DOES ARTIFICIALLY INCREASED SPEECH RATE HELP?

Dr Jan van Doorn
The University of Sydney

Modifications to the acoustic properties of natural speech signals have for some time been mooted as ways of studying intelligibility of disordered speech. An earlier study of acoustic characteristics of dysarthric speech in cerebral palsy showed that syllable length was increased, while relative syllable duration was preserved when compared with normal speech. Hence it was proposed that a global increase of speech rate may bring segmental and suprasegmental features closer to normal and thereby improve intelligibility. The intelligibility of the speech from three speakers with cerebral palsy was determined for three conditions - unmodified, increased rate, and increased rate with pauses removed. Using 50 sentences that were part of the Assessment of the Intelligibility of Dysarthric Speech test, intelligibility was judged by 40 listeners who transcribed what they heard. Results showed that the overall mean intelligibility measures (% words correct) were not significantly different in any of the conditions. However, results for individual speakers were interesting, showing trends that may have implications for future work in the ongoing search to understand the relationship between acoustic features of dysarthric speech and its intelligibility.

BACKGROUND

Synthetic alterations of natural speech have been recognised as being potentially useful for studying disordered speech (Nickerson & Stevens, 1980)). They can be used to look at perceptual and acoustic factors that affect intelligibility (Maassen & Povel, 1985; Maassen, 1986; Osberger & Levitt, 1979), to identify strategies that are useful clinically for improving intelligibility (Hammen, Yorkston & Minifie, 1994; Yorkston, Hammen, Beukelman &Traynor, 1990), and to carry out real time processing that improves speech intelligibility (eg the "The Speech Enhancer™"). The purpose of this study was to use synthetic modifications to look at the impact of changing articulation rate on the intelligibility of dysarthric speech.

The characteristics of dysarthric speech have been studied extensively perceptually (Darley, Aronson & Brown, 1969) acoustically (Caruso & Burton, 1987; Weismer & Martin, 1992; Zeigler & von Cramon, 1986) and physiologically (Kent, Netsell & Bauer, 1975; Kent & Netsell, 1978). Slow speech rate was identified in Darley et al.'s (1969) perceptual study as a consistent feature of most dysarthric speech. This was reflected in acoustic studies that identified longer word duration and more pausing, (Linebaugh & Wolfe, 1984) and in Kent et al.'s (1975) cineradiographic study as slower articulator movements.

Speech rate is characterised by two components, articulation rate and number and length of pauses. Articulation rate is defined as word or syllable duration, or syllables per second in those periods of speech where there are no pauses. In normal speakers, speech rate is slowed mainly by increases in pause length and frequency (Goldman-Eisler, 1961). There is only a small component of change related to articulation rate. However, for dysarthric speech, acoustic and physiological evidence points to slower articulation rate being a major contributor to slow speech rate.

By using acoustic modification of the speech signal it is possible to examine the effect of these two components separately and in combination. Sheard & van Doorn (1998) and Gulek & Rochet (1996) have examined the influence of pausing on the intelligibility of severely dysarthric speech and the results were inconclusive. Artificial insertion of pauses at phrase boundaries appeared to make some small improvements but not pauses after each word. Articulation rate can be altered by using algorithms such as SOLA (Roucos & Wilgus, 1985) and STRAIGHT (Kawahara, 1998) that speed up or slow down speech without altering its pitch. A study by Hammen, Yorkston & Minifie (1994) used this procedure to test the effect of slowed articulation rate on intelligibility of dysarthric speech in patients with Parkinson's Disease. A slowed articulation rate was chosen because clinicians frequently instruct patients with dysarthric speech to further slow down their already slow speech. The even slower speech is thought to encourage more precise articulation by the speaker, and to provide extra processing time for the listener. The study found no significant difference between habitual and artificially slowed articulation rates.

It could be argued that it would have been preferable to artificially increase articulation rate to achieve improved intelligibility. This would have the effect of producing an acoustic signal that contains faster formant transition rates that more closely resemble those in normal speech. The importance of rate of formant transitions in perception of normal speech is well established, and it has been demonstrated that the slope of second formant transitions is highly correlated with intelligibility of dysarthric speech (Kent, Kent, Weismer, Sufit, Brooks, & Rosenbek, 1989). Also, on the basis of an earlier study of acoustic characteristics of dysarthric speech in cerebral palsy that found that relative syllable duration was similar to that of normal speech, increasing speech rate artificially would be likely to produce prosodic timing that is more similar to normal speech. Prosodic timing has also been implicated in speech intelligibility (Tajima, Port & Dalby, 1997). Hence the aim of this study was to investigate the effect of artificial increase of speech rate on the intelligibility of speakers with cerebral palsy whose speech was moderately dysarthric.

