Cochlear implant speech recognition with speech maskers[a)]

Ginger S. Stickney[b)] and Fan-Gang Zeng
University of California, Irvine, Department of Otolaryngology—Head and Neck Surgery, 364 Medical Surgery II, Irvine, California 92697-1275

Ruth Litovsky
University of Wisconsin, Department of Communicative Disorders, Madison, Wisconsin 53706

Peter Assmann
University of Texas at Dallas, School of Behavioral and Brain Sciences, Box 830688, Richardson, Texas 85083-0688

(Received 28 May 2003; revised 20 April 2004; accepted 16 May 2004)

Speech recognition performance was measured in normal-hearing and cochlear-implant listeners with maskers consisting of either steady-state speech-spectrum-shaped noise or a competing sentence. Target sentences from a male talker were presented in the presence of one of three competing talkers (same male, different male, or female) or speech-spectrum-shaped noise generated from this talker at several target-to-masker ratios. For the normal-hearing listeners, target-masker combinations were processed through a noise-excited vocoder designed to simulate a cochlear implant. With unprocessed stimuli, a normal-hearing control group maintained high levels of intelligibility down to target-to-masker ratios as low as 0 dB and showed a release from masking, producing better performance with single-talker maskers than with steady-state noise. In contrast, no masking release was observed in either implant or normal-hearing subjects listening through an implant simulation. The performance of the simulation and implant groups did not improve when the single-talker masker was a different talker compared to the same talker as the target speech, as was found in the normal-hearing control. These results are interpreted as evidence for a significant role of informational masking and modulation interference in cochlear implant speech recognition with fluctuating maskers. This informational masking may originate from increased target-masker similarity when spectral resolution is reduced. [DOI: 10.1121/1.1772399]

PACS numbers: 43.66.Dc, 43.66.Ts [KWG] Pages: 1081–1091

I. INTRODUCTION

Speech recognition by cochlear-implant users has improved significantly over the past decade as a result of advances in technology, with scores averaging 70%–80% for sentences in quiet. However, the ability of most implant users to understand speech in noisy environments remains quite poor. In general, cochlear-implant listeners require much higher target-to-masker ratios than normal-hearing listeners to achieve similar levels of performance on speech recognition tasks in noise (Kessler *et al.*, 1997; Dorman *et al.*, 1998; Zeng and Galvin, 1999). Poor performance in noise in cochlear-implant listeners is due, at least partially, to the limited number of electrodes that can be safely inserted into the cochlea and the spectral mismatch from the warped frequency-to-electrode allocation (Shannon *et al.*, 2001; Friesen *et al.*, 2001; Garnham *et al.*, 2002). When normal-hearing subjects listened to an eight-channel implant simulation, their speech recognition scores for sentences embedded in steady-state speech-shaped noise dropped from 100% correct in quiet to 55% correct at +2 dB signal-to-noise ratio, and to 16% correct at −2 dB signal-to-noise ratio (Dorman *et al.*, 1998). Compared to eight channels required to achieve maximum speech recognition in quiet, 12 or more channels were required to achieve maximum performance for speech in noise (Dorman *et al.*, 1998; Fu *et al.*, 1998). Potential masking effects and mechanisms contributing to the poor performance by cochlear-implant subjects in noise are described below.

Energetic masking is thought to be a peripheral masking phenomenon that occurs when energy from two or more sounds overlaps both spectrally and temporally, thereby reducing signal detection. Studies on speech intelligibility in noise with hearing impaired and cochlear-implant listeners have typically used energetic maskers. When steady-state speech-spectrum-shaped noise (SSN), one of the most effective energetic maskers of speech, is presented as a masker, the difference in the speech recognition threshold between normal-hearing and hearing-impaired listeners ranges from 2 to 5 dB (e.g., Glasberg and Moore, 1989; Plomp, 1994). Much larger differences of 7–15 dB can be found when the background fluctuates in intensity (Duquesnoy, 1983; Takahashi and Bacon, 1992; Eisenberg *et al.*, 1995). When the masking noise is speech, dips in intensity can occur during brief pauses between words or during the production of low-energy phonemes such as stop consonants. Spectral dips can occur in the valleys between formant peaks, or at low frequencies during the production of fricative sounds. Normal-

[a)]Portions of this work were presented at the 25th ARO Annual Midwinter Research Meeting, St. Petersburg Beach, FL, 2002.
[b)]Electronic mail: stickney@uci.edu

hearing listeners have the ability to take advantage of these brief intensity and/or spectral dips to produce improved speech intelligibility and release from masking. In contrast, hearing-impaired listeners show little or no masking release with fluctuating background sounds (Duquesnoy, 1983; Festen, 1987; Festen and Plomp, 1990; Hygge *et al.*, 1992). In addition to their higher thresholds in quiet and, consequently, poorer signal-to-noise ratios during the dips of the masker, the lack of masking release and the poorer performance of hearing-impaired listeners has been attributed to a loss of spectral and temporal resolution (Festen and Plomp, 1990).

More recent studies, however, have found that fluctuating maskers do not always allow masking release even in normal-hearing listeners. For example, some studies have found greater masking with a single competing talker than with SSN in normal-hearing individuals (Brungart, 2001; Hawley *et al.*, 2004). It is believed that the poorer performance with single-talker maskers is due to a combination of ''energetic masking'' (resulting from overlap of the target and masker in the auditory periphery) and a second type of masking called ''informational masking'' [resulting from competition between the target and masker at more central stages of auditory processing (e.g., Brungart, 2001)].

Traditionally, informational masking had been defined as a higher-level masking phenomenon that arises from masker uncertainty in detection tasks (Pollack, 1975; Watson *et al.*, 1976). In terms of speech perception, the temporal and spectral pattern in a competing voice is much less predictable than in a SSN masker. This variation might make it more difficult for the listener to develop certain knowledge about the competitor, and hence the listener may experience more difficulty segregating the target from competing sounds. Another component of informational masking with single-talker maskers can be attributed to the linguistic nature of the masker. Brungart (2001) observed that when listeners were asked to identify the closed-set number and color categories of a target phrase masked by a simultaneous phrase from the same corpus of materials, they often reported one of the words in the masker phrase, as opposed to a random response. Finally, informational masking is believed to occur when the listener is unable to segregate the target's components from those of the similar sounding masker. Thus temporal and spectral similarities between the masker and target appear to play a role in informational masking (Arbogast *et al.*, 2001; Brungart, 2001; Oh and Lufti, 2000; Kidd *et al.*, 2001).

Informational masking can be reduced by introducing a cue that reduces the similarity between the target and masker. For example, when two voices compete, it is easier to understand one voice if the competing voice has a different pitch, or occupies a different fundamental frequency (F0) range (Assmann and Summerfield, 1990; Bird and Darwin, 1998; Brokx and Nooteboom, 1982). Because cochlear-implant users do not receive a strong sensation of pitch through either the temporal or the place coding mechanism (Zeng, 2002), they would have great difficulty separating voices with similar F0's. In addition, due to the relatively small number of spectral channels in cochlear implants, the formants and their transitions are not well defined. Qin and Oxenham (2003) found that normal-hearing subjects listening to speech processed through a 4-, 8-, or 24-channel cochlear-implant simulation had difficulty segregating a target sentence from a competing voice. Normal-hearing subjects listening to simulated implant processing and cochlear-implant users may therefore experience more informational masking with a single-talker masker than has been found with normal-hearing subjects listening to natural, unprocessed speech.

