

Surface electromyographic control of a novel phonemic interface for speech synthesis

Meredith J. Cler, Alfonso Nieto-Castañón, Frank H. Guenther, Susan K. Fager & Cara E. Stepp

To cite this article: Meredith J. Cler, Alfonso Nieto-Castañón, Frank H. Guenther, Susan K. Fager & Cara E. Stepp (2016) Surface electromyographic control of a novel phonemic interface for speech synthesis, *Augmentative and Alternative Communication*, 32:2, 120-130, DOI: 10.3109/07434618.2016.1170205

To link to this article: <http://dx.doi.org/10.3109/07434618.2016.1170205>



Published online: 04 May 2016.



Submit your article to this journal [↗](#)



Article views: 60



View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH ARTICLE

Surface electromyographic control of a novel phonemic interface for speech synthesis

Meredith J. Cler^a, Alfonso Nieto-Castañón^b, Frank H. Guenther^{b,c}, Susan K. Fager^d and Cara E. Stepp^{b,c}

^aGraduate Program for Neuroscience – Computational Neuroscience, Boston University, MA, USA; ^bDepartment of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA; ^cDepartment of Biomedical Engineering, Boston University, Boston, MA, USA; ^dInstitute for Rehabilitation Science and Engineering, Madonna Rehabilitation Hospital, Lincoln, NE, USA

ABSTRACT

Many individuals with minimal movement capabilities use AAC to communicate. These individuals require both an interface with which to construct a message (e.g., a grid of letters) and an input modality with which to select targets. This study evaluated the interaction of two such systems: (a) an input modality using surface electromyography (sEMG) of spared facial musculature, and (b) an onscreen interface from which users select phonemic targets. These systems were evaluated in two experiments: (a) participants without motor impairments used the systems during a series of eight training sessions, and (b) one individual who uses AAC used the systems for two sessions. Both the phonemic interface and the electromyographic cursor show promise for future AAC applications.

ARTICLE HISTORY

Received 19 June 2015
Revised 2 March 2016
Accepted 21 March 2016
Published online 2 May 2016

KEYWORDS

Surface electromyography;
phonemic interface; training;
Guillain-Barré syndrome

Introduction

Individuals who use AAC devices due to motor impairment may use two complementary systems to communicate: an interface with which to construct a message (e.g., a grid of letters on a computer screen) or an input modality with which to select targets (e.g., a head tracker or sip-and-puff system). Innovations are needed in both of these areas, particularly to allow individuals with minimal movement capabilities to access flexible, robust, synthesized speech output. This study evaluated both an input modality that leverages surface electromyography (sEMG) of spared facial musculature, and a novel AAC interface consisting of phonemic targets.

sEMG as an AAC input modality

Individuals who communicate via AAC can use a variety of input modalities to control a high-tech AAC device. Depending on the degree of motor impairment, input modalities range from mechanical switches to devices employing hand movements (touchscreen, typical mouse); head movements (e.g., Williams & Kirsch, 2008); eye movements (e.g., Frey, White, & Hutchison, 1990; Higginbotham, Shane, Russell, & Caves, 2007); tongue movements (e.g., Huo, Wang, & Ghovanloo, 2008); or brain signals (e.g., Guenther & Brumberg, 2011; Orhan et al., 2012; Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). Individuals with some remaining muscle control (e.g., individuals with spinal cord injuries) may benefit from a cursor control system that leverages this spared function.

The electrical activity generated by spared muscles can be detected via electrodes placed on the surface of the skin.

This technique, sEMG, provides a robust neural signal that can then be translated to cursor movements, leading to full 2D control of a computer (Choi, Rim, & Kim, 2011; Cler & Stepp, 2015; Larson, Terry, Canevari, & Stepp, 2013; Vernon & Joshi, 2011; Williams & Kirsch, 2008). sEMG signals can be translated to cursor movements with position-based algorithms, in which activation of one muscle corresponds to the cursor position in the horizontal direction and activation of a different muscle corresponds to vertical cursor position or velocity-based algorithms, in which muscle activation corresponds to cursor velocity. Although position-based systems use fewer muscles than the velocity-based system utilized in this study, they require constant, graded muscle activation. Velocity-based systems allow users to maintain cursor position when relaxing, requiring only brief bursts of muscle activity to move the cursor.

sEMG cursors provide benefits not available with eye tracking, head tracking, or mechanical switches. Many eye-tracking systems require high illumination, stable head positions, and complete control over eye movements (Beukelman, Fager, Ball, & Dietz, 2007), as do many camera-based head-tracking systems (although some head-tracking systems do not require specific lighting or position; e.g., Williams & Kirsch, 2008). sEMG systems do not require any lighting nor does the user need to be directly in front of the computer screen, as with many eye- and head-tracking systems. In addition, many eye-tracking systems have an inherent speed limitation caused by the selection method: Users must dwell over a target for a set amount of time, introducing a speed/accuracy trade-off. Reducing the dwell-time increases accidental selections, while increasing the dwell-time reduces the number of possible selections per minute (Majaranta, MacKenzie, Aula, & Riih , 2006). The sEMG system

presented here utilizes a brief, independent muscle contraction to click, reducing accidental selections and increasing speed compared to the dwell-time selection.

Individuals with very high spinal cord injuries may not be able to control their head position well enough to use head-tracking systems because some required muscles (e.g., sternocleidomastoid) are innervated by cervical nerves. Studies have shown that sEMG can capture activity in hemiparetic muscles that are innervated but do not support movement (Saxena, Nikolic, & Popovic, 1995). This suggests that individuals who do not have enough muscle strength to produce limb or face movements that can control a mechanical interface or are recognized by a camera-based device may still be able to produce sufficiently reliable muscle activity for sEMG control. Mechanical switch options such as head switches or sip-and-puff devices are popular and accessible to many individuals but are often slower than direct selection access methods and require cognitive and motor processes that can be challenging (Beukelman & Mirenda, 2013; Fager, Bardach, Russell, & Higginbotham, 2012). Thus, sEMG cursor control may be a reasonable alternative or adjunct to many of these input options.

