Evaluation of a Speech Recognition Prototype for Speakers with Moderate and Severe Dysarthria: A Preliminary Report

SUSAN K. FAGERa*, DAVID R. BEUKELMANb, TOM JAKOBSc and JOHN-PAUL HOSOMd

aInstitute for Rehabilitation Science and Engineering, Madonna Rehabilitation Hospital, Lincoln, Nebraska, bUniversity of Nebraska, Lincoln, cInvoTek Inc., Alma, AR and dOregon Health and Science University, Portland, OR, USA

This study described preliminary work with the Supplemented Speech Recognition (SSR) system for speakers with dysarthria. SSR incorporated automatic speech recognition optimized for dysarthric speech, alphabet supplementation, and word prediction. Participants included seven individuals with a range of dysarthria severity. Keystroke savings using SSR averaged 68.2% for typical sentences and 67.5% for atypical phrases. This was significantly different to using word prediction alone. The SSR correctly identified an average of 80.7% of target stimulus words for typical sentences and 82.8% for atypical phrases. Statistical significance could not be claimed for the relations between sentence intelligibility and keystroke savings or sentence intelligibility and system performance. The results suggest that individuals with dysarthria using SSR could achieve comparable keystroke savings regardless of speech severity.

Keywords: Speech recognition; Dysarthria; Supplemented speech; Voice recognition

INTRODUCTION

A range of computer access options for people with severe physical conditions and limited movement capabilities has been developed over the years, including head tracking, eye tracking, ergonomically-designed keyboards and mouse options, one-handed keyboards, onscreen keyboards, scanning software applications, and automatic speech recognition (ASR) technology (Beukelman, Garrett, & Yorkston, 2007). While progress in developing new access technologies is encouraging, there is an ongoing need to design technologies that are efficient, minimally fatiguing, and preferred by individuals with severe physical conditions. Many people with moderate and severe dysarthria retain some natural speaking abilities, which they use to communicate with familiar people about predictable messages. While ASR technology has been useful for those with spinal cord injuries (SCI) and typical speech who experience reduced keyboard efficiency due to limited hand movement, the usefulness of this technology for individuals with severe physical impairments and co-occurring dysarthria has been limited.

If ASR technology could recognize the speech of people with dysarthria, it might be an effective interface access strategy for computer text generation. However, for those with severe physical impairments and co-occurring dysarthria, attempts at functional ASR use with standard, commercially available technology (e.g., Dragon NaturallySpeaking) have been either largely unsuccessful or successful for only a limited dictionary of words (Coleman & Meyers, 1991; Ferrier, Jarrell, Carpenter, & Shane, 1992; Ferrier, Shane, Ballard, Carpenter, & Benoit, 1995; Fried-Oken, 1985; Kotler & Thomas-Stonell, 1997; Raghavendra, Rosengren, & Hunnicutt, 2001).

Standard, Commercially Available ASR Technology and Dysarthria

In order to understand ASR technology, one must understand the basic terminology that
describes different variations of these systems. ASR can be described as either discrete utterance or continuous speech based; as well as speaker dependent, speaker independent, or speaker adaptive (Rosen & Yampolsky, 2000; Venkata-giri, 2002). Discrete-utterance-based ASR requires the speaker to insert pauses between words when dictating. Continuous speech-based systems allow the user to dictate a continuous stream of speech without slowing speech rate or inserting unnatural pauses. Speaker-dependent ASR requires that the technology be trained to recognize each individual's voice. With these systems, the need to train every word to be recognized often results in significantly reduced vocabulary sets. Speaker-independent systems have been developed based on a standard speech template that does not require individual training. These systems are commonly used in automated telephone menus. Speaker-adaptive systems develop and refine the recognition templates while in use. These systems do not require extensive training before they can be used; however, extremely low recognition rates are common during the first few uses. If the speaker diligently corrects all misrecognitions, the system eventually adapts to the speaker's unique voice and recognition rates begin to increase.

