Difficulties in Automatic Speech Recognition of Dysarthric Speakers and Implications for Speech-Based Applications Used by the Elderly: A Literature Review

Article *in* Assistive technology: the official journal of RESNA · June 2010

DOI: 10.1080/10400435.2010.483646 · Source: PubMed

Alex Mihailidis
University of Toronto

DIFFICULTIES IN AUTOMATIC SPEECH RECOGNITION OF DYSARTHRIC SPEAKERS AND THE IMPLICATIONS FOR SPEECH-BASED APPLICATIONS USED BY THE ELDERLY: A LITERATURE REVIEW

Victoria Young[1,2], MHSc, and Alex Mihailidis[1,2,3], PhD

[1]Intelligent Assistive Technology and Systems Lab, Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, Ontario.

[2]Institute of Biomaterials & Biomedical Engineering, University of Toronto, Toronto, Ontario.

[3]Correspondence author and reprint requests: 416.946.8573 (tel) and 416.946.8570 (fax).

ACKNOWLEDGMENTS

Funding support for this research has provided through the National Sciences and Engineering Research Council of Canada (NSERC Postgraduate Doctoral Fellowship), the Toronto Rehabilitation Institute’s Student Scholarship Fund, and the University of Toronto.

ABSTRACT

Automatic speech recognition is being used in a variety of assistive contexts, including home computer systems, mobile telephones, and various public and private telephony services. Despite their growing presence, commercial speech recognition technologies are still not easily employed by individuals who have speech or communication disorders. While speech disorders in older adults are common, there has been relatively little research on automatic speech recognition performance with older adults. However, research findings suggest that the speech characteristics of the older adult may, in some ways, be similar to dysarthric speech. Dysarthria, a common neuro-motor speech disorder, is particularly useful for exploring automatic speech recognition performance limitations because of its wide range of speech expression. This paper presents a review of the clinical research literature examining the use of commercially available speech-to-text automatic speech recognition technology by individuals with dysarthria. The main factors that limit automatic speech recognition performance with dysarthric speakers are highlighted and then extended to the elderly using a specific example of a novel, automated, speech-based personal emergency response system for older adults.

Key Words: Automatic speech recognition, dysarthria, speech-to-text, older adult, personal emergency response.

INTRODUCTION

Automatic speech recognition (ASR) is the process by which a machine (e.g., computer) is able to recognize and act upon spoken language or utterances. An ASR system typically consists of a microphone unit, computer, speech recognition software, and some form of audio/visual/action output. A popular ASR application is the automatic conversion of speech to text, which has the potential to increase work output efficiency and improve access to and control of various computer applications, such as word processing, email, dictation and document retrieval. By using speech as input, ASR applications bypass or minimize the more traditional manual input methods (e.g., keyboard, mouse), making it useful as an alternative input method for people with severe physical or neuro-motor disabilities (DeRosier & Farber, 2005; Koester, 2004). Unfortunately, ASR technology performance becomes limited with users having moderate to severe communication disorders, which may also occur with physical and neuro-motor disabilities (Deller, Hsu & Ferrier, 1988; Havstam, Buchholz & Hartelius, 2002; Wade & Petheram, 2001). ASR performance may be affected by many factors including the technology design, type and quality of speech input, the surrounding environment and user characteristics.

This paper reviews the clinical research literature exploring the factors that affect ASR performance with dysarthric speakers using commercial speech-to-text applications. The implications of the review findings for the design of ASR applications for the elderly are then discussed using the specific example of a novel, automated, speech-based personal emergency response system for older adults.

BACKGROUND

Types of Automatic Speech Recognition Systems

There are basically three categories of ASR systems differentiated by the degree of user training required prior to use: (1) speaker dependent, (2) speaker independent, and (3) speaker adaptable ASR. Speaker dependent ASR requires speaker training or enrollment prior to use and the primary user trains the speech recognizer with samples of his or her own speech. These systems typically work well only for the person who trains it. Speaker independent ASR does not require speaker training prior to use. The speech recognizer is pre-trained during system development with speech samples from a collection of speakers. Many different speakers will be able to use this same ASR application with relatively good accuracy if their speech falls within the range of the collected sample; but ASR accuracy will generally be lower than achieved with a speaker dependent ASR system. Speaker adaptable ASR is similar to speaker independent ASR in that no initial speaker training is required prior to use. However, unlike speaker independent ASR systems, as the speaker adaptable ASR system is being used, the recognizer gradually adapts to the speech of the user. This 'adaptation' process further refines the system's accuracy. A few types of speaker adaptable ASR systems exist differing with respect to how the adaptation is implemented. The reader is referred to the ASR review paper by Rosen and Yampolsky (2000) for further information.

ASR technologies also vary by the type of input that they can handle: (1) isolated/discrete word recognition, (2) connected word recognition, and (3) continuous speech recognition (Jurafsky & Martin, 2008; Noyes & Starr, 1996; Rabiner & Juang, 1993; Rosen & Yampolsky, 2000; Venkatagiri, 2002). Discrete word recognition requires a pause or period of silence to be inserted between words or utterances. Connected word recognition is an extension of discrete word recognition and requires a pause or period of silence only after a group of connected words have been spoken. For continuous speech recognition an entire phrase or complete sentences can be spoken without the need to insert pauses between words or after sentences.

Dysarthria and Older Adult Speech

Dysarthria, a neuro-motor speech disorder, may arise secondary to diseases such as Parkinson's, Alzheimer's, multiple sclerosis, and amyotrophic lateral sclerosis; disorders such as right hemisphere syndrome or dementia; or following traumatic brain injury or stroke (LaPointe, 1994). Several types of dysarthria exist, each of which has different expressed speech characteristics. Typically, dysarthria is classified according to the site of lesion and degree of neurological damage; however, in the literature reviewed, dysarthria is loosely classified based on the degree of disorder severity as measured by speech intelligibility and articulation. For example, mild, moderate and severe classifications were used as opposed to site of lesion. In the clinic, dysarthria is mainly assessed subjectively based on human listener perceptual measures of articulation and speech intelligibility (or comprehension) (Kayasith, Theeramunkong & Thubthong, 2006; Yorkston, Beukelman & Bell, 1988). Common clinical assessment tools include the Computerized Assessment of Intelligibility of Dysarthric Speech (CAIDS) (Yorkston, Beukelman & Traynor, 1984), the Franchay Dysarthria Assessment (Enderby, 1983), or the Swedish Dysarthria Test (Lillvik, Allemark, Karlström & Hartelius, 1999).

Age-related voice (speech) deterioration may begin around 60 years of age, but is highly dependent on the overall individual's health and well being (Ramig, 1994, p. 494). A comparison between the characteristics of older adult and dysarthric speech suggests that similarities exist between them. Key characteristic expressions of each type of speech have been summarized in Table 1. For the older adult naturally-aged voice, increased frequency of breathing may lead to intra-word pauses; decreased muscle efficiency, increased tissue stiffness and a dry laryngeal mucosa could affect vocal tract resonance, phonation and speech articulation; and slower cognitive function may reduce rate of speech (Gorham-Rowan & Laures-Gore, 2006; Linville, 2002; Zraick, Gregg & Whitehouse, 2006).