Phonemic interfaces

People who use AAC generally select pictographic symbols, icons, or letters for message construction. Pictographic symbols may not support generative language as they can constrain selections to words or phrases pre-loaded into the device by caregivers, and some pictographic systems do not translate into English (Alant, Life, & Harty, 2005), limiting interactions with others. Letters require individuals who use the system to be literate, which reduces accessibility for individuals who are not (Beukelman & Mirenda, 2013; Koppenhaver & Yoder, 1992; Smith, 2001). If individuals want to communicate via synthesized speech, the synthetic speech output of devices using orthographic input must rely on complex letter-to-sound rules, which can fail on proper names, non-words, or any words not contained in the device's dictionary; this is particularly problematic in English and other languages with opaque orthographies, in which there is not a one-to-one relationship between sounds and letters. Ideally, synthesized speech should contain exactly the sounds that the user intends to produce. Thus, instead of letters, the user could directly select phonemes, contrastive units of sound in a particular language that differentiate meaning. Phonemes carry more information per unit than letters, allowing users to produce flexible synthesized speech quickly. For example, among a preliminary list of suggested AAC messages for individuals with amyotrophic lateral sclerosis (ALS; list from Beukelman & Gutmann, 1999), 96% of the messages have fewer phonemes than letters, with an average of 20% fewer phonemes than letters, a substantial savings in the number of required selections. Similarly, 97% of a large list of vocabulary appropriate for adults who use AAC (Bryen, 2008) have fewer or equivalent phonemes than letters, with an average selection savings of 14%. Finally, phonemes allow individuals who use AAC to have complete control over the sounds produced by the speech synthesizer instead of relying on text-to-speech (TTS) rules.

Most existing phonemic interfaces display a reduced set of phonemes on the screen (e.g., Black, Waller, Pullin, & Abel,

2008; Trinh, Waller, Vertanen, Kristensson, & Hanson, 2012). This either requires individuals to make several choices before a phoneme is selected (time-consuming and requiring a series of motor actions for each selection), or the system must disambiguate intended selections based on prior selections, similar to the T9 texting system (e.g., Kushler, 1998), which limits individuals to only the words contained in its dictionary. To provide individuals who use AAC with full control over their computer-synthesized voice while maximizing speed and reducing motor effort, we have produced a phonemic keyboard in which the full set of English phonemes is displayed and available to select at all times.

Current investigation

We have previously shown that participants without motor impairments could select targets using facial muscle contractions recorded via sEMG electrodes placed upon the facial skin (Cler & Stepp, 2015). Muscle contractions during attempted facial gestures (e.g., left smile, eyebrow raise, wink) were translated, in real-time, into cursor movements (e.g., left, up, click; see Figure 1 for electrode placement). In that study, user performance in selecting targets on an alphabetic interface improved with training (Cler & Stepp, 2015). In a different study, the facial sEMG system was used by participants without motor impairments to produce speech by selecting phonemes on an onscreen phonemic AAC interface during one session (Cler, Nieto-Castanon, Guenther, & Stepp, 2014). What has not yet been studied is the effect of training on performance with the sEMG cursor and a phonemic interface; furthermore, previous work with sEMG cursor control and phonemic interface has been restricted to studies with individuals without motor impairments. We hypothesize that improvements in performance will be seen in healthy individuals and in individuals who use AAC, both due to motor

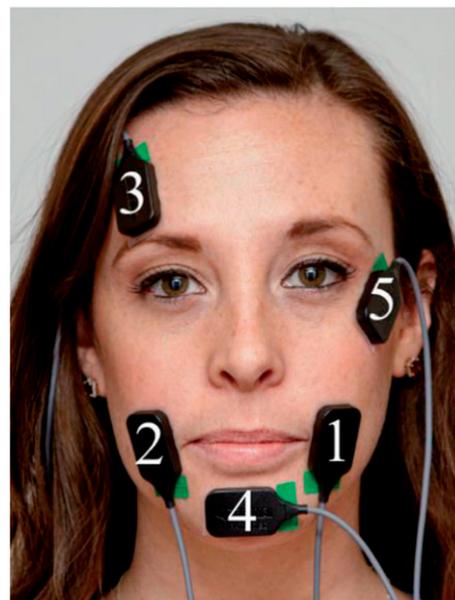


Figure 1. Placement of sEMG sensors for training study. Placements correspond to gesture and cursor actions: (1) Left half smile, cursor moves left; (2) right half smile, cursor moves right; (3) Eyebrow raise, cursor moves up; (4) Chin contraction, cursor moves down; (5) Wink, cursor clicks.

learning during the sEMG cursor control (as in Cler & Stepp, 2015) and due to faster visual search as participants become familiar with the layout of the phonemic targets.

The purpose of this study was to evaluate the combined sEMG-controlled selection system and phoneme-based interface by 10 individuals without motor impairments over multiple days (training study). We also completed a case study showing the ability of one user with severe paralysis to use the system over two sessions. In the training study, performance during typical mouse control indicated how well the user could use the onscreen interface itself, whereas performance with the sEMG cursor control represented both the familiarity of participants with the interface and their ability to control the sEMG cursor. We hypothesized that study participants would show increased speed and accuracy over the training sessions using both typical mouse control and the sEMG cursor control. In the case study, we hypothesized that the individual would be able to use the sEMG cursor without modifications to the sEMG system.

Method

Participants

Adults without Motor Impairments. Ten adults without motor impairments participated. All were native speakers of American English and reported no history of speech, language, or hearing disorders. Participants were university students who did not have previous experience with sEMG research and were not familiar with phonemic keyboards or transcription. The participants (four male; six female) had a mean age of 21.4 years ($SD=2.8$). All participants provided written consent in compliance with the university's institutional review board.

Adult who uses AAC. An individual with severe paralysis (S1) participated in the study. Prior to the onset of Guillain-Barré syndrome (GBS) at age 60, S1 was employed in agriculture, having achieved a bachelor's degree. S1 is a native English speaker and was proficient with standard computer technology (e.g., word processing, photo-editing). Following the onset of GBS, S1 has resided exclusively in a rehabilitation hospital and has received three formal speech evaluations with ongoing AAC treatment and support throughout the course of his illness, via the hospital's comprehensive AAC and assistive technology program. Initially, S1 was non-speaking and completely locked in until approximately 2 months post GBS onset. S1 had limited oculomotor control and significant dry eye issues that precluded the use of eye gaze technology. As such, S1 relied on partner-dependent scanning with a letter board and signaled yes/no with left/right eye movements. S1 then regained limited head movement (left/right) and could control a switch, and began using a DynaMyte^{TM1} with switch scanning as well as an onscreen keyboard with word prediction and pre-stored messages for urgent needs. Within a year post onset, S1 regained some voicing capabilities and then used

natural speech with the scanning speech-generating device (SGD) as a back-up. As speech and head control capabilities progressed by 1.5 years post onset, S1 advanced to a direct access, head tracking system.

