Augmentative and Alternative Communication, June 2004 VOL. 20 (2), pp. 103–112

Taylor & Francis
healthsciences

Telephone Communication: Synthetic and Dysarthric Speech Intelligibility and Listener Preferences

KATHRYN D. R. DRAGER*, KATHERINE C. HUSTAD†, and KERRY L. GABLE

Pennsylvania State University, Pennsylvania, USA

For many individuals with severe speech intelligibility challenges, deciding whether to use one's own oral dysarthric speech or synthesized speech when communicating on the telephone is a difficult process. In the present study, we investigated the intelligibility of dysarthric speech and two types of synthesized speech over the telephone and in quiet conditions, and listeners' preferences for these speech types. Participants heard short sentences in one of three types of speech: dysarthric female speech, DECTalk® Beautiful Betty, and MacinTalk™ Victoria, high quality. They were asked to transcribe sentences for a measure of intelligibility, and to answer questions to indicate their preferences for speech type and their preferences for synthesizer. The overall intelligibility of DECTalk® and MacinTalk™ were significantly higher than the intelligibility of the dysarthric speech. Additionally, the intelligibility of the speaker with dysarthria decreased significantly when listeners were presented with the materials over the telephone. The two synthesized speech types (DECTalk® and MacinTalk™) did not differ significantly in intelligibility between the two conditions. On average, listeners tended to be more comfortable when listening to synthesized speech than when listening to dysarthric speech. Listener preferences for synthesizers were dependent on the condition: most listeners in the Speaker Group preferred the DECTalk® female voice to the MacinTalk™ female voice, while listeners in the Telephone Group were equally distributed.

Keywords: Augmentative and alternative communication (AAC); Synthesized speech; Dysarthric speech; Intelligibility; Telephone

INTRODUCTION

Many individuals who use augmentative and alternative communication (AAC) systems have the ability to use their own natural speech, even though intelligibility may be markedly compromised as a result of dysarthria. These individuals often employ multiple modes of communication, including a combination of gestures, facial expressions, natural speech, low-tech AAC strategies, and voice output AAC systems to enhance their listeners' ability to understand them. Decisions regarding which mode(s) to use during a given interaction are often context- and partner-specific. In face-to-face interaction with familiar partners, for example, natural speech and gestures may be sufficient for successful transmission of meaning between speaker and listener. In other situations, however, such as noisy environments or with unfamiliar partners, AAC systems that provide voice output via speech synthesizer may be necessary in order to ensure successful exchange of meaning.

One communication challenge facing many people who use AAC is use of the telephone. Communicating via telephone differs from face-to-face interaction in several important ways. First, partners are unable to see one another and therefore gestures and facial expression are not effective as part of a multi-modal communication system. Consequently, listeners are required to rely exclusively on the acoustic signal in order to understand the speaker. In addition, the bandwidth available for transmission of the acoustic

*Corresponding author. Department of Communication Sciences and Disorders, Penn State University, 110 Moore Building, University Park, PA 16802, USA. Tel: 814-863-6247. E-mail: kdd5@psu.edu
†Currently affiliated with the Department of Communicative Disorders and Waisman Center, University of Wisconsin – Madison, Madison, WI, USA

ISSN 0743-4618 print/ISSN 1477-3848 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/07434610410001699681

signal via telephone line is reduced. Specifically, the frequencies transmitted by a telephone line are limited to those between 400 Hz and 3400 Hz (Nilsson & Kleijn, 2001). In face-to-face communication, frequencies that carry speech-related information range from 100 Hz to 5000 Hz (Borden, Harris, & Raphael, 2003). Thus, the reduced bandwidth of the telephone line eliminates information at the high and low ends of the speech spectrum, a phenomenon that is not a problem for speakers without disabilities in most circumstances. Finally, additional noise is imposed on the speech signal from the sound field in which the signal was produced (background noise that becomes amplified over the telephone), the telephone line itself, and the sound field in which the signal is received by the listeners (background noise in the room where a listener is located). Thus, the acoustic information available to listeners over the telephone is somewhat degraded relative to what would be present in a face-to-face interaction. These factors seem likely to influence the intelligibility of any communication mode (natural or synthesized) that individuals with disabilities may use over the telephone.

In the present study, we investigated synthesized and dysarthric speech over the telephone and in quiet conditions. Specifically, we examined whether there were differences in the intelligibility of dysarthric speech and two types of synthesized speech over the telephone and in quiet conditions. In addition, listeners' preferences for these speech types were also of interest.

Intelligibility of Synthesized Speech

Speech intelligibility has been defined broadly as the accuracy with which an acoustic signal is conveyed by a speaker and recovered by a listener (Kent, Weismer, Kent, & Rosenbek, 1989; Yorkston & Beukelman, 1980; Yorkston, Strand, & Kennedy, 1996). Several studies of synthesized speech intelligibility have been comparisons of one or two kinds of synthetic speech with non-disordered natural speech (Luce, Feustel, & Pisoni, 1983; Hoover, Reichle, VanTasell, & Cole, 1987; Mitchell & Atkins, 1988; and others). Synthesized speech is typically associated with lower levels of intelligibility than non-disordered natural speech, which contains many redundant acoustic and visual cues that facilitate listeners' ability to recognize sounds and determine contrasts between words. To accurately perceive synthesized speech, listeners need more processing resources than is necessary to perceive non-disordered natural speech. This has been attributed to the absence of visual and acoustic cues available in the synthesized signal (Duffy & Pisoni, 1992). The increased need for processing resources is often realized as a decrease in intelligibility and an increase in response latencies.