{INSERT TABLE 1 HERE}

LITERATURE REVIEW

Literature Review Search Method

The literature review search method was limited to English language journal articles using Scholar's Portal (1960-2009), Ovid-Medline (1950-2008), PubMed and Google Scholar.

Two sets of keyword groups were used in the database searches:

1. ("speech recognition" OR "voice recognition" OR "speech technology") AND ("disab*" OR "communication disorder");
2. ("speech recognition" OR "voice recognition" OR "speech technology") AND ("dysarthr*" OR "elder*" OR "senior" OR "older adult").

The inclusion criteria used to identify the final review articles are listed below:

1. Dysarthria and ASR – specifically a speech-to-text application or clinical/lab test;
2. Older Adults and ASR – specifically computer applications, speech corpora, new acoustic model, or clinical/lab test;
3. ASR usability/user perspectives;
4. Review of ASR and communication disorders or disability.

Eleven papers were found for inclusion criteria 1, three papers for inclusion criteria 2, three papers for inclusion criteria 3, and six papers for inclusion criteria 4.

Introduction to the Literature Review

In order for ASR technology to perform well with dysarthric speech, it must be able to handle speech variability caused by any of the possible characteristic expressions of dysarthria (e.g., poor articulation, disfluencies, intra-word pauses, non-speech sounds). Research literature suggests that greater speech variability often correlates with increasing severity of dysarthria (Blaney & Wilson, 2000; Doyle et al., 1997; Ferrier, Shane, Ballard, Carpenter & Benoit, 1995). In turn, increasing severity of dysarthria often correlates with decreasing degrees of speech intelligibility (Doyle et al.; Ferrier et al.). Since the majority of current ASR algorithms rely to some degree on pattern matching, speech consistency (or similarity) is also important. The literature reviewed explores the relationship between ASR recognition performance as measured by accuracy (% of words correct, divided by total number of words used) and the severity of dysarthria as measured by intelligibility, various sources of speech variability, and perceptual measures of intelligibility and consistency.

The research studies presented in the literature review have been conducted with different types of ASR technologies, research subjects, test vocabulary (e.g., words, phrases, and sentences), environments, training time, and test protocols; therefore, the detailed results cannot be directly compared with each other. As well, commercial speech-to-text ASR technologies are continually evolving. However, despite the fact that ASR technologies and their recognition rates have continued to improve over the years for non-disordered adult speech, revolutionary changes have not occurred in recent years and ASR performance still does not yet equal that of the human auditory system (Benzeghiba et al., 2007). Therefore, it seems probable that the literature review results cited here will continue to provide, for the near future, a good general overview of the main challenges faced by dysarthric speakers when using speech-to-text ASR applications. The current state and future trends of research in this area should also be revealed.

Degree of Dysarthria, Speech Intelligibility & ASR Accuracy

Blaney & Wilson (2000), Thomas-Stonell, Kotler, Leeper & Doyle (1998), and Raghavendra, Rosengren & Hunnicutt (2001) found speech recognition accuracy to be consistently and significantly lower for individuals with moderate to severe dysarthria compared to individuals without dysarthria (herein called 'controls'). Whereas, individuals with mild dysarthria all obtained slightly lower or similar speech recognition accuracy compared to the controls.

Raghavendra et al. (2001) examined ASR accuracy results obtained from four dysarthric speakers (mild, moderate, severe and profoundly severe) and one control speaker, using a speaker dependent, discrete word, whole-word pattern matching ASR system (Infovox RA) and a speaker adaptable, discrete word, phoneme-based ASR system (Swedish Dragon Dictate). Degree of dysarthria was determined using the Swedish Dysarthria Test. For both the Infovox RA and the Swedish Dragon Dictate systems, the average accuracy ratings over three sessions of use, was highest for the control and mildly dysarthric speakers, followed by the moderately dysarthric, then severely dysarthric, and finally the profoundly severe dysarthric speaker. Generally, all speakers achieved higher accuracy ratings using Swedish Dragon Dictate (74%-97%) over Infovox RA (28%-97%).

Accuracy ratings achieved by six dysarthric speakers (two mild, two moderate, and two severe) against six control speakers were examined by Thomas-Stonell et al. (1998) using a speaker adaptable, discrete word ASR system (IBM Voice Type) with sentence input. Degree of dysarthria was determined using CAIDS. After five sessions the highest accuracy ratings were achieved by the controls (93%) and mildly dysarthric speakers (88%), followed by the moderately (75%) and severely dysarthric speakers (77%).

Blaney & Wilson (2000) observed the accuracy results from one mildly and two moderately dysarthric speakers compared with six controls. Degree of dysarthria was determined using the Frenchay Dysarthria Assessment. In general, after five user sessions, the recognition accuracy was again observed to be lower for the moderately dysarthric speakers (66% and 81%), compared to the mildly dysarthric speaker (88%) and the majority of the controls (91-94% and 78%). The one control speaker with the lower accuracy score (78%) was a native speaker of the test language used in the study; however, there were fluctuations in accent and speech patterns as a result of having spent significant time abroad.

In terms of speech intelligibility (as measured using CAIDS), studies by Doyle et al. (1997), Ferrier et al. (1995), and Thomas-Stonell et al. (1995) demonstrated significant correlation with speech recognition accuracy ratings. Specifically, higher intelligibility scores (controls and mildly dysarthric speakers) tended to produce higher ASR recognition accuracy rates, while lower intelligibility scores (moderate to severely dysarthric speakers) tended to produce lower ASR recognition accuracy rates. This was true in the majority of cases with a few exceptions. One out of ten subjects from Ferrier et al. and one out of six subjects from Doyle et al., both with severe dysarthria and low intelligibility scores, obtained better speech recognition ratings than individuals with moderate dysarthria and higher intelligibility scores. Their accuracy ratings in fact, reached levels similar to those achieved by the mildly dysarthric and control speakers. As a result of the deviant cases, Ferrier et al. (1995) concluded that speech intelligibility measures cannot be reliably used as a clinical guideline to definitively predict one's level of success (high accuracy) with speech-to-text ASR applications.

Human Speech Perception versus ASR

The two separate instances in Ferrier, et al. (1995) and Doyle, et al. (1997), where the severely dysarthric speakers were unintelligible to a casual listener, but were sufficiently consistent so that an ASR system could recognize them with relatively high accuracy, lead Ferrier, et al. to hypothesize that when speech intelligibility reaches moderate to severe levels of dysarthria, speaker adaptable, discrete word ASR technologies (e.g., Dragon Dictate) might out-perform the human listener in recognizing dysarthric speech.