Greater informational masking can also occur when the temporal modulation properties of competing sounds are similar. Support for this conclusion comes from psychophysical studies investigating modulation interference (Yost *et al.*, 1989). In two more recent studies, weaker masking effects have been observed with unmodulated maskers compared to modulated noise with modulation rates similar to those found in natural speech, but only when the target speech was band-limited (Kwon and Turner, 2001), presented to implant listeners, or presented as an implant simulation (Nelson *et al.*, 2003). Based on these results, it appears that when the noise was modulated at speechlike rates, it became perceptually indistinguishable from the spectrally limited, processed speech.

The present study investigated speech recognition in cochlear-implant and normal-hearing listeners using sentences masked by either SSN or one of three competing voices with varying degrees of temporal and spectral similarity to the target speech. It was hypothesized that cochlear-implant users would experience greater difficulties with competing single-talker speech than SSN because of the greater role of informational masking when spectral resolution is reduced. The masking effects were evaluated in normal-hearing listeners as a function of the number of noise bands in a cochlear-implant simulation (experiment 1) and in cochlear-implant listeners (experiment 2). The main objective was to determine if, and under what circumstances, listeners with reduced spectral information exploit temporal and/or spectral differences to segregate competing speech stimuli and thereby reduce informational masking.

II. EXPERIMENT 1: SPEECH RECOGNITION BY NORMAL-HEARING SUBJECTS WITH SINGLE-TALKER AND SSN MASKERS

A. Methods

1. Listeners

Three groups of 25 young native English speakers (five subjects for each of the five channel conditions) were recruited from the Undergraduate Social Sciences Subject Pool at the University of California, Irvine. All subjects reported normal hearing. Subjects received course credit for their participation.

2. Test materials

Subjects listened to IEEE sentences (Rothauser *et al.*, 1969) in two conditions: unprocessed and vocoder-processed. All sentences in this study consisted of a subset of the 72 phonetically balanced lists of ten sentences (five keywords each) that were recorded by Hawley *et al.* (1999). The target sentences were spoken by a male talker in the presence

of either steady-state, speech-spectrum-shaped noise (SSN), or a different sentence. Three different talkers and the SSN generated from that talker were used as the maskers for each of three groups of 25 listeners. The competing sentence could be spoken by the same male talker as the target sentence (mean F0=108 Hz), a different male talker (mean F0 =136 Hz), or a female talker (mean F0=219 Hz). The F0 values were estimated using a Matlab implementation of the TEMPO algorithm (Kawahara *et al.*, 1999). The same competing sentence (''Port is a strong wine with a smoky taste'') was used throughout testing to avoid confusion of the target and masker sentences when the same male was used as the masker. The SSN maskers were constructed by filtering white noise with the masker sentence's long-term spectral envelope derived via a 20-order autocorrelation LPC analysis. The LPC approach removed the harmonicity and periodicity associated with F0 while producing the same long-term spectrum as the masker sentence. The masker and target had the same onset, but the masker's duration was longer than all target sentences. The level of the masker was set to approximately 65 dB SPL (Brüel & Kjær 2260 Investigator sound level meter; Brüel & Kjær Type 4152 artificial ear) and the level of the target varied around 65 to 85 dB SPL depending on the target-to-masker ratio (TMR).

3. Signal processing

The unprocessed sentences and SSN were scaled to the same root-mean-square value prior to reducing the target sentence attenuation (TDT-II PA4) to allow the following TMR conditions: +20, +15, +10, +5, and 0 dB. The target sentence was then mixed with the masking signal (TDT-II SM3). In the cochlear-implant simulation, the combined target and masking signal was processed by a real-time noise-excited vocoder (DSP sound card: Turtle Beach FIJI; Motorola DSP chip: DSP56311EVM). The mixed signal was preemphasized using a first-order Bessel IIR filter with a cutoff frequency of 1200 Hz and processed into 1, 2, 4 or 8 frequency bands using sixth-order elliptical IIR filters based on the Greenwood map (Greenwood, 1990). The bandpass filter cutoff frequencies for the two-channel simulation were 300, 2009, and 10 000 Hz. For the four-channel simulation, the cutoff frequencies were 300, 840, 2009, 4536, and 10 000 Hz. Cutoff frequencies for the eight-channel processor were 300, 519, 840, 1314, 2009, 3032, 4536, 6748, and 10 000 Hz. The envelope from each band was extracted by half-wave rectification followed by low-pass filtering using second-order Bessel IIR filters at a 500-Hz cutoff frequency. The envelope was then used to modulate a white noise carrier processed by the same bandpass filter used for the original analysis band. The envelope-modulated noise from each band was then combined and delivered through headphones.

4. Procedure

The stimuli were presented monaurally to the right ear through headphones (Sennheiser HDA 200), with subjects seated in an IAC sound booth. Prior to testing, subjects were presented with three practice sessions of ten sentences each. The subjects typed their response using the computer keyboard and were encouraged to guess if unsure. Subjects were instructed to correct typos and avoid misspellings. Their responses were collected and automatically scored by the percentage of the keywords correctly identified. In the first practice session, subjects listened to unprocessed sentences in quiet at an average level of 65 dB SPL. Correct identification of at least 85% of the sentence key words was required to partake in the test session. Approximately 5% of the subjects were disqualified based on this performance criterion. The second and third practice sessions were used to familiarize listeners with the specific masking and channel condition that they were assigned to in the test session. Separate practice sessions were used for single-talker and noise maskers. In both of the latter practice sessions, two sentences were presented for each of the five TMR conditions used in the actual experiment: 0, 5, 10, 15, and 20 dB. No score was calculated for these two practice sets.

In the test session, each group of 25 subjects listened to one of the three competing talkers in one session, and the corresponding SSN in a separate session, with the order of sessions randomized. Each session took approximately 30 min to complete. Within each group of 25 listeners, there were five subjects for each of the randomly assigned speech processing conditions (one, two, four, or eight channels, or the unprocessed speech). Each subject was presented with one talker and speech processing combination, but received all five TMRs. There were ten randomized sentences (five keywords each) for each TMR, for a total of 50 sentences for the single-talker masker and another 50 sentences for the SSN masker. Results were scored in terms of the percentage of keywords correctly identified at each TMR. Results with one and two channels were close to 0% and were not included in the statistical analyses.

Because of the ceiling and floor effects for the natural and four-channel conditions, respectively, the TMR conditions were extended for a second group of ten subjects (five subjects each for the four-channel and natural speech conditions). For the four-channel condition, higher TMRs were added as well as a quiet condition, producing the following five conditions: 15, 20, 25, and 30 dB TMR and ''in quiet''. For natural speech, lower TMRs were added (−10 and −15 dB TMR), producing five conditions: 5, 0, −5, −10, and −15 dB TMR. Only the different male talker and the SSN generated from that talker were used as maskers, otherwise the previous procedures and sentences were retained.

B. Results

1. Speech recognition in noise as a function of the number of channels

Using HINT sentences, higher scores around 20% have been observed with a two-channel simulation (Friesen *et al.*, 2001). The lower performance with the IEEE sentences used here was most likely due to their reduced contextual information (Nittrouer and Boothroyd, 1990; Rabinowitz *et al.*, 1992).

