Clear speech perception in acoustic and electric hearing[a)]

Sheng Liu[b)]
Hearing and Speech Research Laboratory, Department of Biomedical Engineering, University of California, Irvine

Elsa Del Rio
Hearing and Speech Research Laboratory, Department of Biology, University of California, Irvine

Ann R. Bradlow
Phonetics Laboratory, Department of Linguistics, Northwestern University, Evanston, Illinois

Fan-Gang Zeng[b)]
Hearing and Speech Research Laboratory, Departments of Otolaryngology, Biomedical Engineering and Cognitive Sciences, University of California, Irvine

(Received 22 May 2003; revised 17 June 2004; accepted 6 July 2004)

When instructed to speak clearly for people with hearing loss, a talker can effectively enhance the intelligibility of his/her speech by producing ''clear'' speech. We analyzed global acoustic properties of clear and conversational speech from two talkers and measured their speech intelligibility over a wide range of signal-to-noise ratios in acoustic and electric hearing. Consistent with previous studies, we found that clear speech had a slower overall rate, higher temporal amplitude modulations, and also produced higher intelligibility than conversational speech. To delineate the role of temporal amplitude modulations in clear speech, we extracted the temporal envelope from a number of frequency bands and replaced speech fine-structure with noise fine-structure to simulate cochlear implants. Although both simulated and actual cochlear-implant listeners required higher signal-to-noise ratios to achieve normal performance, a 3–4 dB difference in speech reception threshold was preserved between clear and conversational speech for all experimental conditions. These results suggest that while temporal fine structure is important for speech recognition in noise in general, the temporal envelope carries acoustic cues that contribute to the clear speech intelligibility advantage. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1787528]

PACS numbers: 43.71.Es, 43.71.Gv, 43.71.Ky [PFA] Pages: 2374–2383

I. INTRODUCTION

Speech recognition in noise is an extremely challenging task, particularly for hearing-impaired listeners. One way to alleviate this difficulty is to speak ''clearly'' as opposed to ''conversationally'' to these individuals. Previous studies have reported an advantage of 10–20 percentage points in intelligibility for clear speech over conversational speech for a range of speech materials and listening conditions and for a variety of listener populations. Speech stimuli for which a clear speech intelligibility advantage has been reported include isolated syllables (Chen, 1980; Gagne *et al.*, 2002), words (Gagne *et al.*, 1994), nonsense sentences (Picheny *et al.*, 1985; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Helfer, 1997; Krause, 2001) and meaningful sentences (Bradlow and Bent, 2002; Bradlow *et al.*, 2003). Listening conditions include white noise (Chen *et al.*, 1980; Gagne *et al.*, 1995; Uchanski *et al.*, 1996; Bradlow and Bent, 2002; Gagne *et al.*, 2002; Bradlow *et al.*, 2003), speech-spectrum-shaped noise (Krause, 2001; Krause and Braida, 2002), cafeteria noise (Schum, 1996), multitalker babble (Helfer, 1997; 1998; Ferguson and Kewley-Port, 2002), and reverberation (Payton *et al.*, 1994).

Listeners who demonstrated a benefit from clear speech include normal-hearing young adults (Chen *et al.*, 1980; Gagne *et al.*, 1994; 1995; 2002; Uchanski *et al.*, 1996; Helfer, 1997; Krause, 2001; Krause and Braida, 2002; Bradlow and Bent 2002), elderly adults (Payton *et al.*, 1994; Helfer, 1998; Ferguson and Kewley-Port, 2002), hearing-impaired adults (Picheny *et al.*, 1985; Uchanski *et al.*, 1996; Krause, 2001; Ferguson and Kewley-Port, 2002; Krause and Braida, 2002), children with and without learning disabilities (Bradlow *et al.*, 2003), and non-native listeners with normal hearing although these last three listener groups showed smaller benefits from clear speech than other listener populations (Bradlow and Bent, 2002; Bradlow and Pisoni 1999).

The intelligibility difference between clear and conversational speech is related to specific acoustic-phonetic characteristics (Picheny *et al.*, 1986; 1989; Moon and Lindblom, 1994; Bond and Moore, 1994; Bradlow *et al.*, 1996; Uchanski *et al.*, 1996). In comparison to conversational speech, clear speech has a generally slower speaking rate (Picheny *et al.*, 1985; Moon and Lindblom, 1994; Bradlow *et al.*, 1996) and larger temporal envelope fluctuations, i.e., greater temporal amplitude modulations (Payton *et al.* 1994; Krause and Braida, 2002; Krause and Braida, 2004). Several studies

[a)]Portions of this work were presented at the 26th Midwinter Meeting of the Association for Research in Otolaryngology, Daytona Beach, Florida, 2003.

[b)]Corresponding authors: University of California, 364 Med Surge II, Irvine, CA 92697; electronic mail: fzeng@uci.edu; sliu@uci.edu

have systematically analyzed the role of the decreased speaking rate in producing the clear speech intelligibility advantage (Picheny *et al.*, 1989; Uchanski *et al.*, 1996; Krause, 2001; Krause and Braida, 2002). Picheny *et al.* (1989) used digital signal processing to uniformly increase the clear speech rate or decrease the conversational speech rate in the range of 100 to 200 wpm without changing the voice pitch and found that both manipulations significantly degraded speech intelligibility. They attributed the degraded intelligibility to signal processing artifacts because using the same digital signal processing techniques to restore the processed sentence to the original rate did not restore the original intelligibility. Uchanski *et al.* (1996) used a nonuniform time-scaling method to change the phonetic segments within a sentence to reflect the previously measured segmental durational differences between clear and conversational speech. Although this method was generally less harmful to intelligibility than the uniform-scaling method, Uchanski *et al.* found that the normal-rate speech produced by nonuniformly altering segment duration of the original slow-rate clear speech had lower intelligibility than the unprocessed conversational speech. Taking a different approach, Krause and Braida (2002) instructed talkers to produce natural clear and conversational speech at various rates and were able to demonstrate the clear speech advantage even at the fast speaking rate. They concluded that the slow speaking rate in clear speech is not necessary for maintaining the high intelligibility.

Increasing the speaking rate also altered other inherent acoustic properties, one of which was the temporal amplitude modulation index. Payton *et al.* (1994) found that clear speech at a slow speaking rate has greater temporal amplitude modulations than conversational speech. It was unknown whether these greater temporal amplitude modulations were a result of the slower rate in the clear speech. However, Krause and Braida (2004) found greater temporal amplitude modulations in naturally produced clear speech at a normal speaking rate, suggesting that these greater temporal modulations do not have to be associated with a change in the speaking rate and may directly contribute to the clear speech intelligibility advantage.