Sy and Horowitz (1993) examined the relationship between human perceived measures of speech intelligibility, phoneme differences and ASR performance. Their study compared the results from one moderately dysarthric speaker and one control speaker. Thirty-eight listeners with no hearing dysfunction provided perceptual measures of speech intelligibility. An isolated-word, speaker dependent, speech recognition system (developed in the lab), based on dynamic time warping was used. Perceptual measures for intelligibility and ASR recognition accuracy were derived based on an exponential cost function based on phoneme numbers. The results showed that perceptual measures of intelligibility for dysarthric and normal speech could be evaluated consistently by human listeners. Thus individuals with and without disordered speech were evaluated in the same way suggesting that individuals with non-disordered speech could be used effectively as control subjects. In general, at high and moderately-high intelligibility levels, ASR accuracy measures were found to be lower than listeners' measures of intelligibility 92.5% of the time. Sy and Horowitz concluded that their ASR system was "not very useful for computer access and communication" (p.1295). Overall, no correlation was found between the perceptual measures of speech intelligibility and ASR accuracy at the word level; however, some correlation was found at the phoneme level for the phoneme confusion errors (e.g., between consonants, vowels, etc.).

Thomas-Stonell, et al. (1998) examined non-expert listeners' perceptual measures of dysarthric speech intelligibility for mild, moderate and severe dysarthric classifications and found good correlation with speech recognition accuracy ratings. However, in Doyle, et al. (1997), using the same ASR technology as in Thomas-Stonell et al. (the IBM VoiceType), the controls achieved the highest accuracy ratings, followed by the moderately dysarthric speakers as expected, but accuracy variability and overlapping boundaries occurred among the mildly and severely dysarthric groups. One of the two severely dysarthric subjects obtained accuracy ratings similar to the controls and one of the two mildly dysarthric subjects achieved only a moderate accuracy rating. The perceptual measures of speech intelligibility were consistent with the subjects' degree of dysarthria as assessed by the CAIDS (e.g., controls had higher measures than mildly dysarthric subjects, who had higher measures than moderately dysarthric subjects, who had higher measures than severely dysarthric subjects). To remove the possibility of communication via situational context or non-verbal message cues in these studies, perceptual measures of speech intelligibility in Doyle et al. (1997) and Thomas et al. (1998) were based on single-words presented out of context.

The variability observed in these study findings suggests that ASR recognition accuracy could be affected by more than just level of speech intelligibility.

Perceptual Measures of Speech Consistency

Despite the fact that dysarthric speech is characterized by reduced speech intelligibility and increased speech variability, consistency of speech, rather than good articulation and high intelligibility, is also important for obtaining good speech recognition accuracy (Noyes and Frankish, 1992; Noyes and Star, 1996). Due to time limitations and equipment access issues, however, clinicians often have difficulty obtaining quantitative acoustical measures of an individual's speech consistency (Thomas-Stonell et al., 1998). Therefore, clinical decisions tend to be made based on the clinician's perceptual judgment of the patient's speech consistency. These perceptual measures might then be used to determine a patient's potential success with using ASR technology. The study by Thomas-Stonell et al. (1998) indicated that for individuals with mild to severe dysarthria using speaker-adaptable ASR software no significant correlation existed between the user's ASR recognition accuracy ratings and the listeners' perceptual measures of speech consistency. The study findings led researchers to conclude that perceptual measures of speech consistency should not be used to determine ASR technology suitability, but rather, clinicians should allow the user to trial an ASR system prior to making a final judgment. However, when speech consistency was controlled for in the statistical calculations, speech intelligibility was also no longer found to correlate with ASR accuracy.

In contrast, Kayasith et al. (2006) proposed a new measure of speech consistency as an alternative method of assessing degree of dysarthria that could also be used to predict one's ability to use ASR with high accuracy rates. This measure of speech consistency, called the speech consistency score (SCS), was defined as a ratio of speech similarity over dissimilarity. This study compared the SCS results against degree of dysarthria measured by articulation and speech intelligibility test results, in addition to the accuracy ratings obtained using different types of ASR algorithms. The results from this study demonstrated that the SCS could be used to evaluate degree of dysarthria and was able to predict ASR accuracy with less error than the other measures (intelligibility and articulation).

Dysarthric Speech Variability

Sy and Horowitz (1993) further explored the possibility that patterns of articulation errors in dysarthric speech, defined as “slight differences in the timing or placement of a speech sound [phoneme]” (p. 1282), could be grouped according to spectral features (e.g., quality of voice, manner and place of articulation, quality of vowels/consonants, etc.). The study observed that the majority of the dysarthric subject’s articulation errors were based on speaker confusion with consonant pairs and vowel pairs. Consonant pair errors were almost always related to the *fricative* or *stop* manners of articulation and were mostly *alveolar* (tongue tip positioned on alveolar ridge) or *labial* (lips) places of articulation. Confusion with vowel pairs was found to relate to vowels articulated mostly in front of the mouth as opposed to the back. These types of articulation errors were consistent with the findings of the subject’s physical expression of motor dysfunction (e.g., airflow control issues and lower jaw would move to the left). The researchers noted that many of their test words required front of the mouth articulation during pronunciation; but also, that many English language consonants naturally use front of the mouth articulation.

Blaney and Wilson (2000) acoustically analyzed the speech from controls, and mildly and moderately dysarthric speakers to examine the specific acoustic features for sources of dysarthric speech variability. The acoustic measures (i.e., voice onset time (VOT), vowel duration (VD), fricative duration (FD), vowel formant (e.g., peak in frequency spectral envelop) frequency F1/F2 and word stem duration) were applied to 32 words/tokens including minimal-pairs (e.g., pat/bat, sheep/cheap) and "mono, bi and polysyllabic" words (e.g., let, letter, lettering). Variability of acoustic measures was determined using the mean, standard deviation (SD), and coefficient of variability (CV = SD/mean) values. Moderately dysarthric speakers demonstrated greater variability, compared to controls, over all acoustic features measured (e.g., VOT, VD, FD, vowel formants), minimal-pair categories were not preserved (i.e., acoustic space merged and minimal-pair contrasts were violated) and timing discrepancies were observed for the word stem durations. Lower recognition accuracy scores thus reflected higher acoustic measure variability. Words with two and three syllables tended to have higher errors than words with one syllable. Mildly dysarthric speakers produced similar acoustic measures as the controls, except for word stem segmental timing inconsistencies and modified timing or shifts between the category boundaries of phonemic contrasts (e.g., voiced/voiceless). A separate study by Raghavendra et al. (2001) found similar results with variability in timing and duration, in addition to pauses and slow speech.