Figure 1 shows speech recognition results as a function of the TMR for three types of maskers: same male talker as the target (top row), different male talker than the target (middle row), or a female talker (bottom row). These were

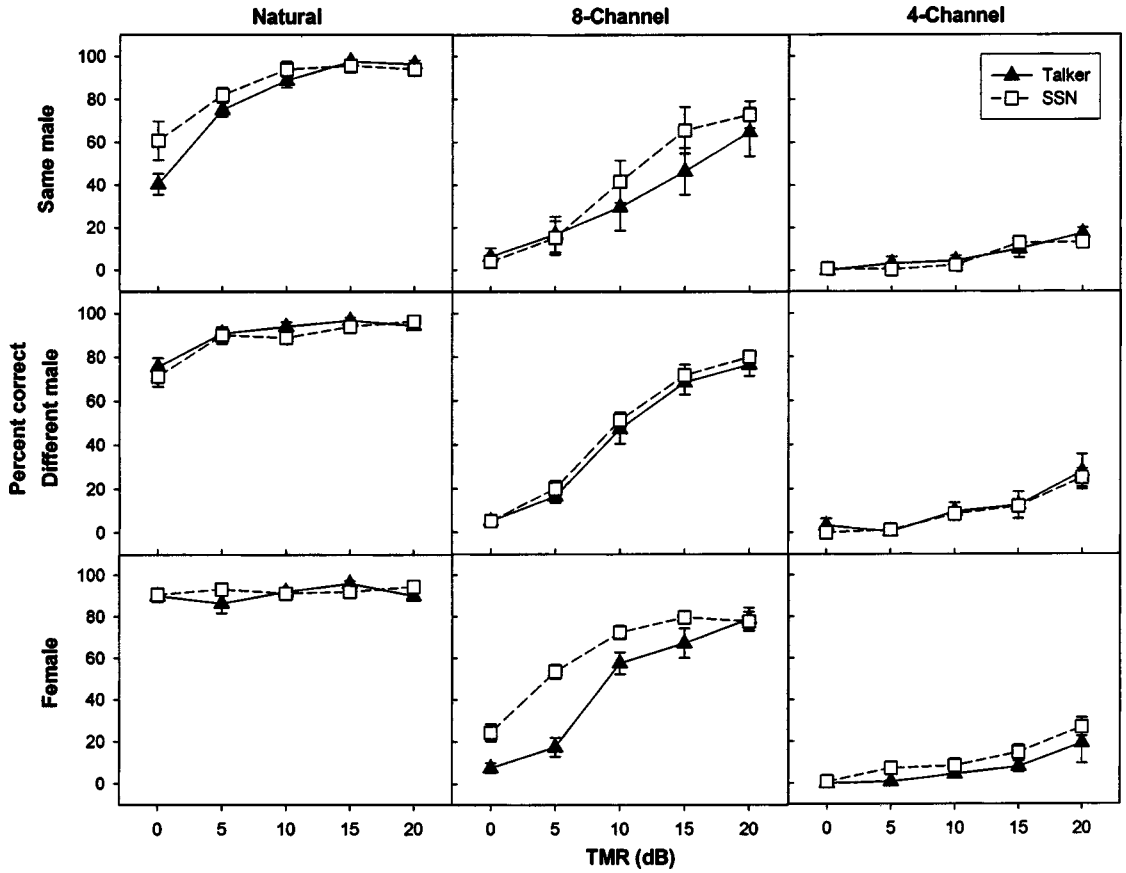

FIG. 1. Speech recognition performance of normal-hearing subjects listening to either a single-talker masker (filled triangles) or steady-state, speech-spectrum-shape noise, SSN (unfilled squares), as a function of the target-to-masker ratio (TMR). Separate panels show performance as a function of the number of channels (columns) and talker (rows) used for the masker. Error bars represent the standard error of the mean.

each presented either as natural speech (left column) or an eight-channel (middle column) or four-channel (right column) cochlear-implant simulation. Within each panel, results are compared for SSN (open squares) and single-talker (filled triangles) maskers. The SSN maskers had a constant intensity over time, whereas the single-talker masker was a sentence which varied in intensity and spectral content over time. Note that, similar to there being three single-talker maskers, there were three SSN maskers that were each generated from one of the three talkers. A mixed design ANOVA was performed with talker and channel as between-subjects variables and the TMR as a within-subjects variable. The general finding was that the cochlear-implant simulation produced significantly poorer performance than the natural condition [$F(2,36)=711.13$, $p<0.001$]. A *posthoc* Scheffé analysis (Scheffé, 1953), collapsed across masker type, showed significant differences among the three processing conditions ($p<0.001$). Higher performance was found with more channels, and natural speech produced the best performance. Performance generally increased as a function of the TMR [$F(4,33)=145.06$, $p<0.0001$], and there was a significant interaction between TMR and processing [$F(8,66)=27.42$, $p<0.001$]. Figure 1 demonstrates that natural speech maintained fairly high levels of intelligibility down to 0 dB TMR. Although performance with natural speech dropped to 40% at lower TMRs, it remained generally high compared with all other conditions. In the 8-channel condition, even the highest TMR of 20 dB resulted in less than 80% correct responses.

Of particular interest in the present findings was the significantly different performance between the SSN and the single-talker masker [$F(1,36)=36.00$, $p<0.001$] and the significant interaction between the masker type and the processing [$F(2,36)=5.21$, $p<0.01$]. A simple effects analysis revealed that single-talker maskers produced lower performance than noise maskers with natural speech [$F(1,12)=6.405$, $p<0.05$] and eight channels [$F(1,12)=8.404$, $p<0.05$], but not with four channels ($p=0.43$). For natural speech this occurred only when the masker was the same talker as the target. This was most apparent when the masker and target were at similar levels, suggesting that the same male masker produced some informational masking. Performance with the other two talkers was at or near ceiling and therefore failed to show an effect of masker. Likewise performance with four channels was so low for most TMR conditions that there was no masker effect. For the eight-channel condition, only the female talker showed a difference across masker types, with the single-talker masker producing lower performance than the noise masker. The combined results contributed to a main effect of talker [$F(2,36)=10.78$, $p<0.001$], a significant interaction of talker×masker [$F(2,36)=4.27$, $p<0.05$], and a significant four-way interaction of talker×masker×processing×TMR [$F(16,101)=2.65$, $p<0.01$].

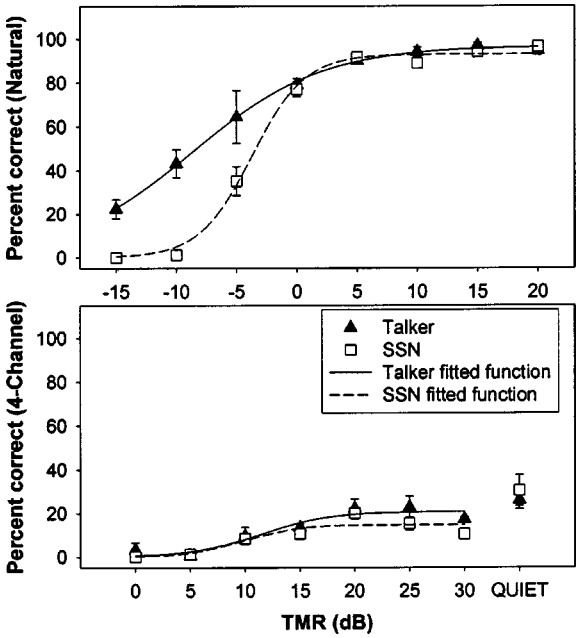

FIG. 2. Speech recognition performance with extended target-to-masker ratios (TMRs) for the natural speech (upper panel) and four-channel conditions (lower panel). Performance is shown for single-talker (filled triangles) and SSN maskers (unfilled squares). The data in this figure was for the "different male" masker condition only, since this was the only talker used for the "extended TMR" group.