Other evidence suggests that temporal amplitude modulation may play an important role in speech intelligibility in general. First, studies have shown that speech remains intelligible when the temporal fine structure is removed but the temporal amplitude modulations are preserved in a small number of broad frequency bands. The preserved temporal modulations were used to modulate band-limited white noise or biphasic impulses and delivered to normal-hearing or cochlear-implant subjects. Both groups of the subjects were able to achieve a high level of speech perception at least in quiet (Van Tasell *et al.*, 1987; Wilson *et al.*, 1991; Rosen, 1992; Shannon *et al.*, 1995; Dorman *et al.*, 1997). Second, the temporal amplitude modulation index has been used to predict speech intelligibility in terms of the Speech Transmission Index or STI (Houtgast and Steeneken, 1985; Payton and Braida, 1999). The speech-based STI has been shown to be an accurate predictor of intelligibility and used to predict the clear speech intelligibility advantage (Krause and Braida, 2004), suggesting that physical differences in temporal amplitude modulation exist between clear and conversational speech. Third, several studies have shown that speech intelligibility is reduced when temporal amplitude modulations are decreased. The modulations were digitally decreased by compressing the amplitude of peaks and/or expanding the amplitude of troughs of the extracted temporal envelope, or naturally reduced when speech signals passed through auditory systems of patients with auditory neuropathy, a disorder associated with impaired processing of amplitude modulations. Either the digital or natural means of decreasing temporal amplitude modulations degraded speech intelligibility (e.g., Hou and Pavlovic, 1994; Noordhoek and Drullman, 1997; Zeng *et al.*, 1999b)

To extend previous studies of the perception and acoustic analysis of clear speech, this study used a different group of subjects (cochlear-implant subjects) and different speech processing strategies. The first goal of this study was to measure clear and conversational speech perception as a function of signal-to-noise ratio in order to derive psychometric functions that could provide a complete characterization of the clear speech advantage over a range of signal-to-noise ratios. We used measures such as the speech reception threshold (Dirks *et al.*, 1982) and speech dynamic range (Zeng *et al.*, 2002) to quantify the clear speech advantage. The second goal was to measure the relative contributions of temporal envelope and fine structure to the clear speech advantage. We extracted the temporal envelope from a number of frequency bands and replaced speech fine-structure with noise fine-structure. We conducted four experiments to achieve these goals. Experiment I measured clear and conversational speech perception as a function of signal-to-noise ratio in normal-hearing listeners. Experiment II measured the speech recognition performance in quiet with the speech processed to contain temporal envelope cues in 2, 4, 8, or 16 frequency bands. Experiment III measured the speech recognition performance with an eight-band processor as a function of signal-noise-ratio in normal-hearing listeners. Finally, Experiment IV measured the performance as a function of signal-to-noise ratio in cochlear-implant listeners.

II. METHODS

A. Subjects

Twenty-seven normal-hearing listeners were recruited from the Undergraduate Social Science Subject Pool at the University of California, Irvine. Eleven subjects (divided into two groups) participated in Experiment I, five in Experiment II, and eleven in Experiment III. To evaluate the effect of individual difference in clear speech production across talkers, Experiment I consisted of five subjects tested on sentences produced by a female talker and six subjects tested on sentences produced by a male talker. Experiments II and III only presented speech material produced by the female talker. More subjects were tested in Experiment III in the case of the female talker because of the greater variability in their performance. None of the subjects reported any speech and/or hearing impairment. All were native English speakers and received course credit for their participation.

TABLE I. Biographical and audiological information for cochlear implant subjects.

Subject	Age (yrs)	Onset age	Deaf dur. (yrs)	Etiology	CI use (yrs)	Device	Strategy
S1	67	46	17	unknown	2	CII	CIS
S2	49	4	0	unknown	4	CI	MPS
S3	70	40	2	unknown	4	CI	CIS
S4	68	63	1	sudden	4	N24	ACE
S5	40	18	3	otosclerosis	1	N24	SPEAK
S6	45	35	0	trauma	10	N22	SPEAK
S7	25	23	unknown	unknown	3	Med-El	CIS+
S8	39	9	28	unknown	2	Med-El	CIS

Eight cochlear-implant subjects participated in Experiment IV. Experiment IV only presented speech material produced by the female talker. Table I details the implant subjects' biographical and audiological information. These subjects were 25–70 years old and all were post-lingually deafened with 1–10 years of implant use. Three used the Clarion device, another three used the Nucleus device, and the remaining two used the Med-El device. All used envelope-based strategies including continuous interleaved sampling (CIS), advanced combined encoder (ACE), multiple pulsatile stimulation (MPS), and spectral peak (SPEAK).

B. Stimuli

Stimuli consisted of a total of 144 sentences recorded in clear and conversational styles. These sentences were modified from the original Bamford-Kowl-Bench (BKB) sentences used for British children (Bench and Bamford, 1979). A male and a female adult talker recorded these sentences with a sampling rate of 16 kHz in a sound-treated room in the Phonetics Laboratory of the Department of Linguistics at Northwestern University, Evanston, Illinois (Bradlow *et al.*, 2003). Except for the four original lists consisting of a total of 64 sentences from the female talker, the remaining sentences from the female talker and all the sentences from the male talker were processed in the Hearing and Speech Research Laboratory at University of California, Irvine. The breath noise in the original recordings was removed by a 150-Hz, tenth-order Butterworth high-pass filter (Cool Edit Pro™ 2.0). All sentences were normalized to have the same long-term rms level and then stored in a Microsoft Windows PCM wav file. A total of 144 sentences was processed to result in 18 lists each consisting of 8 sentences. The sentences in each list were either clear or conversational speech.

In Experiment I, the original sentences were individually mixed with a speech-spectrum shaped noise at signal-to-noise ratios from −20 to 20 dB in 5 dB steps. The speech-spectrum shaped noise was produced independently for the female and male talkers by filtering white noise with a tenth-order linear predictive coding (LPC) spectral envelope derived from combined clear and conversational speech sentences from each talker. Sentences from the female and male talkers were presented to two groups of listeners, five listeners in the female condition and six in the male condition. In the remaining experiments, only the female sentences were used. In Experiment II, the original sentences were processed to preserve the temporal envelope cues (Fig. 1, see also Shannon *et al.*, 1995). The stimuli were first divided into several spectral bands (2, 4, 8, or 16) via band-pass filters with their cut-off frequencies calculated from the Greenwood map which purportedly maps each equally distanced cochlear partition into a corresponding physical frequency range (Greenwood, 1990). The output of each band-pass filter was then full-wave rectified and low-pass filtered at 400 Hz to extract the temporal envelope. The envelope was multiplied by a white noise, and then band-pass filtered again by a filter that was identical to the analysis filter in the first stage. The filtered band-limited signals were finally summed to form a synthesized signal that contained the original sentence's temporal envelope but no fine structure cues. In Experiment III, the same original sentences in quiet and noise as in Experiment I were processed via an eight-band processor to test speech perception in noise. In Experiment IV, the same stimuli as in Experiment I were used for cochlear implant users.

