The Journal of the Acoustical Society of America

MAY 08 2007

Talker intelligibility differences in cochlear implant listeners

Tim Green; Sotira Katiri; Andrew Faulkner; Stuart Rosen

Check for updates

J. Acoust. Soc. Am. 121, EL223–EL229 (2007)
https://doi.org/10.1121/1.2720938

View Online

Export Citation

ASA
ACOUSTICAL SOCIETY
OF AMERICA

Talker intelligibility differences in cochlear implant listeners

Tim Green, Sotira Katiri, Andrew Faulkner, and Stuart Rosen
*Department of Phonetics and Linguistics, University College London, 4 Stephenson Way,
London NW1 2HE, United Kingdom*
tim@phon.ucl.ac.uk, katiri@gmail.com, andy@phon.ucl.ac.uk, stuart@phon.ucl.ac.uk

Abstract: People vary in the intelligibility of their speech. This study investigated whether across-talker intelligibility differences observed in normally-hearing listeners are also found in cochlear implant (CI) users. Speech perception for male, female, and child pairs of talkers differing in intelligibility was assessed with actual and simulated CI processing and in normal hearing. While overall speech recognition was, as expected, poorer for CI users, differences in intelligibility across talkers were consistent across all listener groups. This suggests that the primary determinants of intelligibility differences are preserved in the CI-processed signal, though no single critical acoustic property could be identified.

PACS numbers: 43.66.Ts, 43.71.Gv [QJF]
Date Received: October 9, 2006 **Date Accepted:** March 8, 2007

1. Introduction

While it is well established that individual talkers can make their speech more intelligible by using a "clear" rather than a "conversational" speaking style (e.g., Ferguson and Kewley-Port, 2002; Krause and Braida, 2002, 2004; Picheny *et al.*, 1985, 1986; Uchanski *et al.*, 1996), comparatively little research has investigated the acoustic-phonetic properties related to differences in intelligibility *across* talkers. Initial studies with relatively few talkers implicated factors such as word and vowel duration, size of vowel space, and fundamental frequency (F0) range (Bond and Moore, 1994; Bradlow *et al.*, 1996). Hazan and Markham (2004) conducted a more extensive study using single word materials from 45 talkers. Two measures, the total energy in the 1–3 kHz region, and word duration, together accounted for about half the variability in intelligibility. Interestingly, profiles of individual high and low intelligibility talkers revealed considerable differences in the patterning of various acoustic-phonetic measures for talkers of similar intelligibility. Thus, it appears that while, at least for normally-hearing listeners, talker intelligibility is very consistent across listeners, high intelligibility can result from various combinations of characteristics.

A further important issue concerns the extent to which intelligibility will be similarly affected across different listening populations. As might be expected, hearing-impaired listeners benefit from talkers using a clear, as opposed to a conversational, speaking style (e.g., Payton *et al.*, 1994; Picheny *et al.*, 1985; Uchanski *et al.*, 1996). Cochlear implant (CI) users might also be expected to benefit from clear speech. However, while modern CI systems typically allow good speech perception, at least in quiet, the auditory information provided by an implant differs markedly from that available in normal hearing. For example, CI processing provides only weak cues to F0 (Green *et al.*, 2002, 2004); allows very limited spectral resolution (Friesen *et al.*, 2001); and typically involves distortion of normal frequency-place mappings (Faulkner *et al.*, 2006; Shannon *et al.*, 1998). These differences raise the possibility that factors that have been suggested to contribute to intelligibility differences for normally-hearing listeners, such as F0 range and the size of the vowel space, may not operate in the same way for CI listeners.

Despite this Liu *et al.*, (2004) found that the advantage for clear over conversational speech produced by a single female talker was similar for normally-hearing listeners, CI users, and normally-hearing listeners presented with acoustic simulations of implant processing.