The TMR conditions were subsequently extended in the four-channel and natural speech conditions to examine differential amounts of masking from single-talker and noise maskers without confounding floor and ceiling effects, respectively. Because independent sample *t*-tests showed no differences between standard and extended TMR subject groups with similar TMR conditions ($p \geqslant 0.06$), data from the two groups were averaged. Figure 2 shows pooled results with the different male masker for the natural speech (top) and the four-channel (bottom) processing conditions. A three-parameter sigmoid function was fitted with the solid line representing the fit to the single-talker masker data and the dashed line to the SSN data (Zeng and Galvin, 1999). The sigmoid function was well fit to the natural speech ($r^2 = 0.99$) and natural SSN ($r^2 = 0.99$) data, producing the estimated speech reception threshold (i.e., the TMR required to produce a score of 50%) of -9 dB for the single-talker masker and -4 dB for the SSN masker. The fit to the four-channel data was reasonable (single talker: $r^2 = 0.90$; SSN: $r^2 = 0.81$), producing the estimated speech reception threshold of 11 dB for the single talker masker and 10 dB for the SSN masker. For the natural speech data, note that from 0 to -10 dB TMR the difference in intelligibility between the single-talker and SSN masker increased. Although extending the TMRs below 0 dB produced significantly poorer performance for the SSN than for the single-talker masker [$F(1,4) = 21.92$, $p < 0.01$] in the natural speech condition, extending the TMRs above 20 dB in the four-channel condition produced no difference in performance between the two masker types ($p = 0.07$) and no further improvement in speech recognition scores. In the natural speech condition (for TMRs$<$0 dB), the better performance with the single-talker masker than with the noise masker indicated a release from masking, a phenomenon that was never observed in cochlear-implant simulations.

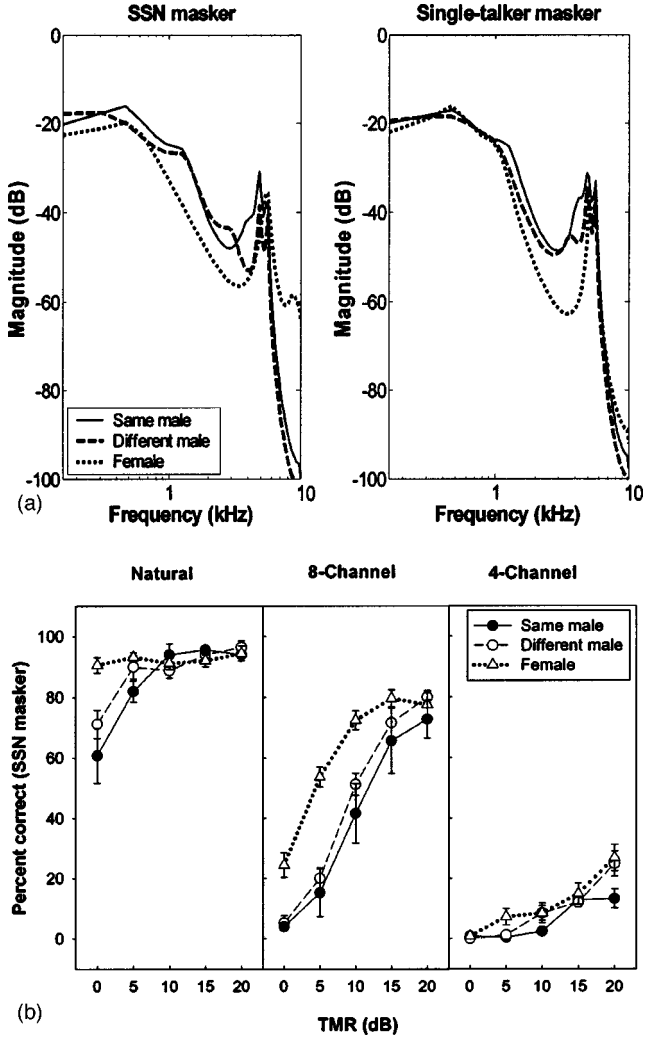

FIG. 3. (a) The long-term average spectral shape of the two male ("same male:" solid line; "different male:" dashed line) and female talkers (dotted line). Similar spectral shapes were found for the steady-state, speech-spectrum-shaped noise masker, SSN (left panel), and the single-talker masker (right panel). (b) Speech recognition performance with each steady-state, speech-spectrum-shaped noise masker (SSN) as a function of the target-to-masker ratio (TMR) and number of channels (columns). Results for the "same male" SSN as the masker are shown as filled circles with solid lines, unfilled circles and dashed lines are used for the "different male" SSN, and unfilled triangles with dotted lines are used for the "female" SSN.

2. Comparison of SSN maskers across talkers

For a closer inspection of the spectral energetic maskers, Fig. 3(a) shows the long-term SSN amplitude spectrum (left panel) and speech spectrum (right panel) for the same male (solid line), different male (dashed line), and female talkers (dotted line). The spectral shapes of the two male voices were fairly similar in comparison to the female voice, which had less energy at intermediate frequencies between 2 and 5 kHz but more energy at higher frequencies. To facilitate comparison, Fig. 3(b) replots the SSN data from Fig. 1 and contrasts speech recognition performance with different talkers in the same panel for both natural (left panel) and simu-

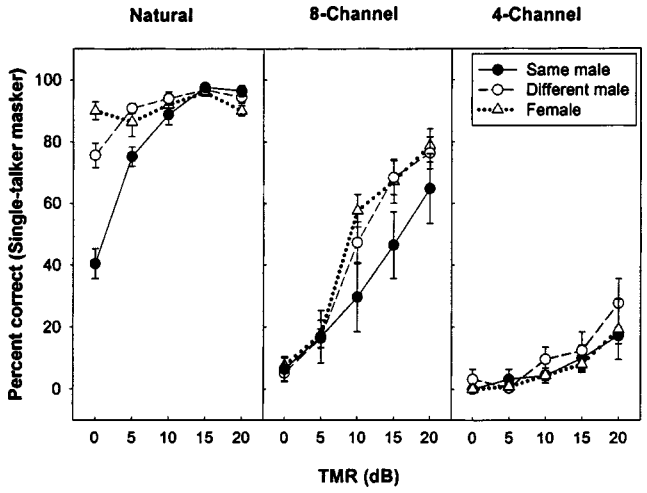

FIG. 4. Speech recognition performance with each competing talker as a function of the target-to-masker ratio (TMR) and number of channels (columns). Results for the ''same male'' talker as the masker are shown as filled circles with solid lines, unfilled circles and dashed lines are used for the ''different male'' talker, and unfilled triangles with dotted lines are used for the ''female'' talker.

lation conditions (two right panels). A simple effects analysis focusing on the three SSN maskers collapsed across channels revealed a significant effect of talker [$F(2,36)=16.58$, $p<0.001$]. A *posthoc* Scheffé analysis showed significant differences between male and female masker (same male: $p<0.001$; different male: $p<0.01$), but no differences between the two male maskers ($p=0.16$). These results are consistent with greater differences in the long-term average spectral shape for the female SSN relative to the male SSN maskers. Different from the female SSN, the two male SSN maskers provide similar amounts of spectral energetic masking.