At the time of testing, S1 was 8 years post-onset of GBS and 68 years old. S1 relied primarily on oral speech during the day, and was 74% intelligible at the sentence level on the Speech Intelligibility Test (Yorkston, Beukelman, Hakel, & Dorsey, 2007) due to flaccid dysarthria characterized by imprecise articulation, low volume, decreased respiratory support for speech, and decreased lingual and labial range-of-motion. For the past 7 years, S1 has also used the AccuPoint^{TM 2} head tracking system to support access to an onscreen keyboard and pre-stored messages, along with AccuKeys² for computer control and to type messages into email. S1 also used a head switch to gain the attention of care staff and a partner-assisted letter board if natural speech was not available, which was periodically the case when ventilation was required at night or because of a temporary illness.

Testing took place at the rehabilitation hospital in S1's room. Informed consent was obtained in compliance with the university's institutional review board: All consent documentation was read aloud, the participant provided verbal consent, and then provided written consent using a pen controlled with the mouth. An advocate also signed to indicate the participant's consent.

Procedures

Training Study. The participants without motor impairments completed eight training sessions within 14 days. The first session lasted up to 90 min and consisted of a pretest, sEMG portion, and posttest. All subsequent sessions lasted 40–60 min and contained only the sEMG portion and posttest. Pre- and posttests consisted of the participant using a typical mouse to produce a series of 15 words on an onscreen phonemic interface. The sEMG portions consisted of skin preparation, placement of five single differential electrodes (see Figure 1), a brief calibration period, and use of the sEMG cursor control system to produce a series of 45 words using the same onscreen phonemic interface.

Each trial in the pretest, posttest, and sEMG portions began with the aural and visual presentation of a novel word. Aural stimuli were generated with the phonemic interface offline and visual stimuli were phonemic spellings of the word; for example, if the target stimulus was "neighbor" participants heard it synthesized by the interface and saw it displayed phonemically as n-ay-b-er. All stimuli were American English words containing four phonemes (e.g., "among", "measure", and "group", presented as uh-m-uh-ng, m-eh-zsh-er, and g-r-oo-p). Session stimuli lists were generated to approximate a consistent distribution of phonemes across sessions; lists contained 15 words (pretest and posttest with typical mouse) or

¹DynaMyte is a product of Tobii DynaVox, Inc. of Danderyd, Sweden. <http://www.tobiidynavox.com/>

²AccuPointTM and AccuKeysTM are products of Invotek of Alma, AR. <http://www.invotek.org/>

45 words (sEMG portion) and were randomly ordered for each participant. After each stimulus was presented, participants selected the given phonemes with the phonemic interface using the predefined control modality (either a typical mouse or their sEMG signals; details below) to control the cursor. When participants finished selecting phonemic targets, the selected phonemes were synthesized, providing auditory feedback to the user. During the sEMG portion, participants also received an online estimate of their speed and accuracy, represented by information transfer rate (ITR; described more fully below in Data Collection and Analysis section) for each trial as feedback and were verbally encouraged to increase this score.

Case Study. One individual with severe paralysis participated for two sessions on two consecutive days. The first session involved placing sEMG sensors to find sensor positions and facial gestures that produced strong and independent signals. During the second session, S1 used the sEMG cursor system to control two different communication interfaces to produce text and speech: an alphabetic keyboard for six blocks of five trials and then the phonemic interface for one block of four trials. For both systems, the interface visually prompted S1 with a series of letters or series of phonemes to type. The phonemic interface synthesized the selected phonemes as auditory feedback. The alphabetic interface did not produce auditory feedback, but instead provided a real-time estimate of the user's ITR. Using 10cm visual analog scales (VAS) anchored by statements ranging from *I couldn't control the system at all* to *I could completely control the system*, S1 completed a questionnaire with 10 questions regarding the sEMG cursor control system and the phonemic interface.

Technology

sEMG Cursor Control. The sEMG cursor control system used in these studies provided continuous cursor control as described previously (Cler et al., 2014; Cler & Stepp, 2015). Five sEMG electrodes were placed to record activity of muscles activated during attempted facial gestures (see Figure 1). Facial electrode locations and gestures were chosen to correspond with the velocity of the cursor; for example, when participants contracted muscles in the left cheek and mouth, the cursor moved left (see Cler and Stepp, 2015 for details). Participants controlled both the direction of the cursor, via relative activation of the different sensors; and the speed of the cursor, by the relative magnitude of the activation. Thus, participants could choose to move in only one direction (e.g., up, by raising only their eyebrows and leaving the rest of their facial muscles at rest), but could also move in any 360° direction by combining facial gestures. A small muscle activation corresponded to a slow cursor movement, and a large muscle activation corresponded to a fast cursor movement. Cursor velocity based on these signals was calculated every 100ms, leading to a smooth but responsive cursor trajectory. The sEMG signals were preamplified and filtered using

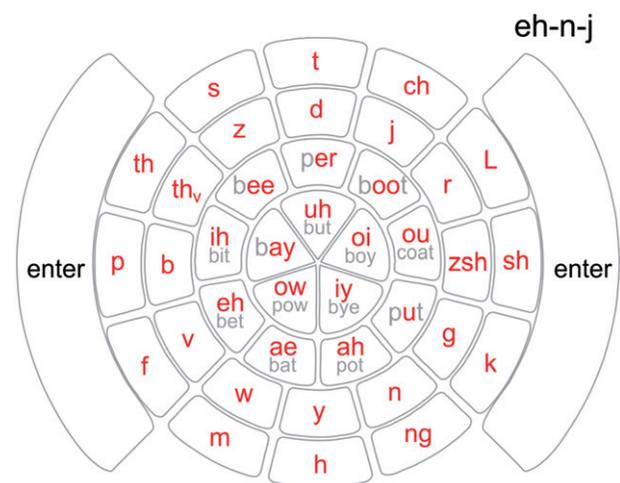


Figure 2. Phonemic interface. Note that arrangement of phonemes is related to articulatory features, so related sounds (e.g., /t/and/d/) are located at the same angle but different radii. Phonemes that have been selected are displayed on the top right of the interface (see “eh-n-j”).

Bagnoli™ 2-channel handheld EMG systems,³ and recorded digitally with National Instruments hardware and custom Python software at 1000 Hz (Cler et al., 2014).

Phonemic Keyboard and Speech Synthesizer. The phonemic keyboard and speech synthesizer used in this study (see Figure 2) were originally developed for use with a variety of inputs, including a finger on a touch screen computer, a typical mouse, or any alternative input method. Users select phonemes arranged in a circular layout based roughly on articulatory features (manner and place of articulation). For example, phonemes that are differentiated only by voicing (e.g., /f/and/v/) are located at the same angle but different radii. As none of the participants had experience with phonemic transcription or the International Phonetic Alphabet (IPA), phonemes were labeled with an English approximation. Vowels and diphthongs were labeled with an example word to clarify their sound (see Figure 2).