Since implant processing eliminates much spectral detail and temporal fine structure, this suggests that the primary cues contributing to the clear speech advantage were carried by variations in duration, temporal envelope, or relatively gross spectral differences. However, the properties that distinguish between clear and conversational speech may vary between different talkers and may not map straightforwardly onto the properties that determine across-talker differences in intelligibility. The present study focuses on such across-talker differences, examining whether differences in intelligibility observed in normally-hearing listeners are maintained in cochlear implant listeners and acoustic simulations of implant processing.

2. Methods

2.1 Stimuli

Stimuli were taken from the UCL Talker database (Markham and Hazan, 2002). Two male adults, two female adults, and two female schoolchildren were selected. One in each pair had high intelligibility and one low, based on mean single word error rates calculated by Markham and Hazan (2002).

Recordings of 108 individual words were assigned to six lists of 18 words each, based on mean error rates across the six talkers. To confirm equivalence of intelligibility across lists, error rates were submitted to a two-way ANOVA with talker and list as factors. As expected, this analysis showed a significant effect of talker [$F(5,612)=16.67, p<0.001$], but importantly neither the main effect of list nor the interaction were significant [$F\mathrm{s}<1$]. In order to allow an adequate speech sample for perceptual attunement all single words were concatenated to the carrier phrase “And now please say” recorded from the appropriate talker.

Intelligibility in connected speech was evaluated using 20 semantically unpredictable sentences (SUS) (Benoit *et al.*, 1996). These sentences, each containing four key words, provide no semantic contextual cues so that each word of the sentence is unpredictable, e.g., “The front press scores the saint.” Sentence material was available only for the two male talkers.

2.2 Speech processing

Noise-excited vocoding (Shannon *et al.*, 1995) was implemented in Matlab and comprised the following steps: analysis bandpass filtering (sixth-order Butterworth IIR, three orders per upper and lower side) to divide the spectrum into four or eight bands; half-wave rectification and low-pass filtering (fourth-order Butterworth, 400 Hz) to extract the amplitude envelope for each band; modulation of a noise carrier by each envelope; output filtering matching the initial analysis filtering; adjustment of rms level at filter outputs to match the original analysis outputs; summation across channels. Analysis filters covered the range 100 Hz–5 kHz with spacing based on equal basilar membrane distance (Greenwood, 1990). Frequency responses crossed 3 dB down from the pass-band peak.

In an attempt to avoid ceiling effects, unprocessed stimuli for normally hearing listeners were presented in twenty-talker babble at a signal-to-noise ratio (SNR) of +6 dB, as in Hazan and Markham (2004). For each utterance, a section of noise of equivalent duration was selected at random from the 15 s available. Calculations of signal and noise power were performed over the entire length of the speech utterance and the noise. After summation, all stimuli were normalized to the same rms level. No noise was added to vocoded stimuli, or those presented to CI listeners.

2.3 Participants

Six users of Clarion cochlear implants took part. Three had C2 implants and used the Hi-Res processing strategy. The remaining three had C1 implants and included one user each of the continuous interleaved sampling (CIS), paired pulsatile sampler (PPS), and simultaneous analog stimulation (SAS) processing strategies. Their ages ranged from 32–77 (mean 61) and all had at least four years experience of implant use. Eighteen female adults with normal hearing also participated. Their ages ranged from 21–46 (mean 25). None had any history of hearing deficit.

2.4 Design and procedure

Testing was carried out under computer control. Cochlear implant users were tested using their normal speech processor settings in a sound-proofed room. Unprocessed words and sentences were presented via loudspeaker (QUAD PRO-63) at an individually-determined comfortable level.

Normally-hearing listeners were randomly assigned to one of three groups tested with different types of stimuli: vocoded speech with either four or eight channels, or speech-in-babble. Stimuli were presented via Sennheiser HD 540 headphones in a quiet room at a comfortable listening level fixed for all listeners.

In single word tests, each of the six word lists for each talker was presented to a different participant. Each participant heard one list from each talker, a total of 108 words, presented in random order. Participants heard six practice stimuli, one for each talker. Practice stimuli were processed in the same way as those about to be presented but consisted of words not contained in the main test. In sentence tests each participant heard ten sentences spoken by each of the two male talkers. With this constraint, the choice of talker for each sentence and the order of presentation were random. Because no other SUS sentences were available from these two talkers, the six practice sentences were similar sentences spoken by a female talker.