Several factors improve the intelligibility of synthesized speech. One of these factors is linguistic context (Drager & Reichle, 2001b; Hoover et al., 1987; Marics & Williges, 1988; Mirenda & Beukelman, 1987, 1990; Oshrin & Siders, 1987; Slowiaczek & Nusbaum, 1985). Another factor that improves the intelligibility of synthesized speech is repeated exposure (McNaughton, Fallon, Tod, Weiner, & Neisworth, 1994; Schwab, Nusbaum, & Pisoni, 1985; Venkatagiri, 1994). However, intelligibility of synthesized speech is negatively affected to a greater extent than non-disordered natural speech in listening situations such as background noise (Fucci, Reynolds, Bettagere, & Gonzales, 1995; Koul & Allen, 1993), or when the listener's attention is divided (Drager & Reichle, 2001a).

In previous studies, the intelligibility of synthesized speech typically has been compared with that of intelligible, non-disordered natural speech. A problem with this approach is that the speech of many individuals with physical disabilities is often compromised because of dysarthria. The intelligibility of an individual's natural speech plays a crucial role in determining appropriate communication strategies (natural speech, AAC systems, or combination of each).

Intelligibility of Dysarthric Speech

The speech intelligibility of individuals who have dysarthria is typically reduced compared with non-disordered speech (Darley, Aronson, & Brown, 1969). The extent to which intelligibility is diminished in individuals with dysarthria is extremely variable. Factors that impact intelligibility include the constellation of speech subsystems involved, and the nature and severity of the impairment. In addition, variables such as the listener and situational/environmental context have an important effect on speech intelligibility (Hustad & Beukelman, 2001; Hustad, Beukelman, & Yorkston, 1998; Kent, 1993; Kent et al., 1989). In some speakers with dysarthria, prosody, stress, and intonation are relatively preserved, resulting in a signal that sounds more natural and contains more acoustic redundancy than synthesized speech. In addition, other, deliberate supplemental AAC interventions such as the use of iconic gestures (Garcia & Dagenais, 1998; Hustad & Garcia, 2002), alphabet cues, topic cues, and combined topic and alphabet cues (Beliveau, Hodge, & Hagler, 1995; Hustad & Beukelman, 2001; Hustad, Jones, & Dailey, 2003)

have been shown to dramatically increase the intelligibility of dysarthric speech.

Some of the same variables that influence synthesized speech intelligibility also influence the intelligibility of dysarthric speech. These include context (Garcia & Cannito, 1996; Garcia & Dagenais, 1998; Hustad & Garcia, 2002; Hustad & Beukelman, 2001, 2002), and listener familiarity (DePaul & Kent, 2000; Hustad & Cahill, 2003; Tjaden & Liss, 1995; Yorkston & Beukelman, 1983). Because advances in technology have allowed the intelligibility of higher quality synthesized speech signals to approach that of non-disordered natural speech (Mirenda & Beukelman, 1987, 1990), it can be assumed that the intelligibility of synthesized speech would be higher than that of dysarthric speech, in all but the mildest of cases. Even the best speech synthesizers, however, lack the prosodic, supra-segmental, and naturalness features present in the human speech signal (Kent & Read, 2002). Although synthesized speech may be more intelligible than moderately or severely dysarthric speech, these naturalness parameters may affect the preferences of listeners for hearing synthesized versus natural (dysarthric) speech.

Telephone Communication

Communicating intelligibly over the telephone often is an issue of great concern for individuals who have dysarthria. Currently, individuals with severe speech intelligibility challenges may: (a) use their own (dysarthric) speech to communicate, (b) use a voice output communication aid (VOCA) with synthesized or digitized speech output, (c) have someone else make the phone call, or (d) use a telecommunication relay service (TRS) (Silverman & Schauer, 1996).

Only recently have researchers begun to focus on synthesized speech and telephone use. Riley and Fries (2000), for example, compared the intelligibility of DECTalk™ and digitized speech (on a communication device) over the telephone. Results showed that digitized speech was more intelligible (96.5%) than synthesized speech (90.8%). These researchers did not compare the intelligibility of synthesized speech over the telephone with other conditions (e.g., in a free-field), however, to examine the effect of the telephone on intelligibility.