When using a typical input method (e.g., finger), the interface is configured to start building a new string of phonemes each time the finger is placed on a phoneme. A string of phonemes continues to be built as the finger slides without lifting and ends when the finger is lifted from the final phoneme. When using an alternative input (e.g., eye tracker, facial sEMG control), users select each phoneme individually by clicking. Once a string of phonemes is selected and the user wishes the utterance to be synthesized, the user selects the enter key and the utterance is synthesized via a concatenative process (i.e., short pre-recorded segments of speech – in this case, diphones to account for coarticulation – are merged) and played over the computer's speaker. Although the interface typically contains delete buttons, these were disabled for this experiment.

³Bagnoli™ 2-channel handheld EMG system is a product of Delsys, Inc. of Natick, MA. <http://www.delsys.com/>

Data collection and analysis

Information transfer rate (ITR) was used to measure performance during each trial. ITRs encapsulate both the speed at which a user selected a series of targets and the accuracy of those selections as compared to the prompt. For experimental purposes, users were not able to make any self-corrections or revisions. ITRs were calculated in bits per minute using Wolpaw's method (Wolpaw et al., 2000), which uses bits per selection (calculated in Equation 1). In this equation, n is the number of potential targets on the screen (38) and a is the accuracy of each trial. Accuracy was estimated from 0 to 1 using the algorithm provided by Soukoreff and MacKenzie (2001) to calculate the minimum distance between the prompt and selected phonemes. The result of Equation (1) in bits/selection is converted to ITR in bits/min by multiplying Equation (1) by the selection rate: the number of selections divided by the time the user took to select the series of phonemes.

$$\text{bits/selection} = \log_2 n + a \times \log_2 a + (1 - a) \times \log_2 \frac{1 - a}{n - 1} \quad (1)$$

ITRs were calculated with custom MATLAB^{TM4} and Python code. Statistical analysis was performed using Minitab^{TM5}. ITRs from the typical mouse control and sEMG cursor control in the training study were tested separately for normality with Kolmogorov–Smirnov tests and then analyzed in separate one-factor repeated measures analyses of variance (ANOVAs) to examine the effect of session (1–8 for sEMG cursor control and pretest plus 1–8 for the typical mouse control) on the outcome measure of mean session ITR, with subject as a random factor. To determine significance, paired t-tests were run between consecutive and one-away-pairs of sessions separately for each input (i.e., sEMG session 1 was compared to sEMG sessions 2 and 3). The Holm-Bonferroni method was used to correct for multiple comparisons. In addition, verbal responses from the case study participant were transcribed. Responses to the visual analog scale were measured continuously from 0 to 10 and converted to percent agreement by multiplying by 10.

Results

Training study

All participants in the training study were able to interact with the phonemic interface using both typical mouse control and facial sEMG control. Their information transfer rates improved through the series of training sessions (see Figure 3). The ITRs generated with the typical mouse were normally distributed (KS test, $p > 0.15$), as were the ITRs generated with the sEMG cursor (KS test, $p > 0.15$). The one-factor, repeated measures ANOVA that examined the effect of training session on

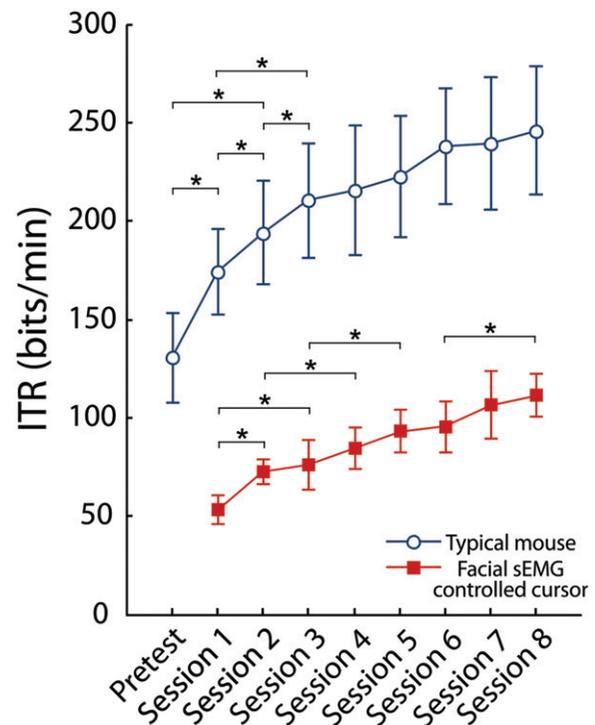


Figure 3. Results of training study in individuals without motor impairments. Mean information transfer rates (ITRs) achieved with typical mouse control of the phonemic interface and speech synthesizer are shown in empty circles. Mean ITRs achieved with facial sEMG control of the phonemic interface and speech synthesizer are shown in filled squares. Error bars represent standard deviations. Statistical significance determined with paired t-tests and Holm-Bonferroni multiple comparison correction is shown via asterisks.

typical mouse ITRs showed a significant effect of session (effect size $\eta_p^2 = 0.86$; $F = 55.0$; $p < 0.001$). The one-factor, repeated measures ANOVA examining the effect of training session on ITRs generated with the sEMG cursor also showed a significant effect of session (effect size $\eta_p^2 = 0.95$; $F = 46.7$; $p < 0.001$). Figure 3 indicates the results of the paired t-tests with multiple comparisons corrected for with the Holm-Bonferroni method.

Although ITR incorporates both speed and accuracy in one measure, it can be useful to look at these separately. Accuracy varied somewhat among subjects, but accuracy during even the first session with the sEMG cursor was high, ranging from 85 to 98% with a mean of 94%. Accuracy during the final session ranged from 88 to 100%, with a mean of 97%. During the first session, participants took an average of 24.9 s to spell one word with a range of 19.2–30.1 s, and by the final session, each word took an average of 11.6 s (range: 9.5–14.0 s). This is an average of 2.9 s per phoneme selected.

Case study

The first session of the case study involved finding an appropriate set of gestures and sEMG sensor locations to provide robust, independent signals for the sEMG cursor system. The final sensor configuration is shown in Figure 4. Most of the sensors were placed in locations utilized in the training study (Figure 1). However, a clear sEMG signal was not detected during an eyebrow raise, perhaps due to muscle damage from GBS. S1 regularly operated a mechanical head switch with a consistent shoulder shrug. During the second session,

⁴MATLABTM is a product of MathWorks, Inc. of Natick, MA. <http://www.mathworks.com/>

⁵MinitabTM is a product of Minitab, Inc. of State College, PA. <https://www.minitab.com/>

then, an electrode was placed on S1's trapezius muscle instead of the frontalis, so that a slight shoulder raise would move the cursor vertically. S1 was able to use the remainder of the same electrode placements and gestures as the participants in the training study.