METHOD

Speakers

The speakers were three males with cerebral palsy, aged 44, 46 and 52 years, whose speech was dysarthric, but was not affected by poor voice quality. Their speech was judged perceptually by a speech language pathologist to be moderately intelligible.

Speech samples

Fifty 5-word sentences (containing 5 to 11 syllables) were selected from the Assessment of Intelligibility of Dysarthria test (Yorkston & Beukelman, 1981). The set of selected sentences contained all speech sounds of Australian English, in proportions that generally reflected the normal balance of these sounds in the language.

Recording procedure

The fifty sentences were recorded onto DAT tapes for each speaker using a Sony TCD-10PRO digital tape recorder. The recordings were made in a quiet room at the subjects' place of work.

Speech manipulations

A set of 450 files (50 sentences x 3 speakers x 3 conditions) was generated for the listening experiment. All sentences for all speakers were recorded at 10kHz and saved as digital files on a Kay CSL4100 speech analysis system. A modified file for each sentence was created by reducing the overall duration to two thirds of its original duration. The percentage reduction was based on an average ratio of utterance duration for normal speakers to cerebral palsied speakers obtained from an earlier study on the acoustic characteristics of cerebral palsied speech (McKilligan, van Doorn & Pitt, 1994). The reduction was achieved using a time scaling function (RATESYN) on the CSL4100. The exact algorithm for this function was not specified by the manufacturer. A second modification was then made to the shortened file by removing any pauses between words that were longer than 150ms so that the effect of artificially reducing the articulation rate could be examined without any interference from inclusion of pauses.

Listening tapes

Four different listening audiocassettes were generated from the sets of files. The sentence combinations on each tape were based on a Latin squares design. Forty-five of the fifty sentences were used to create the tapes. Each sentence was represented once on each tape. There were fifteen sentences in each of the three conditions (unmodified, shortened, and shortened with pauses removed). Each speaker was equally represented on each tape. The order of the sentences was randomly generated, with the proviso that consecutive sentences must not have been spoken by the same speaker.

Procedure

Four different groups with 10 listeners in each group were selected to listen to the tapes. The listeners were all undergraduate students in speech pathology who were not experienced in listening to dysarthric speech. Each person listened to the tape through headphones in a quiet room with the playback volume set at a comfortable level. The person was asked to listen to each sentence and write down what they heard. Each sentence was presented once only before the listener was asked to transcribe it. There was a sufficient pause on the tape for the listeners to finish their transcription before the next sentence was presented.

Measurement of intelligibility
From the orthographic transcriptions, a score of number of words correct was calculated for each sentence in each condition across all listeners, and then expressed as a percentage of the total number of words. The means of the intelligibility scores across the three conditions were compared using analysis of variance.

RESULTS

Table 1 shows the overall means for the intelligibility in each condition. Analysis of variance indicated that the intelligibility scores for the time manipulated conditions were not significantly different from those of the original speech nor were they different from each other.

Table 1. Mean intelligibility scores and standard deviations (% words correct) for three speech rate conditions.

Unmodified	Rate Increased	Rate Increased & pauses removed
66.0(22.3)	62.6 (25.7)	61.8 (20.5)

Effect of speaker
The data were re-examined for the three speakers separately. The results of that analysis are shown in Table 2. This table indicates that the result was speaker dependent with intelligibility of speech from speakers 1 and 3 showing improvements when rate was increased, but that for speaker 2 deteriorated. None of the differences reached statistical significance under analysis of variance.