Nakamura, Arima, Sakamoto, and Toyota (1993) investigated listeners' preferences for synthesized speech over the telephone. In this investigation, researchers contacted participants over the telephone, using a voice-output device to ask common questions (i.e., "*Is _ at home?*"). After the telephone call was completed, the researcher phoned the participant again using his natural voice and asked the participant about his or her impressions and feelings during the previous telephone call. Results showed that male participants reported feelings of anxiousness, while female participants reported more positive experiences. These results are consistent with previous research findings that show that listeners prefer human speech to synthesized speech, and rate human speech as more natural (Crabtree, Mirenda, & Beukelman, 1990; Mirenda & Beukelman, 1990; Ratcliff, Coughlin, & Lehman, 2002). To date, there is no published research comparing listeners' judgments or perceptions of synthesized speech versus natural, dysarthric speech over the telephone. The added component of communication over the telephone may also have an impact on listener attitudes. The first utterance of the phone conversation is critical, as some listeners may hang up on the caller if there is too long a delay, or if the communication mode is considered unintelligible or unacceptable. For speakers with dysarthria, this preference may be one factor in the decision about which communication mode to use in any given situation.

There is no published research on the intelligibility of dysarthic or synthesized speech or listener preferences for dysarthic or synthesized speech over the telephone; to make appropriate recommendations regarding the use of dysarthric speech or synthesized speech via this communication mode, it is important to ascertain two things. First, what are the preferences of listeners for each of these speech types? Second, what is the intelligibility of each speech type over the telephone compared to normal conditions?

METHOD

Participants and Selection Criteria

Forty listeners participated in the study. Listeners were recruited through posted flyers or solicitations in university classrooms and were compensated for their time. All participants were adults who self-reported to be native English-speaking with no history of speech and language, psychiatric, or neurologic impairments. All participants reported themselves to be in good health and to have no intellectual or cognitive problems. Hearing was tested and was found to be within normal limits and no participant had more than incidental exposure to synthesized speech or more than 1 h of interaction with an individual with dysarthic speech. Participants were randomly assigned to two groups, with 20 participants in each group. In the Telephone Group, participants listened to sentences through a telephone handset.

In the Speaker Group, participants listened to sentences through free-field speakers.

The Telephone Group consisted of 11 female and 9 male participants. The mean age of participants was 24.3 years (range 18–40 years). The Speaker Group consisted of 11 female and 9 male participants. The mean age of participants in this condition was 20.8 years (range 18–27).

Materials

The participants in both groups listened to sentences from the Hearing in Noise Test (HINT) (Nilsson, Soli, & Sullivan, 1994). The HINT consists of semantically and syntactically complete sentences ranging from four to seven words in length that are phonemically matched and balanced. The HINT was selected as stimuli because of its well controlled psychometric properties and because it is an accepted set of speech stimuli for research focusing on intelligibility of dysarthric speech (see Beukelman, Fager, Ullman, Hanson, Logemann, 2002; Hustad & Cahill, 2003). Four different sentence lists were employed for this study, for a total of 40 sentences. Each list of the 40 sentences was generated in three types of speech: dysarthric female speech, DECTalk® Beautiful Betty, and MacinTalk™ Victoria, high quality. Three of the lists were used for intelligibility tasks, while sentences from the fourth list were used for the preference tasks.

A 45 year old woman with spastic-athetoid dysarthria secondary to cerebral palsy produced the dysarthric speech sample. Her intelligibility was approximately 85% as measured by the Sentence Intelligibility Test (SIT) (Yorkston, Beukelman, & Tice, 1996), thus resulting in a clinical designation of mild dysarthria. The following perceptual features characterized her speech: hypernasality, nasal air emission, imprecise consonants, and intermittent breathiness. The speaker indicated that she used speech as her primary mode of communication. She also indicated that she frequently had difficulty communicating over the telephone.

For this experiment, the speech sample from the woman with dysarthria was recorded on digital audio tape (DAT) (44.1 kHz sampling rate; 16 bit quantization) in a double walled soundproof room with the speaker wearing a headmounted microphone. The DECTalk® sample was generated from a Multivoice™ external speech synthesizer, in Beautiful Betty, a commonly used female voice. The MacinTalk™ sample was generated from a Macintosh PowerBook 190cs, in Victoria, high quality. Speech from both synthesizers was digitized onto a personal computer (44.1 kHz sampling rate; 16 bit quantization) using Sound Forge 4.1 (computer software). The sentences were produced at approximately 180 words per minute, which was the dysarthric speaker's natural rate of speech, and the default setting for speech rate for the synthesized sentences. All sentences were peak amplitude normalized and then digitally recorded to compact discs.

Procedure

Because many individuals with motor impairments use speakerphones for telephone conversations, a speakerphone was used for the telephone transmission in this investigation. For the Telephone Group, stimulus CDs were played through a loud speaker, which was placed approximately 0.6 meters from a speakerphone. The presentation level of the speech stimuli, as measured from the speakerphone was 65 dB SPL, a level selected to approximate a conversational loudness. The speech signal was transmitted via telephone line into an adjacent sound-treated room where participants listened to the sentences through a telephone handset. For both groups, listeners participated individually.

For the Speaker Group, the stimulus CDs were played for participants in a sound-treated room via stereo speakers in a free field environment. Participants were seated approximately 1 meter from each speaker and stimuli were presented at a level of 65 dB SPL.