For each channel condition, a Scheffé analysis along with an examination of the means showed that the natural speech condition did not exhibit any differences across SSN maskers [$F(2,12)=3.36$, $p=0.07$] due to the ceiling effect. For the eight-channel condition, the highest performance was obtained with female SSN (female versus same male: $p<0.01$; female versus different male: $p<0.05$) and no difference was found between the two male SSN maskers ($p=0.56$). For the four-channel condition, the highest scores were again obtained for the female SSN, but significant differences were only found between the most intelligible (i.e., the female SSN) and the least intelligible (i.e., the same male SSN) masker ($p<0.01$).

3. Comparison of single-talker maskers

Figure 4 shows speech recognition with single-talker maskers as a function of channel condition and talker. Simple effects analyses of single-talker maskers, averaged across processing conditions, revealed a significant effect of talker [$F(2,36)=5.17$, $p<0.05$]. A *posthoc* Scheffé analysis revealed that the lowest performance occurred with the same male single-talker masker ($p<0.05$), and no significant differences were found between the different male and female maskers ($p=0.99$).

An analysis of each channel condition showed that the differences across single-talker maskers stems exclusively from the natural speech condition [$F(2,12)=11.54$, $p<0.01$], which showed the lowest performance with the same male single-talker masker (Scheffé: $p<0.01$) and no difference between the other two single-talker maskers ($p=0.98$). In contrast, no significant talker differences were found across any of the single-talker maskers for the eight- and four-channel conditions. This interesting finding is unlike that found with each of the SSN maskers, and an explanation is offered in Sec. IV.

III. EXPERIMENT 2: SPEECH RECOGNITION BY COCHLEAR-IMPLANT SUBJECTS WITH SINGLE-TALKER AND SSN MASKERS

A. Methods

1. Listeners

Five postlinguistically deafened users of the Nucleus cochlear implant participated in this experiment (Table I). All cochlear-implant subjects were native English speakers with 5 to 13 years of experience with their device.

2. Test materials

The cochlear-implant subjects listened to only the unprocessed, natural sentences, which included the same target and masker sentences used for the normal-hearing listeners in experiment 1. As with the normal-hearing listeners, there was no repetition of the test material.

3. Procedure

The stimuli were passed through a Cochlear Corporation Audio Input Selector (AIS) connected to the subjects' speech processor. The AIS attenuates the analog output from the TDT or soundcard before it is delivered to the speech processor. Prior to testing, subjects listened to sentences and were asked to adjust the level of the AIS to a comfortable

TABLE I. Subject demographics.

Subject	Age	Implant	Speech strategy	Duration of hearing loss (years)	Duration of deafness (years)	Duration of implant use (years)
CI1	45	Nucleus-22	SPEAK	<1	<1	10
CI2	51	Nucleus-22	SPEAK	<1	5	12
CI3	60	Nucleus-22	SPEAK	51	13	11
CI4	69	Nucleus-22	SPEAK	43	5	13
CI5	68	Nucleus-24	SPEAK	<1	17	5

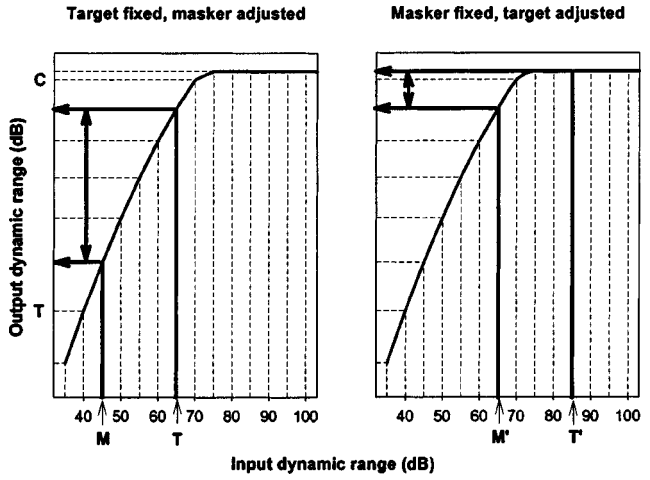

FIG. 5. Schematic of a compression input-output function. The input dynamic range (*x* axis) is shown as a function of the output dynamic range (*y* axis) for the two procedures used in experiment 2. The left panel shows the schematic for the procedure where the target (T) is fixed at a comfortable loudness and the masker (M) is adjusted from below that level to produce each target-to-masker ratio (TMR). The right panel shows the function for when the masker (M′) was fixed at the comfortable loudness level and the target (T′) was raised above that level to produce each TMR (right panel). Note the much greater range of TMRs (arrows along the *y* axis) available to the listener when both the masker and target are kept at or below the upper comfort level (C). Above the C-level the sound would be peak-clipped.

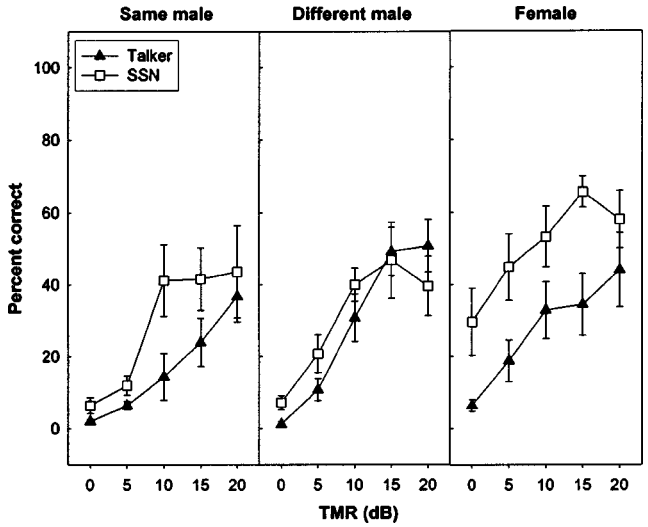

FIG. 6. Speech recognition performance for cochlear-implant subjects listening to a single-talker masker (filled triangles) or steady-state, speech-shaped noise masker, SSN, (unfilled squares). Results are shown as a function of the target-to-masker ratio (TMR) and talker (columns) used for the masker. The cochlear-implant data shown here (experiment 2) is with the ''fixed masker'' procedure (the same procedure used for the normal-hearing listeners) for a better comparison of implant and normal-hearing data. Error bars represent the standard error of the mean.

listening level. If the maximum setting was reached on the AIS, the subject then adjusted the sensitivity of their speech processor to reach a comfortable loudness.

One potential confound when testing subjects who use amplification devices is the possibility of presenting stimuli at intensities where compression occurs (Stone and Moore, 2003). This is particularly problematic for experiments that attempt to deliver a sound within the very narrow dynamic range of electric hearing. Figure 5 demonstrates the effect of compression using a schematic input–output function. In the Nucleus device, used by all five cochlear-implant subjects in this study, the input dynamic range is only 30 dB (User Manual, The Nucleus 22 Channel Cochlear Implant System, p. 4-SP). This means that compression will limit the sound input to a 30-dB range to fit within the cochlear-implant user's dynamic range, defined within the boundaries of the patient's threshold (T-level) and upper comfort level (C-level).