Table 2. Mean intelligibility scores and standard deviations (% words correct) in three speech rate conditions for individual speakers

	Unmodified	Rate increased	Rate increased & pauses removed
Speaker 1	61.1 (24.1)	66.9 (22.0)	69.6 (18.4)
Speaker 2	71.2 (27.8)	54.3 (27.1)	58.4 (20.9)
Speaker 3	61.0 (23.4)	66.7 (26.2)	61.7 (22.9)

DISCUSSION

The results of this study infer that achieving an artificial increased articulation rate by globally decreasing duration did not make any significant difference to speech intelligibility. This was true for both "increased rate" conditions, with and without the removal of any long pauses that had been present in the original speech. Although the results for individual speakers did not reach statistical significance, the trends are of interest. The speech intelligibility for two of the speakers improved when the rate was increased, while it deteriorated for the third speaker. The intelligibility score of the unmodified speech of the third speaker (71%) was considerably higher than that of the other two speakers (61%). This difference in intelligibility of unmodified speech may have been a contributing factor to the interspeaker differences, and should be taken into account in the design of future experimental investigations.

While the present study found no significant change in intelligibility when speech rate was artificially increased, Hammen et al (1994) found no significant change when speech rate was artificially slowed.

The combination of the results from the studies implies that "articulation rate" may not be the major rate component that influences intelligibility of dysarthric speech. The "pausing" component may have greater significance. In the present study pausing was only manipulated after articulation rate had been modified. It was not treated as an experimental condition independent of articulation rate. Interestingly though, for those two speakers where rate increase improved intelligibility, removing long pauses further improved the intelligibility for one while for the other it produced intelligibility scores similar to those for unmodified speech. It will be necessary to see if there were significant differences in the number of pauses that were present in the unmodified speech for these two speakers before an explanation of the contradictory results can be offered. Other studies on the impact of the "pausing" component of speech rate on intelligibility of dysarthric speech have also been inconclusive (Gulek & Rochet, 1996; Sheard & van Doorn, 1998). Maybe, then, speech rate alone is not a major contributor to intelligibility of dysarthric speech. There are other physical features, both segmental and suprasegmental, that have been described as contributing to speech intelligibility. In particular, several fundamental frequency (F_o) features have been related to intelligibility eg reduced range of Fo variability (Bunton, Kent, Kent, & Rosenbeck, in press), and flattened fundamental frequency contours (Laures & Weismer, 1999; Wingfield, Lombardi & Sokol, 1984)

The need for a model that relates intelligibility of disordered speech to its acoustic characteristics, and ultimately to its production has been articulated by several authors (eg Nickerson & Stevens, 1980; Kent, Weismer, Kent & Rosenbek, 1989; Weismer & Martin, 1992). Studies such as the one reported in this paper are slowly building up the picture for dysarthric speech. Although the results from this study did not confirm the preliminary findings from McKilligan, van Doorn & Pitt (1994) that artificially increasing speech rate would improve intelligibility of dysarthric speech, the differential findings for the three different speakers in this study may hold a clue to further investigation. To pursue the challenge articulated twenty years ago by Nickerson & Stevens (1980) to search for "models that would relate intelligibility to the physical properties of speech" it is proposed that the individual speaker results from the present study be re-examined. It is hoped that an acoustic analysis of the segmental and suprasegmental features of the original and modified speech for each speaker will reveal differences that can explain the disparate intelligibility results from speaker to speaker.

ACKNOWLEDGEMENTS
The author wishes to acknowledge the financial support of Cable & Wireless Optus for the conduct of this study. The author also wishes to thank the Spastic Centre of New South Wales for their assistance in finding suitable participants. Finally, particular thanks go to those people who so willingly participated in the study.

References

Bunton, K., Kent, R., Kent, J., & Rosenbeck, J. (in press) "Perceptuo-acoustic assessment of prosodic impairment in dysarthria", Clinical Linguistics and Phonetics

Caruso, A.J. & Burton, E.K. (1987) "Temporal acoustic measures of dysarthria associated with amytrophic lateral sclerosis", Journal of Speech and Hearing Research 30, 80-87.

Darley, F., Aronson, A., & Brown, J. (1969) "Differential diagnostic patterns of dysarthria", Journal Of Speech and hearing research 12, 246-269.

Goldman-Eisler F. (1961) "The significance of changes in the rate of articulation", Language and Speech 4, 171-175.

Gutek, J.M. & Rochet, A.P. (1996) "Effect of insertion of interword pauses on the intelligibility of dysarthric speech". In D.A. Robin, K.M. Yorkston & D.R. Beukelman, Disorders of Motor Speech, pp 105-120, (Paul H. Brookes, Baltimore)

Hammen, V.L., Yorkston, K.M. & Minifie, F.D. (1994) "Effects of temporal alterations on speech intelligibility in Parkinsonian Dysarthria", Journal of Speech and Hearing Research 37, 244-253.