All participants were asked to perform three tasks. For the first task, intelligibility, participants were asked to listen to 10 different sentences of each speech type (dysarthric, DECTalk®, and MacinTalk™) and to transcribe each sentence. The order of presentation of speech type was counterbalanced across listeners, and each sentence was presented only once. In addition, sentence lists were counterbalanced so that different lists were associated with the three speech types for different listeners. Each list contained an approximately equal number of words-per-sentence. The CD playback was paused between each sentence to allow sufficient time for transcription. Participants raised their hands to indicate that they were ready for the next sentence. Each listener transcribed 30 sentences, 10 of each speech type.

After completing the intelligibility task with three lists of sentences, each listener completed a series of qualitative questions that targeted their preference for speech type (dysarthric or synthesized) and preference for synthesizer (DECTalk® or MacinTalk™). For the preference for speech type questions, each participant was presented

with a novel single sentence from the HINT in each of the three speech types. Participants were instructed to listen to the speech samples and were asked to respond to a series of four qualitative questions using a Likert-type scale (1–10). The questions required the participants to rate their comfort level, understanding, perceptions of competence of the speaker, and their willingness to interact with the individual when she used her own speech to communicate. For the preference for synthesizer question, each participant was presented with a single sentence produced by each of the two synthesized speech types: DECTalk® and MacinTalk™. Listeners were asked to choose which voice they preferred and provide a brief explanation of why.

Data Analysis

For the intelligibility tasks, the dependent variable was intelligibility, as measured by response accuracy (percent of words in each sentence transcribed correctly). A word was considered correct if it was an exact phonemic match to the corresponding word in the sentence. The dependent variable was intended to account for the number of words that were accurately conveyed to the listener, independent of understanding or comprehension. Therefore, word order was not considered in calculating correct responses. Misspellings were considered correct only if they resulted in an exact phonemic match to the target word. The mean intelligibility for each speech type for each group was calculated. The data were analyzed using a 2×3 mixed-factor analysis of variance (ANOVA). The between-subjects factor was Participant Group – Telephone Group and Speaker Group. The within-subjects factor was Speech Type – dysarthric speech, DECTalk®, and MacinTalk™.

For the preference for speech type task, the dependent variable was response on the Likert-type scale. The mean rating on each question for each group was calculated. Differences between groups were tested using a Mann–Whitney U test. For the preference for synthesizer task, the dependent variable was response on the forced-choice question. The number of listeners who preferred each synthesizer in each group was tallied. A nonparametric binomial test was used to determine if there was a difference between groups.

Reliability

To evaluate the reliability with which the experimenter scored intelligibility, an independent judge scored intelligibility for 20% of the participants' responses (six sentences for each participant). Using the formula: percent agreement = total number of words agreed on/(total number of words agreed on + total number of words disagreed about) *100, the reliability was 97% (range 90–100%), which indicated good reliability for the scoring of the dependent measure.

RESULTS

Intelligibility

Figure 1 summarizes the mean percent intelligibility (with error bars representing one standard deviation) for each speech type for each group. Overall, the mean intelligibility (percent of words correctly transcribed) for participants in the Speaker Group was 95.8%. The mean intelligibility for participants in the Telephone Group was 87.7%. The difference between these two conditions (main effect for Group) was statistically significant, $F(1,38) = 26.99$, $p < 0.0001$. Intelligibility was significantly higher for listeners in the Speaker Group than in the Telephone Group. The main effect for Speech Type was also statistically significant, $F(1,76) = 113.95$, $p < 0.0001$. Follow-up analyses of group differences were accomplished with three paired-samples t-tests in which the alpha level was adjusted using the Bonferroni procedure. The overall intelligibility of DECTalk® (98.0%) and MacinTalk™ (96.7%) were significantly higher than the intelligibility of the dysarthric speech (80.6%), $t(39) = 8.41$, $p < 0.0001$ and $t(39) = 7.60$, $p < 0.0001$ respectively. Intelligibility of the MacinTalk™ and DECTalk® synthesizers did not differ from one another, $t(39) = -1.92$, $p = 0.062$.

The Group × Speech Type interaction was also significant, $F(1,76) = 34.49$, $p < 0.0001$. The nature of the interaction was explored with three

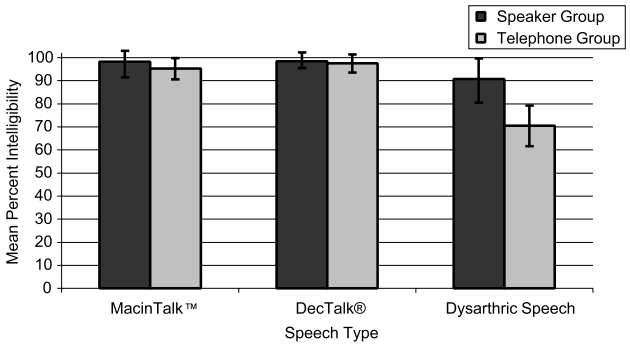

FIGURE 1 Intelligibility of synthesized and dysarthric speech by speech type and listening condition (error bars represent 1 SD).

independent-samples *t*-tests, in which the alpha level was again adjusted using the Bonferroni procedure. Results showed that intelligibility for the Speaker Group was significantly higher than for the Telephone Group when listening to dysarthric speech, $t(38) = -6.67$, $p < 0.0001$. Neither of the two synthesized speech types differed significantly in intelligibility between the two conditions [DECTalk®: $t(39) = -0.87$, $p = 0.392$, MacinTalk™: $t(39) = -1.66$, $p = 0.105$].

