal the performance trajectory when participants are completely naïve to AAC and novel access methods and then use the interface over many sessions. These results suggest that 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 (i.e., 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. The results in individuals with motor impairments show initial reactions and user impressions, indicating that future research into phonemic interfaces in this population is warranted.

Phonemic Input as Compared to Other Interfaces

The comments made by participants with motor impairments can roughly be grouped into three segments that will drive future development and implementation studies. First, all participants agreed that they would improve with practice and made comments to that effect ("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"). This suggests that none of the participants thought that the interfaces were so complex that they were tempted to abandon them entirely. Next, while most participants preferred prediction, those who did not may have still benefitted from it ("I don't remember what sound that makes…oh well, I'll pick it anyway"). Finally, participant comments suggested that they agreed that phonemic input was flexible and usable ("As I use it more, I can see that you could get it to do the dictation just as you would want"; "I'm getting used to the sounds now"), even if they did not like it immediately. These give insight into possible roadblocks to implementation and indicate that training should include both the possible benefits and drawbacks to the interfaces.

Future research is needed to directly compare communication rates of phonemic input to orthographic input within participant. However, we can compare the results of Study 1 to previous results using the same access method (participants without motor impairments using an sEMG cursor). A previous study using this access method with an alphabetical interface saw communication rates of 29 selections/min during the final (fourth) session (Cler & Stepp, 2015). This is very similar to the rate of 30.1 selections/min seen in the optimized/predictive group in the final (12th) 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 (four phonemes per word and no spaces), whereas the 29 letters/min on the alphabetic interface represents 4.8 wpm (five letters plus space per word). 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.

Effects of Interface Optimization

The first study (in individuals without motor impairments) evaluated the effects of interface optimization and found that 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. Previous work suggested that an optimized interface should show 30.9% increase in communication rate, assuming ideal motor access and ideal phoneme selection. Our empirical results suggest that optimization improved communication rates by 23.0% (random/predictive group) and 10.2% (optimized/predictive group) in the final session. The reason for the discrepancies between these improvements and the theoretical improvements are 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, their communication rates would not be increased over someone using the random layout.

Interface was not a significant contributor to communication rate during early 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/searching 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.

Effects of Prediction

Both of the studies in this article assessed the effects of phonemic prediction. In the study of individuals without motor impairments, prediction had a significant effect on communication rate. In the study of participants with motor impairments, participants generally preferred the prediction, but our study was not designed to assess quantitative difference in communication rates between the predictive and static versions of the interface.

Teaching Users the Phonemic System

The prediction had the effect of drawing the users' attention to phoneme labels that they may not have chosen otherwise. One unanticipated effect was that the prediction seemed to teach the users which sounds to choose. These effects can be elucidated by comparing the dictionary series of phonemes for each prompt to what the participants actually selected. Mismatching selections could have many different sources: a motor error (i.e., clicked a target accidentally), a 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). One benefit of phonemic interfaces is that participants may choose a variety of targets to create a message and thus are protected from certain types of "errors" (Cler et al., 2016). Swapping /θ/ and /ð/ or even /aɪ/ and /eɪ/ will result in messages that are still intelligible. However, if the prediction's main effect was to teach the user to use the interface dictionary's phonemic system, then users may benefit more from the motor-based optimization and perhaps make faster selections as the cognitive effort of choosing the intended target is lessened. We illustrate the possible influence of prediction on prompt-to-selection mismatches with three different examples.

Of all of the 20,849 trials, 3,962 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, whereas 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[4] and calculated the types

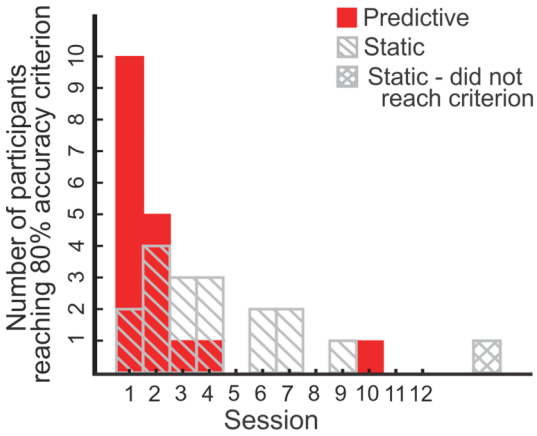
[4]We considered trials with only a [TH] or [DH] in the prompt and excluded those with both for simplicity. Of the remaining trials, 1,079 were correct [TH] trials, 1,429 were correct [DH] trials, 62 were [TH] prompts with [DH] selected, 2,309 were [DH] prompts with [TH] selected, and 277 could not be 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]; and 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.

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 (i.e., 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.

Finally, we can illustrate the effect of prediction by evaluating 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, as shown in Figure 2. Any mismatches are unlikely to be 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. These errors varied across time and by cohort: Figure 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 (gray-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.

The possible accuracy increases provided by prediction are not explicitly incorporated in the main results in Figure 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

Figure 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 reach criterion in the first two sessions. Gray striped bars: static groups. Note that these participants take longer to reach criterion, and one participant never reaches criterion (gray checked box).