The Fatigue Factor

Physical and psychological fatigue can affect the voice, mind and body, and are known to cause degradation in ASR technology performance. Individuals with speech disorders are known to be more easily and frequently fatigued than those individuals without speech disorders, thus the effect of fatigue is more pronounced for those with a greater severity of speech impairment and lower intelligibility (Ferrier et al., 1995; Noyes & Frankish, 1992; Noyes & Star, 1996; Rosen & Yampolsky, 2000). Fatigued speech is more variable and less consistent than non-fatigued speech and may be misrecognized in speaker dependent ASR systems or may cause voice drifting in speaker adaptable ASR systems. Voice drifting occurs when the ASR system starts to adapt to altered or fatigued speech, thereby increasing the possibility of misrecognition when non-fatigued speech is used. For these reasons, Ferrier et al. (1995) suggested that fatigue must be accounted for during clinical trials with ASR technology. Researchers, for example, could ensure that frequent breaks are taken during a trial session and also limit the length of time the ASR is used per day. Discrete word ASR may also induce vocal fatigue as a result of the insertion of pauses between words (Olson, Cruz, Izdebski & Baldwin, 2004). Pause insertion increases the physical effort required for voicing, especially during repeated forced glottal closure (Kambeyanda, Singer & Cronk, 1997; Ramig, 1994).

Misuse/Abuse Factors

Olson et al. (2000) examined users who have misused/abused their voices through continual use of ASR technologies despite vocal fatigue. Muscle tension dysphonia (MTD), a condition caused by improper closure of the vocal folds resulting from inappropriate muscle tension, was observed in five patients 2 to 8 weeks after starting to use discrete word ASR technologies (e.g., Dragon Dictate and IBM Via Voice). Reported symptoms of MTD caused by ASR technology use included hoarseness, increasing strain and voice fatigue, pain, and even the inability to voice (aphonia). Surprisingly, when the subjects spoke in a natural speaking voice they did not experience MTD. Only when they started to speak using their "computer voice" or with computer speech, did dysphonia appear. Long term voice therapy, including training against the use of monotonous and lower pitched voices, as well as, limiting speech by inclusion of voice breaks during training, was found to improve the symptoms for the majority of subjects. Although, use of continuous ASR technologies are not immune to causing user fatigue, and misuse could still lead to MTD, some individuals were able to use continuous ASR systems for several hours longer before fatigue or dysphonia symptoms would re-occur.

Other Personal Factors

Ferrier et al. (1995) concluded that "personal motivation, educational level, manual dexterity, reading and writing skills and visual factors," (p. 173) were some of the 'other factors' that could determine one's ultimate success at using ASR technology. In the study by Havstam et al. (2002) motivation was clearly a very important factor in the successful application of ASR technology for one profoundly dysarthric individual with severe motor impairment and cerebral palsy. The goal of this study was to determine if ASR could be used to augment an existing "switch access" writing system (Bliss system) used by the participant (Havstam et al.). Researchers initially questioned whether this subject should be included in the study due to his poor health condition and limited speech – the subject could only speak three functional words. In the end, the subject was included in the study and successfully demonstrated that ASR technology (i.e., the Swedish version of Dragon Dictate) could be used successfully by an individual with profound disability and dysarthria. The study results showed that compared to his original system, using ASR improved the computer access efficiency by 40% with just a few words. In terms of everyday use, researchers concluded that individuals with similar degrees of severe dysarthria and motor impairment would likely still not be able to function completely independently with the ASR technology. External factors must also be considered such as background noise and a supportive network of individuals willing to help users with the technology.

System and User Voice Training

Decreasing speech intelligibility and increasing severity of dysarthria were found to lengthen the time required to complete training routines for speaker dependent and adaptable ASR systems, and achieve stable, possibly higher, recognition accuracy (Hird & Hennessey, 2007; Ferrier et al., 1995; Kotler & Thomas-Stonell, 1997; Raghavendra et al. 2001). A 'stable state' was defined differently in the various research studies, but in general, the definition used by Kotler and Thomas-Stonell is a good starting point. Stability was defined to be the point at which 10% or less variation in recognition accuracy was achieved over three consecutive training sessions, after completing four initial training sessions. Given the fact that adaptable ASR systems were used primarily in these studies, it seems reasonable that speech with greater variability requires more time for adaptation because speech that is less disordered should match more closely to the non-disordered speech samples of the ASR acoustic model. The recognition accuracy trend was found to consistently resemble a steep incline of rapid improvement after the first training session, with subsequent sessions marked by decreasing gradual improvements, leading eventually to stability or the maximum recognition accuracy achievable.

Blaney and Wilson (2000) and Doyle et al. (1997) concluded in their studies that after five training sessions using IBM VoiceType and Dragon Dictate, none of the dysarthric speakers had yet achieved their stability point. Thomas-Stonell et al. (1998) found that after five sessions with IBM VoiceType, the results from the last two sessions were similar for the sentence tests. Ferrier et al. (1995) found that for mildly, moderately and severely dysarthric speakers, the maximum gains in recognition accuracy were achieved within the initial four training sessions using Dragon Dictate. In this study, participants performed a total of eight sessions in an attempt to achieve stability at 80% speech recognition accuracy over three continuous trial sessions.

80% accuracy was achieved by all mild and moderate dysarthric speakers, but only one of the four subjects in the severe and moderate/severe category reached this final goal.

The study by Kotler and Thomas-Stonell (1997) examined specifically the number of training sessions required in order to reach a stable state using IBM VoiceType. They also explored whether voice training would have an effect on the maximum recognition accuracy rate. For discrete word recognition, results indicated that at least six sessions were required to reach stability at 72% recognition accuracy (less than 70 words used in this trial). For ‘words in sentences’ (herein referred to as ‘sentences’), three sessions were required for stability at 90% recognition accuracy. The difference in the number of sessions required for stability was attributed to the fact that when sentences were used for training, the ASR system had more chances to adapt to the speaker’s voice. As well, sentences contain several words that might provide context as to what the other words could be. For discrete word ASR systems, context is not provided by other words. These results are further supported by another study by Thomas-Stonell et al. (1998).

Kotler and Thomas-Stonell (1997) demonstrated that discrete word voice training was effective and useful in reducing certain types of articulation errors. The study concluded that if stability could not be achieved by the guideline proposed previously, six and three sessions for words and sentences respectively, speech training should be used to further improve recognition accuracy. Support for using voice therapy was provided by Hird and Hennessey (2007) who examined fifteen dysarthric speakers and different types of voice training. They demonstrated that physiological therapy (i.e., respiration training, elongation phonation practice with biofeedback) was effective in improving voice resonance and producing more consistent speech.

Acceptable and Achievable Recognition Accuracy Rates

To achieve acceptable recognition rates within a reasonable clinical assessment period, Raghavendra et al. (2001) suggested at least three sessions would be needed. However, only the mildly and moderately dysarthric speakers could achieve successful accuracy results. The severely dysarthric speakers could achieve relatively high recognition accuracy but might do better with more training. Individuals with severe and profoundly severe dysarthria would likely need continued assistance with using the ASR system (e.g., error correction, modification of training word lists).

