Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners

Peggy B. Nelson[a]
Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455

Su-Hyun Jin
Department of Communication Disorders, Department of Otolaryngology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 5 February 2003; revised 12 February 2004; accepted 16 February 2004)

Previous work [Nelson, Jin, Carney, and Nelson (2003), J. Acoust. Soc. Am **113**, 961–968] suggested that cochlear implant users do not benefit from masking release when listening in modulated noise. The previous findings indicated that implant users experience little to no release from masking when identifying sentences in speech-shaped noise, regardless of the modulation frequency applied to the noise. The lack of masking release occurred for all implant subjects who were using three different devices and speech processing strategies. In the present study, possible causes of this reduced masking release in implant listeners were investigated. Normal-hearing listeners, implant users, and normal-hearing listeners presented with a four-band simulation of a cochlear implant were tested for their understanding of sentences in gated noise (1–32 Hz gate frequencies) when the duty cycle of the noise was varied from 25% to 75%. No systematic effect of noise duty cycle on implant and simulation listeners' performance was noted, indicating that the masking caused by gated noise is not only energetic masking. Masking release significantly increased when the number of spectral channels was increased from 4 to 12 for simulation listeners, suggesting that spectral resolution is important for masking release. Listeners were also tested for their understanding of gated sentences (sentences in quiet interrupted by periods of silence ranging from 1 to 32 Hz as a measure of auditory fusion, or the ability to integrate speech across temporal gaps. Implant and simulation listeners had significant difficulty understanding gated sentences at every gate frequency. When the number of spectral channels was increased for simulation listeners, their ability to understand gated sentences improved significantly. Findings suggest that implant listeners' difficulty understanding speech in modulated conditions is related to at least two (possibly related) factors: degraded spectral information and limitations in auditory fusion across temporal gaps. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1703538]

PACS numbers: 43.71.Ky, 43.66.Ts, 43.72.Dv, 43.66.Dc [PFA] Pages: 2286–2294

I. INTRODUCTION

Cochlear implant users commonly report difficulty understanding speech in noisy situations. It may be that device limitations, weak temporal pitch cues, the nature of pulsatile cochlear implant processing itself, or the lack of good spectral representation of speech results in difficulty segregating a target signal from a background noise. Nelson, Jin, Carney, and Nelson (2003) reported that listeners with cochlear implants did not demonstrate masking release from modulated maskers. Normal-hearing listeners obtain between 5 and 20 dB improvements in masked speech recognition when listening in gated noise compared to steady-state maskers of the same intensity (e.g., Bacon *et al.*, 1998). Implant listeners, in contrast, showed no difference in masked speech recognition performance when maskers were steady or gated noise, indicating that they did not take advantage of temporal gaps in gated noise. In addition, normal-hearing listeners who listened to four-band simulations of implant processing (after Shannon, Zeng, Kamath, Wygonski, and Ekelid, 1995, which greatly reduced speech spectral information) also failed to demonstrate masking release. These findings were in sharp contrast to well-established results from normal-hearing listeners (e.g., Gustafsson and Arlinger, 1994), who show significant masking release for gated maskers. Results from implant listeners were significantly poorer than those reported for listeners with cochlear hearing loss as well (e.g., Bacon, Opie, and Montoya, 1998). Nelson *et al.* (2003) reported that listeners with implants were significantly affected by background noise, even at favorable signal-to-noise ratios (SNRs), and did not demonstrate a benefit from short temporal ''glimpses'' of the quiet signal at any gate frequency, even when those temporal gaps in the noise were as long as 250 ms. Specific implant device-related processing characteristics (e.g., CIS versus SPEAK processing) did not seem to explain this finding, as listeners with three different processor types all showed the same reduced masking release. The authors proposed several possible explanations for the findings, including modulation masking, reduced auditory stream segregation or fusion abilities, and limited spectral resolution abilities of implant and simulation listeners. Some evidence in the literature suggests that implant listeners may show

---

[a] Send correspondence to Peggy Nelson, Ph.D., 115 Shevlin Hall, Minneapolis, Minnesota 55455; Phone: (612) 625-4569; Fax: (612) 624-7586; E-mail: nelso477@umn.edu

significant modulation masking when speech signals are presented in background noise. Others have proposed that implant listeners have a difficult time segregating signals because of their reduced spectral resolution.