A number of studies have documented trials of standard, commercially available ASR technology by speakers with dysarthria. Many of the earliest reports describe accuracies of 40–96% using discrete-utterance, speaker-dependent systems on limited vocabulary sets (alpha-numeric codes and small lists of words from standard reading passages), with individuals with mild dysarthria achieving the highest accuracies (Coleman & Meyers, 1991; Ferrier et al., 1992; Ferrier et al., 1995; Fried-Oken, 1985; Kotler & Thomas-Stonell, 1997). Some have compared the performance of speaker-adaptive and speaker-dependent systems using small vocabulary sets and have found that for those with mild dysarthria, recognition rates were high on both technologies; however, for those with severe dysarthria, the results varied considerably between participants (Raghavendra et al., 2001). The researchers noted the differences in the speech characteristics demonstrated by the participants with severe dysarthria as an explanation for the differences in accuracies.

One of the few studies that have examined the use of a discrete utterance versus a continuous system included an individual with mild dysarthria due to traumatic brain injury (TBI) and a typical speaker (Hux, Rankin-Erickson, Manasse, & Lauritzen, 2000). Both the control speaker and the speaker with mild dysarthria were able to achieve higher recognition rates with the continuous speech system. Other studies with individuals who have mild dysarthria using continuous speech ASR technology have produced similar results (Manasse, Hux, & Rankin-Erickson, 2000).

What is clear from these reports is the difficulty that standard, commercially available ASR technology has in recognizing the speech of individuals with moderate to severe dysarthria. Rosen and Yampolsky (2000) provided an extensive review of ASR technology and described the affect that acoustic characteristics of dysarthria have on recognition accuracy. Blaney and Wilson (2000) identified intra-speaker variability across several acoustic parameters as a reason for low recognition rates for individuals with moderate dysarthria. While some levels of ASR accuracy with standard, commercially available technology have been reported, it should be noted that this usually involved speakers who had mild dysarthria and tasks that included very limited vocabulary sets. No reports have documented the functional, daily use of standard, commercially available ASR by individuals with severe dysarthria to support written communication needs.

The authors of this study informally surveyed 10 clinical specialists who were not involved in the current study, but who served people with dysarthria in regional clinics, to determine if these specialists had served individuals with moderate to severe dysarthria who routinely used ASR technology. Routine use was defined as use of ASR for communication interaction, word processing, email, and/or Internet access. None of the clinical specialists reported having individuals with moderate to severe dysarthria who routinely used ASR in the past 5 years. Any reports of routine use of standard, commercially available ASR technology related to individuals who had very mild or no dysarthria.

**ASR Adapted for Dysarthria**

Due to the challenge individuals face using standard, commercially available ASR, extensive efforts have been underway to develop ASR technology based on models of dysarthric speech (Chen & Kostov, 1997; Omar, Morales, & Cox, 2009; Patel & Roy, 1998; Polur & Miller, 2005; Sawhney & Wheeler, 1999; Wan & Carmichael, 2005). The STARDUST (Speech Recognition for People with Severe Dysarthria) and SPECS (Speech-driven Environmental Control Systems) projects developed ASR based on models of dysarthric speech (Green et al., 2003; Hatzis et al., 2003; Hawley et al., 2003). These projects targeted the development of a speech-controlled environmental control system (ECS) for severe dysarthria (Hawley, 2002; Hawley et al., 2007;
The ENABL (Enabler for Access to Computer-Based Vocational Tasks with Language and Speech) project adapted an ASR to the speech of individuals with dysarthria to control an engineering design system (Magnuson & Blomberg, 2000; Rosengren, 2000; Talbot, 2000). This system utilized sentence-level commands to control the engineering design program. For a male participant with dysarthria, words correctly identified by the ASR increased from 22.7 to 53%.

Researchers associated with the Rehabilitation Engineering Research Centre for Communication Enhancement (AAC-RERC) in the United States have also developed an acoustic model based on the speech of individuals with a range of dysarthria severity due to amyotrophic lateral sclerosis (ALS). Data collection for the development of the model included recordings of participants reading the numbers 0 through 9. Initial tests with the prototype model resulted in 86.24% word recognition for a female with mild dysarthria, 70.27% for a female with moderate dysarthria, 75.28% for a male with moderate dysarthria, and 51.25% for a female with severe dysarthria (Caves, Boemler, & Cope, 2007).