A comparison between three major ASR technologies: Microsoft Dictation, Dragon Naturally Speaking 3.0, and Voice Pad Platinum, was conducted by Hux, Rankin-Erickson, Manasse, Lauritsen et al. (2000) with individuals with dysarthria over five user sessions. Dragon Naturally Speaking and Microsoft Dictation are speaker adaptable, continuous word ASR applications and Voice Pad Platinum is a speaker adaptable, discrete word ASR application. This study compared results from one mildly dysarthric subject to one control speaker. The results found that Dragon Naturally Speaking produced the highest accuracy ratings for the dysarthric speaker which was approximately 65% accuracy. The study suggests that this accuracy rate could only be considered acceptable, if at all, by individuals with higher degrees of dysarthria and with upper limb disabilities for whom no other input options may be available. Researchers acknowledged, however, that only a minimal degree of training was performed. Additional training options were available but not performed.

Fried-Oken et al. (1985) assessed ASR accuracy using a discrete word, speaker dependent ASR for two individuals, both mildly dysarthric with concomitant severe physical disability. Subject 1 was quadriplegic and Subject 2 had a spinal cord injury and traumatic brain injury. Results showed that discrete word ASR yielded 45 to 60% accuracy after 273 utterances for subject 1, and 79-96% accuracy after 173 utterances for subject 2.

Results obtained by Kotler and Tam (2002) indicated speech variability and lower recognition accuracy rates amongst highly intelligible individuals using ASR technologies deployed outside of the clinic with various speech tasks. In this study, the researchers reported from previous clinical experience that an average ASR accuracy of 74% was obtained for individuals without speech disorders and 57% accuracy for individuals with speech impairments (degree and type of speech impairment was not mentioned). In this study, six individuals with intelligible speech and physical disabilities, two of whom had minimal speech impairments, were followed using discrete word ASR software (e.g., VoiceType, VoiceType2, and Dragon Dictate) in their homes. ASR accuracy rates ranging from 62 to 84% were obtained for a variety of speech tasks including dictation, numbers, name/address, and letter composition.

Usability

In a study by Kotler and Tam (2002), user perceptions on the use of discrete word ASR technologies were obtained from six physically disabled individuals with intelligible speech. ASR technology limitations included the time it takes to make corrections, the system's susceptibility to noise, the lack of confidentiality that occurs as a result of speaking out loud, the potential risk of having voice related health problems, and the lack of support readily available to help with various applications.

Hux et al. (2002) noted that, when using adaptable ASR systems, the user must be taught to turn off the microphone when interjecting with non-voice features (e.g., sneezing, throat clearing, laughing), otherwise these features would be recognized as speech.

Koester (2006) found that using the commands “scratch that”, “undo” or “erase” to remove a mistake made by the ASR system, obtained lower performance ratings compared to using the correction commands “correct“, “fix” or “edit”. Essentially, uncorrected errors degrade the ASR acoustic model, affecting the final user performance results (Koester, 2006). Correcting the errors would also improve the ASR performance. The importance of providing adequate and proper training is supported by DeRosier and Farber (2005) who observed, ”absence or presence of training….may have an influence on the psychosocial impact and satisfaction scores reported by individuals with disabilities” (p.131).

Havstam, et al. (2002) and Noyes, et al. (1989) noted that situations of repeated speech misrecognition or the inability on the part of the user to consistently produce the desired output can result in feelings of irritability and frustration. Unfortunately, in these cases, increasing irritability and frustration further compounds the problem and could result in continued speech or voicing variations – thus lack of consistency in the speech output.

Although high recognition accuracy is typically the goal for the majority of adult users and designers of ASR systems (e.g., 90-100%) (Noyes, et al., 1989; Rosen, & Yampolsky, 2000); individuals with disabilities, in general, are satisfied with the assistive benefits of ASR technologies even with lower accuracy rates and other accompanying usability difficulties (DeRosier & Farber, 2005; Noyes & Starr, 1996).

DISCUSSION

The literature reviewed shows that from the early 1990’s to the early 2000’s the general overall ASR performance trends or patterns revealed have remained similar despite improvements in ASR technology, differences in research study protocols, study subjects and category of ASR system used. Overall, the study findings indicate a similar trend of decreasing ASR performance with decreasing speech intelligibility and increasing severity of dysarthria and speech variability. Mildly dysarthric speakers should be able to use existing commercial ASR technologies and still achieve good ASR performance compared to individuals without speaking disorders. Moderately to severely or profoundly severe dysarthric speakers, on the other hand, have tended to achieve lower ASR performance with the commercial, speech-to-text ASR applications. The small number of dysarthric individuals found to deviate from these trends, however, suggests that a greater complexity exists in not only how dysarthria might be measured but also how ASR performance is assessed in the presence of so many internal and external factors of influence.

The studies reviewed showed that ASR performance could be improved to a certain extent with increased user and system training, but that accuracy rates were negatively affected by increasing user fatigue, frustration, and user error. The studies exploring ASR performance with speech consistency did not reveal consistent trends and no generalizations can be made. A significant relationship does appear to exist between speech intelligibility and consistency, however, and future research will hopefully clarify this association.

Even though many factors were identified as influencing speech-to-text ASR performance with dysarthric speakers, the key factors of importance included the user's fatigue level, the type of input, the type and category of ASR technology employed (i.e., adaptive, dependent or independent, continuous or discrete, small or large vocabulary ASR), and also the amount of user and system training provided.

Dysarthric speakers demonstrated less difficulty speaking isolated or discrete words rather than continuous sentences; thus a speaker with moderate, severe or profoundly severe dysarthria might perform better using a discrete word, speech-to-text ASR system. In terms of category of ASR system, speaker dependent and adaptable systems were shown to provide better results for the individuals with dysarthria; however, there was the potential for increased fatigue. With respect to system training, while a longer training time was found to be beneficial for moderately to severely dysarthric speakers using speaker dependent and adaptable ASR systems, it was also very time consuming for both the clinician and end-user. A considerable amount of motivation and patience was required, especially for individuals with profoundly severe dysarthria. Over the years, commercial speech-to-text ASR applications have been developed with increasingly larger vocabulary; which, although very useful for the end-user with non-disordered speech, may actually increase overall training time required and possibly decrease system usability for a dysarthric speaker.

Different applications of ASR technology are also more robust than others at handling specific characteristics of dysarthric speech. For example, reduced word timing and duration issues found in dysarthric speech may be easily removed by editing in speech-to-text technologies (Ferrier et al., 1995). However, in 'action' output ASR systems, for example, environmental control units used to control household devices, these types of disfluences cannot be 'edited-out' and may cause an error or non-response, possibly adding to user frustration.

It should be noted, that what constitutes successful utilization of an ASR technology differs depending on the evaluator, the end user, the specific application, and the system performance. Therefore, as revealed in the literature, depending on the perceived benefits gained from using the ASR technology and the available alternative options, individuals with physical disabilities and speech disorders may still find ASR technologies with lower accuracy rates acceptable to use.