The right panel of Fig. 5 shows that when the masker level is fixed at 65 dB and the target level is adjusted above the level of the masker to produce each TMR, there would only be 5 dB of head room to effectively change the target-to-masker ratios. Thus, compression limits the intended 0–20 dB TMR range to 5 dB or less in cochlear-implant users, possibly producing a plateau or drop in performance (due to peak clipping) as the TMR is increased. On the other hand, when the target level is fixed at 65 dB and the masker level is adjusted from a lower intensity to produce the 0–20 dB TMR range (left panel), there would be about 25 dB of leg room which would minimize the compression effect. Both compressive (masker fixed; target increased above the most comfortable loudness level) and less-compressive techniques (target fixed; masker increased from below the comfortable loudness level) were used and compared in experiment 2, and all masker and talker conditions were performed with both techniques.

After establishing the appropriate level, cochlear-implant subjects were presented with a practice session in quiet. They were asked to type their response into the computer and were encouraged to guess if unsure. Because cochlear implant users typically have a wide range of performance variability, no formal minimum performance requirement was set other than that the subjects have some open-set, speech understanding with auditory cues alone. Scores for the cochlear implant subjects on the practice session ranged from 78% to 92%. For the test session, unlike the normal-hearing subjects, cochlear-implant listeners participated in all single-talker and SSN test sessions over a period of several days, with each session lasting between 1 and 2 h per day. All testing was performed with subjects seated in an IAC sound-booth. Results were scored in percent correct for the following TMRs: 0, 5, 10, 15, and 20 dB TMR.

B. Results

1. Speech recognition performance by cochlear-implant subjects

Figure 6 shows performance as a function of the TMR for single-talker and SSN maskers. The results in the figure are from the ''fixed masker'' data to provide an easier comparison with the normal-hearing listeners who used the same procedure. Average speech recognition performance in quiet was 83%, determined by the IEEE practice sentences. Consistent with previous studies, sentence recognition performance in noise by cochlear-implant subjects was between that obtained with a 4- to 8-channel simulation in normal-hearing listeners (Friesen *et al.*, 2001; Garnham *et al.*, 2002).

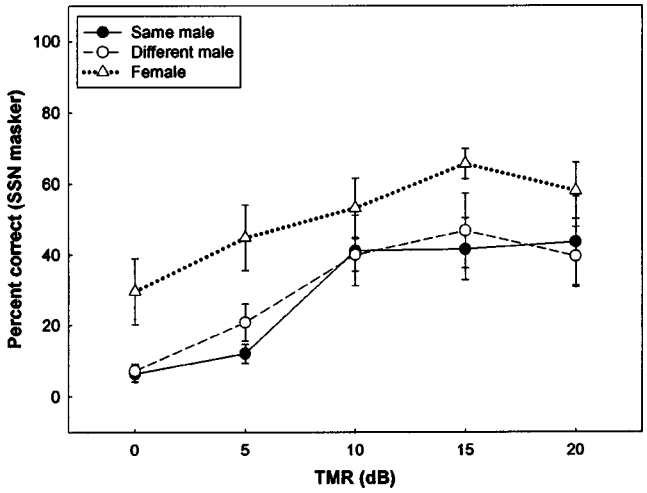

FIG. 7. Cochlear-implant data with the steady-state, speech-shaped noise (SSN) masker as a function of the TMR and talker used as the masker. Results for the ''same male'' SSN as the masker are shown as filled circles with solid lines, unfilled circles and dashed lines are used for the ''different male'' SSN, and unfilled triangles with dotted lines are used for the ''female'' SSN.

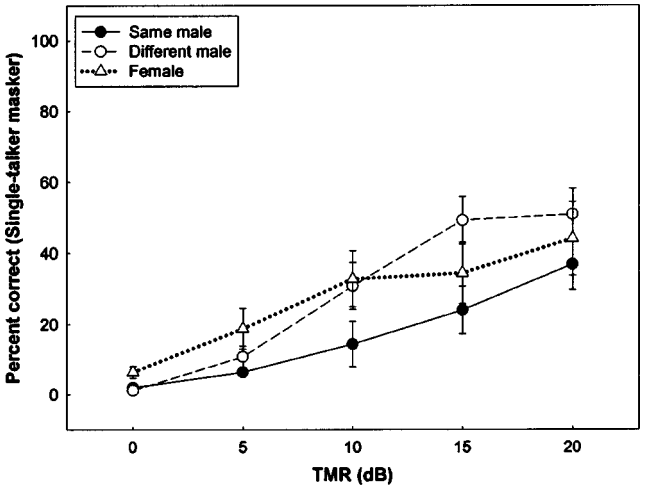

FIG. 8. Cochlear-implant data showing performance with single-talker maskers as a function of the TMR and competing talker. Results for the ''same male'' talker as the masker are shown as filled circles with solid lines, unfilled circles and dashed lines are used for the ''different male'' talker, and unfilled triangles with dotted lines are used for the ''female'' talker.

The average performance by cochlear-implant subjects was slightly better than normal-hearing subjects listening to a four-channel simulation, but only the highest performing cochlear-implant subjects were able to obtain similar levels of speech understanding as the normal-hearing subjects listening to an 8-channel simulation. Although Fig. 6 shows a trend for speech recognition performance in noise to improve with increasing TMRs, a repeated measures ANOVA failed to demonstrate a significant main effect of TMR for the cochlear-implant users ($p=0.19$). However, Bonferonni pairwise comparisons (Dunn, 1961) demonstrated a significant drop in performance from a 10 to 5 dB TMR ($p<0.01$), but no significant decrements in performance from a 5 to 0 dB TMR or improvements above a 10 dB TMR. As found in normal-hearing subjects listening to the eight-channel simulation, there was an effect of masker [$F(1,4)=11.92$, $p<0.05$], with single-talker maskers again producing lower performance (32.9%) than SSN maskers (41.8%). There was also a main effect of talker when collapsed across both masker types [$F(2,3)=10.77$, $p<0.05$]. This is discussed in more detail in the following two sections.

2. Comparison of SSN maskers across talkers

Figure 7 replots the SSN data from Fig. 6 to contrast performance with each of the three SSN maskers as a function of the TMR. Bonferonni pairwise comparisons confirmed significant differences between the female SSN masker and the different male SSN [$F(1,4)=29.92$, $p<0.01$], but no differences between the two male SSN maskers ($p=0.62$). These results are similar to those found with the 8-channel simulation in experiment 1, and closely follow the differences in the long-term average speech spectrum shown in Fig. 3(a).

3. Comparison of single-talker maskers

Figure 8 contrasts speech recognition performance with single-talker maskers as a function of the TMR and talker. No significant differences were found across single-talker maskers (Bonferonni pairwise comparisons: $p=0.26$). Thus, even though the cochlear-implant users were able to use gross spectral differences to segregate the steady-state female noise masker from the temporally fluctuating male speech target, the spectral differences were not sufficient for cochlear-implant users to segregate two fluctuating speech sounds. This result is similar to the normal-hearing subjects listening to the eight-channel simulation.

4. Effects of compression on performance in noise

Figure 9 compares the effect of compression on speech recognition in noise with either the masker fixed (filled triangles) or the target fixed (open triangles) at the most comfortable loudness for single-talker (upper panel) and SSN (lower panel) maskers. The ANOVA showed no main effect of method (i.e., masker fixed or target fixed), but there was a significant interaction between the method used and the TMR [$F(4,1)=1574.27$, $p<0.05$]. Paired *t*-tests confirmed significant differences between the two methods only at the highest TMRs for single-talker maskers (i.e., 15 and 20 dB) and at a 20 dB TMR for SSN maskers ($p<0.01$ for all three results). When the masker was fixed and the target was increased above the most comfortable loudness (Fig. 9: filled triangles), performance either reached a plateau or decreased at high TMRs. This effect was most evident for single-talker maskers whose peak amplitudes could exceed the upper comfort level and become peak-clipped. In contrast, performance generally improved with higher TMRs when the target was fixed and the masker was kept at or below the comfortable loudness level (open triangles). This pattern of results was consistent with the prediction shown in the schematic compression function (Fig. 5).