Using the electrode locations determined in the first session, S1 was able to interact with both an alphabetic interface and the phonemic interface during the second session. S1 was able to use the facial sEMG system to achieve mean ITRs of 12.9 bits/min with an alphabetic interface over 30 trials, and 19.4 bits/min with the phonemic interface over four trials. Figure 5 shows ITRs for blocks of trials with both systems. Although variability was high during these blocks, S1

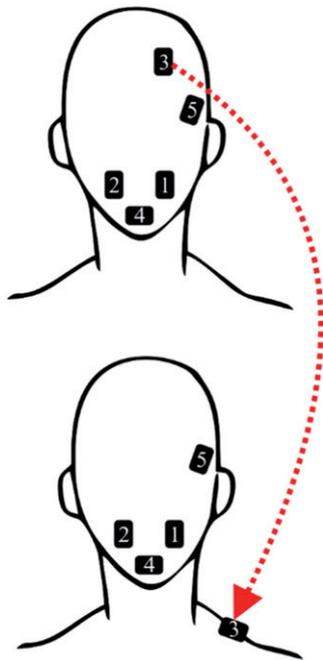


Figure 4. Altered placement of sEMG sensors for case study. All placements/gestures the same as in the training study (Figure 1) except for sensor 3, in which a shoulder raise (rather than eyebrow raise) corresponded to vertical movement of cursor.

improved across the duration of the session, whether using the alphabetic keyboard (see Figure 5, right) or phonemic interface.

S1 also provided verbal feedback about both systems during and following the experiment and completed a researcher-generated questionnaire about the systems. Results from the questionnaire in addition to unprompted statements are shown in Table 1. S1 expressed an interest in using the phonemic interface more and was excited about the concept of selecting phonemes. When asked if performance would improve with practice, S1 said, *Of course. I would probably get very fast with this one once I figure out where all the phonemes are.* S1 rated both "I could completely control the system" and "This system was more flexible than my typical communication device" very highly (94% and 82% agreement, respectively). S1 further noted that the systems felt slower and more tiring than the head-tracker (77% and 87%, respectively); however, S1 was quick to note that the comparison was difficult, as S1 had used the head-tracker for 7 years and was very proficient with it.

Discussion

Training effects

The mean session ITRs produced by participants increased with training for participants from the training study and case study. Participants in the training study increased ITRs from a group mean of 53 bits/min during Session 1 to 111 bits/min during Session 8 using sEMG control. Typical mouse control increased from an average of 130 bits/min during the pretest to 246 bits/min during Session 8. Interestingly, ITRs produced with the typical mouse increased during early sessions (1–3), but appear to show a ceiling effect and stabilize after Session 5 (Figure 3). ITRs generated with the sEMG cursor, however, showed longer-lasting significant increases. This suggests that additional training sessions may provide further ITR increases when using sEMG cursor control. Similarly, the case study

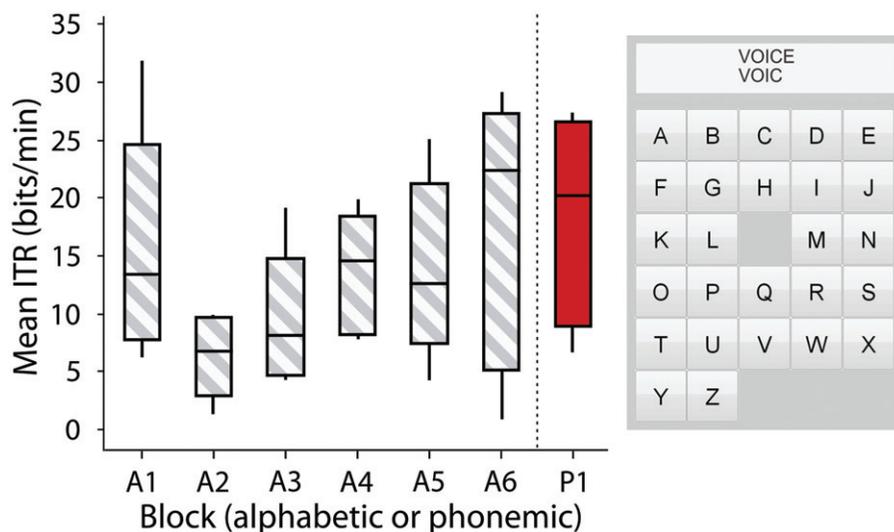


Figure 5. Left: Mean ITRs from blocks of trials in case study. Blocks of trials in which the participant used the sEMG cursor control system to interact with an alphabetic keyboard are blocks A1–A6 (striped); each block consisted of five trials. The block of trials in which the participant used the sEMG cursor to control the phonemic interface is block P1 (filled); this block consisted of four trials. Note that all trials took place during one session. Right: The alphabetic keyboard used in the case study.

Table 1. Questionnaire results.

Category	Statement	Agreement	Additional remarks
sEMG cursor + alphabetic keyboard	I could completely control the system	67%	
	It got easier as the session went on	80%	
sEMG cursor + phonemic interface	I could completely control the system	94%	
	It got easier as the session went on	94%	
sEMG vs. head-tracker	I liked using these systems	53%	<i>It's a hard choice, because I'm used to the AccuPoint [head-tracker]</i>
	I could see myself using this in my daily life	58%	
	This felt slower than my typical communication	77%	
	This felt harder than my typical communication	96%	
	This was more tiring than my typical communication	87%	
Phonemic interface	This system was more flexible than my typical communication device	82%	<i>More intuitive the way you move [referring to the cursor movements required to use the circular layout] I wish I could play with this some more I'm really excited about the concept [Vowels in the center] is a stroke of genius</i>
	[Free answer]	–	

Table 2. Comparisons to other communication systems.