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 (i.e., 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 user's selection history into the prediction algorithm. Because the participants often used nondictionary 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.

In Individuals With Motor Impairments

Within-participant assessments of prediction enabled us to gather participant preferences and reactions. Following the results of Study 1, an expected effect of the prediction is in helping participants learn the location and identity of various targets. However, alternating blocks of predictive and static interfaces in the within-participant design of Study 2 means that any learning effects would likely carry over to the static blocks and thus be washed out. As a result, the

trends in Figure 5 that show no block-to-block changes in communication rates with predictive interfaces versus static interfaces were expected. Our main outcomes were the ratings of the interfaces, which will help drive further interface development and implementation.

Verisimilitude of Participants Acting as Model of Motor Impairments

Although the main outcome of this study was feasibility and user feedback, we were also able to assess communication rate and thus evaluate the success of our participants without motor impairments acting as a model of participants with motor impairments. Participants without motor impairments used the sEMG cursor to produce communication rates of 5.0–15.6 phonemes/min on the first day ($M = 9.5$, $SD = 2.2$). Participants with motor impairments used a variety of access methods to produce communication rates of 2.8–11.6 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. Although 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. The four participants with motor impairments who completed 2 days of sessions had a mean percent increase of 23.8% from Day 1 to Day 2 (range: −27.0 to 69.1). The 36 participants without motor impairments had a mean percent increase from Day 1 to Day 2 of 41.6% (range: −12.9 to 82.9). Thus, although both groups saw improvement with training, the participants without motor impairments made larger strides on average. This is likely due to varying factors: (a) 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, P5) or used new-to-them access methods due to the recency of their injuries (P2, P5). (b) The pattern of motivation might be quite different between the groups; for example, participants without motor impairments were sufficiently motivated by the automated feedback provided as a score after each trial. The participants with motor impairments were more motivated by accuracy and would spend many seconds searching for the "correct" target, even when prompted to just do their best and go quickly. This resulted in less overall practice if they only completed a few trials. We also asked them to rate the interfaces in between each block, which may have interfered both with the learning process and emphasized that their motivation was to evaluate the interfaces rather than produce messages quickly. (c) Finally, the cognitive load of the task may have differentially affected the participants with motor impairments. They were generally older than the participants without motor impairments and more likely to deal with other effects of their impairments, including chronic pain and medication side effects.

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/?" and "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.

Although 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 and perhaps unfair. 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.

Finally, our study in people who use AAC was limited and preliminary and was primarily designed as a measure of feasibility and to gather feedback on our design. Our methods for soliciting feedback were informal (i.e., not a given set of qualitative research questions), and our VAS scales were purpose built and not validated. Much future

work is needed to refine the interfaces, determine who may or may not benefit from such an interface, and establish what training may be needed.

Applications to Nonphonemic Interfaces

Some of these advances may be applied to orthographic or symbol-based interfaces. Orthographic interface layouts 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, weighted Voronoi diagrams, 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; furthermore, this algorithm implementation (via pyvoro) is fast enough that it is usable with even overtrained access methods (e.g., typical mouse and touch screen input) without a noticeable delay.

Conclusions

These studies empirically assessed the effects of computational optimization and prediction on communication rates generated by participants with and 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 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. Assessments in participants with motor impairments revealed that the participants could use the interfaces to generate messages and that most participants preferred the interface with prediction. Future work is needed to validate these novel methods of optimization and prediction for AAC and to translate these results into clinical practice.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant F31 DC014872 (awarded to G. J. C.) and National Science Foundation Grants 1452169 (awarded to C. E. S.) and 1247312 (awarded to J. M. V.). The authors would like to thank several people for their help with these studies and article. First, thank you to Tabatha Sorensen for helping to recruit participants; Jaime Kim, Rebecca Glover, and Tiffany Peters for helping G. J. C., K. R. K., and J. P. N. to run participants; and Andreas Singer for recording the message bank. Thank you to Jay Bohland, Frank Guenther, and Chris Moore for providing input on the design of the study in participants without motor impairments. Thank you also to Kathleen Nagle for consultations about the visual analog scale interpretation.

References

Anton, F., Mioc, D. & Gold, C. M. (1998). Dynamic additively weighted Voronoi diagrams made easy. In *Canadian Conference on Computational Geometry (CCCG)*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beddoes, M. P., & Hu, Z. (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

Beukelman, D. R., Fager, S. K., 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., & Gutmann, M. (1999). *Generic message list for AAC users with ALS*. Lincoln: University of Nebraska–Lincoln. Retrieved from https://cehs.unl.edu/documents/secd/aac/vocablists/ALS_Message_List1.pdf

Beukelman, D. R., & Mirenda, P. (2013). *Augmentative and alternative communication: Supporting children and adults with complex communication needs* (4th ed.). Baltimore, MD: Brookes.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. Sebastopol, CA: O'Reilly Media, Inc. https://doi.org/10.1017/CBO9781107415324.004

Black, R., Waller, A., Pullin, G., & Abel, E. (2008). *Introducing the PhonicStick: Preliminary evaluation with seven children*. Paper presented at the 13th Biennial Conference of the International Society for Augmentative and Alternative Communication, Montréal, Canada.