Kwon and Turner (2001) showed that normal listeners identifying consonant-vowel (CV) stimuli processed by implant simulations display increased susceptibility to high-frequency modulation masking when CV syllables are presented in gated noise. They hypothesized that cochlear implant users may be negatively affected by modulation masking because listeners with reduced spectral resolution have increased perceptual weighting for natural amplitude modulations in speech recognition (e.g., Hedrick and Jesteadt, 1996; Hedrick and Carney, 1997). In these studies, the amplitude cues were more important for hard-of-hearing (Hedrick and Jesteadt, 1996) and cochlear implant (Hedrick and Carney, 1997) listeners than for normal-hearing listeners. Thus, the modulated noise may directly interfere with the listener's perception of amplitude envelope cues at both the syllabic and segmental levels. Kwon and Turner evaluated the effects of modulated noise on listeners' understanding of spectrally impoverished signals using 12-band implant simulations. They found that under certain conditions, listeners obtained some masking release. However, a modulated high-frequency masker sometimes reduced consonant identification when compared to identification with an unmodulated masker. They concluded that high-frequency modulated maskers can cause interference in consonant recognition that may offset any benefit provided by the masking release. Apparently in the presence of high-frequency modulated noise, modulation masking caused consonant recognition errors by simulation listeners, and this effect may hold true for cochlear implant users.

In addition to their proposed susceptibility to modulation masking, implant listeners may demonstrate reduced ability to segregate auditory streams. Chatterjee and Galvin (2002) have shown that implant listeners can perform simple auditory stream segregation tasks. Implant listeners in their study were able to segregate rhythmic sequences presented on two electrodes. Greater electrode separation resulted in improved stream segregation, but listeners were able to stream sequences presented to the same electrode pair when those sequences had differing temporal envelopes. Results suggested that implant listeners can segregate streams of information when they differ in place (frequency) and/or modulation rate.

Grimault, Bacon, and Micheyl (2002) showed that normal-hearing listeners can segregate auditory streams on the basis of temporal fluctuations alone, in the absence of other spectral or temporal cues. Listeners generally segregated sequences into two streams when the difference in amplitude modulation (AM) rate between their noise carriers was greater than an octave. These results suggest that listeners (including cochlear implant users) may be able to successfully segregate sound streams based on the AM characteristics of each stream, regardless of the spectral content of the two sound stream carriers, as long as the modulation frequencies of the two sounds streams are at least an octave apart.

Qin and Oxenham (2003) recently reported that implant simulation listeners have significantly less making release when understanding speech in modulated versus steady noise. They compared normal-hearing listeners' speech recognition in steady speech-shaped noise, modulated speech-shaped noise, and single-talker speech interference under conditions of 4-, 8-, and 24-band implant simulations (after Shannon *et al.*, 1995). They found that increasing the number of spectral channels significantly improved simulation listeners' performance in modulated noise, but that even for simulation signals with 24 bands of spectral information, listener performance was significantly poorer than for the original speech signals. They hypothesized that spectral information is critical for segregation of speech from modulated noise, and that low-frequency pitch cues (that are reduced in simulation speech signals) are important for speech segregation.

In the current set of experiments, the various factors that may be related to reduced masking release in implant users were investigated, including informational masking, spectral resolution, and auditory fusion/segregation. Informational masking has been defined as threshold elevation due to non-energetic factors such as signal uncertainty, masker–stimulus similarity, or distraction from extraneous sounds (informational maskers) that do not physically mask the target signals (see Wightman *et al.*, 2003 and Durlach *et al.*, 2003). If implant users perceive the modulated maskers and speech signals as similar, the maskers may provide significant informational, rather than energetic, masking. The role of spectral resolution in masking release was proposed by our previous study and was investigated further by Qin and Oxenham (2003). Both studies hypothesize that reduced spectral information is related to reduced masking release from modulated noise. Auditory fusion and segregation of sentences from noise have also been proposed as important factors in understanding masking release. Qin and Oxenham also theorized that a strong pitch cue is necessary for normal segregation of speech from noise, and that implant listeners may not perceive the needed strong pitch cue.

These hypotheses are tested further with groups of normal and implant listeners. Listeners in the current experiments included cochlear implant users as well as listeners with normal hearing sensitivity who were tested for their understanding of speech that has been modified to reduce spectral cues (after Shannon *et al.*, 1995). We first investigated the roles of informational versus energetic masking in gated noise tasks (Experiment 1). In Experiment 1a, listeners were tested for their understanding of speech in gated noise (as in Nelson *et al.*, 2003) as the duty cycle of the noise was varied from 25% to 50% to 75%. Punch (1978) demonstrated that normal-hearing listeners show a systematic increase in masking as masker duty cycle increases from 25% to 50% to 75%, corresponding to the physical increase in the duration of the masker. Nelson *et al.* (2003) hypothesized that the spectrally impoverished signals experienced by implant and simulation listeners caused listeners to rely on amplitude envelope cues that were easily disrupted by gated noise. The disruption from gated noise might be due to the energetic masking of the gated masker (as observed in Punch), but the disruption may be exaggerated in implant and simulation listeners who use the amplitude envelope cues for understanding. If listeners' performance in gated noise is directly related to the proportion of time that the gated noise is ''on,'' then performance is likely determined by the physical masking of the speech by the noise (energetic masking). If in the current study, however, the duty cycle of the noise has little effect on the amount of masking release, then it is likely that the effect of the gated masker was to disrupt the fusion of the auditory stream of information (causing ''informational masking''), rather than to physically mask the energy of the speech. Results of the investigation of masker duty cycle may shed some light on the role of informational masking in cochlear implant users.

