APPLICATION OF SPEECH RECOGNITION/CLARIFICATION TECHNOLOGY TO DYSARTHRIC SPEECH

D.R. Sainani and D G. Jamieson
The Institute for Research on Assistive Devices
The University of Western Ontario
London, ON N6G 1H1

The aim of this study is to develop and evaluate a communications system which uses speech recognition as a method of speech input and clarification to produce intelligible synthesized speech as the output.

Specifically, this study will address the differences between the intelligibility ratings of unclarified (natural) dysarthric speech and speech clarified with a speech recognition and voice output system. To do this, differences in the intelligibility of words versus phrases across both the clarified and unclarified speech conditions will be investigated. Further, differences in the intelligibility ratings by naive versus experienced listeners will be investigated.

BACKGROUND

The ability to produce speech that is understandable to normal listeners is denoted as the intelligibility of speech. Dysarthric speakers produce speech characterized by distorted vowels and imprecise consonants, and so, often have low speech intelligibility, (SI). This inhibits all forms of verbal communication. Even though some dysarthric speakers are difficult or impossible to understand, they are still capable of producing a variety of sounds and word attempts with some level of consistency[1,2]. Coleman and Meyers[2] found dysarthric speakers to have relatively consistent speech, even though it was less intelligible than normal speakers. If dysarthric speakers are consistent, and capable of differentiating sounds, then speech recognition technology coupled with voice synthesizers provides a potential solution to the problem of low speech intelligibility.

Speech recognition systems fall basically into two categories, speaker dependent and speaker independent. These two categories are further subdivided into two classes, namely continuous and discrete. Typically, the most expensive systems are the speaker independent continuous recognition systems which require no training period and exert no restrictions on the user. Less expensive and more widely available systems, are the speaker dependent discrete systems. It is this type of system that is employed in this study.

Studies have demonstrated that speech recognition systems recognize words of impaired (dysarthric) speakers better than human listeners. Stevens and Berstein[3] found that the speech recognition system recognized single words for five deaf speakers more accurately than did human listeners. Human listeners recognized 5% to 74% of the single word utterances while a speech recognition system recognized 75% to 99%. Carlson and Berstein[4] tested 50 disabled speakers comprising of hearing impaired and cerebral palsy individuals (dysarthric speakers). Word recognition was compared between naive human listeners and the speech recognition system and was found to be better for the speech system. This type of technology appears to have great potential for dysarthric speakers.

Intelligibility of synthesized speech has also been studied. Green, Logan and Pisoni[5] evaluated the intelligibility of eight text-to-speech systems. Naive listeners transcribed words produced by the speech synthesizers resulting in an average intelligibility for synthesized speech systems of 85.7% with DECtalk achieving the best performance of 96.7%. Mitchell and Atkins[6] found intelligibility rates of 63% and 66% for the Echo II Plus and EvalPac respectively. The intelligibility of words and sentences of natural and synthesized speech was compared by Miranda and Beukelman[7]. They found the intelligibility of three DECtalk voices to be statistically equivalent to natural speech. In general, sentences were found to be more intelligible than single words. It appears that with specific synthesized speech devices, replacing unintelligible speech with intelligible synthesized speech is a viable alternative.

METHOD

Subjects

Upto ten individuals with dysarthric speech will be recruited for inclusion in the study. The subjects will be over ten years of age and have a differential diagnosis of Dysarthria confirmed by qualified speech language pathologists. For inclusion in this study, the subjects' speech must be rated at greater than 20% intelligibility on the Computerized Assessment of Speech Intelligibility (CAIDS, Yorkston,K., Beukelman,D., Traynor,C., 1984) by naive listeners.

Procedure

The subject's SI will be determined by the transcription method of rating the CAIDS test. The subjects undergo two 1 hour training sessions to use the speech recognition

system and then train the system to recognize their voice. The subjects will then proceed to the testing session where they will imitate a standardized list of phonetically balanced words and phrases with and without the speech recognition/clarification system. The sessions will be videotaped and the naive and experienced listeners will review the videotapes at a later date. Intelligibility ratings of the subjects based on the listeners experiences will be computed by comparing what was transcribed by the listeners to what was said by the subject.

Equipment

For the purposes of this study, the DragonWriter 1000 speech recognition development system is being used as the speech recognition system. Software application programs have been developed in conjunction with the speech recognizer to train the subject on how to use the system as well as enabling the user to train the system to his or her voice. Development is underway on a software application that will present phonetically balanced words and phrases to the subject so that he/she may repeat them for the speech recognition system when prompted by the system. The DragonWriter 1000 speech recognition board takes the vocalizations of the user and turns them into text. These text strings are then output to the voice synthesizer, which in this case is DECtalk. DECtalk is a text to speech device which includes a choice from 9 predefined voice patterns including male, female and children's voices and one user definable voice. For this study, voices will be chosen that reflect the age and character of the individual.

SUMMARY

By training the speech recognition system to interpret the utterances of the dysarthric speaker, and converting these instructions into synthesized speech, it is hoped that the whole process of communication for a dysarthric speaker will improve. Currently, there are problems with the communication rate of contemporary devices. There is a need to find ways to enable users of such a system to communicate in a more timely and effective manner. By converting speech via speech recognition technology, into highly intelligible synthesized speech, it is hoped that some of these issues can be addressed.

REFERENCES

1. Neilson, P., & O'Dwyer, N.J. (1984). Reproducibility and variability of speech muscle activity in athetoid dysarthria of cerebral palsy. Journal of Speech and Hearing Research, 27, 502-517.
2. Coleman, C.L. & Meyers, L.S. (1991). Computer Recognition of the Speech of Adults with Cerebral Palsy and Dysarthria. Augmentative and Alternative Communication. 7, 34-42.
3. Stevens, G., & Berstein, J. (1985). Intelligibility and machine recognition of deaf speech. Proceedings RESNA 8th Annual Conference. Washington DCARESNA, 308-310.
4. Carlson, G.S. & Berstein, J. (1987). Speech recognition of impaired speech. In R.D. Steel & W. Gerry Eds., Proceedings of the 10th Annual Conference on Rehabilitation Technology. Washington DC:RESNA. 165-167.
5. Greene, B.G. Logan, J.S. & Pisoni, DEB. (1986). Perception of synthetic speech produced automatically by rule: Intelligibility of eight text to speech systems. Behavior Research Methods* Instruments & Computers, 18, 100-107.
6. Mitchell P.R. & Atkins, C.P. (1988). A comparison of the single word intelligibility of two voice output communication aids. Augmentative and Alternative Communication, 5, 84-88.
7. Miranda, P. & Beukelman, D.R. (1987). A comparison of speech synthesis intelligibility with listeners from three age groups. Augmentative and Alternative Communication, 3, 120-128.

ACKNOWLEDGEMENTS

This work is supported by a grant from the Ontario Rehabilitation Technology Consortium, Ontario, Canada..