Supplemented Speech Recognition Prototype Development

Preliminary work has been completed by some of the authors on a prototype application based on models of dysarthric speech that will be described in this report. This prototype application, called supplemented speech recognition (SSR), is a “hybrid” approach to ASR that incorporates first-letter identification as well as word prediction, which make it potentially more efficient for speakers with moderate to severe dysarthria versus technology that relies on the speech signal alone. While some commercially available ASR technologies incorporate word prediction (e.g., SpeakQ2) the addition of the first-letter cue makes this system unique. Additionally, using first-letter cues (i.e., alphabet supplementation) is a strategy that many individuals with dysarthria are familiar with, as they use it to compensate for decreased intelligibility (Hanson, Yorkston, & Beukelman, 2004; Hanson, Beukelman, Heidemann, & Shutts-Johnson, 2009; Hustad, Auker, Natale, & Carlson, 2003; Hustad & Beukelman, 2001; Hustad, Jones, & Dailey, 2003).

The following includes a description of (a) the development of the ASR technology incorporated by the SSR system, (b) the word prediction algorithm that was employed and how it interacts with the ASR, (c) the user interface, and (d) how the entire prototype system functions for an individual using the system.

ASR technology incorporated by SSR

The speaker-dependent, discrete-utterance SSR system developed for this work was trained on each individual's voice using a single recording of each of 211 words. The researchers were cognizant of the time and effort individuals with dysarthria require to train speaker-dependent ASR technology; therefore, an effort was made to reduce the number of words required to train the system to recognize the individual’s speech. The list of 211 words was chosen to be the smallest subset possible of the initial target list of 500 words, while covering a maximum number of phonemes in different phonetic contexts. A phoneme-level Hidden Markov Model (HMM) was employed with words represented as a sequence of context-independent phonemes. The pronunciation of each word was made as general as possible, to allow for phonetic variation among the participants with dysarthria. Training of the entire system on 211 words required about 10 min on a 2 GHz computer.

Recognition was performed using a total vocabulary of 500 words, restricted to those words beginning with the letter indicated by the user. The list of 500 words included the 211 words used in training. More detail about the HMM/ANN framework and feature set can be found in the description of the baseline system described by Hosom for the task of forced alignment (Hosom, 2009). This general approach has yielded state-of-the-art results on tasks such as connected digit recognition (Hosom, Cole, & Cosi, 1998).

While the system described in this prior publication is for speaker-independent processing, the system described here is speaker-dependent because training data comes from a single speaker.

Word prediction in SSR

The word prediction component to the SSR was developed custom to the project by Invotek, Inc.,
using a bigram language model. The following is a description of how the word prediction algorithm worked in conjunction with the SSR system described above: First, the word prediction software generated two lists based on the last word in the text being generated and the first letter of the target word that was entered by the user. This first list was based on a 5,000-word database that included the probability of every word in the database, given the previous word that was generated in the text. A second list, based on the first letter of the target word that was typed, was then generated. This second list was a subset of the first list and only contained the words known to the SSR system. The SSR system then attempted to match the acoustic signal of the spoken word produced by the individual with dysarthria with the list produced by the SSR system. The best match was entered automatically into the text in the message window and the next five most probable alternative words were displayed on the word-prediction buttons.

The SSR interface

During this preliminary investigation, the participants with dysarthria did not interact with the user interface (other than providing the spoken target word). The following is a description of the interface, to allow the reader to understand how the prototype system functioned: The user interface of the prototype system consisted of an on-screen keyboard in alphabetized layout, a message window and five word-prediction buttons that were positioned along the right-hand side of the message window (Figure 1). The interface was limited to five word-prediction buttons, that were vertically aligned based on previous reports of the cognitive demands related to scanning word-prediction buttons, in an attempt to reduce these demands for individuals using the system (Koester & Levine, 1998; Lesher, Moulton, & Higginbotham, 1999). The interface was similar in design to many interface display options in commercially available speech generating devices (e.g., Vmax by Dynavox Mayor-Johnson).

Functioning of SSR

The SSR software prototype system functioned in the following manner: The first letter of the target word of each sentence was typed using the onscreen keyboard on a touch-screen tablet computer. This letter was displayed in the message window. The participant would then hear a beep, which was the auditory prompt for him or her to speak the target word. The word that the system interpreted to be the closest match produced by the participant would then appear in the message window, and the next five probable choices would be displayed alphabetically on the word-prediction buttons to the right of the message window. If the word was correctly displayed in the message window, this procedure would be repeated for the next word in the sentence. However, if the word in the message window was not correct, but the correct word was available on one of the word-prediction buttons, that word would be selected and it would replace the incorrect word in the message window. If the correct word was not displayed in the message window or on any of the word-prediction buttons, subsequent letters of the target word would be entered until the correct word appeared on one of the word-prediction buttons or until the participant had typed it in its entirety. Once the correct word was in the message window this process would be repeated for subsequent words within the target sentence.