3. Results

3.1 Single words

Due to the binomial nature of the outcome measure (proportion correct), a logistic regression was used to determine the dependence of word identification performance upon talker type (male, female, or child), intelligibility (high or low), and processing condition (CI, four channel vocoding, eight channel vocoding or babble). Logistic regression also has the advantage of minimizing floor and ceiling effects. Model fitting proceeded from a fully saturated model ($3 \times 2 \times 4$) with methods appropriate for overdispersion applied (Collett, 2003, pp. 206–210). Terms that were not significant at the $p < 0.05$ level were excised sequentially using changes in deviance. There were no significant interactions, but all three main effects were significant ($p < 0.05$).

The significant effect of talker type reflected poorer performance with the child talkers. Averaged across the different processing conditions mean performance with the male talkers was 62.7% and 47.5% for the high intelligibility and low intelligibility talkers, respectively, while the corresponding figures were 62.7% and 47.7% for the female talkers and 56.3% and 42.1% for the child talkers. Although the interaction between talker type and processing condition was not significant, the tendency for poorer performance with the child talkers was more pronounced in the two vocoded conditions.

Figure 1(a) plots performance (averaged across talker type) with high intelligibility talkers against that with low intelligibility talkers for each individual listener. Nearly all listeners showed better performance with the high intelligibility talkers (most points lie above and to the left of the diagonal). For normally-hearing listeners, overall performance levels are clearly highest in the babble condition, lowest with four channel vocoding and intermediate with eight channel vocoding. Individual CI users' performance was quite widely spread within the range covered by the two vocoded conditions. The advantage for high over low intelligibility talkers appears broadly similar in all four processing conditions, reflecting the absence of any interaction between the two factors.

In order to assess the influence of the two major determinants of intelligibility differences identified by Hazan and Markham (2004), single word recognition scores were first averaged across listeners for each combination of talker and processing condition and then normalized by processing condition to the overall mean (Fig. 2). Both mean word duration and mean energy in the $1-3$ kHz region were significantly correlated with normalized word recognition (r=0.419, p=0.021 and r=0.592, p=0.001, respectively). Linear regressions showed that in each case the proportion of the variance accounted for was not significantly increased by allowing

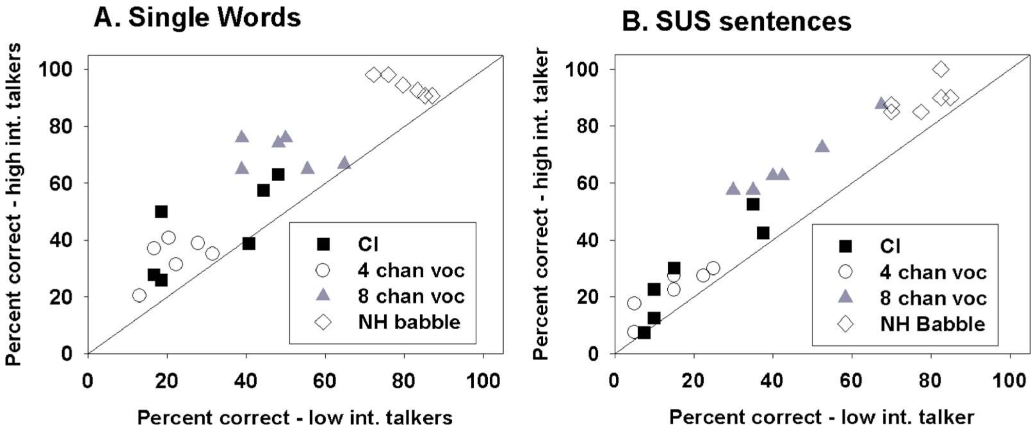

Fig. 1. Speech perception performance with high intelligibility talkers plotted against that with low intelligibility talkers for each individual listener. (A) Single word recognition averaged across talker type; (B) performance on key words in SUS sentences (male talkers only). Diagonal lines represent identical performance for high and low intelligibility talkers.

separate slopes for each processing condition, compared to a single slope. Thus, there was no evidence that the dependence of word recognition on either duration or energy differed across processing conditions.