In recent literature, increasing focus has been placed more on the custom design and development of ASR systems for individuals with dysarthria, instead of using existing commercial ASR technologies. ASR applications custom designed for dysarthric speakers have generally achieved better overall speech recognition performance compared to those observed for commercial speech-to-text ASR systems (Hasegawa-Johnson, Gunderson, Penman & Huang, 2006; Hawley et al., 2007; and Polur & Miller, 2005). Another research direction with some positive results involves using the acoustic-phonetic characteristics or spectral transformations of disordered speech to account for the speech variability prior to speech recognition (Hosom, Kain, Mishra, van Santen, Fried-Oken and Staehely, 2003).

Similar to the findings with dysarthric speakers, research exploring the use of custom developed ASR systems for the elderly or older adult suggests that an ASR system trained specifically with older adult speech tends to perform better (higher accuracy) then when trained with non-older adult speech (Anderson, Liberman, Bernstein, Foster, Cate & Levin, 1999; Baba, Yoshizawa, Yamada, Lee & Shikano, 2004; Wilpon & Jacobsen, 1996). By applying the literature review findings with dysarthric speakers to the elderly, we will examine how these results may help in the development of a novel, automated, speech-based personal emergency response system (PERS) for the older adult.

CASE STUDY

An Automated, Speech-based Personal Emergency Response System

Older adults, 65 years of age and older, are at a higher risk of experiencing medical complications during an emergency situation as a result of co-morbidities, poly-pharmacy, possible functional/cognitive impairment and/or general fragility (Gibson, 2006; Hwang &

Morrison, 2007; Salvi, Morichi, Grilli, Giorgi, De Tommaso & Dessi-Fulgheri, 2007). Therefore, it is essential that emergency assistance be provided, as promptly as possible, to increase chances for a full recovery (Handschu, Poppe, Rauß, Neundörfer & Erbguth, 2003; Rosamond, Evenson, Schroeder, Morris, Johnson & Brice, 2005). Unfortunately, older adults may not immediately recognize the severity of an emergency situation, may not ask for assistance, and/or may be unable to obtain assistance when needed (e.g., injured and alone) (Fogle et al., 2008; Rosamond et al., 2005). Personal emergency response systems (PERS) are often installed in the home of older adults to provide them with immediate access to 24 hour, emergency assistance. PERS usage has been shown to ease caregiver and user anxiety, support aging-in-place (aging at home), and minimize overall healthcare costs (Mann, Belchior, Tomita, & Kemp, 2005; Porter, 2005).

A traditional PERS is activated by pressing a body-worn, wireless, panic button (e.g. a necklace or watch). This "assistance required" signal is instantly transmitted to an emergency call centre where an emergency responder contacts the subscriber either through their PERS speaker-phone or telephone. The emergency responder subsequently contacts emergency services, care providers, family and/or friends to provide immediate assistance as required.

Studies show that less than half of PERS owners actually use their system and many older adults who might benefit from having a PERS do not own one (Mann et al., 2005; Porter, 2005). Reasons for non-use include cost; feelings of stigmatization and burden from wearing 'the button'; fear of bothering caregivers or responders, institutionalization, and/or loss of independence; and an inability to push the button (e.g., not wearing it, too fragile) (Porter, 2005). With this push button system, the majority of calls to the emergency call centres are also false alarms (accidental). This inefficiency may stress already limited emergency resources and may also result in loss of work-time for the care provider (Mann et al., 2005; Porter, 2005).

Automated, speech-based PERS interfaces may improve the basic push button PERS's overall system efficiency and usability; leading to increased PERS adoption (Mihailidis, Tam, McLean & Lee, 2005; McLean, Young, Boger & Mihailidis, 2009). Eliminating the need to wear a button should decrease feelings of stigmatization and burden; the addition of speech activation should improve usability; and enabling call cancellation should support user autonomy and decrease the occurrence of false alarms; hence improve system efficiency.

Implications for speech-based applications for older adults

To gain a better understanding of ASR performance limitations when used by an older adult in an emergency situation, we are interested in the speech characteristics of the older adult in stressful states of distress. Literature suggests that voice disorders are common in older adults (Roy, Stemple, Merrill & Thomas, 2007) and in an emergency or stressful situation human speech may become altered, if not already, to the point of impairment or disorder, either as a result of a medical trauma, disease or strong emotion (Devillers & Vidrascu, 2007; Fogle et al., 2008; Handschu et al., 2003; Hansen & Patil, 2007; LaPointe, 1994, p. 359). In addition, the characteristics of the naturally aged voice have been found to be less easily recognized by commercial ASR systems that are often designed for a non-disordered, specific accent, younger adult age group (Lippman, 1997). Clinical research literature that examines the performance and use of ASR specifically by older adults is also limited.

The fact that psychological/stress factors can influence speech recognition performance is also of particular relevance for this type of ASR application. The ASR must take into account the possibility that speech may be altered from normal speech patterns. As well, the user dialogue must minimize the possibility of user frustration and irritability.

One strategy for dealing with stress-related reduction in accuracy is to reduce the size of the vocabulary required for speech recognition. As well, the words recognized should be simple with as few syllables as possible (e.g., words with 1-2 syllables) to minimize error. Designers might consider using isolated or discrete word recognition as opposed to continuous, large vocabulary speech recognition.

Another strategy would be to minimize the length of the training period required before using the ASR. A long user training period would be particularly problematic in the case of the elderly (e.g., 85 years of age and older), who may be fragile or mildly cognitively impaired and easily fatigued. The automated, speech-based PERS should thus be configured to minimize the amount of training time required and should be designed to be more robust to voice drifting resulting from fatigue, distress, or natural speech variation over time.

To reduce user frustration and irritability the designers might consider developing a user dialogue that is easy to use and that responds to the user's request as quickly as possible, while still ensuring an effective and efficient system. Given that an ASR system is more likely to result in lower accuracy ratings for moderate to profoundly severe dysarthric speakers, a potential design principle for the automated, speech-based PERS may also be to default to a live emergency response operator if severely disordered or unrecognizable speech is detected.

Following the research trend towards custom designed ASR systems for individuals with dysarthric speech, research comparing an ASR system trained with 'older adult and/or dysarthric speech only' and 'older adult, dysarthric and adult speech combined' in various environmental conditions, might also be beneficial in the development of the ASR for an automated, speech-based PERS. Alternatively, if the source of older adult disordered speech variability could be accounted for, as in the study by Hosom et al. (2003), then the ASR system might also be pre-calibrated to account for these variations before being processed by the PERS.

In terms of the best category of ASR system to use for a hands-free PERS, the system should be able to work with multiple users with minimal or no training, thus an independent ASR system would be necessary. If the system was adaptable, given that the older adult voice may change during times of stress, the PERS may not work as well as intended.