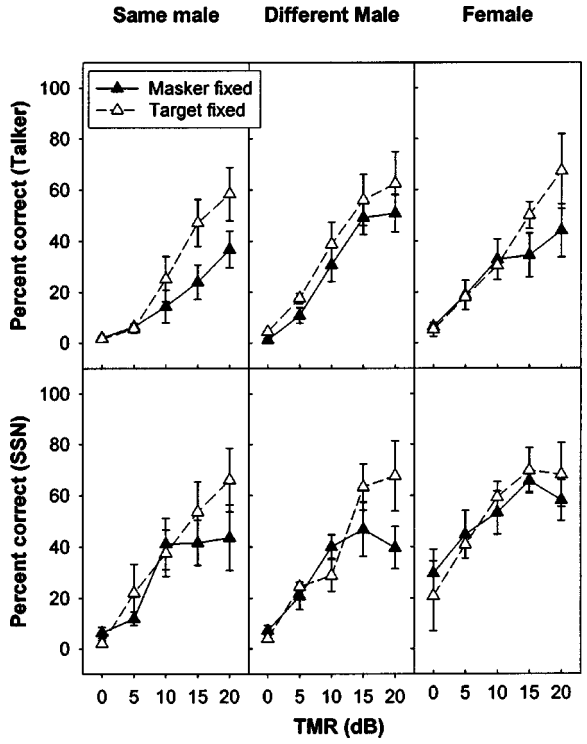

FIG. 9. Comparison of the two procedures (i.e., ''masker fixed'' or ''target fixed'') used to examine the effects of compression in cochlear-implant users. Performance (*y* axis) is shown as a function of the target-to-masker ratio (*x* axis). The data is plotted separately for single-talker (upper panels) and SSN maskers (lower panels). Data for each talker is shown in columns.

IV. GENERAL DISCUSSION

Although normal-hearing subjects listening to natural speech were able to maintain high levels of intelligibility for TMRs down to 0 dB, performance for cochlear-implant listeners and normal-hearing subjects listening to spectrally degraded speech declined sharply with the addition of noise. Even at high TMRs, performance was dramatically reduced from that obtained in quiet. Another difference among the subject groups was in their ability to take advantage of the temporal dips in the fluctuating masker to glimpse portions of the target sentence, allowing higher performance with a single-talker masker than SSN, and indicating a release from masking. As seen previously at low TMRs (Brungart, 2001; Peters *et al.*, 1998; Qin and Oxenham, 2003), normal-hearing subjects listening to natural speech at TMRs less than 0 dB showed a 20% improvement in score with the different male, single-talker masker over the steady-state noise masker. However, the higher performance with single-talker maskers relative to SSN maskers was not found in the simulation or cochlear-implant results presented here.

Regardless of whether sentences were presented in quiet or in noise, it was very difficult for normal-hearing listeners to recognize words from sentences with the four-channel simulation. It is likely that with only four coarse spectral channels, listeners found the low-context, processed IEEE target sentence too difficult to understand, in which case it was treated as ''noise'' in a similar manner as the SSN masker. In the study by Nelson *et al.* (2003), who also used IEEE target sentences, four-channel sentence recognition performance in noise showed very small differences between a steady and modulated noise masker at all of the gate frequencies tested. Only a very slight amount of masking release was found in the +16 and 0 dB TMR conditions, and none with the +8 dB TMR. These results suggest that with four or fewer functional channels, listeners may have more difficulty distinguishing between the target speech and the masker, regardless of whether the masker is steady-state noise or a modulating masker. It should be pointed out, however, that Qin and Oxenham (2003) found greater masking effects for single-talker than noise maskers with as few as four channels. One possible explanation for this result is that Qin and Oxenham used excerpts from a speech passage as a masker that varied from trial to trial whereas this study used a single sentence that was repeated each trial. Thus the fixed masker used here might have reduced the degree of informational masking.

In contrast with the four-channel simulation, greater masking was observed for single-talker than noise maskers with the eight-channel simulation and with cochlear-implant listeners. The most plausible explanation for these effects is that in addition to energetic masking, the single-talker masker contained meaningful information, thereby acting as a higher-level (informational) masker. Informational masking comprises at least three factors: stimulus uncertainty, target-masker similarity (which relates to modulation interference), and linguistic masking. Since the same masking sentence was used repeatedly, stimulus uncertainty most likely played a minor role. Linguistic masking, on the other hand, might have been an important factor with a single-talker masker. That is, informational masking is possible when the speech masker (e.g., reversed speech or a foreign language) is unintelligible (possesses no semantic content), but carries language-based context. This would parallel observations by Hawley *et al.* (2004) and Freyman *et al.* (2001) who found similar informational masking patterns with speech (content) and time-reversed speech (context). Although not intelligible, time-reversed speech maskers preserve some of the phonetic properties of natural speech which may become confused with those of the target sentence. Last, masking introduced by target-masker similarity might have contributed to the poorer performance found with competing speech compared to noise maskers. It is well known that the spectral and intensity variations in the speech masker are much more difficult for the listener to track and ignore when the target has similar variations (Arbogast *et al.*, 2002; Brungart, 2001; Kwon and Turner, 2001). The acoustic redundancy available to normal-hearing listeners provides a greater range for spectral-temporal discriminations. This makes auditory streaming (i.e., the process by which sound elements are grouped into auditory objects) an easier task by allowing the listener to first identify the masker separate from the target and then to use this information as an aid to glimpse unmasked portions of the target sentence. When the spectral detail is not available, the results presented here and elsewhere (Arbogast *et al.*, 2002; Kwon and Turner, 2001; Qin and Oxenham, 2003) suggest that it may be more difficult to perform the first step for auditory streaming, which is to identify which fluctuating sound is the target and which is the masker. As indicated in a recent study in our laboratory,

speaker identification is a very difficult task for cochlear-implant listeners, most likely because pitch and temporal fine structure cues are not adequately coded (Kong *et al.*, 2003). In support of this, the cochlear-implant users in the present study and the normal-hearing subjects listening to the simulations (both here and in the study by Qin and Oxenham) showed no greater benefit of one talker over another, even when the masker and target speech differed in gender.

With steady-state noise, the masker is more predictable, and obvious differences in temporal envelope between the masker and speech target make segregation a much easier task. Although the female single-talker masker failed to show a significant improvement over that obtained with the same male masker, there was a benefit from gross spectral differences when the masker was SSN (i.e., higher performance was observed with the ''female SSN'' over the ''same male SSN''). These results demonstrate that cochlear-implant users can segregate competing sounds much more easily if it is a simple, temporally based, ''steady vs. fluctuating'' distinction, but not when both masker and target are fluctuating and the listener is forced to rely on the coarse spectral information from their speech processor to segregate a female single-talker masker from a male speech target. This suggests that modulation interference may limit performance when spectral information is reduced.