The 1–3 kHz energy measure accounted for 35.1% of the variance in normalized word recognition. The addition of word duration did not significantly increase the proportion of variance accounted for as the two predictors were strongly correlated for the six talkers used here (r=0.865, p<0.001). Note, though, that these two properties were uncorrelated across Hazan and Markham's (2004) complete set of talkers.

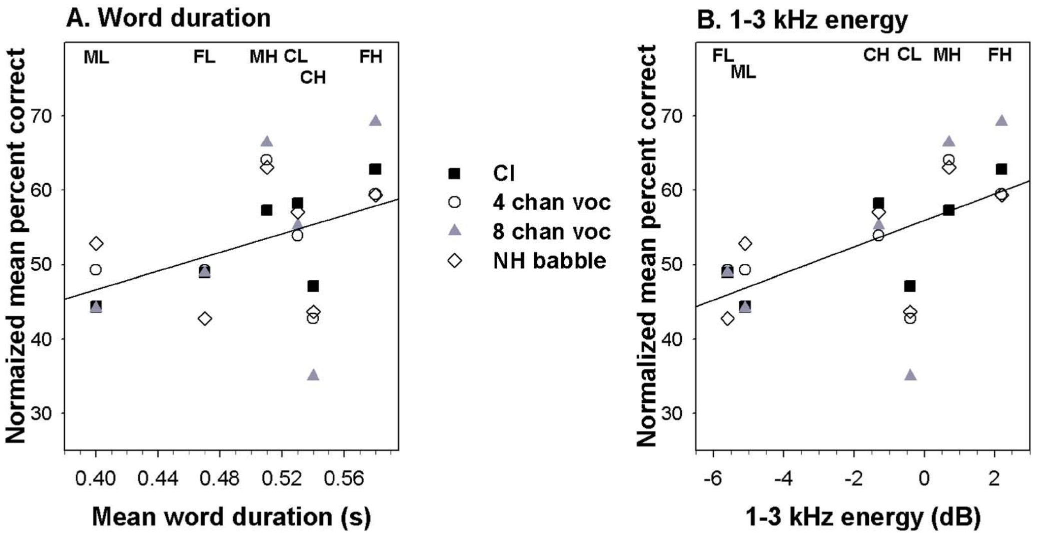

Fig. 2. Normalized mean single word recognition for each processing condition plotted against (A) word duration and (B) amount of energy in the 1–3 kHz range. Talkers are identified by the text symbols at the top of each plot (e.g., MH=high intelligibility male talker). Mean duration and energy values are taken from the data of Hazan and Markham (2004). Best fitting regression lines are shown in each case.

3.2 Semantically unpredictable sentences

A similar logistic regression analysis was applied to scores on correct key words [Fig. 1(b)], to assess the effects of intelligibility and processing condition. The interaction between the two factors was not significant, but both main effects were ($p < 0.001$). Overall performance varied across processing condition in a similar fashion to that seen with single word recognition.

4. Discussion

The key finding was that, while speech recognition performance varied substantially with processing condition, differences in intelligibility across talkers were apparent for all the different listening groups. Consistent with the findings of Liu *et al.* (2004) for a single talker employing either clear or conversational speaking styles, the present results suggest that intelligibility differences across different talkers are largely preserved despite the degradation of the speech signal associated with CI processing.

Although Markham and Hazan (2002) reported that the mean intelligibility scores for the talkers that we selected varied little across the different types of talker (male, female, and child), in the present study there was a significant effect of talker type. While there were no significant interactions with other factors, this effect appears to be primarily attributable to poorer performance with vocoded speech from the two girl talkers. Poorer vowel recognition for girl talkers compared to men, women, and boys has previously been observed in CI users (Loizou *et al.*, 1998) but we found word recognition performance for CI listeners to be unaffected by talker type. Most importantly, in the present context, the differences between the high and low intelligibility talkers within each pair were unaffected by whether the talkers were male adults, female adults, or children.