CONCLUSIONS

Current commercial speech-to-text ASR systems are designed specifically for a mainstream, non-speech disordered adult population; thus purposely excluding individuals with speaking disorders. The literature reviewed demonstrates the numerous challenges faced by moderately to severely dysarthric speakers in achieving good ASR performance, including type and category of ASR application, amount of system and user training, motivation, fatigue, frustration, error, and the surrounding environment. Possible areas for future research include exploring the relationship between speech intelligibility and consistency in relation to ASR, and identifying specific sources of acoustic-phonetic variation and whether this can be accounted for in the pre-speech recognition stage. Resent research is moving away from using commercial ASR applications towards the development of custom designed ASR systems for individuals with speech disorders.

Given the similarity between older adult speech and dysarthric speech, these review findings may also be useful in the design of ASR systems used by the elderly, such as the novel, automated, speech-based PERS for older adults. Using the literature findings, a small vocabulary, small syllable, isolated word ASR system may be a good starting point for the novel PERS ASR. Future research areas to explore might include determining the best category and type of ASR and the best speech training set to use (e.g., older adult and/or dysarthric speech; a combination of older adult, dysarthric and adult speech); as well as, determining the degree of intelligence and appropriate dialogue. If the challenges faced by individuals with speaking disorders when using ASR applications can be better understood, ASR technology developers might then be able to discover new ways for overcoming or accommodating these difficulties. Perhaps then, future commercial ASR applications could be developed that work for any individual, regardless of whether a communication disorder exists or not.

REFERENCES

Anderson, S., Liberman, N., Bernstein, E., Foster, S., Cate, E., & Levin, B. (1999). Recognition of elderly speech and voice-driven document retrieval. *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99, 15-19 March 1999,* Phoenix, AZ, USA. IEEE; Signal Process. Soc, 145-8 vol.1.

Baba, A., Yoshizawa, S., Yamada, M., Lee, A., & Shikano, K. (2004). Acoustic models of the elderly for large-vocabulary continuous speech recognition. *Electronics and Communications in Japan, Part 2,* 87(7), 390-397.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, D., Mertins, A., Ris, C., Rose, R., Tyagi, Vo., and Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. Speech Communication, 49, 763-786.

Blaney, B., & Wilson, J. (2000). Acoustic variability in dysarthria and computer speech recognition. *Clinical Linguistics & Phonetics,* 14(4), 307-327.

Deller Jr., J.R., Hsu, D., & Ferrier, L.J. (1988). Encouraging Results in the Automated Recognition of Cerebral Palsy Speech. *IEEE Transactions on Biomedical Engineering,*35(3), 218-220.

DeRosier, R., & Farber, R.S. (2005). Speech recognition software as an assistive device: a pilot study of user satisfaction and psychosocial impact. Work (Reading, Mass.), 25(2), 125-134.

Devillers, L., & Vidrascu, L. (2007). Real-life emotion recognition in speech. In C. Müller (Ed.), *Speaker Classification I, LNAI:* (p34-42). Heidelberg: Springer-Verlag Berlin.

Doyle, P. C., Leeper, H. A., Kotler, A. L., Thomas-Stonell, N., O'Neill, C., Dylke, M.C., & Rolls, K. (1997). Dysarthric speech: A comparison of computerized speech recognition and listener intelligibility. *Journal of Rehabilitation Research and Development*, 34(3), 309-316.

Enderby, P.M. (1983). *Frenchay Dysarthria Assessment.* San Diego, California: College-Hill Press.

Ferrier, L., Shane, H., Ballard, H., Carpenter, T., & Benoit, A. (1995) Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. *Augmentative and Alternative Communication*, 11, 165-175.

Fogle, C.C., Oser, C.S., Troutman, T.P., McNamara, M., Williamson, A.P., Keller, M. et al. (2008). Public education strategies to increase awareness of stroke warning signs and the need to call 911. *Journal of Public Health Management Practice*, 14(3), E17-E22.

Fried-Oken, M. (1985). Voice recognition device as a computer interface for motor and speech impaired people. *Archives of Physical Medicine and Rehabilitation*, 66(10), 678-681.

Gibson, M.J. (2006). We can do better: lessons learned for protecting older persons in disasters. *In AARP Report.* Washington:.*AARP*.

Gorham-Rowan, M.M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders,* 39, 171-184.

Handschu, R., Poppe, R., Rauß, J., Neundörfer, B., & Erbguth, F. (2003). Emergency calls in acute stroke. *Stroke,* 34, 1005-1009.

Hansen, J.H.L., & Patil, S., (2007). Speech under stress: Analysis, modeling and recognition. *In C. Müller (Ed.) Speaker Classification I, LNAI 4343 (p.*108-137). Berlin Heidelberg: Springer-Verlag.

Hasegawa-Johnson, M., Gunderson, J., Penman, A., & Huang, T. HMM-based and SVM-based recognition of the speech of talkers with spastic dysarthria. *2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE Cat. no. 06CH37812C); 14-19 May 2006, Toulouse, France.*

Havstam, C., Buchholz, M., & Hartelius, L. (2003) Speech recognition and dysarthria: a single subject study of two individuals with profound impairment of speech and motor control. *Logopedics Phoniatrics Vocology,* 28(2), 81-90.

Hawley, M.S., Enderby, P., Green, P., Cunningham, S., Brownsell, S., Carmichael, J. et al. (2007). A speech-controlled environmental control system for people with severe dysarthria. *Medical Engineering & Physics*, 29, 586-593.

Hird, K., & Hennessey, N.W. (2007). Facilitating use of speech recognition software for people with disabilities: A comparison of three treatments. *Clinical Linguistics & Phonetics*, 21(3), 211-226.

Hosom, J.P., Kain, A.B., Mishra, T., van Santen, J.P.H., Fried-Oken, M., & J. Staehely. (2003). Intelligibility of modifications to dysarthric speech. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings,* 1, 924-927.

Hux, K., Rankin-Erickson, J., Manasse, N., & Lauritsen, E. (2000). Accuracy of three speech recognition systems: Case study of dysarthric Speech. *Augmentative and Alternative Communication*, 16, 186-196.

Hwang, U., & Morrison, R.S. (2007). The Geriatric Emergency Department. *Journal of the American Geriatrics Society*, 55, 1873-1876.

Jurafsky, D. & Martin, A. (2008) Speech and Language Processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2nd Edition. Upper Saddle River, N.J.: Prentice Hall.

Kambeyanda, D., Singer, L., & Cronk, S. (1997). Potential problems associated with use of speech recognition products. *Assistive Technology*, 9, 95-101.

Kayasith, P., Theeramunkong, T., & Thubthong, N. (2006). Recognition rate prediction for dysarthric speech disorder via speech consistency score. *Lecture Notes in Computer Science,* 885-889.

Koester, H. H. (2004). Usage, performance, and satisfaction outcomes for experienced users of automatic speech recognition. *Journal of Rehabilitation Research and Development,* 41(5), 739-754.