C. Speech analysis

Due to a computer memory limitation, only 64 of the 144 sentences were concatenated to produce a long running

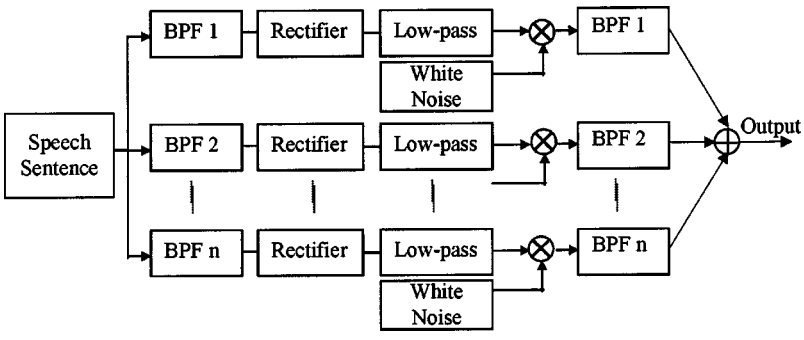

FIG. 1. Digital signal processing used to process original sentences to simulate cochlear implants in experiments II and III.

speech file for each of the four conditions, namely, the female clear and conversational speech, and the male clear and conversational speech. A 1/3-octave modulation spectrum was calculated as a function of modulation frequency for these concatenated sentences (Steeneken and Houtgast, 1980; Payton and Braida, 1999; Krause, 2001; Krause and Braida, 2004). The modulation spectrum was derived by first filtering the concatenated sentences into octave bands centered at 125 to 4000 Hz in octave steps, then squaring the band-limited signals, and finally low-pass filtering at 60 Hz to obtain the slow-varying intensity envelope. The intensity envelope signal was down-sampled to 200 Hz and a 1/3-octave modulation spectrum representation was calculated by summing all spectral components over a 1/3-octave interval with the center frequencies (or modulation frequencies) ranging from 0.4 to 20 Hz. The modulation spectrum representation was normalized by the averaged intensity within the original octave band to obtain an index reflecting the amount of temporal modulation. Therefore, for each different octave band with center frequencies between 125 and 4000 Hz, the modulation indexes were produced as a function of the modulation frequencies ranging from 0.4 to 20 Hz.

D. Procedures

Normal-hearing subjects listened to the stimuli monaurally presented via a Sennheiser HDA 200 headphone in an independent atomic center (IAC) sound-treated booth. The speech presentation level was always at 65 dBA. The noise level was varied to produce different signal-to-noise ratios. The cochlear-implant subjects listened to the stimuli monaurally presented via direct connection. The speech presentation level was adjusted to fit the individual subject's most comfortable level.

To avoid a sentence repetition effect on intelligibility, sentences were used only once for a given subject over the course of the entire experiment. To avoid a presentation order effect, in each testing session clear and conversational speech sentences were mixed together and presented in random order. To counteract a task-learning effect, the experimental conditions were conducted in the order of decreasing signal-to-noise ratios from 20 to −20 dB and the number of bands from 16 to 2. To minimize the effect of differences in inherent difficulty among the sentence lists, each subject was presented with 7 conversational speech lists and 7 clear speech lists that were randomly selected from a total of 18 sentence lists. Finally, to familiarize the subjects with the test materials and procedures, a short session with 5 sentences in quiet was conducted for each experiment.

For formal data collection, the subjects were asked to type the sentence presented via a keyboard and were instructed to double-check the spelling before entering the answer. A computer program automatically calculated the recognition accuracy score based on the number of the key words correctly identified. Each experimental condition had 8 sentences containing three or four keywords each and took about 5 min to finish. The reported result was the averaged score from these 8 sentences. Experiment I had a total of 28 conditions including 2 talkers, 2 speaking styles, and 7 signal-to-noise ratios, Experiment II had 8 conditions (1 talker X 2 speaking styles X 4 bands), Experiment III had 12 conditions (1 talker X 2 speaking styles X 6 signal-to-noise ratios), and Experiment IV had 14 conditions (1 talker X 2 speaking styles X 7 signal-to-noise ratios).

E. Data analysis

The percent correct scores (PC) as a function of signal-to-noise ratio (SNR) from experiments I, III, and IV were fitted with a three-parameter sigmoid function (Zeng and Galvin, 1999a)

$$\mathrm{PC} = \frac{S}{1 + e^{-(\mathrm{SNR} - a/b)}}, \tag{1}$$

where S indicates the asymptotic performance, a is the intercept corresponding to the SNR at which performance is 50% of the asymptotic performance, and b is a parameter related to the slope. The actual slope at the 50% of the asymptotic performance can be derived

$$\mathrm{Slope} = \frac{S}{4b}. \tag{2}$$

In addition, the speech reception threshold (SRT) corresponding to the 50% correct score can be derived

$$\mathrm{SRT} = a - b \ln\left(\frac{S}{50} - 1\right). \tag{3}$$

Finally, the dynamic range (DR), defined as the dB difference between the signal-to-noise ratios producing 10 and 90% of the asymptotic performance, can be derived

$$\mathrm{DR} = b\left[\ln\left(\frac{S}{10\%*S} - 1\right) - \ln\left(\frac{S}{90\%*S} - 1\right)\right]. \tag{4}$$

III. RESULTS

A. Speech analysis

Figure 2 shows wave-form examples from the female (left panels) and male (right panels) talkers in both clear (top panels) and conversational (bottom panels) speech styles. First, note that, regardless of the talker, clear speech has longer overall duration and contains longer and more frequent interword pauses than conversational speech. Second, note that this difference in the temporal patterns of clear and conversational speech appears to be smaller for the male talker than for the female talker.

Table II shows the mean and standard deviation of the overall duration for clear and conversational speech from both talkers. On average, for the female talker, clear speech was 2.2 times longer than conversational speech, whereas for the male talker, clear speech was only 1.5 times longer than conversational speech. Similar ratios for the standard deviation were also observed (see also Bradlow *et al.*, 2003). We should point out that the observed differences between the two talkers may reflect individual differences rather than a gender difference per se.