An additional experiment using gated noise (Experiment 1b) tested the hypothesis that greater spectral resolution leads to increased masking release. Our previous work (Nelson *et al.*, 2003) showed that normal listeners obtain little to no masking release when they are listening to speech that has reduced spectral information. The relationship between the amount of spectral information and the amount of masking release is still unknown, however. If listeners with greater numbers of spectral channels are more successful in understanding speech in gated noise, it may be presumed that masking release in gated noise is directly related to the amount of signal spectral information available to the listeners. Qin and Oxenham (2003) recently demonstrated that increasing the number of spectral channels of implant simulations improves, but does not fully restore, masking release. In the current study this is explored further. Listeners with normal hearing sensitivity were presented with original full-spectrum speech stimuli, or with 12- or 4-band simulations of implant processing.

Experiment 2 focuses on the ability of implant and simulation listeners to understand interrupted sentences. When sentences are gated or interrupted by periods of silence, instead of noise, listeners who understand key words in sentences in the presence of the silent interruptions are presumed to have successfully grouped or fused the auditory information into a continuous speech stream. Research on understanding interrupted sentences was done by Miller and Licklider (1950) and by Bashford and Warren (1987), who showed that words and sentences interrupted by silent gaps were less intelligible than continuous speech. It was presumed that this task reflects auditory grouping or fusion (also called ''induction'' by Warren, 1999.) In Experiment 2, listeners were tested for their ability to understand sentences (in quiet) that had themselves been gated, or interrupted by silence. Listeners in Experiment 2a included implant users, normal-hearing listeners, and simulation listeners.

If listeners are more successful at fusing interrupted sentences with increasing spectral information, we presume that auditory fusion by implant listeners can be improved by increasing the amount of spectral information, or the number of spectral channels, available to the implant simulation listeners. In Experiment 2b, normal-hearing listeners were tested for their understanding of gated sentences using implant simulations with either 4 or 12 spectral bands.

II. METHODS

A. Stimuli

Speech stimuli consisted of IEEE sentence materials (IEEE, 1969) spoken by five male and five female talkers. Stimuli were recorded and digitized using a 20 kHz sampling rate. In addition, noise-vocoded simulations of the IEEE sentences were generated (after Shannon *et al.*, 1995) that simulate the processing strategies of cochlear implants. For the initial four-band implant simulations in Experiment 1a, speech stimuli from two talkers were filtered into four frequency bands (100–300, 300–500, 500–1700, and 1700–6000 Hz). The amplitude envelope was extracted from each band using the envelope extraction algorithm provided in Cool Edit Pro™. That envelope was then used to modulate the amplitude of narrowband noises having the same bandwidths. The resulting amplitude-modulated noise bands were recombined with their original relative amplitudes.

For Experiment 1, noise maskers were generated on the fly using the Tucker–Davis waveform generator (TDT WG1). The noise was passed through a Rane 30-band equalizer so that the spectrum of the resulting noise matched the long-term spectrum of the IEEE sentences. Noise maskers were presented either continuously (steady), or gated using 2 ms $\cos^2$-ramped gating that was also implemented on the fly.[1] Gate frequencies ranged from 1 to 32 Hz, while the duty cycle of the noise was systematically altered. Three duty cycle conditions were tested: 75% (the noise was on 75% of each cycle, followed by 25% silence), 50% (noise and silence each comprised 50% of each cycle), and 25% (the noise was on for only 25% of each cycle, followed by a 75% silent interval). The duration of the resulting noise bursts ranged from 8 ms (32 Hz gate frequency, 25% duty cycle) to 750 ms (1 Hz gate frequency, 75% duty cycle). Whereas the 75% condition provided the most physical masking of the stimuli, all three duty cycles produced noise bursts that presumably disrupt the envelope of the speech signal. Signal-to-noise ratios (SNRs) were $+16$, $+8$, 0, -8, or -16 dB, depending upon the listener and condition. Stimuli for Experiment 1b were either full-spectrum natural speech or speech processed through implant simulators (after Shannon *et al.*, 1995) that consisted of 4 to 12 bands. Filter settings are given in Table I. All stimuli for Experiment 1b were presented at -4 dB SNR.