V. SUMMARY AND CONCLUSIONS

Consistent with previous studies, these results demonstrate that most cochlear implant listeners can use the envelope information delivered through the speech processor to follow conversations in quiet environments, but their performance deteriorates substantially with background noise, particularly if the noise is also speech. Although not directly examined in the present study, the potential contributors to poor performance in noise include reduced spectral resolution and the lack of fine structure information in current cochlear implants. Additionally, the compression mechanism used in cochlear implants may often produce less than optimal TMRs and poorer speech recognition performance as demonstrated here. The cochlear-implant listener is therefore at a serious disadvantage for speech understanding in realistic listening environments, such as classrooms, restaurants, and other social gatherings that require more acoustic redundancy than current cochlear implants provide.

(i) Similar to the results of Qin and Oxenham (2003), the current study found poorer performance with a competing talker than SSN for normal-hearing subjects listening through an implant simulation, and the converse was observed when normal-hearing listeners heard natural, unprocessed speech stimuli.

(ii) The results presented here extend those of Qin and Oxenham by including data from actual cochlear-implant listeners who, like normal-hearing subjects listening to the 8-channel simulation, demonstrated lower scores with a competing talker than with SSN.

(iii) The combined effect of informational and energetic masking most likely contributed to the poorer performance with speech maskers. Informational masking appears to have a strong influence on spectrally impoverished speech since listeners have great difficulty distinguishing the components of the target speech from the masking speech. As a result, portions of the masker may become perceptually integrated with the target.

ACKNOWLEDGMENTS

We are very grateful for the time and dedication our cochlear-implant listeners have offered for this study. We also acknowledge Sheetal Desai, Michael Vongphoe, and Charlotte Guo for their assistance in data collection, Kaibao Nie and Sheng Liu for generating spectrograms [Figs. 3(a)], and John Wygonski for implementing the cochlear-implant simulation software. This work was supported in part by grants from the National Institutes of Health (F32 DC05900 to GSS and 2R01 DC02267 to FGZ).

Arbogast, T., Mason, C., and Kidd, G. (**2002**). ''The effect of spatial separation on informational and energetic masking of speech,'' J. Acoust. Soc. Am. **112**, 2086–2098.

Assmann, P. F., and Summerfield, Q. A. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Bird, J., and Darwin, C. J. (**1998**). ''Effects of a difference in fundamental frequency in separating two sentences,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (London, Whurr), pp. 263–269.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perception of simultaneous voices,'' J. Phonetics **10**, 23–36.

Brungart, D. S. (**2001**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Dorman, M. F., Loizou, P. C., and Tu, Z. (**1998**). ''The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processor with 6–20 channels,'' J. Acoust. Soc. Am. **104**, 3583–3585.

Dunn, O. (**1961**). ''Multiple comparisons among means,'' J. Am. Stat. Assoc. **56**, 52–64.

Duquesnoy, A. J. (**1983**). ''Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons,'' J. Acoust. Soc. Am. **74**, 739–743.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). ''Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing,'' J. Speech Hear. Res. **38**, 222–233.

Festen, J. M. (**1987**). ''Explorations on the difference in SRT between a stationary noise masker and an interfering speaker,'' J. Acoust. Soc. Am. **82**, S4.

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). ''Spatial release from informational masking in speech recognition,'' J. Acoust. Soc. Am. **109**, 2112–2122.

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.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Garnham, C., O'Driscoll, M., Ramsden, R., and Saeed, S. (**2002**). ''Speech understanding in noise with a Med-El COMBI 40+ cochlear implant using reduced channel sets,'' Ear Hear. **23**, 540–552.

Glasberg, B. R., and Moore, B. C. (**1989**). ''Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech,'' Scand. Audiol. Suppl. **32**, 1–25.

Greenwood, D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Hawley, M. L., Litovsky, R. Y., and Colburn, S. H. (**1999**). ''Speech intelligibility and localization in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.
Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). ''The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer,'' J. Acoust. Soc. Am. **115**, 833–843.
Hygge, S., Ronnber, J., Larsby, B., and Arlinger, S. (**1992**). ''Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds,'' J. Speech Hear. Res. **35**, 208–215.
Kawahara, K., Masuda-Katsuse, I., and de Cheveigne, A. (**1999**). ''Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds,'' Speech Commun. **27**, 187–207.
Kessler, D. K., Osberger, M. J., and Boyle, P. (**1997**). ''CLARION patient performance: an update on the adult and children's clinical trials,'' Scand. Audiol. Suppl. **47**, 45–49.
Kidd, G., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2001**). ''Informational masking in listeners with sensorineural hearing loss,'' J. Assoc. Res. Oto. **3**, 107–119.
Kong, Y.-Y., Vongphoe, M., and Zeng, F.-G. (**2003**). ''Independent contributions of amplitude and frequency modulations to auditory perception. II. Melody, tone, and speaker identification,'' abstract from the Twenty-sixth ARO Midwinter Meeting, Daytona, FL, p. 213.
Kwon, B.-J., and Turner, C. W. (**2001**). ''Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?'' J. Acoust. Soc. Am. **110**, 1130–1140.
Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (**2003**). ''Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners,'' J. Acoust. Soc. Am. **113**, 961–968.
Nittrouer, S., and Boothroyd, A. (**1990**). ''Context effects in phoneme and word recognition by young children and older adults,'' J. Acoust. Soc. Am. **87**, 2705–2715.
Oh, E. L., and Lufti, R. A. (**2000**). ''Effect of masker harmonicity on informational masking,'' J. Acoust. Soc. Am. **108**, 706–709.
Peters, R. W., Moore, B. C., and Baer, T. (**1998**). ''Speech reception thresholds in noise with and without spectral and temporal dips for hearing impaired and normal hearing people,'' J. Acoust. Soc. Am. **103**, 577–587.
Plomp, R. (**1994**). ''Noise, amplification, and compression: Considerations of three main issues in hearing aid design,'' Ear Hear. **15**, 2–12.
Pollack, I. (**1975**). ''Auditory informational masking,'' J. Acoust. Soc. Am. **57**, S5.
Qin, M. K., and Oxenham, A. J. (**2003**). ''Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers,'' J. Acoust. Soc. Am. **114**, 446–454.
Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (**1992**). ''Relations among different measures of speech reception in subjects using a cochlear implant,'' J. Acoust. Soc. Am. **92**(4 Pt. 1), 1869–1881.
Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). ''I.E.E.E. recommended practice for speech quality measurements,'' IEEE Trans. Audio Electroacoust. **17**, 227–246.
Scheffé, H. (**1953**). ''A method for judging all contrasts in the analysis of variance,'' Biometrika **40**, 87–104.
Shannon, R. V., Galvin, III, J. J., and Baskent, D. (**2001**). ''Holes in Hearing,'' J. Assoc. Res. Oto. **3**, 185–199.
Stone, M. A., and Moore, B. C. J. (**2003**). ''Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task,'' J. Acoust. Soc. Am. **114**, 1023–1034.
Takahashi, G. A., and Bacon, S. P. (**1992**). ''Modulation detection, modulation masking, and speech understanding in noise in the elderly,'' J. Acoust. Soc. Am. **35**, 1410–1421.
Watson, C., Kelly, W., and Wroton, H. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1186.
Yost, W. A., Sheft, S., and Opie, J. (**1989**). ''Modulation interference in detection and discrimination of amplitude modulation,'' J. Acoust. Soc. Am. **86**, 2138–2147.
Zeng, F.-G., and Galvin, III, J. J. (**1999**). ''Amplitude compression and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.
Zeng, F.-G. (**2002**). ''Temporal pitch in electric hearing,'' Hear. Res. **174**, 101–106.