Unsurprisingly, overall speech perception was highest for the normally-hearing listeners presented with unprocessed stimuli in babble. The better performance in noise-excited vocoder conditions with eight channels than with four can be attributed to the greater degree of spectral resolution in the former case. In the majority of cases the performance of CI users was similar to that in the four channel condition. It should be noted, though, that because there are many aspects of electrical hearing that cannot be emulated in acoustic simulations, this cannot be taken as a measure of the degree of spectral resolution available to the implant users in this study.

The present data set is too limited to allow definitive conclusions regarding the impact of the various processing conditions on possible factors underlying across-talker intelligibility differences. However, the fact that broadly comparable differences between high and low intelligibility talkers were observed in all processing conditions and for all talker types suggests that, for this talker set, the primary factors determining intelligibility differences were largely unaffected by the manipulations involved in simulated and actual implant processing. In addition to properties dependent on a high level of spectral resolution or fine structure temporal information, this would appear to rule out a major role for F0-related factors.

The main factor identified by Hazan and Markham (2004), mean energy in the $1-3$ kHz region, accounted for 35.1% of the variance in normalized single word recognition in the present study. Word recognition was also quite strongly correlated with mean word duration, but inclusion of this factor in the regression did not significantly increase the proportion of the variance accounted for. It might, perhaps, have been expected that the much reduced spectral resolution associated with CI processing would have resulted in an increase in the contribution to intelligibility differences of temporal properties, such as word duration, relative to spectral properties. However, in the present limited data set there was no evidence of any difference in the role of either $1-3$ kHz energy or word duration across the different processing conditions.

One aspect of the speech signal that would be expected to be well preserved by implant processing is low-frequency modulation of the amplitude envelope and it has been suggested that this temporal information plays an important role in determining within-talker intelligibility differences between clear and conversational speech. Using techniques developed in prior speech intelligibility research (Payton and Braida, 1999; Steeneken and Houtgast, 1980), Liu *et* *al.* (2004) calculated envelope modulation spectra from concatenated sentence material in octave bands with center frequencies ranging from 125–4000 Hz. For their female talker, in all octave bands, modulation index values were larger and peaked at lower modulation frequencies (1–3 Hz) for clear compared to conversational speech. For the male talker, a similar pattern was present above 2000 Hz, but there was little difference between clear and conversational speech in the lower octave bands.

Using the methods of Liu *et al.* (2004) envelope modulation spectra were derived from recordings of a read passage (around 2 min) for the adult talkers in the present study (the required material was not available for the children). Separate spectra were obtained for speech in quiet and in the conditions in which speech was presented to listeners in the present study (i.e., in babble or noise-vocoded to four or eight channels).[11] In general, there was little difference in modulation spectra between the high and low intelligibility talkers in most octave bands. Only in the 2 kHz band for the female talkers was there consistent evidence of greater low frequency modulation for the high intelligibility talker. On this evidence, it does not appear that the modulation spectra capture an essential feature responsible for across-talker differences in intelligibility. However, as noted by Hazan and Markham (2004), there is much heterogeneity in the patterning of acoustic-phonetic features for talkers of similar intelligibility. Similarly, it is noteworthy that the child talkers in the present study had very similar measures of both word duration and 1–3 kHz energy, despite the large difference in intelligibility between them. Thus, it is possible that low frequency amplitude modulation is a contributing factor in the intelligibility of some talkers, but not all.

On the basis of the present results it would appear that, while across-talker intelligibility differences are similar in normal hearing and actual and simulated electric hearing, there is no single property that is critical in determining intelligibility differences in implant users. Instead, it seems likely that implant processing may adequately preserve a number of different properties that contribute to intelligibility differences. However, this conclusion needs to be tested further with research employing a larger talker set and incorporating a more detailed investigation of variation in the contribution of possible determinants of talker intelligibility, both across different CI users and between CI users and normally-hearing listeners.