Figure 3 shows the modulation spectra to further quantify the differences between clear and conversational speech. First, note that clear speech generally has a larger modulation

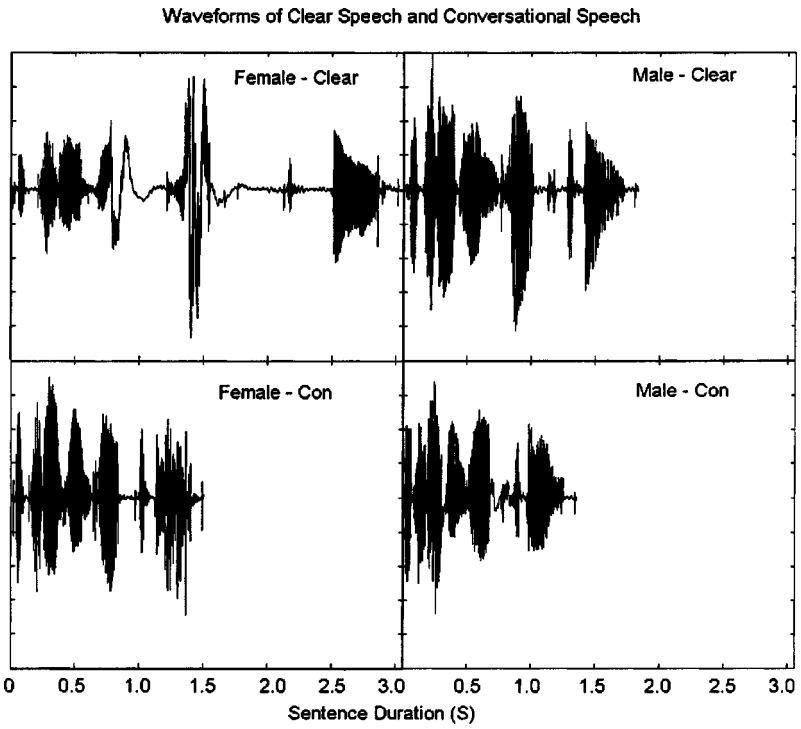

FIG. 2. Sample wave forms of clear (top panels) and conversational (bottom panels) speech by a female (left panels) and male (right panels) talker. The sentence was: *The children dropped the bag*.

index than the conversational speech, particularly for frequency bands with higher center frequency and for the female talker. A simple pairwise *t*-test across 64 sentences shows that the larger modulation index in clear speech is significant in all subbands ($p<0.05$) for the female talker but only in high-frequency bands (center frequencies $>=2000$ Hz) for the male talker. Second, note that regardless of the speech style, the bands with high center frequencies produce larger modulation indexes than the bands with low center frequencies (pairwise *t*-test, $p<0.05$), possibly reflecting the acoustic differences between consonants and vowels. Third, the overall modulation spectra have a lower global peak for clear speech (1–3 Hz) than conversational speech (2–6 Hz). Considering the overall duration difference between clear and conversational speech (Fig. 2 and Table I), these peaks most likely reflect the overall syllable rate. On the other hand, the small but distinct peaks at lower modulation frequencies (e.g., 0.5–0.8 Hz) most likely reflect the overall sentence rate. For example, the averaged sentence duration for the male conversational speech was 1.3 s, corresponding nicely to the 0.8-Hz peak in the modulation spectrum. Unfortunately, the averaged sentence duration for the female clear speech was too long (3.3 s) to be displayed as a peak (0.3 Hz) in the present modulation spectrum since it was beyond the range of the analyzed modulation frequency.

TABLE II. Average duration and standard deviation of 144 clear and 144 conversational speech sentences in both female and male talkers (These data are also reported in Bradlow *et al.* 2003).

Talker	Clear speech (s)	Conversational speech (s)
Female	3.32±0.45	1.47±0.19
Male	1.97±0.27	1.31±0.14

B. Experiment I: Speech perception in normal-hearing subjects

Figure 4 shows percent correct scores as a function of signal-to-noise ratio obtained with both female (left panel) and male (right panel) talkers for clear (open circles) and conversational (closed triangles) speech. The most significant finding is that clear speech produced higher intelligibility than conversational speech for both female and male talkers [$F(1,11)=105.09$, $p<0.05$]. The clear speech advantage can be viewed by examining both the percent correct scores at a given signal-to-noise ratio and the SRT difference. For example, the percent correct score at -5 dB was 86.0 and 60.2% for the female clear and conversational speech, respectively. Similarly, the score was 80.7 and 55.5% for the male clear and conversational speech, respectively. The SRT difference between the clear and conversational speech was 3.1 and 2.2 dB for the female and male talker, respectively. There was no significant difference between the female and male talkers [$F(1,11)=0.24$, $p>0.05$].

As expected, the percent correct score increased as a function of signal-to-noise ratio [$F(6,66)=430.05$, $p<0.05$]. A significant interaction between speech style and signal-to-noise ratio was also observed [$F(6,66)=19.43$, $p<0.05$]. The interaction reflected a significant difference in performance between clear and conversational speech at the intermediate signal-to-noise ratios, but no significant difference at low and high signal-to-noise ratios due to the floor and ceiling effect, respectively.

C. Experiment II: Speech perception with reduced spectral cues in quiet

Figure 5 shows percent correct scores as a function of the number of bands for clear (open circles) and conversa-

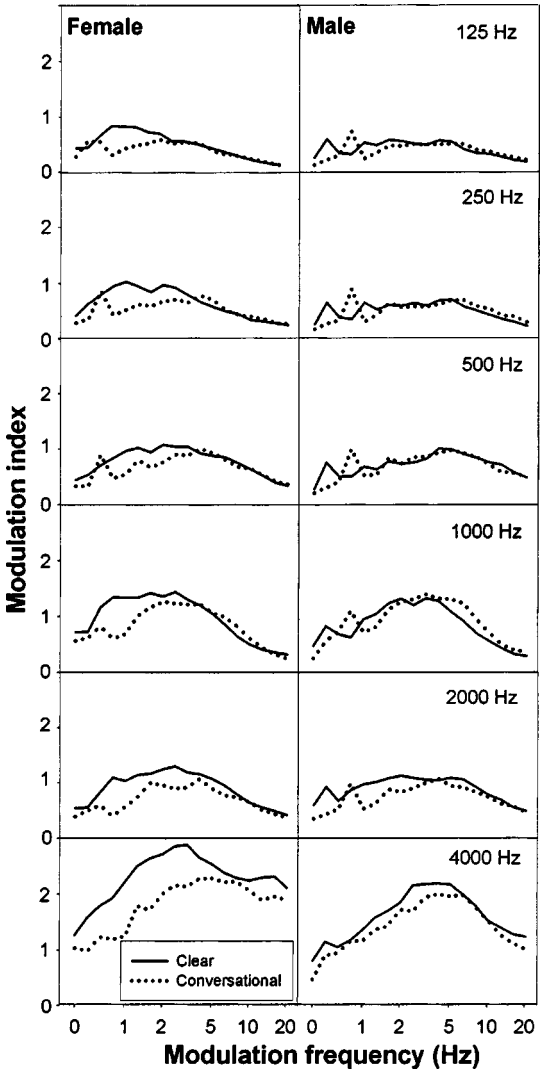

FIG. 3. Modulation spectra showing modulation index (*y*-axis) as a function of modulation frequency (*x*-axis) for the female (left panels) and male (right panels) talkers. The modulation spectra were measured in octave bands from 125 to 4000 Hz with the solid line representing clear speech and the dotted line representing conversational speech.