Purpose

The purpose of this project was to investigate the performance of the SSR prototype system for speakers with a range of dysarthria severity. The specific research questions were:

(a) How does keystroke savings derived from using SSR compare to conventional text generation using word prediction only? (b) How does the combination of ASR and word prediction (provided through the SSR system) impact word recognition accuracy compared to ASR alone, and (c) What are the relationships between keystroke savings and speech intelligibility and SSR performance and speech intelligibility?
METHOD

Participants

Seven speakers with dysarthria participated in this project. One participant had cerebral palsy and the remaining six had a primary diagnosis of traumatic brain injury (TBI). The participants included six males and one female with an age range of 24–51 years. The type of dysarthria each participant exhibited was determined by the lead authors using the Darley, Aronson, and Brown (1969) classification of dysarthria. One participant demonstrated spastic dysarthria and the remaining six participants demonstrated mixed dysarthria. Intelligibility was measured using the Speech Intelligibility Test (SIT, Yorkston, Beukelman, Hakel, & Dorsey, 2007). Intelligibility scores ranged from 16.4–89.1%. Two of the seven participants routinely used augmentative and alternative communication (AAC) devices and strategies to support communication. Table 1 provides the demographic information of all of the participants.

Materials

Software

The software designed for the project was the SSR prototype system described previously. The software interface included an onscreen keyboard, message window, and word-prediction buttons (Figure 1).

Hardware

A Toshiba Satellite R20 4 tablet computer with an Intel Pentium 2 GHz 5 processor hosted the software. The laptop touch screen was used to identify first letters of target words. An Andrea NC-7100 USB head mounted microphone was used to capture the audio signals from the participants with dysarthria.

Target stimuli

The target stimuli consisted of 17, non-standardized sentences and phrases developed from the 500 words in the SSR master dictionary.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Diagnosis</th>
<th>Age</th>
<th>Dysarthria type</th>
<th>Sentence intelligibility</th>
<th>AAC to support communication?</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>CP</td>
<td>35</td>
<td>Spastic</td>
<td>16.4%</td>
<td>Yes</td>
</tr>
<tr>
<td>P2</td>
<td>TBI</td>
<td>42</td>
<td>Mixed-spastic</td>
<td>45.0%</td>
<td>No</td>
</tr>
<tr>
<td>P3</td>
<td>TBI</td>
<td>48</td>
<td>Mixed</td>
<td>50.0%</td>
<td>Yes</td>
</tr>
<tr>
<td>P4</td>
<td>TBI</td>
<td>38</td>
<td>Mixed-flaccid</td>
<td>74.6%</td>
<td>No</td>
</tr>
<tr>
<td>P5</td>
<td>TBI</td>
<td>51</td>
<td>Mixed</td>
<td>77.3%</td>
<td>No</td>
</tr>
<tr>
<td>P6</td>
<td>TBI</td>
<td>24</td>
<td>Mixed</td>
<td>82.7%</td>
<td>No</td>
</tr>
<tr>
<td>P7</td>
<td>TBI</td>
<td>45</td>
<td>Mixed</td>
<td>89.1%</td>
<td>No</td>
</tr>
</tbody>
</table>

Note. CP = cerebral palsy; TBI = traumatic brain injury.

Figure 2. Average keystroke savings for each speaker participant (ordered by sentence intelligibility) with researcher-controlled touchscreen computer using full SSR on typical and atypical sentences compared the max keystroke savings possible using only word prediction on typical sentences and atypical phrases.
(Appendix A); 10 of these sentences were typical in that they were grammatically correct as judged by the researchers. The typical sentence set had a predictability index of 50% using the assessment reported by Cannito, Suiter, Chorna, Beverly, Wolf, Watkin, and Pfeiffer (2008). One of the side effects of a small dictionary of common words for the ASR was that, in typical sentences, the words were often easily predicted by the word prediction language model alone. To get a better sense of the SSR performance, an atypical stimulus set of seven phrases and incomplete sentences was developed. The researchers judged this set to be phrases and incomplete sentences that were consistent with spoken conversation. Many individuals with moderate to severe dysarthria rely on AAC strategies to support their communication. These conversational phrases were chosen, in addition to the typical sentences, because in the future some individuals may use the SSR to support spoken communication.