Kawahara, H., de Cheveigne, A., & Patterson R. (1998) "An instantaneous-frequency-based pitch extraction method for high quality speech transformation: Revised TEMPO in the STRAIGHT suite", Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Dec. 1998.

Kent R.D., Kent J.F., Weismer, G., Sufit, R.L., Brooks, B.R. & Rosenbek, J.C. (1989) "Relationships between speech intelligibility and the slope of the second formant transitions in dysarthric subjects", Clinical Linguistics and Phonetics 4, 347-358.

Kent, R.D., Netsell, R., & Bauer, L.L. (1975) "Cineradiographic assessment of articulatory mobility in the dysarthrias", Journal of Speech & Hearing Disorders 40, 467-480.

Kent, R.D. & Netsell, R. (1978) "Articulatory abnormalities in athetoid cerebral palsy", Journal of Speech and Hearing Disorders 43, 353-373.

Kent, R.D., Weismer, G., Kent, J.F., & Rosenbek, J. (1989) "Toward intelligibility testing in dysarthria", Journal of Speech & Hearing Disorders 54, 482-489.

Laures, J. & Weismer, G. (1999) "The effects of a flattened fundamental frequency on intelligibility at the sentence level", Journal of Speech, Language and Hearing Research 42, 1148-1156.

Linebaugh, C.W. & Wolfe, V.E. (1984) "Relationships between articulation, rate, intelligibility and naturalness in spastic and ataxic speakers." In M.R. McNeil, J.C. Rosenbek, & A.E. Aronson (eds), The Dysarthrias: Physiology, Acoustics, Perception, Management, pp 197-205, (College-Hill Press, San Diego).

Maassen, B. & Povel, D.J. (1985) "The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech", Journal of the Acoustic Society of America 78, 877-886.

Maassen, B. (1986) "Marking word boundaries to improve the intelligibility of the speech of the deaf", Journal of Speech and Hearing Research 29, 227-230.

McKilligan, C., van Doorn, J. & Pitt, S. (1994) "The intelligibility of speech in cerebral palsy: The effects of manipulating the acoustic speech signal". In R. Togneri (ed) Proceedings of the Fifth Australian International Conference on Speech Science and Technology, Perth. pp 492-497.

Nickerson, R.S. & Stevens, K.N. (1980) "Approaches to the study of the relationship between intelligibility and physical properties of speech". In J.D. Subtelny (ed), Speech Assessment and Speech Improvement for the Hearing Impaired, pp 338-364, (The Alexander Graham Bell Association for the Deaf: Washington).

Osberger, M.J. & Levitt, H. (1979) "The effect of timing errors on the intelligibility of deaf children's speech", Journal of the Acoustic Society of America 66, 1316-1324.

Roucos, S. & Wilgus, A.W. (1985) "High quality time scale modification of speech", Proceedings ICASSP'85, 493-496.

Sheard, C.S. & van Doorn, J. (1998) "Effects of pauses on the intelligibility of dysarthric speech" Proceedings of FONETIK 98, 78-81.

Tajima K., Port R., & Dalby J. (1997) "Effects of temporal correction on intelligibility of foreign-accented English", Journal of Phonetics 25, 1-24.

van Doorn, J. (1983) "Speech waveforms in cerebral palsy - An acoustic analysis", Journal of the Acoustic Society of Australia 11, 17-23.

Weismer, G. & Martin, R.E. (1992) "Acoustic and perceptual approaches to the study of intelligibility". In R.D. Kent (ed), Intelligibility in Speech Disorders, 67-118, (John Benjamins: Amsterdam/Philadelphia).

Wingfield, A., Lombardi, L., & Sokol, S. (1984) "Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation", Journal of Speech and Hearing Research 27. 128-134.

Yorkston, K.M., Hammen, V.L., Beukeiman, D.R., & Traynor, C.D. (1990) "The effect of rate control an the intelligibility and naturalness of dysarthric speech", Journal of Speech and Hearing Disorders 55, 550-560

Ziegler, W. & von Cramon, D.R. (1986) "Spastic dysarthria after acquired brain damage: an acoustic study", British Journal of Disorders of Communication 21, 173-187