tional (closed triangles) speech with the female talker in quiet only. Both clear and conversational speech perception increased from essentially 0% with 2 bands to 100% with 8 and 16 bands [$F(3,15)=10.21$, $p<0.05$], but clear speech produced significantly better overall performance than conversational speech [$F(1,5)=539.95$, $p<0.05$]. A post-hoc analysis indicated that this overall difference was due to a 35 percentage point advantage for clear speech over conversational speech in the 4-band condition only [$F(1,5)=81.39$, $p<0.05$].

D. Experiment III: Speech perception with reduced spectral cues in noise

Figure 6 shows percent correct scores for the eight-band clear (open circles) and conversational (closed triangles) speech as a function of signal-to-noise ratio in normal-hearing subjects. Similar to the natural stimuli (Fig. 4 in Experiment I), both the speech style [$F(1,11)=351.82$, $p<0.05$] and the signal-to-noise ratio [$F(6,66)=60.72$, $p<0.05$] were significant factors. The eight-band clear speech had a 32.4 percentage point advantage over the eight-band conversational speech (70.9% vs. 38.5%) at the −5 dB signal-to-noise ratio. The corresponding difference in SRT was 3.2 dB, essentially identical to the 3.1 dB difference in the natural signals. Different from the natural signals, no significant interaction was observed between speech style and signal-to-noise ratio [$F(6,66)=3.96$, $p>0.05$], indicating a more or less parallel shift in the overall performance from the clear speech to the conversational speech. In fact, no floor or ceiling effect was observed as there were still about 10 and 15 percentage point differences between clear and conversational speech at the −10 and 15 dB signal-to-noise ratios, respectively.

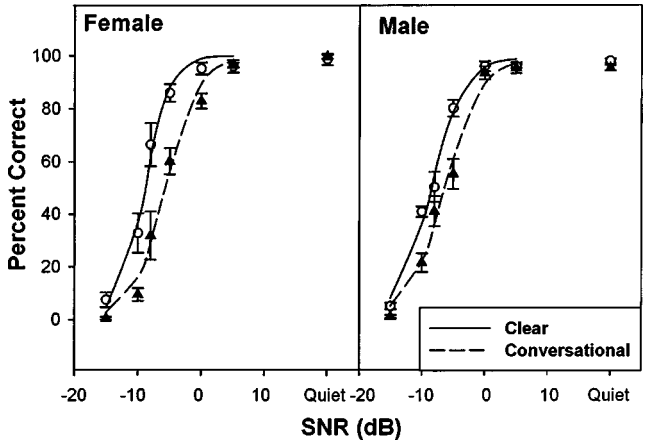

FIG. 4. Results of Experiment I showing percent correct scores as a function of signal-to-noise ratios in 11 normal-hearing subjects for the original clear (open circles) and conversational (closed triangles) speech produced by the female (left panel) and the male (right panel) talker, respectively, five subjects for female and six for male talker experiments. The solid line represents the best fitting of a sigmoid psychometric function for clear speech and the dashed line represents the best fit for conversational speech.

E. Experiment IV: Speech perception in cochlear-implant listeners

Figure 7 shows percent correct scores for clear (open symbols) and conversational (closed symbols) speech as a function of signal-to-noise ratio in eight cochlear-implant subjects. The left panel shows averaged data and standard deviations from five good users whose intelligibility scores were 75% or higher for conversational speech in quiet. The right panel shows averaged data from three relatively poor users whose scores were 60% or lower for conversational speech in quiet. The reason for dividing them into two groups was that we intended to derive globally useful parameters such as speech reception threshold; for example, if a user's score was less than 50%, it would be meaningless and theoretically impossible to derive the 50%-correct speech reception threshold.

Despite large individual variability, Fig. 7 shows an apparent clear speech advantage. Even taking all cochlear implant subjects into account, the pattern of results obtained with cochlear-implant subjects was remarkably similar to that obtained with the eight-band simulation in normal-hearing subjects: both the speech style [$F(1,8)=50.37$,

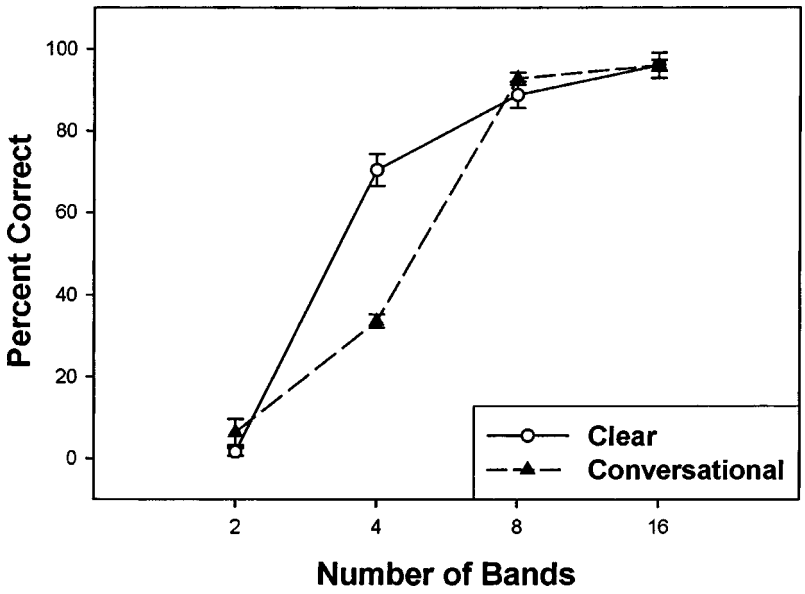

FIG. 5. Results of Experiment II showing percent correct scores as a function of the number of frequency bands (*x*-axis) in five normal-hearing subjects for the processed clear (open circles) and conversational (closed triangles) speech produced by the female talker only. The solid line represents the best fitting of a sigmoid psychometric function for clear speech and the dashed line represents the best fit for conversational speech.