Preference for Speech Type

Mean listener responses to preference questions are presented graphically in Figure 2. Descriptive results showed that the mean rating of listener comfort across groups was 3.80 (mean response for Speaker Group = 4.75; mean response for Telephone Group = 2.85). The mean rating of listener understanding was 2.08 across groups (mean response for Speaker Group = 2.10; mean response for Telephone Group = 2.05). The mean rating for competence was 3.38 (mean response for Speaker Group = 3.45; mean response for Telephone Group = 3.30). For each of these three questions, listener responses indicated a clear preference for synthesized speech over dysarthric speech.

Finally, the mean rating for listener willingness to communicate with the speaker using her own speech was 3.50 (mean response for Speaker Group = 2.80; mean response for Telephone Group = 4.20). It is important to note that the lower numbers reflected greater willingness. These results suggest that listeners were very willing to communicate with the speaker using her own speech.

The Mann-Whitney *U* test revealed that the difference between groups was significant for the comfort question only, $Z = -2.258$, $p = 0.024$. Participants in the Telephone Group indicated a higher level of comfort with synthesized speech compared to the dysarthric speech than did participants in the speaker condition. For the remaining questions, there were no significant differences between participants' responses in the speaker and the telephone conditions.

Preference for Synthesizer

In a forced-choice response for preference for synthesizer, 16 listeners in the Speaker Group (80%) preferred DECTalk®, while 4 listeners in this group (20%) preferred MacinTalk™. In contrast, 10 listeners in the Telephone Group (50%) preferred DECTalk® and 10 listeners (50%) preferred to listen to MacinTalk™. A nonparametric binomial test revealed a significant result for the Speaker Group ($p = 0.012$). Significantly more participants in this group

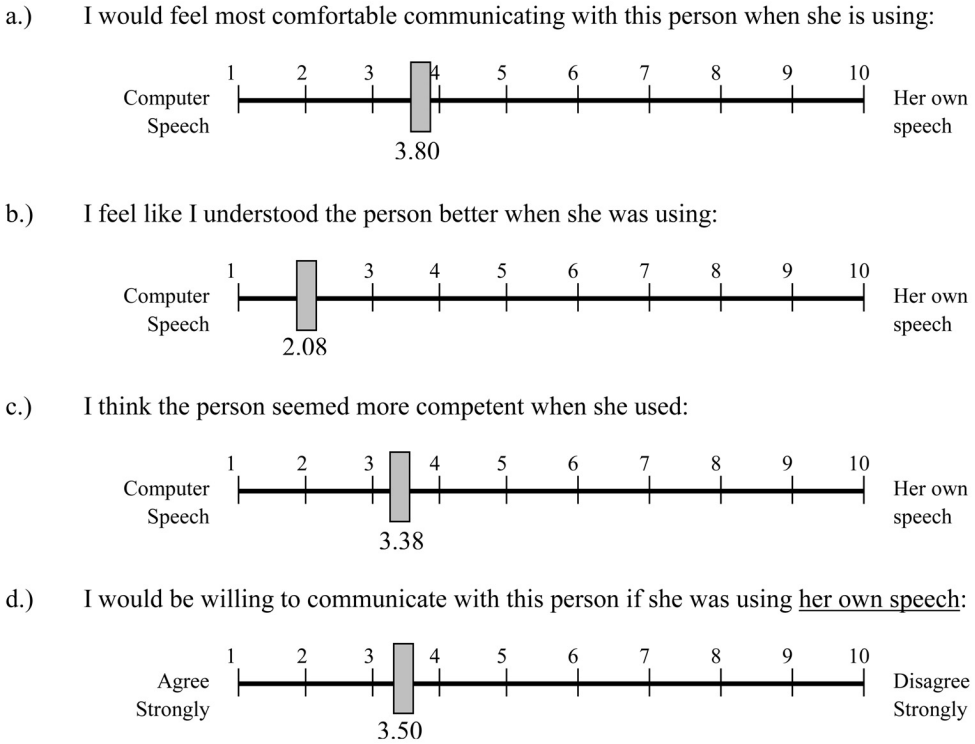

FIGURE 2 Results of Likert-type scale questions for preference for speech type.

preferred DECTalk® than preferred MacinTalk™.

DISCUSSION

In the present study, we compared the intelligibility of synthesized (DECTalk® and MacinTalk™) and dysarthric speech in two different listening conditions: over the telephone and in a sound field. Results showed that intelligibility did not differ between the two listening conditions for either of the two speech synthesizers. In contrast, intelligibility did decrease significantly (by about 20%) for the speaker with dysarthria in the telephone condition. Also of interest were ratings of listener preference for speech type (dysarthric, DECTalk®, MacinTalk™) and listener preference for synthesizer (DECTalk® versus MacinTalk™). Results showed that listeners, particularly those who were in the Telephone Group, seemed to prefer the synthesized speech, but were very willing to communicate with the speaker when she used her natural speech. Finally, results showed that listeners preferred the DECTalk® voice in the sound field (speaker) condition, and preferred both synthesizers equally in the telephone condition. Findings are discussed in detail below.