Procedures
The researcher controlled the touch screen, in order to control for participants’ varying levels of manual and visual capability and reduce their learning time. Individuals with dysarthria often present with a myriad of visual and motor deficits. For example, in the current study, three of the participants demonstrated visual field cuts, one required switch-scanning to use an AAC device, two required head-tracking technology to control a computer cursor, and one required changes in select time on a touch-screen device due to ataxic limb movements. At the time of the experiment, the prototype system did not accommodate all of the potential access needs that individuals with severe motor and visual impairments may require (alternative access methods such as head/eye control, switch-scanning capabilities, modifications to touch-screen sensitivity, etc.). Additionally, substantial learning and practice may be required in order for an individual to become competent using a system that requires accurate first-letter identification as well as the ability to use word prediction. While these are important considerations to be addressed in future studies, the cognitive, visual, and motor requirements of using such a system pose potentially confounding variables. The goal of this preliminary work was to examine how the system performed using real-time dysarthric speech.

Seven individuals with dysarthria participated in the first part of the project. The researcher presented each word of the target sentence orally. The researcher typed the first letter of the target word. Next, the participant heard a beep as a prompt to say the target word. The participant spoke the word, and the SSR placed the most probable word into the message window and provided the next five probable word options on the word-prediction buttons. The researcher then did one of the following: (a) proceeded to the next target word in the sentence if the word was accurately displayed in the message window, (b) selected the target word from the word prediction list, or (c) spelled the word using the onscreen keyboard portion of the interface. A measure of keystroke savings was calculated during this task. The percentage of time during which the target word appeared in the message window and on the
word-prediction buttons was also calculated and used to represent SSR system performance. For the second portion of the project, the researcher entered each target sentence/phrase with the ASR turned off (no speech recognition). The word prediction was active and when target words appeared on the word-prediction buttons, they were selected. A measure of keystroke savings was calculated during this task.

**Measurements**

**Keystroke savings**

Keystroke savings of SSR used the following formula: \((X - Y)/X\), whereas \(X\) represented the total number of keystrokes required to enter a word plus a word-terminating character (e.g., space); and \(Y\) represented the total number of keystrokes required to enter the word in the SSR (Higginbotham, 1992; Koester & Levine, 1998; Magnuson & Hunnicutt, 2002; Trnka, Yarrington, McCoy, & Pennington, 2005; Venkatagiri, 1994). In the SSR prototype, the following scenarios could occur:

**Scenario 1:** The SSR placed the correct target word in the message window. In this instance, only one keystroke would be entered (the first letter of the word that was to be spoken) and the SSR identified an accurate match. It then placed that word in the message window, and the individual using the system would then proceed to the next word. For the target word ball, the keystroke savings formula would yield the following result:

\[
\frac{B - A - L - L}{X} = 5 \text{ keystrokes (X)} \\
\text{SSR} = \text{required 1 keystroke (Y)} \\
\frac{(X-Y)}{X} = \frac{5-1}{5} = \frac{4}{5} = 80\%
\]

**Scenario 2:** The SSR did not place the correct target word into the message window but the correct word was available on one of the word-prediction buttons. The individual using the system would have to enter a keystroke (the first letter of the word that was to be spoken) as well as perform the equivalent of another keystroke (selection of the word from the word prediction list). If the target word again was ball, the keystroke savings formula would yield the following result:

\[
\frac{B - A - L - L}{X} = 5 \text{ keystrokes (X)} \\
\text{SSR} = \text{required 2 keystrokes (Y)} \\
\frac{(X-Y)}{X} = \frac{5-2}{5} = \frac{3}{5} = 60\%
\]

**Scenario 3:** The SSR did not place the correct target word into the message window or the word-prediction buttons, so that the word had to be spelled either in its entirety or until it appeared on a word-prediction button. If the target word appeared on a word-prediction button after the “a” was typed in the example ball, the individual using the system would have entered one keystroke (the first letter of the word to be spoken), plus one keystroke for the “a”, and one more keystroke to select the target word from the word prediction list. The keystroke savings formula would then yield the following result:

\[
\frac{B - A - L - L}{X} = 5 \text{ keystrokes (X)} \\
\text{SSR} = \text{required 3 keystrokes (Y)} \\
\frac{(X-Y)}{X} = \frac{5-3}{5} = \frac{2}{5} = 40\%
\]

**SSR system performance**

To provide a better understanding of the potential benefits of the individual components of the SSR, the percentage of target stimulus words that were displayed in the message window was calculated and compared to the percentage of target stimulus words that were displayed in the message window or on the word-prediction buttons.