System	ITR range (bits/min)	Reference examples
Other sEMG systems (continuous muscle control)	5–51	Choi et al. (2011); Larson et al. (2013); Vernon and Joshi (2011); Williams and Kirsch (2008)
Non-invasive BCIs	1–24	Nijboer et al. (2008); Sellers et al. (2006); Wolpaw et al. (2002)
Invasive BCIs	5–69	Brunner et al. (2011); Guenther and Brumberg (2011); Hill et al. (2006); Simeral et al. (2011)
Head tracking	78–174	Epstein et al. (2014); Williams and Kirsch (2008)
Alphabetic control of identical sEMG cursor (with training)	70–121	Cler and Stepp (2015)
Eye tracking (includes predictive methods)	60–222	Frey et al. (1990); Higginbotham et al. (2007); Liu et al. (2012); Majoranta et al. (2006)
Mechanical switch (includes predictive methods)	96–198	Higginbotham et al. (2007)

participant appeared to improve throughout this single session, whether interacting with the alphabetic interface or the phonemic interface (Figure 5), likely due to increasing familiarity with the novel interface and input modality.

Comparisons to other alternative input modalities

sEMG control of the phonemic interface produced ITRs that are comparable to other methods available to individuals with motor impairments who typically use AAC (see Table 2). Participants produced a mean ITR of 53 bits/min in Session 1 and 111 bits/min during Session 8. This shows an improvement in ITRs compared to other sEMG cursor control systems (Choi et al., 2011; Larson et al., 2013; Vernon & Joshi, 2011; Williams & Kirsch, 2008), which range from 5 bits/min to 51 bits/min. Many of these sEMG systems use a position-based algorithm, in which activation of one muscle corresponds to the cursor position in the horizontal direction, and activation of a different muscle corresponds to vertical cursor position. Although this method uses fewer muscles than the velocity-based system shown here, users must then use constant graded muscle activation to remain still; a relaxed position corresponds to one corner of the screen, rather than remaining in the location where the user last moved it as with a typical mouse or the sEMG cursor utilized in the current study. This constant muscle activation is likely more fatiguing, as users cannot rest between movements while retaining their position.

Other input modalities available to individuals with minimal movement capabilities include eye tracking, head tracking, brain-computer interfaces (BCIs), and mechanical switch options. These input modalities can produce ITRs in a wide range, often due to the employment of language or letter predictive algorithms. Invasive BCIs range from 5.4–69 bits/min (Brunner, Ritaccio, Emrich, Bischof, & Schalk, 2011; Guenther & Brumberg, 2011; Hill et al., 2006; Simeral, Kim, Black, Donoghue, & Hochberg, 2011), whereas non-invasive BCIs can range from 1.8–24 bits/min (Nijboer et al., 2008; Sellers, Krusienski, McFarland, Vaughan, & Wolpaw, 2006; Wolpaw et al., 2002). Eye tracking can produce ITRs in the range of 60–222 bits/min (Frey et al., 1990; Higginbotham et al., 2007; Liu et al., 2012; Majoranta et al., 2006), and ITRs produced with head tracking devices range from 78–174 bits/min (Epstein, Missimer, & Betke, 2014; Williams & Kirsch, 2008). Mechanical switches can produce ITRs in the range of 96–198 bits/min. While the ITRs produced by the sEMG cursor are well within these ranges (111 bits/min), future directions include incorporating predictive methods, which have been shown to increase ITRs by as much as 100% (Liu et al., 2012).

The case study participant had 7 years of experience using a head tracking system, but noted that it was difficult to communicate with care staff at night. Individuals with complex medical needs may require ventilation that can limit natural speech. Camera-based cursor control systems require that the specific muscles that control eye or head movements are spared. Performance using these systems is also degraded with head movement or changes in ambient lighting (Higginbotham

et al., 2007). The sEMG cursor control discussed here, however, does not require any specific lighting, as user intent is captured through surface electromyography rather than by a camera.

Although the sEMG sensors in these studies were placed on the face of the participants without motor impairment and the face and shoulders of the case study participant, the system can be used with any arrangement of five sensors that target independent muscle activation. This suggests that the sEMG cursor could be a viable option for individuals who do not have the head or eye control required for other devices, but do retain control of other trunk, limb, or facial muscles (as in some disease trajectories in ALS or multiple sclerosis, or in diseases characterized by paresis). Finally, sEMG can capture activity in muscles that are innervated but do not support movement as in hemiparesis (Saxena et al., 1995); it has also been shown that 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 that may not be detectable by mechanical interfaces or camera-based devices. Although this is not relevant to all possible users, those whose illnesses are characterized by paresis (e.g., incomplete spinal cord injury) may not have enough muscle strength to make hand movements, but may still be able to produce reliable and detectable sEMG signals in order to control the sEMG cursor with these same muscles. Thus, sEMG-based cursors may be a viable option, either in place of or in addition to other input devices. Further study is needed to determine the sEMG cursor's efficacy and reliability in a variety of potential populations.

Comparisons to other interfaces

Trinh et al. (2012) showed that individuals without motor impairments could produce speech using a touchscreen phonemic interface and synthesizer in which participants were presented with a subset of phonemes at a time, rather than all phonemes simultaneously. Thus, choosing one phoneme to synthesize took one to three intermediate selections using a touchscreen. After two training sessions, participants without motor impairments produced ITRs of 24 bits/min when the system did not incorporate prediction and 60 bits/min when a predictive language model was incorporated.⁶ After two training sessions with our novel phonemic interface, participants produced ITRs of 76 bits/min with the sEMG cursor and 210 bits/min with a typical mouse. Although it is difficult to compare these interfaces as they have been tested with different input modalities, these results suggest that an interface in which the entire set of phonemes is available for selection at all times may still facilitate robust ITRs. The methods employed by Trinh et al. may provide additional

benefits to individuals with very little residual motor control, or those who have additional cognitive impairments.

Although it is not our intent to directly compare performance on this phonemic interface to performance on alphabetic interfaces, some conclusions can be drawn from ITRs produced in this study versus those produced in a similar previous study (Cler & Stepp, 2015). In Cler and Stepp (2015), participants without motor impairments used a similar sEMG cursor to interact with a grid of alphabetic targets over four training sessions. During the fourth and final session, participants achieved mean ITRs between 96 bits/min and 135 bits/min (group mean = 121 bits/min; $SD = 12$ bits/min). In the present study, participants using the phonemic interface and a similar sEMG cursor achieved mean ITRs between 66 and 99 bits/min (group mean = 85; $SD = 11$ bits/min) after four sessions. However, participants achieved mean ITRs between 92 bits/min and 130 bits/min in their eighth training session (group mean = 111 bits/min; $SD = 12$ bits/min). While a direct comparison is not feasible given other experimental variables present in the Cler and Stepp study (e.g., different parameters used in the sEMG cursor; different algorithms for calculating accuracy for ITRs due to the different interfaces), it does suggest that although participants using a phonemic interface do initially show lower performance than those using an alphabetic interface, participants reached similar levels of performance after a small number of additional sessions. This is promising for future work using phonemic interfaces, as it suggests that individuals become quickly proficient with the novel phonemic targets.