Kotler, A.L, & Tam, C. (2002). Effectiveness of using discrete utterance speech recognition software. *Augmentative and Alternative Communication*, 18, 137-146.

Kotler, A.L., Thomas-Stonell, N. (1997). Effects of speech training on the accuracy of speech recognition for an individual with a speech impairment. *Augmentative & Alternative Communication*, 13, 71-80.

LaPointe, L.L. (1994). Neurogenic disorders of communication. In F.D. Minifie (Ed.), *Introduction to communication sciences and disorders:* Chapter 9 (pp. 351-397). Singular Publishing Group: McNaughton & Gunn.

Lillvik, M., Allemark, E., Karlstrom, P., & Hartelius, L. (1999). Intelligibility of dysarthric speech in words and sentences: development of a computerised assessment procedure in Swedish. *Logopedics Phoniatrics Vocology*, 24(3), 107-119.

Linville, S.E. (2002). Source characteristics of aged voice assessed from long-term average spectra. *Journal of Voice*, 16(4), 472-479.

Lippmann, R.P. (1997). Speech recognition by machines and humans. *Speech Communication*, 22, 1-15.

Mann, W.C., Belchior, P, Tomita, M. & Kemp, B.J. (2005). Use of personal emergency response systems by older individuals with disabilities. *Assistive Technology*, 17, 82-88.

McLean, McLean, M., Boger, J., & Mihailidis, A. (2009 in press). Development of an Automated Speech Recognition Interface for Personal Emergency Response Systems. *International Journal of Speech Technology.*

Mihailidis, A., Tam, T., McLean, M. and Lee, T. (2005). An intelligent health monitoring and emergency response system. In: From Smart Homes to Smart Care, Eds: S. Giroux, H. Pigot, International Conference on Smart Homes and Health Telematics (ICOST), Sherbrooke, Canada, 272-281.

Noyes, J., & Starr, A. (1996). Use of automatic speech recognition: current and potential applications. *Computing & Control Engineering Journal*, (October), 203-208.

Noyes, J. M., Haigh, R., & Starr, A. F. (1989). Automatic speech recognition for disabled people. *Applied Ergonomics,* 20(4), 293-298.

Noyes, J. M. & Frankish, C. (1992) Speech recognition technology for individuals with disabilities, *Augmentative and Alternative Communication*, 8(December), 297-303.

Olson, D. E. L., Cruz, R.M., Izdebski, K, & Baldwin, T. (2004) Muscle tension dysphonia in patients who use computerized speech recognition systems. *ENT-Ear, Noise & Throat Journal,* 83(3), 195-198.

Porter, E.J. (2005, Oct). Wearing and using personal emergency response systems. *Journal of Gerontological Nursing,* 26-33.

Polur, P.D., & Miller, G.E. (2005). Effect of high-frequency spectral components in computer recognition of dysarthric speech based on a mel-cepstral stochastic model. *Journal of Rehabilitation Research and Development, 42*(3), 363-371.

Raghavendra, P., Rosengren, E., & Hunnicutt, S. (2001). An investigation of different degrees of dysarthric speech as input to speaker-adaptive and speaker-dependent recognition systems. *Augmentative and Alternative Communication, 17*(4), 265-275.

Ramig, L.O. (1994). Voice Disorders. In F.D. Minifie (Ed.), *Introduction to communication sciences and disorders:* Chapter 12 (pp. 481-519). Singular Publishing Group: McNaughton & Gunn.

Rosamond, W.D., Evenson, K.R., Schroeder, E.B., Morris, D.L., Johnson, A.M., & Brice, J.H. (2005). Calling emergency medical services for acute stroke: A study of 9-1-1 tapes. *Prehospital Emergency Care,* 9(1), 19-23.

Rosen, K., & Yampolsky, S. (2000). Automatic speech recognition and a review of its functioning with dysarthric speech. *Augmentative and Alternative Communication, 2000, 16, 1, Mar, 16*(1), 48-60.

Roy, N., Stemple, J., Merrill, R.M., & Thomas, L. (2007). Epidemiology of voice disorders in the elderly: preliminary findings. *The Laryngoscope,* 117: 628-633.

Salvi, F., Morichi, V., Grilli, A., Giorgi, R., De Tommaso, G., & Dessi-Fulgheri, P. (2007). The elderly in the emergency department: a critical review of problems and solutions. *Internal and Emergency Medicine*, 2(4), 292-301.

Sy, B. K., & Horowitz, D. M. (1993). A statistical causal model for the assessment of dysarthric speech and the utility of computer-based speech recognition. *IEEE Transactions on Biomedical Engineering, 40*(12), 1282(17)-1299.

Thomas-Stonell, N., Kotler, A., Leeper, H. A., & Doyle, P. C. (1998). Computerized speech recognition: Influence of intelligibility and perceptual consistency on recognition accuracy. *AAC: Augmentative and Alternative Communication, 14*(1), 51-56.

Wade, J., Petheram, B., & Cain, R. (2001). Voice recognition and aphasia: Can computers understand aphasic speech? *Disability & Rehabilitation, 23*(14), 604-613.

Wilpon, J.G., & Jacobsen, C.N. (1996). A study of speech recognition for children and the elderly. *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, 7-10 may 1996, Atlanta, GA, USA. Signal Process. Soc. IEEE,* 349-52 vol. 1.

Yorkston, K.M., Beukelman, D.R., & Traynor, C.D. (1984). *Computerized assessment of intelligibility of dysarthric speech.* Tigard, OR: C.C. Publications.

Yorkston, K.M., Beukelman, E.R., & Bell, K. (1988). Clinical Management of Dysarthric Speakers. Massachusetts : Little, Brown and Company – College-Hill Publications.

Zraick, R.I., Gregg, B.A., & Whitehouse, E.L. (2006). Speech and voice characteristics of geriatric speakers: A review of the literature and a call for research and training (Tutorial). *Journal of Medical Speech-Language Pathology,* 14(3), 133-142.

TABLES

Table 1: Similarities in older adult and dysarthric speech characteristics.

Older Adult Speech (Gorham-Rowan & Laures-Gore, 2006; Linville, 2006; Zraick et al., 2006)	**Dysarthric Speech** (LaPointe, 1994; Yorkston et al., 1988)
• changes in fundamental frequency or pitch	• sudden changes in pitch
• articulation imprecision (e.g., longer voice-onset time, longer duration of vowels and consonants)	• poor articulation • dysfluencies (e.g., sound, syllable or word repetitions)

• increased respiration frequency
  ○ intra-word pauses
	• breathiness, phonatory control difficulties
	  ○ intra-word pauses
	  ○ non-speech sounds

• slower pace	• inconsistent speech rate (e.g., unsteady, slow or sudden variability)
• increased voice perturbations (e.g., tremor, spectral noise, hoarseness)	• hypernasality, involuntary noises (e.g., coughing, laughing, saliva, grunts, lip smacking)
• decreased voice intensity	• reduced/increased loudness