**Statistical analysis**

Data were described using descriptive statistics (means). Significance was tested using the Wilcoxon Test. Relationships were examined using Spearman’s Rank Order Correlation.

**RESULTS**

The results include an investigation of the keystroke savings using word prediction only compared to using the full SSR; the accuracy by which the SSR identified the target word (SSR system performance); and the relationship between keystroke savings, SSR performance, and the sentence intelligibility of participants. Data are reported for the typical sentences and atypical phrases.

**Keystroke Savings with Word Prediction Only**

For typical sentences with the researcher optimally controlling the keyboard, the keystroke savings were 59.6% with the word prediction only (no ASR was employed). For atypical phrases, the keystroke savings were 44% with the word prediction only (no ASR was employed). These percentages represent the maximum keystroke savings possible using the SSR prototype.
Keystroke Savings with full SSR (ASR + Word Prediction)

Keystroke savings using the SSR for typical sentences averaged 68.2% (range = 64.8–70.1%) for all speaker participants when the researcher controlled the touch-screen computer optimally. Keystroke savings for atypical phrases averaged 67.5% (range = 60.4–73.9%) for all speaker participants with the touch-screen computer optimally controlled by the researcher. Figure 2 displays the individual keystroke savings results on each stimulus set for each speaker participant with the touch-screen computer controlled by the researcher. There was a significant difference using the full SSR (Mdn = 67.8) compared to word prediction only (Mdn = 59.6) for typical sentences ($Z = 2.366$, $p = 0.022$). There was also a significant difference using the full SSR (Mdn = 64.32) compared to word prediction only (Mdn = 44) for atypical phrases ($Z = 2.197$, $p = 0.035$).

SSR System Performance

The word inserted into the message window of the SSR system reflected the system’s most probable choice. For typical sentences, an average of 50.7% of the target stimulus words were displayed in the message window (range = 31.9–75.2%). For atypical phrases, an average of 51.2% of the target stimulus words were displayed in the message window for atypical phrases (range = 27.4–73.4%).

As reported above, the most probable word was inserted into the message window on the SSR interface. Words that appeared on the word-prediction buttons represented the next five most probable word choices ordered alphabetically on the SSR interface. To examine the performance of the full SSR, the percentage of target stimulus words that appeared either in the message window or on the word-prediction buttons was examined. For typical sentences, an average of 80.7% of the target stimulus words were displayed either in the message window or on one of the word-prediction buttons (range = 67.4–97.5%). For atypical phrases, an average of 82.8% of the target stimulus words appeared either in the message window or on one of the word-prediction buttons (range of 69.4–94.4%).

Relationships between Keystroke Savings and Intelligibility, SR System Performance, and Intelligibility

The relationships between keystroke savings and speech intelligibility did not reach statistical significance for typical sentences ($r(5) = .750$, $p = .052$), or atypical phrases ($r(5) = .607$, $p = .148$). There was not a statistically significant relationship between the percentage of target stimulus words that were displayed in the message window and intelligibility for typical sentences ($r(5) = .750$, $p = .052$), but there was a significant relationship between intelligibility and atypical phrases ($r(5) = .757$, $p = .049$). There were no statistically significant relationships for the percentage of target stimulus words displayed either in the message window or on word-prediction buttons, or for intelligibility on typical sentences ($r(5) = .714$, $p = .071$) or atypical phrases ($r(5) = .500$, $p = .253$). A review of these results suggests that, for speakers with dysarthria with a range of speech severity, the keystroke savings and system performance of the SSR prototype was similar across the different stimulus sets.

DISCUSSION

This preliminary investigation examined the impact of an assistive writing prototype designed for individuals with a range of dysarthria severity. This unique system incorporated a variety of components with encouraging results. The specific components in this prototype system included ASR for dysarthric speech, first-letter identification (alphabet supplementation), and word prediction algorithms. To date, most assistive technology writing options have focused primarily on these components separately. Therefore, it is not possible to directly compare the current results to other ASR technologies based on dysarthric speech, because they do not include first-letter identification in addition to word prediction. The combining of these strategies into a hybrid approach may be a more efficient strategy compared to ASR or word prediction alone for speakers with moderate to severe dysarthria.