$p<0.05$] and signal-to-noise ratio [$F(6,48)=81.99$, $p<0.05$] were significant factors with no significant interaction between them [$F(6,48)=2.89$, $p>0.05$]. The averaged asymptotic performance from the five good users was 95.4% correct for clear speech and 88.8% for conversational speech. The averaged asymptotic performance from the three poor users was 62.8% correct for clear speech and 49.6% for conversational speech. For the good users, the intelligibility difference between clear and conversational speech was the smallest (five percentage points) at the −10 dB signal-to-noise ratio and the largest (35 percentage points) at −5 dB. In contrast, the smallest intelligibility difference for the poor users was zero due to the floor effect at −10 and −5 dB and the largest was 30 percentage points at 0 dB.

IV. DISCUSSION

A. Macroanalysis of the perceptual data

Table III summarizes three fitting parameters and two derived parameters for the perceptual data from Experiments I, III, and IV [see Eqs. (1)–(4)]. For Experiment IV, only the averaged data from five good users were included for discussion. Except for the male conversational speech condition in Experiment I where the asymptotic performance approached perfect level, clear speech always produced higher asymptotic performance (*S*), lower speech reception thresholds (SRT), and a steeper slope (*b*) than conversational speech. Note, however, that the relative difference in SRT between clear and conversational speech in the natural condition (3.1 dB) was closely preserved in both simulated (3.2 dB) and actual (4.2 dB) cochlear implant conditions. Note also that both simulated and actual cochlear-implant listeners produced similar asymptotic performance, slope, and dynamic range.

Detailed comparisons revealed several additional differences between clear and conversational speech perception. First, a simple multiplication of the slope and the SRT difference would convert the SRT difference into the traditionally measured clear speech advantage in percentage points.

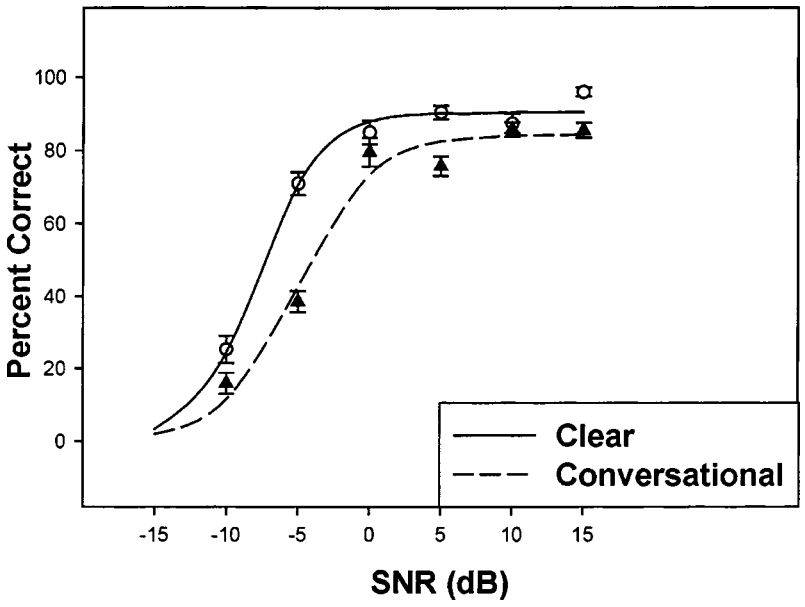

FIG. 6. Results of Experiment III showing percent correct scores as a function of signal-to-noise ratios in 11 normal-hearing subjects for the eight-band processed clear (open circles) and conversational (closed triangles) speech. The solid line represents the best fitting of a sigmoid psychometric function for clear speech and the dashed line represents the best fit for conversational speech.

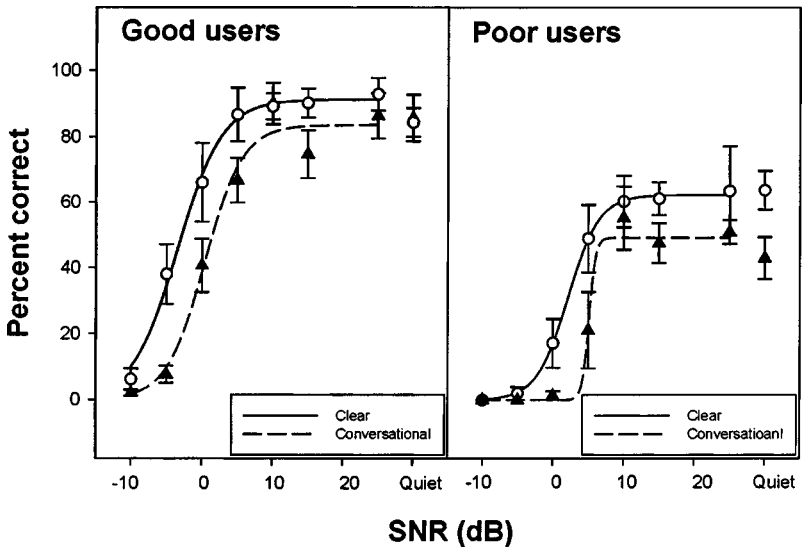

FIG. 7. Results of Experiment IV showing percent correct scores as a function of signal-to-noise ratios in cochlear-implant subjects for the original clear (open circles) and conversational (closed triangles) speech. The left panel represents data from five good cochlear implant users and the right panel represents data from three relatively poor users. The solid line represents the best fitting of a sigmoid psychometric function for clear speech and the dashed line represents the best fit for conversational speech.

The estimated clear speech advantage in percentage point difference was 29.2, 28.8, and 37.8 for the normal hearing, simulated, and actual cochlear implant good users conditions, respectively. This result implies that cochlear-implant listeners may benefit more from clear speech than normal-hearing listeners. Second, compared with the natural speech in normal-hearing listeners, the overall psychometric function was shifted toward a higher signal-to-noise ratio for the simulated (+2 dB) and actual (+6 dB) cochlear-implant experiments. The greater shift in the actual cochlear implant condition was consistent with previous studies using the hearing in noise test (HINT) sentences (e.g., Dorman *et al.*, 1997; Friesen *et al.*, 2001). Finally, note the slightly better performance [$F(1,5)=11.31$, $P<0.05$] with the 25 dB signal-to-noise ratio condition than the quiet condition in the good cochlear implant user group, indicating that a low-level noise might improve speech performance (Zeng *et al.*, 2000; Collins., 1999).

B. Microanalysis of the perceptual data: Individual differences

Here, the clear speech advantage is discussed at a more detailed level by examining who benefited from clear speech and where the benefit occurred. Figure 8 shows the improvement in percentage points for clear speech relative to conversational speech as a function of the percent correct score for conversational speech by normal-hearing (top panel), simulated (middle panel), and actual (bottom panel) cochlear-implant conditions. The closed symbols represent individual data obtained in noise and the open circles represent data obtained in quiet. The minus 45° diagonal line represents the theoretical maximum of the clear speech advantage. For example, if conversational speech perception had already reached a 100% performance level, then the largest improvement that clear speech could reach would be at the same level, resulting in a zero percentage point improvement. This was true for the natural condition in quiet (see the right most open circle on the top panel).