Selecting phonemes may provide additional benefits not represented in the speed and accuracy of selection captured by ITR. For example, the ITR calculations shown here only consider selections per minute, rather than a more direct comparison of actual semantic or phonemic information per minute. Individual words found on published vocabulary lists appropriate for adults who use AAC (Bryen, 2008) show an average of 14% savings in selections, while a preliminary and condensed list of AAC messages suggested for ALS users (Beukelman & Gutmann, 1999) showed a 20% savings in the number of selections that would be necessary to produce each word or phrase. Finally, phonemes allow individuals who use AAC to have complete control over the sounds produced by the speech synthesizer, instead of relying on text-to-speech (TTS) rules.

The particular arrangement of phonemes by articulatory features may provide benefits that are also not reflected in ITRs. The relation of the layout of the phonemes to the configuration of the vocal tract when producing speech may provide additional cues when users are learning the locations of phonemic targets. In addition, phonemes are paired based on voicing, such that / f / and / v / are neighboring, as are other pairs such as / p / and / b / and / θ / and / ð /. If a user intended to select the phonemes / b / / g / / n / and instead selected / p / / g / / n / due to a precision error (e.g., accidentally selecting / p / instead of its neighbor / b /), the synthesized output would still likely be perceivable as "begin" in context, unlike if a user were trying to type b-e-g-i-n on a typical keyboard and instead typed n-e-g-i-n (e.g., accidentally selecting n while attempting to select its neighbor b). The ITR

⁶It is important to note, however, that the participants were prompted with an aural phrase, rather than also being shown a specific sequence of phonemes as in the current studies. The additional mental processing time to translate an aural phrase to phonemes would likely result in lower ITRs.

calculation considers these errors equivalent. However, if intelligibility were the outcome measure, the phonemic interface would likely be more intelligible in these cases, and thus could be considered to be more tolerant of certain types of precision errors. Future studies will examine these intelligibility benefits in more detail, as well as the selection benefits and limitations of using phonemic input instead of alphabetic input.

Limitations and implications for AAC access and interface design

These results have implications for future study of innovative AAC devices. Ease and speed of device use can increase significantly over time, particularly when devices require novel motor actions; thus, experimental protocols intending to produce reports of usability and speed of use should include training. sEMG cursor control represents a viable option for individuals with widespread paralysis and some residual muscle control. Although future work should further explore its usability more widely in the target population, this initial case study shows promise. In addition, the benefit of sEMG as a control modality that works at night (in low light and when individuals may use ventilators) should be specifically investigated. To be clinically viable, the selection of electrode locations for individual users will need to be streamlined; in this case study, the researchers spent one session attempting to find gestures and sensor locations that were easily repeatable by the participant and provided a strong sEMG signal. Future work will be directed toward devising a protocol that incorporates both the expertise of physical and/or occupational therapists along with machine learning in order to select gestures and sensor locations quickly and automatically. Finally, this approach is not dependent on any particular sEMG sensor; future improvements in sEMG technology would be compatible with our sEMG cursor software and interface, including new hardware that is smaller and less obtrusive.

Results also suggest that this phonemic interface is promising for further evaluation and development. In future evaluations of the phonemic interface, participants who use AAC would ideally complete a variety of tasks, including the fully prompted protocol employed here, structured productions with aural prompting only, open-ended responses to questions, and self-generative tasks. Incorporating a predictive language model based on what is known about phonotactic probability and neighborhood density would assist in language generative tasks. Predictive language models may increase these ITRs by as much as 100% (Trinh et al., 2012), allowing individuals to produce their intended message as quickly and accurately as possible, while minimizing possible fatigue. These studies would help elucidate the amount and type of training needed for phonemic communication interfaces, as well as the comparative benefits of phoneme-based prediction and orthographic word prediction.

Further work is needed to determine if there are benefits in ITRs or ease of use with a variety of phonemic target arrangements, as seen here with the unanticipated error

tolerance. Future development is also needed to produce phonemic interfaces that give individuals who use AAC full control over their synthesized voice, both in terms of speed and accuracy and in terms of producing speech with user-defined prosody. Evaluations of future systems will also include additional ratings of system usability by participants who use AAC, and will incorporate measures of speed and accuracy that compare the amount of semantic and/or phonemic information conveyed per minute rather than just selections per minute to fully capture the benefits and limitations of phonemic interfaces as compared to alphabetic interfaces.

Finally, although our case study indicates that both of these technologies (i.e., the sEMG cursor and phonemic interface) show promise for further development, more evaluation in this population is required. For example, time constraints meant that the individual with severe paralysis was only able to participate in two sessions, rather than the eight sessions completed by the participants without motor impairments. The first session consisted entirely of modifying the sensor locations rather than using the interfaces, further limiting our ability to evaluate the expected improvements with training. Whereas the participants without motor impairments used a typical mouse as an input modality for brief sessions, this modality was not available to the user with motor impairments. Use of a typical mouse may have aided learning. Future work is planned to evaluate these technologies in this population over time and to incorporate feedback and suggestions from individuals with minimal movement capabilities.

Conclusion

In this paper, we have presented an evaluation of two complementary AAC components: a phonemic interface and a cursor control system using sEMG to capture residual muscle activity from innervated and spared facial muscles. Results of a training study in 10 healthy participants over eight sessions showed high performance of the phonemic interface, both when controlled with a typical mouse and with the sEMG-controlled cursor. Participants had statistically significant improvements in ITRs over time with the typical mouse until Session 3 and had a trend of increased ITRs through Session 6. Conversely, the participants showed statistically significant increases in ITRs with the sEMG-controlled cursor throughout all training sessions, suggesting possible further improvements with additional training. A case study involving one individual with minimal movement capabilities showed initial feasibility of the facial sEMG-controlled cursor in this population, and the user provided positive feedback about the phonemic interface. Both studies show promise for the phonemic interface and sEMG-controlled cursor in future AAC applications.

Acknowledgements

The authors wish to thank Carolyn Michener for her assistance with data collection.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Funding information

This research is supported by NSF grant 1452169 and NIH grants 5T90DA032484-04 and R01 DC002852.