Intelligibility

Results of the present study suggest that synthesized speech is more robust with regard to intelligibility than the speech of the speaker with dysarthria, when the bandwidth of available speech frequencies is reduced. Indeed, the intelligibility of the speaker with dysarthria in the present study decreased by more than 20% when her speech was presented over the telephone, compared to face-to-face interactions. Interestingly, this speaker, who had mild dysarthria (91% intelligible) in face-to-face interactions, would be considered to have a moderate dysarthria over the telephone (70% intelligible). This decrease in intelligibility would likely result in some difficulty in telephone communication, especially if the content of her message was unpredictable. It seems logical that this difference in intelligibility can be attributed, in large part, to the restricted bandwidth of telephone transmission, that is extreme high and low frequencies carrying speech-related information are removed from the signal presented to the listener over the telephone. This effect may be particularly pronounced for the speaker who participated in this study because of her gender. That is, female speech tends to have more energy in the higher frequencies—specifically females have higher formant frequencies and larger formant bandwidths than male speakers (Kent & Read, 2002). The loss of some high frequency information via telephone transmission likely compromised her intelligibility to a greater extent than might be the case for a male speaker or when her speech was presented to listeners in the sound field via speaker.

The finding, in the present study, that speech intelligibility was reduced over the telephone is particularly troubling for speakers with dysarthria. This result likely explains some of the difficulty that individuals with motor speech disorders frequently report when using the telephone. The speaker in the present study had a mild dysarthria. The effect of diminished intelligibility over the telephone may be more or less pronounced for individuals with more or less severe dysarthria.

In contrast, MacinTalk™ and DECTalk® did not differ significantly in intelligibility between speaker and phone conditions. In fact, the intelligibility for both was near 100%. The two speech synthesizers also did not differ significantly in intelligibility from each other. Hustad, Kent, and Beukelman (1998) found that with the exception of one of the DECTalk® voices being vastly superior (“Perfect Paul”), DECTalk® and MacinTalk™ synthesizers were comparable for word level intelligibility. In the present study, these two synthesizers are also comparable for sentence intelligibility for the default female voices.

For individuals with dysarthria attempting to decide which oral communication mode to use over the telephone, results of the present study suggest that the use of synthesized speech via a VOCA may be a viable alternative to using dysarthric speech to ensure optimal intelligibility. The decision to use synthesized speech, however, does not necessarily mean that it need be used exclusively. In particular, synthesized speech may be beneficial in certain contexts, while natural (dysarthric) speech may be sufficient for others. As an example, when the telephone conversation includes urgent or time-dependent messages, where it is critical that the message be transmitted as quickly as possible, synthesized speech may be more efficient than dysarthric speech. Similarly, synthesized speech may be used for introductory messages for unfamiliar listeners. The more intelligible synthesized speech may be used to orient the listener to the speaker and the topic of the conversation. Once established, the individual may then rely on natural speech for the remainder of the discussion. Additionally,

synthesized messages may be prepared to help resolve communication breakdowns that may occur as a result of the diminished intelligibility of the dysarthric speech.

Preference for Speech Type

On average, listeners tended to be more comfortable when they listened to synthesized speech than when they listened to dysarthric speech. Listeners also felt that they understood the speaker better when synthesized speech was used, and rated the speaker as somewhat more competent when she used synthesized speech. These results are consistent with intelligibility data, which showed that intelligibility of synthesized speech was higher than that of dysarthric speech across both listening conditions. Regardless of these perceptions, listeners indicated a willingness to communicate with the speaker when she used her own speech.

The effect of telephone listening was only significant for listener comfort level. This was likely the result of the significant decrease in intelligibility of the dysarthric speech over the telephone. Listeners in the Telephone Group were more comfortable listening to the synthesized speech (compared with the dysarthric speech, with which they had significant difficulty understanding) than listeners in the Speaker Group.

Preference for Synthesizer

Despite the absence of intelligibility differences between the two speech synthesizers, listeners in the Speaker Group overwhelmingly preferred the DECTalk® female voice to the MacinTalk™ female voice. Unfamiliar listeners may be more accepting of this voice in typical listening situations. When the sentences were presented over the telephone, however, similar preference ratings for the two speech synthesizers were found. It is important to note, however, that the perceivable difference between the two synthesizers was manifested only in qualitative preference differences, and did not compromise the intelligibility of either system.

Half of the listeners in both conditions who preferred DECTalk® thought the synthesizer was clearer and easier to understand (8 of the 16 listeners in the Speaker Group; 5 of the 10 listeners in the Telephone Group). In contrast, while 31% of listeners in the Speaker Group stated that DECTalk® was more natural and human sounding, 70% of listeners in the Telephone Group stated these same reasons for their preference for MacinTalk™.