Note first that the overall trend of the improvement in all conditions had an inverted ''U'' shaped curve, indicating that clear speech provided the maximal benefit in terms of percentage points when conversational speech scored moderately. When conversational speech scored too high or too low, the benefit that clear speech provided reached a minimal point. In addition, the individual variability increased significantly from the normal-hearing condition to the simulated and actual cochlear-implant conditions. However, both good and poor cochlear implant users clearly derived a significant clear speech advantage. In the quiet condition, the two implant users who scored the lowest with the conversational speech (about 40%, the two leftmost open circles in bottom panel) had an improvement of 19 and 36 percentage points with the clear speech. In noise conditions, many implant users reached or approached the maximal benefits when their conversational speech had scores of 40% correct or above. The greatest benefit of 64 percentage points (the highest filled square) was achieved by a good user who scored merely 10% correct with the conversational speech at −5 dB signal-to-noise ratio. This good user achieved a 96% correct score in conversational speech recognition in quiet (the third open circle from the right).

C. Talker and rate effects

When different talkers are instructed to speak clearly, they may use different strategies to produce clear speech by slowing down the overall speaking rate, by inserting pauses, enhancing consonant intensity, increasing plosive duration, and/or expanding the vowel space. Our acoustic analysis shows that the female and male talker in the present study appeared to use different strategies to produce clear speech. While she had a comparable speaking rate for conversational speech, the female talker had a much slower rate than the male talker in producing clear speech (Table II). Because no statistical difference was observed in intelligibility between the female and male clear speech for the listeners, the present result provides additional evidence for the previously proposed hypothesis that speaking rate is not the most critical acoustic cue responsible for the clear speech advantage (Krause and Braida, 2002). However, Bradlow and Bent (2002) and Bradlow *et al.* (2003) reported that the female

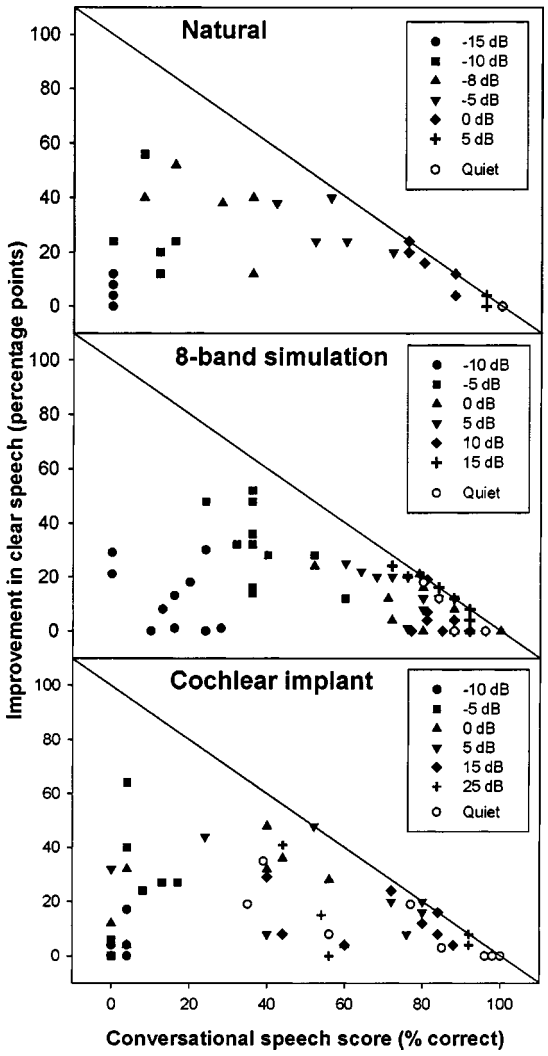

FIG. 8. Individual data showing the clear speech advantage in percentage points (*y*-axis) as a function of the conversational speech score (*x*-axis) in normal-hearing (top panel), simulated (middle panel), and actual (bottom panel) cochlear-implant listeners. The minus 45° line represents the predicted theoretical maximum of the clear speech advantage.

talker's clear speech intelligibility was significantly higher than that of the male talker when the listeners were adult non-native listeners, and children with or without diagnosed learning disabilities. This difference in performance between the present study and the study of Bradlow *et al.* (2003) suggests that different listener populations are sensitive to different clear speech features.

D. Temporal envelope and fine structure

Both the acoustic analyses and the perceptual results in the present study implicate a critical contribution of the temporal envelope to the clear speech advantage. Acoustic analyses of the two speaking styles showed that clear speech had larger temporal modulation indexes than conversational speech for both the female and the male talkers (Fig. 3). This result is consistent with that of Krause and Braida (2002) who showed a similar difference in temporal modulation indexes between clear and conversational speech.

The present study provides direct evidence linking this acoustic difference to the perceptual difference observed between clear and conversational speech. We showed that when the temporal fine structure was removed but the temporal envelope was preserved in the four-band cochlear implant simulation, clear speech produced an intelligibility score that was 35 percentage points higher than the conversational speech in the quiet condition (Fig. 5). More importantly, we showed in both the simulated and actual cochlear implant conditions that the clear speech advantage in noise (3–4 dB SRT difference) was preserved. The consistent clear speech advantage strongly supports the hypothesis that temporal envelope is a major acoustic correlate responsible for the difference in clear and conversational speech perception. On the other hand, the overall shift in the psychometric function in the simulated and actual cochlear implant results indicates that temporal fine structure contributes equally to both clear and conversational speech perception. Together, these data suggest that better encoding of both temporal envelope and fine structure is needed to improve cochlear implant performance in noise.

TABLE III. Comparison of parameters derived from the psychometric function in Experiments I, III, and IV (column 1). The asymptotic performance level "*S*" and the intercept "*a*" were defined by Eq. (1) in the text. The slope, speech-reception-threshold (SRT), and dynamic range (dB) were defined by Eqs. (2), (3), and (4), respectively, in the text.

Experiment	*S* (%)	*a* (dB)	Slope (%/dB)	SRT (dB)	DR (dB)
I-Female-Clear	95.4	−9.1	14.0	−9.0	7.5
I-Female-Conv.	92.9	−6.3	10.6	−5.9	9.6
I-Male-Clear	98.5	−8.6	9.9	−8.5	11.0
I-Male-Conv.	100.0	−6.3	8.6	−6.3	12.8
III-Clear	90.5	−7.8	10.3	−7.3	9.7
III-Conv.	84.3	−5.1	7.8	−4.1	11.9
IV[a]-Clear	95.4	−3.9	9.2	−3.6	11.4
IV[a]-Conv.	88.8	0.0	8.7	0.6	11.2

[a]Because three cochlear implant subjects only achieved an asymptotic performance at 60% correct or below, their data were not used to derive the parameters for experiment IV.