References

- Alant, E., Life, H., & Harty, M. (2005). Comparison of the learnability and retention between blissymbols and cyberglyphs. *International Journal of Language & Communication Disorders, 40*, 151–169. doi: 10.1080/13682820400009980.
- Basmajian, J.V. (1972). Electromyography comes of age. *Science, 176*, 603–609. doi: 10.1126/science.176.4035.603.
- Beukelman, D., & Gutmann, M. (1999). Generic message list for AAC users with ALS, from 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*, 230–242. doi: 10.1080/07434610701553668.
- Beukelman, D.R., & Mirenda, P. (2013). *Augmentative & alternative communication: Supporting children & adults with complex communication needs* (4th ed.). Baltimore: Paul H. Brookes Pub.
- 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, Montreal, CA.
- Brunner, P., Ritaccio, A.L., Emrich, J.F., Bischof, H., & Schalk, G. (2011). Rapid Communication with a “P300” Matrix Speller Using Electroencephalographic Signals (ECoG). *Frontiers in Neuroscience, 5*. doi: 10.3389/fnins.2011.00005.
- Bryen, D.N. (2008). Vocabulary to support socially-valued adult roles. *Augmentative and Alternative Communication, 24*, 294–301. doi: 10.1080/07434610802467354.
- Choi, C., Rim, B.C., & Kim, J. (2011). *Development and evaluation of an assistive computer interface by sEMG for individuals with spinal cord injuries*. Paper presented at the IEEE International Conference on Rehabilitation Robotics, Zurich. doi: 10.1109/icorr.2011.5975386.
- Cler, M.J., Nieto-Castanon, A., Guenther, F.H., & Stepp, C.E. (2014). *Surface electromyographic control of speech synthesis*. Paper presented at the Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 26–30 August, Chicago, IL. doi: 10.1109/embc.2014.6944958.
- Cler, M.J., & Stepp, C.E. (2015). Discrete vs. Continuous mapping of facial electromyography for human-machine-interface control: Performance and training effects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 23*, 572–580. doi: 10.1109/tnsre.2015.2391054.
- Epstein, S., Missimer, E., & Betke, M. (2014). Using kernels for a video-based mouse-replacement interface. *Personal and Ubiquitous Computing, 18*, 47–60. doi: 10.1007/s00779-012-0617-z.
- Fager, S., Bardach, L., Russell, S., & Higginbotham, J. (2012). Access to augmentative and alternative communication: New technologies and clinical decision-making. *Journal of Pediatric Rehabilitation Medicine, 5*(1), 53.
- Frey, L.A., White, Jr K.P., & Hutchison, T.E. (1990). Eye-gaze word processing. *IEEE Transactions on Systems, Man and Cybernetics, 20*, 944–950. doi: 10.1109/21.105094.
- Guenther, F.H., & Brumberg, J.S. (2011). *Brain-machine interfaces for real-time speech synthesis*. Paper presented at the Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA. doi: 10.1109/iembs.2011.6091326.
- Higginbotham, D.J., Shane, H., Russell, S., & Caves, K. (2007). Access to AAC: Present, past, and future. *Augmentative and Alternative Communication, 23*, 243–257. doi: 10.1080/07434610701571058.
- Hill, N.J., Lal, T.N., Schroder, M., Hinterberger, T., Wilhelm, B., Nijboer, F., ... Birbaumer, N. (2006). Classifying EEG and ECoG signals without subject training for fast BCI implementation: comparison of nonparalyzed and completely paralyzed subjects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14*, 183–186. doi: 10.1109/TNSRE.2006.875548.
- 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*, 921–930. doi: 10.1682/jrrd.2007.06.0096.
- Koppenhaver, D.A., & Yoder, D.E. (1992). Literacy issues in persons with severe speech and physical impairments. In R. Ross-Gaylord (Ed.), *Issues and research in special education* (Vol. 2, pp. 156–201). New York, NY: Teachers College Press, Columbia University.
- Kushler, C. (1998). *AAC: Using a reduced keyboard*. Paper presented at the CSUN '98, Northridge, CA.
- Larson, E., Terry, H.P., Canevari, M.M., & Stepp, C.E. (2013). Categorical vowel perception enhances the effectiveness and generalization of auditory feedback in human-machine-interfaces. *PLoS One, 8*, e59860. doi: 10.1371/journal.pone.0059860.
- 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*, 37–42. doi: 10.5405/jmbe.836.
- Majaranta, P., MacKenzie, I.S., Aula, A., & Riih , K.-J. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Universal Access in the Information Society, 5*, 199–208. doi: 10.1007/s10209-006-0034-z.
- Nijboer, F., Sellers, E.W., Mellinger, J., Jordan, M.A., Matuz, T., Furdea, A., ... Kubler, A. (2008). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology, 119*, 1909–1916. doi: 10.1016/j.clinph.2008.03.034.
- Orhan, U., Hild, K.E., Erdogmus, D., Roark, B., Oken, B., & Fried-Oken, M. (2012). *RSVP keyboard: An EEG based typing interface*. Paper presented at ICASSP 2012 – 2012 IEEE International Conference on Acoustics, Speech and Signal Processing, March 25–30, Kyoto, Japan. doi: 10.1109/icassp.2012.6287966.
- Saxena, S., Nikolic, S., & Popovic, D. (1995). An EMG-controlled grasping system for tetraplegics. *Journal of Rehabilitation Research and Development, 32*, 17–17.
- 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*, 242–252. doi: 10.1016/j.biopsycho.2006.04.007.
- Simeral, J.D., Kim, S.P., Black, M.J., Donoghue, J.P., & Hochberg, L.R. (2011). Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array. *Journal of Neural Engineering, 8*, 025027. doi: 10.1088/1741-2560/8/2/025027.
- Smith, M.M. (2001). Simply a speech impairment? Literacy challenges for individuals with severe congenital speech impairments. *International Journal of Disability, Development and Education, 48*(4), 331–353. doi: 10.1080/10349120120094257.
- Soukoreff, R.W., & MacKenzie, I.S. (2001). *Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic*. Paper presented at the CHI'01 extended abstracts on Human factors in computing systems. doi: 10.1145/634067.634256.
- 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 ASSETS'12, Boulder, Colorado. doi: 10.1145/2384916.2384927.
- Vernon, S., & Joshi, S.S. (2011). Brain–muscle–computer interface: Mobile-phone prototype development and testing. *IEEE Transactions on Information Technology in Biomedicine, 15*, 531–538. doi: 10.1109/titb.2011.2153208.

- 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, 485–496. doi: 10.1109/TNSRE.2008.2006216.
- 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, 164–173. doi: 10.1109/tre.2000.847807.
- Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., & Vaughan, T.M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113, 767–791. doi: 10.1016/s1388-2457(02)00057-3.
- Yorkston, K., Beukelman, D.R., Hakel, M., & Dorsey, M. (2007). *Speech Intelligibility Test (SIT) [Computer Software]*. Lincoln, NE: Madonna Rehabilitation Hospital.