Acknowledgments

The authors thank all participants and are grateful to the cochlear implant team at the Royal National Throat Nose and Ear Hospital for help in recruiting implant users. Thanks also to Paul Iverson and Valerie Hazan for assistance in setting up and running the experiment. Our implant work, and Tim Green in particular, is supported by the RNID. Some of the work reported here was performed as part of the UCL Speech Sciences BSc thesis of Sotira Katiri.

References and links

[1] See EPAPS Document No. E-JASMAN-121-502705 for Fig. 3 which shows envelope modulation spectra calculated in quiet for modulation frequencies in the range 0.5–20 Hz in octave bands from 125–4000 Hz. This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).

Benoit, C., Grice, M., and Hazan, V. (**1996**). "The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences," Speech Commun. **18**, 381–392.

Bond, Z. S., and Moore, T. J. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, 325–337.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech: 1. Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Collett, D. (**2003**). *Modelling binary data*, 2nd ed. (CRC Press, Boca Raton, FL), pp. 206–210.

Faulkner, A., Rosen, S., and Norman, C. (**2006**). "The right information can matter more than frequency-place alignment: Simulations of frequency-aligned and upward shifting cochlear implant processors for an electrode array insertion depth of 17 mm," Ear Hear. **27**, 139–152.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). “Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,” J. Acoust. Soc. Am. **110**, 1150–1163.

Green, T., Faulkner, A., and Rosen, S. (**2002**). “Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants,” J. Acoust. Soc. Am. **112**, 2155–2164.

Green, T., Faulkner, A., and Rosen, S. (**2004**). “Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants,” J. Acoust. Soc. Am. **116**, 2298–2310.

Greenwood, D. D. (**1990**). “A cochlear frequency-position function for several species—29 years later,” J. Acoust. Soc. Am. **87**, 2592–2605.

Hazan, V., and Markham, D. (**2004**). “Acoustic-phonetic correlates of talker intelligibility for adults and children,” J. Acoust. Soc. Am. **116**, 3108–3118.

Krause, J. C., and Braida, L. D. (**2002**). “Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility,” J. Acoust. Soc. Am. **112**, 2165–2172.

Krause, J. C., and Braida, L. D. (**2004**). “Acoustic properties of naturally produced clear speech at normal speaking rates,” J. Acoust. Soc. Am. **115**, 362–378.

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F.-G. (**2004**). “Clear speech perception in acoustic and electric hearing,” J. Acoust. Soc. Am. **116**, 2374–2383.

Loizou, P. C., Dorman, M. F., and Powell, V. (**1998**). “The recognition of vowels produced by men, women, boys, and girls by cochlear implant patients using a six-channel CIS processor,” J. Acoust. Soc. Am. **103**, 1141–1149.

Markham, D., and Hazan, V. (**2002**). “UCL Speaker Database,” Speech, Hearing and Language: UCL work in progress **14**, 1–17 (available at http://www.phon.ucl.ac.uk/home/shl14/pdffiles/markhamH.pdf). Viewed 4/1/07.

Payton, K. L., and Braida, L. D. (**1999**). “A method to determine the speech transmission index from speech waveforms,” J. Acoust. Soc. Am. **106**, 3637–3648.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). “Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,” J. Acoust. Soc. Am. **95**, 1581–1592.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). “Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech,” J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). “Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech,” J. Speech Hear. Res. **29**, 434–446.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). “Speech recognition with primarily temporal cues,” Science **270**, 303–304.

Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1998**). “Speech recognition with altered spectral distribution of envelope cues,” J. Acoust. Soc. Am. **104**, 2467–2476.

Steeneken, H. J. M., and Houtgast, T. (**1980**). “A physical method for measuring speech-transmission quality,” J. Acoust. Soc. Am. **67**, 318–326.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). “Speaking clearly for the hard of hearing. 4. Further studies of the role of speaking rate,” J. Speech Hear. Res. **39**, 494–509.