V. CONCLUSIONS

Consistent with previous acoustic studies, the present study shows that clear speech has a slower speaking rate and larger temporal modulation indexes than conversational speech. Also, consistent with previous perceptual studies, the present study finds a significant clear speech advantage in intelligibility, particularly in noise. Through the system atic collection and quantitative analysis of acoustic and perceptual data in both normal-hearing and cochlear-implant listeners, the present study has revealed several findings including:

(1) A quantitative measure of the clear speech intelligibility advantage in terms of the speech reception threshold (SRT) equal to 3.1, 3.2, and 4.2 dB in normal-hearing, simulated, and actual cochlear-implant listeners, respectively (Table III). Taking the slope and dynamic range into account, these SRT differences translate into higher clear speech intelligibility scores of 29.2, 28.8, and 37.8 percentage points, respectively.

(2) A direct relationship between greater temporal modulations and higher intelligibility scores (Figs. 3 and 4). This relationship is validated by the preserved, or even

slightly enhanced, clear speech advantage in both simulated and actual cochlear-implant conditions where primarily temporal envelope cues were available (Figs. 5–7).

(3) A demonstration of the clear speech advantage in both good and poor cochlear-implant users (Figs. 7 and 8). However, a high degree of variability still exists for clear speech perception with some cochlear-implant users achieving the theoretical maximal benefit while others derive a relatively small benefit.

(4) A different role of temporal envelope and fine structure in clear and conversational speech perception. While the temporal fine structure contributes equally to both clear and conversational speech perception, the temporal envelope carries acoustic cues that contribute to the clear speech intelligibility advantage.

ACKNOWLEDGMENTS

The authors thank the Associate Editor, Dr. Peter Assmann and two anonymous reviewers for their helpful comments on earlier drafts of this paper. The authors thank our normal-hearing and cochlear-implant subjects for their time and dedication. This work was supported by a grant from the National Institutes of Health, Department of Health and Human Services (2 RO1-DC02267).

Bench, J., and Bamford, J. (**1979**). *Speech-hearing tests and the spoken language of hearing-impaired children* (Academic Press, London).

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 I: Global and fine-grained acoustic-phonetic talker characteristics,'' Speech Commun. **20**, 255–272.

Bradlow, A. R., and Pisoni, D. B. (**1999**). ''Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors,'' J. Acoust. Soc. Am. **106**, 2074–2085.

Bradlow, A. R., and Bent, T. (**2002**). ''The clear speech effect for non-native listeners,'' J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Erin, H. (**2003**). ''Speaking clearly for learning-impaired children: Sentence perception in noise,'' J. Speech Lang. Hear. Res. **46**, 80–97.

Chen, F. R. (**1980**). ''Acoustic characteristics and intelligibility of clear and conversational speech at segmental level,'' Unpublished master's dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Collins, L. M. (**1999**). ''SR in cochlear implant speech perception,'' Paper presented at 1999 Conference on Implantable Auditory Prostheses, Pacific Grove, CA. (1999).

Dirks, D. D., Morgan, D. E., and Dubno, J. R. (**1982**). ''A procedure for quantifying the effects of noise on speech recognition,'' J. Speech Hear Disord. **47**, 114–123.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channelsof stimulation for signal processors using sin-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

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 channel: comparison of acoustic hearing and cochlear implants,'' *ibid.* **110**, 1150–1163.

Gagne, J. P., Masterson, V., Munhall, K. G., Bilida, N., and Querengesser, C. (**1994**). ''Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech,'' J. Acad. Rehabil. Audiol. **27**, 135–158.

Gagne, J. P., Querengesser, C., Folkeard, P., Munhall, K. G., and Mastern, V. M. (**1995**). ''Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech,'' The Volta Review **97**, 33–51.

Gagne, J. P., Rochette, A. J., and Charest, M. (**2002**). ''Auditory, visual and audiovisual clear speech,'' Speech Commun. **37**, 213–230.

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

Helfer, K. S. (**1997**). ''Auditory and auditory-visual perception of clear and conversational speech,'' J. Speech Lang. Hear. Res. **40**, 432–443.

Helfer, K. S. (**1998**). ''Auditory and auditory-visual recognition of clear and conversational speech by older adults,'' J. Am. Acad. Audiol **9**, 234–242.

Hou, Z., and Pavlovic, C. V. (**1994**). ''Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility,'' J. Acoust. Soc. Am. **96**, 1325–1340.

Houtgast, T., and Steeneken, H. J. M. (**1985**). ''A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria,'' J. Acoust. Soc. Am. **77**, 1069–1077.

Krause, J. C. (**2001**). ''Properties of naturally produced clear speech at normal rates and implications for intelligibility,'' Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

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**, 2165–2173.

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.

Moon, S.-J., and Lindblom, B. (**1994**). ''Interaction between duration, context and speaking-style in English stressed vowels,'' J. Acoust. Soc. Am. **96**, 40–55.

Noordhoek, I. M., and Drullman, R. (**1997**). ''Effect of reducing temporal intensity modulations on sentence intelligility,'' J. Acoust. Soc. Am. **101**, 498–502.

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.

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

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: intelligibility differences between clear and conversational speech,'' J. Speech Hear. Res. **29**, 434–446.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). ''Speaking clearly for the hard of hearing III: intelligibility differences between clear and conversational speech,'' J. Speech Hear. Res. **32**, 600–603.

Rosen, S. (**1992**). ''Temporal information in speech: Acoustic, auditory and linguistics aspects,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Schum, D. J. (**1996**). ''Intelligibility of clear and conversational speech of young and elderly talkers,'' J. Am. Acad. Audiol **7**, 212–218.

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

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 IV: Further studies on speaking rate,'' J. Speech Hear. Res. **39**, 494–509.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, R. D. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Zeng, F. G., and Galvin, J. J. (**1999a**). ''Amplitude mapping and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.

Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**1999b**). ''Temporal and speech processing deficits in Auditory Neuropathy,'' NeuroReport **10**, 3429–3435.

Zeng, F. G., Fu, Q.-J., and Morse, R. P. (**2000**). ''Human hearing enhanced by noise,'' Brain Res. **869**, 251–255.

Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., and Segel, P. (**2002**). ''Speech dynamic range and its effect on cochlear implant performance,'' J. Acoust. Soc. Am. **111**, 377–386.