This preliminary study demonstrated keystroke savings of 59.9% for typical sentences and 44% for atypical phrases using word prediction only (no ASR). While these results were similar to keystroke savings using word prediction in other reports (Lesher, Moulton, & Higginbotham, 1999), they cannot be compared directly to other keystroke savings research due to the differing stimuli and difference in dictionary size. However, keystroke savings were significantly higher when using the full system than for word prediction alone. Using the full SSR, the optimal keystroke savings averaged 68.2% and 67.5% for typical and atypical sentences, respectively.

This hybrid approach promoted a unique, dynamic adjustment in the core dictionary size.
that led to the encouraging results. Coupling ASR with first-letter identification (alphabet supplementation) dynamically decreased the dictionary set from which an ASR system had to match the acoustic signal. This set was further reduced by using word prediction algorithms. Word prediction was included so that the system presented multiple options to maximize the chances of the target word being easily accessible to the speaker. This dynamic reduction utilized several features that are common in assistive technology and strategies with which speakers with dysarthria (i.e., alphabet supplementation) are familiar.

The dynamic reduction in the core dictionary resulted in significant keystroke savings for all participants, regardless of severity of dysarthria. Similar keystroke savings results occurred regardless of the dysarthria severity, as measured by speech intelligibility, for these speakers. A review of these results revealed that for the speakers whose dysarthria was so severe that they could not use their natural speech to meet their daily communication needs, keystroke savings were within a similar range to that of speakers with more functional natural speech abilities. The SSR prototype appeared to provide a significant impact on keystroke savings regardless of speaker intelligibility, compared to ASR alone. While the results were encouraging for this small sample, use of this technology with larger numbers of individuals with dysarthria is required in order to generalize the potential benefit this technology may have to a larger population. In addition, future work with more participants should focus on the impact of dysarthria type on speech recognition performance.

A greater range of text-generating tasks is needed to more fully understand how SSR may support functional written expression and face-to-face communication needs. Although this study used tasks that were much larger and less predictable than the tasks used in previous reports (e.g., environmental control commands used by Hawley et al., 2007, 2009; engineering software commands used by Talbot, 2000; and numbers used by Caves et al., 2007), future research needs to investigate the use of the SSR technology with messages that would be characteristic of routine use to support writing and communication interaction.

Finally, these results, while encouraging, reflect optimal performance in that the researcher controlled the touch-screen computer. The first letter was always identified accurately, and the word prediction-buttons were selected immediately, when the target word was displayed. Due to the cognitive demands associated with using word prediction, individuals using the system may not make the same “optimal” selections during functional use of this technology. In future research, it will be important to require participants to physically control the interface through either touch or head- or eye-tracking as they use the SSR technology. Such investigations would determine if the physical demands of interface control would change speech characteristics in such a way that would interfere with speech recognition. In addition, it would allow an assessment of the impact of the cognitive load required to control the interface on overall system performance.

Acknowledgements

This project was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development under grant number 1R43HD047044-01. The authors thank the participants for their involvement in this project.

Declaration of interest: This work was completed on the first prototype of the Supplemented Speech Recognition System funded under the NIH SBIR grant 1R43HD047044-01. Tom Jakobs of Invotek, Inc. continues to develop this technology. The authors alone are responsible for the content and writing of this paper.

Notes

2 SpeakQ by Quillsoft Ltd., 2416 Queen Street East, Toronto, Ontario M1N 1A2, Canada, http://www.wordq.com

References


SUPPLEMENTED SPEECH RECOGNITION


**APPENDIX A**

**Typical Sentences**

1. My mom and dad leave on Monday.
2. She has been sick since Thursday.
3. They like being married.
4. I’ll be back this Friday.
5. I’ve been on the go.
6. I’m looking for a big watch.
7. You can take a picture with the phone.
8. He is crazy about her.
9. Her kids play well together.
10. I do not have paper with me.

**Atypical Phrases**

1. not to look half bad
2. makes him sure money
3. a make believe movie about Kyle
4. their big idea
5. talk at this meeting
6. anything at that time
7. boy who wants this