Limitations and Future Directions

Although the results of the present study clearly demonstrate that telephone transmission has a marked detrimental effect on intelligibility of dysarthric speech, there are several factors that may have affected these results. Although efforts were made to consider social validity (e.g., use of the speakerphone), the context and sentences were contrived. A typical telephone conversation consists of strings of sentences that are closely related in topic. In the present study, sentences that were unrelated to one another were used. The additional context that a real telephone conversation would provide would likely increase the intelligibility for both types of speech. Additionally, the listeners in the study were all unfamiliar listening partners. Although speaking with unfamiliar listeners is a common situation for individuals with expressive communication difficulties, the results of the study did not consider the conversations that took place with listeners who were familiar with the speaker's natural voice. In the current study, listeners were also allowed to hear each sentence only once. In a true telephone conversation, communication breakdown strategies, such as repeating the message, would be available.

There are also several factors that may have resulted in higher than expected intelligibility outcomes in the present study. For instance, each sentence was played into the speakerphone in a sound protected environment, and each listener was placed in a sound protected environment while listening to the sentences. In reality, the use of a speakerphone in a typical environment would introduce more background noise than that which was present in the current study. Additionally, the listener in a typical environment would also be subjected to background noise and potential outside distracters. In the present study, sentences that were 4–7 words in length were also used. A typical conversation likely consists of sentences that would be both shorter and longer. It is unclear what the effect of sentence length on intelligibility for the two speech types might. Longer sentences may have resulted in a decrease in intelligibility for the synthesized speech samples (Drager & Reichle, 2001b).

Future research in this area is needed into the effects of telephone communication for speakers with different types of dysarthria and different levels of severity, to determine if the results would be consistent. If intelligibility is severely compromised by telephone transmission, strategies to enhance dysarthric speech intelligibility over the telephone must be identified. Additionally, investigation into different synthesizers and different

synthesized voices (e.g., male, child) is warranted to determine the relative impact of the telephone for many individuals who use VOCAs. Other factors that may affect intelligibility include background noise, the type of message, the way in which the message is formulated (prestored ahead of time, or formulated during the course of the conversation), and familiarity of the listener.

CONCLUSION

The present study represents a first look at factors affecting decisions about whether to use dysarthric speech or synthesized speech when communicating on the telephone. From the results of the study, it appears that the use of synthesized speech may be one alternative to increase intelligibility of messages over the telephone. This information on intelligibility and listener attitudes, combined with other critical factors, such as speaker preference for communication mode, may assist speakers with severe speech intelligibility challenges to make this decision.

Acknowledgements

Portions of this paper were presented at the Ninth International Conference of the International Society for Augmentative and Alternative Communication, Washington, DC, August 2000, and the Annual Convention of the American Speech-Language-Hearing Association, Washington, DC, November 2000.

References

Beliveau, C., Hodge, M., & Hagler, P. (1995). Effect of supplemental linguistic cues on the intelligibility of severely dysarthric speakers. *Augmentative and Alternative Communication*, *11*, 176–186.

Beukelman, D., Fager, S., Ullman, C., Hanson, E., & Logemann, J. (2002). The impact of speech supplementation and clear speech on the intelligibility and speaking rate of people with traumatic brain injury. *Journal of Medical Speech Language Pathology*, *10*, 237–242.

Borden, G., Harris, K., & Raphael, L. (2003). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech* (4th edn.). Baltimore: Lippincott, Williams & Wilkins.

Crabtree, M., Mirenda, P., & Beukelman, D. R. (1990). Age and gender preferences for synthetic and natural speech. *Augmentative and Alternative Communication*, *5*, 256–261.

Darley, F. L., Aronson, A. S., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, *12*, 246–269.

DePaul, R., & Kent, R. D. (2000). A longitudinal case study of ALS: Effects of listener familiarity and proficiency on intelligibility judgments. *American Journal of Speech Language Pathology*, *9*, 230–240.

Drager, K. D. R., & Reichle, J. E. (2001a). Effects of age and divided attention on listeners' comprehension of synthesized speech. *Augmentative and Alternative Communication*, *17*, 109–119.

Drager, K. D. R., & Reichle, J. E. (2001b). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC. *Journal of Speech Language and Hearing Research*, *44*, 1052–1057.

Duffy, S. A., & Pisoni, D. B. (1992). Comprehension of synthetic speech produced by rule: A review and theoretical interpretation. *Language and Speech*, *35*, 351–389.

Fucci, D., Reynolds, M. E., Bettagere, R., & Gonzales, M. D. (1995). Synthetic speech intelligibility under several experimental conditions. *Augmentative and Alternative Communication*, *11*, 113–117.

Garcia, J., & Cannito, M. (1996). Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. *Journal of Speech and Hearing Research*, *39*, 750–760.

Garcia, J. M., & Dagenais, P. A. (1998). Dysarthric sentence intelligibility: Contribution of iconic gestures and message predictiveness. *Journal of Speech, Language, and Hearing Research*, *41*, 1282–1293.

Hoover, J., Reichle, J., Van Tassell, D., & Cole, D. (1987). The intelligibility of synthesized speech: Echo II versus Votrax. *Journal of Speech and Hearing Research*, *30*, 425–431.

Hustad, K. C., & Beukelman, D. R. (2001). Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. *Journal of Speech, Language, and Hearing Research*, *44*, 407–510.

Hustad, K. C., & Beukelman, D. R. (2002). Listener comprehension of severely dysarthric speech: Effects of linguistic cues and stimulus cohesion. *Journal of Speech, Language, and Hearing Research*, *45*, 545–558.

Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech Language Pathology*, *12*, 198–208.

Hustad, K. C., & Garcia, J. M. (2002). The influences of alphabet supplementation, iconic gestures, & predictive messages on intelligibility of a speaker with cerebral palsy. *The Journal of Medical Speech Language Pathology*, *10*, 279–285.

Hustad, K. C., Beukelman, D. R., & Yorkston, K. M. (1998). Functional outcome assessment in dysarthria. *Seminars in Speech and Language*, *19*, 291–302.

Hustad, K. C., Jones, T., & Dailey, S. (2003). Implementing speech supplementation strategies: Effects on intelligibility and speech rate of individuals with chronic severe dysarthria. *Journal of Speech, Language, and Hearing Research*, *46*, 462–474.

Hustad, K. C., Kent, R. D., & Beukelman, D. R. (1998). DECTalk and MacinTalk speech synthesizers: Intelligibility differences for three listener groups. *Journal of Speech, Language, and Hearing Research*, *41*, 744–752.

Kent, R. (1993). Speech intelligibility and communicative competence in children. In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing Children's Communication: Foundations for Intervention* (Vol. 2, pp. 223–239). Baltimore: Paul H. Brookes.

Kent, R. D., & Read, C. D. (2002). *Acoustic Analysis of Speech* (2nd edn). Albany, NY: Singular Publishing Group, Inc.

Kent, R., Weismer, G., Kent, J., & Rosenbek, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, *54*, 482–499.

Koul, R. K., & Allen, G. D. (1993). Segmental intelligibility and speech interference thresholds of high-quality synthetic speech in presence of noise. *Journal of Speech and Hearing Research*, *36*, 790–798.

Luce, P., Feustel, T., & Pisoni, D. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*, *25*(1), 17–32.

Marics, M. A., & Williges, B. H. (1988). The intelligibility of synthesized speech in data inquiry systems. *Human Factors*, *30*, 719–732.

McNaughton, D., Fallon, K., Tod, J., Weiner, F., & Neisworth, J. (1994). Effect of repeated listening experiences on the intelligibility of synthesized speech. *Augmentative and Alternative Communication*, *10*, 161–168.

Mirenda, P., & Beukelman, D. R. (1987). A comparison of speech synthesis intelligibility with listeners from three age groups. *Augmentative and Alternative Communication*, *3*, 120–128.

Mirenda, P., & Beukelman, D. R. (1990). A comparison of intelligibility among natural speech and seven speech synthesizers with listeners from three age groups. *Augmentative and Alternative Communication*, *6*, 61–68.

Mitchell, P., & Atkins, C. (1988). A comparison of the single word intelligibility of two voice output communication aids. *Augmentative and Alternative Communication*, *5*(2), 84–88.

Nakamura, K., Arima, M., Sadamoto, A., & Toyota, R. (1993). Telephoning with a voice output device: Listener reactions. *Augmentative and Alternative Communication*, *9*, 251–258.

Nilsson, M., & Kleijn, W. B. (2001). Avoiding overestimation in bandwidth extension of telephony speech. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, *2*, 869–872.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech recognition thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, *95*, 1085–1099.

Oshrin, S. E., & Siders, J. A. (1987). The effect of word predictability on the intelligibility of computer synthesized speech. *Journal of Computer-Based Instruction*, *14*, 89–90.

Ratcliff, A., Coughlin, S., & Lehman, M. (2002). Factors influencing ratings of speech naturalness in augmentative and alternative communication. *Augmentative and Alternative Communication*, *18*, 11–19.

Riley, J. K., & Fries, B. (2000, August). "*Hello, hello!" Synthetic versus natural speech over the telephone*. Poster session presented at the biennial conference of the International Society for Augmentative and Alternative Communication (ISAAC), Washington, DC.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors*, *27*, 395–408.

Silverman, F. H., & Schauer, S. L. (1996). Why do so few persons who are severely speech impaired for reasons other than deafness use telecommunication relay services? *Augmentative and Alternative Communication*, *12*, 59.

Slowiaczek, L. M., & Nusbaum, H. C. (1985). Effects of speech rate and pitch contour on the perception of synthetic speech. *Human Factors*, *27*, 701–712.

Tjaden, K. K., & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics and Phonetics*, *9*, 139–154.

Venkatagiri, H. S. (1994). Effect of sentence length and exposure on the intelligibility of synthesized speech. *Augmentative and Alternative Communication*, *10*, 96–104.

Yorkston, K., & Beukelman, D. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders*, *13*, 15–31.

Yorkston, K. M., & Beukelman, D. R. (1983). The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In W. Berry (Ed.), *Clinical Dysarthria* (pp. 155–164). Austin, TX: Pro-Ed.

Yorkston, K., Beukelman, D., & Tice, R. (1996). *Sentence Intelligibility Test for Macintosh*. Lincoln: Communication Disorders Software. Distributed by Tice Technology Services, Lincoln, NE.

Yorkston, K., Strand, E., & Kennedy, M. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech Language Pathology*, *5*, 55–66.