Cler, G. J. (2018). *Computational optimization and prediction strategies for increasing communication rate in phoneme-based augmentative and alternative communication (AAC)* (Unpublished doctoral dissertation). Boston University, Boston, MA.

Cler, G. J., & Stepp, C. E. (2017). Development and theoretical evaluation of optimized phonemic interfaces. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility—ASSETS '17, 2017,* 230–239. https://doi.org/10.1145/3132525.3132537

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-Castañón, A., Guenther, F. H., & Stepp, C. E. (2014). Surface electromyographic control of speech synthesis. *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014, 2014,* 5848–5851. 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. *Proceedings of the ACL Workshop on Natural Language Processing for Communication Aids*. Madrid, Spain, 37–42.

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. *Proceedings of the Fourth International Conference on Spoken Language Processing, 1996,* 1393–1396.

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.

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

Higginbotham, D. J., Shane, H., Russell, S., & Caves, K. (2007). Access to AAC: Present, past, and future. *Augmentative and Alternative Communication, 23*(3), 243–257. https://doi.org/10.1080/07434610701571058

Kushler, C. (1998). *AAC: Using a reduced keyboard*. Paper presented at the CSUN Conference on Technology for Persons With Disabilities, California State University, Northridge, CA.

Lefcheck, J. S. (2016). 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

Lesher, G., Moulton, B., & Higginbotham, D. J. (1998). Techniques for augmenting scanning communication. *AAC: Augmentative and Alternative Communication, 14*(2), 81–101. https://doi.org/10.1080/07434619812331278236

MacKenzie, I. S., & Zhang, S. X. (1999). The design and evaluation of a high-performance soft keyboard. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1999,* 25–31.

Magnien, L., Bouraoui, J. L., & Vigouroux, N. (2004). Mobile text input with soft keyboards: Optimization by means of visual clues. *International Conference on Mobile Human–Computer Interaction, 2004,* 337–341.

McGuffin, M. J., & Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer–Human Interaction, 12*(4), 388–422.

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

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

Okabe, A., Boots, B., Sugihara, K. & Chiu, S. N. (2009). *Spatial tessellations: Concepts and applications of Voronoi diagrams* (Vol. 501). West Sussex, United Kingdom: Wiley.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science, 6,* 1–36.

Rycroft, C. H. (2009). VORO++: A three-dimensional Voronoi cell library in C++. *Chaos, 19*(4). https://doi.org/10.1063/1.3215722

Schroeder, J. E. (2005). *Improved spelling for persons with learning disabilities*. Paper presented at the 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.

Shoup, J. (1980). Phonological aspects of speech recognition. In W. A. Lea (Ed.), *Trends in speech recognition* (pp. 125–138). New York, NY: Prentice-Hall.

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

Soukoreff, R. W., & MacKenzie, I. S. (2001). Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic. *Proceedings of the Extended Abstracts on Human Factors in Computing Systems, 2001,* 319–320. https://doi.org/10.1145/634067.634256

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., Waller, A., Vertanen, K., Kristensson, P. O., & Hanson, V. L. (2012). *iSCAN: A phoneme-based predictive communication aid for nonspeaking individuals*. Paper presented at the 14th International ACM SIGACCESS Conference on Computers and Accessibility, Boulder, CO.

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

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, 2007,* 195–202. 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. *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), 2007,* 173–176. https://doi.org/10.3115/1614108.1614152

Tuisku, O., Majaranta, P., Isokoski, P., & Räihä, K.-J. (2008). Now Dasher! Dash away!: Longitudinal study of fast text entry by eye gaze. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, 2008,* 19–26.

Vertanen, K., & Kristensson, P. O. (2011). The imagination of crowds: conversational AAC language modeling using crowdsourcing and large data sources. *Proceedings of the Conference*

on Empirical Methods in Natural Language Processing, 2011, 700–711.

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), 2012,* 19–27.

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*. Paper presented at the Conference on Motor Speech, Savannah, GA.

Ward, D. J., & MacKay, D. J. C. (2002). Artificial intelligence: Fast hands-free writing by gaze direction. *Nature, 418,* 838. https://doi.org/10.1038/418838a

Weide, R. (2005). *The Carnegie Mellon pronouncing dictionary [cmudict 0.6]*. Pittsburgh, PA: Carnegie Mellon University.

West, L. J. (1998). *The standard and Dvorak keyboards revisited: Direct measures of speed. Santa Fe Institute working papers.* Santa Fe, CA: Sante Fe Institute. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.8.6886&rep=rep1&type=pdf

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.

Zhai, S., Conversy, S., Beaudouin-Lafon, M., & Guiard, Y. (2003). Human on-line response to target expansion. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2003,* 177–184. New York, NY: ACM.

Zhai, S., Hunter, M., & Smith, B. A. (2002). Performance optimization of virtual keyboards. *Human–Computer Interaction, 17*(2–3), 229–269.