For Experiment 2, sentences in quiet were gated with a 50% duty cycle using 4 ms $\cos^2$-ramped square gating. No noise was present for this experiment. Gate frequencies ranged from 1–32 Hz, resulting in regular bursts of speech that ranged in duration from approximately 14 (32 Hz gate frequency) to 500 ms (1 Hz gate frequency). Stimuli for Experiment 2b were either full-spectrum natural speech or speech processed through implant simulators with 4 and 12 bands, as shown in Table I.

B. Procedures

For all experiments, listeners were seated in the center of a sound-treated chamber. Speech signals were delivered diotically through two Bose 301 speakers (placed at ±45° azimuth) at an overall level of 65 dBA. Speech stimuli were

TABLE I. Filter cutoffs for 4- and 12-band simulations used in Experiments 1b and 2b.

Band	Low-pass cutoff (Hz)	High-pass cutoff (Hz)
4-band simulations:		
1	327	648
2	677	1341
3	1400	2773
4	2898	5739
12-band simulations:		
1	192	356
2	356	550
3	550	775
4	775	1036
5	1037	1342
6	1342	1700
7	1700	2116
8	2116	2602
9	2602	3169
10	3169	3830
11	3830	4600
12	4600	5500

presented in blocks of ten sentences, using all ten talkers in random order for each list. Sentences contained an average of five key words per sentence. All conditions were randomized prior to the beginning of each subject's testing. The listeners responded verbally to each sentence, and the experimenter scored the key words correct for each sentence, circling the correct answers on an answer form. Each listener's results (percent correct keywords) for each condition were later entered into computer database files.

For Experiment 1, keyword identification was evaluated in steady and gated noise. On each trial, the masking noise started first. The sentence began after a random delay that ranged from 10 to 100 ms. The noise was either steady or modulated, as described above. The level of the noise varied depending upon the condition being tested. The listeners in the implant and simulation groups heard the noise at +8- and +16-dB SNR. (Pilot testing with three very successful implant listeners indicated that performance was near 0% for all gate conditions at 0-dB SNR and lower.) Listeners in the normal group heard the noise at 0-, −8-, and −16-dB SNR. All listeners also completed two blocks of sentences in quiet.

For Experiment 2, key-word identification was evaluated for the gated sentences without noise. The sentences were presented either continuously, or were gated on and off, as described above.

C. Subjects

Subjects for the simulation experiments included two groups of eight young adult listeners with normal hearing sensitivity who were not familiar with the stimuli or the implant simulation algorithm. None of the normal-hearing listeners had participated in the previous experiments from Nelson *et al.* (2003). All were between the ages of 19 and 32 years and each participated in only one experiment.

TABLE II. Summary of implant subject characteristics.

Listener	CI/Processor	Age at test	Age at onset of deafness	Age at implantation
N12	Nucleus 22/SPEAK	53	32	42
N14	Nucleus 22/SPEAK	58	49	50
N32	Nucleus 22/SPEAK	34	5	29
C02	Clarion 1.2/CIS	42	18	37
C03	Clarion 1.2/CIS	53	22	49
C05	Clarion 1.2/CIS	47	42	43
C14	Clarion HiFocus/CIS	64	16	60
C15	Clarion HiFocus/CIS	42	33	40
C16	Clarion HiFocus/CIS	48	29	43

Nine postlingually deafened users of cochlear implants were also tested. These were the same nine listeners that participated in the previous experiments (Nelson *et al.*, 2003). A detailed description of the implant users is shown in Table II. Their mean age was 49 years (range: 34 to 64 years), and their average length of deafness prior to implantation was 16 years (range 1 to 44 years). All listeners obtained significant open-set speech recognition from their implants and had worn their implants for more than 2 years (mean: 5 years, range 2 to 11 years). Three listeners used the Nucleus 22 device with the SPEAK processor, three used the Clarion 1.2 device with CIS processing, and three used the Clarion HiFocus device with the CIS processing strategy. Listeners in the implant group used their regular speech processors set to typical sensitivity with no noise reduction. At the beginning of each session, the users set the sensitivity while listening to practice lists, and they were instructed not to change the sensitivity setting.

III. RESULTS AND DISCUSSION

A. Experiment 1a: Understanding speech in gated noise with varying duty cycles

Figure 1 shows the normal-hearing listeners' understanding of sentence keywords in gated noise at duty cycles of 25%, 50%, and 75%. Performance is plotted in percent

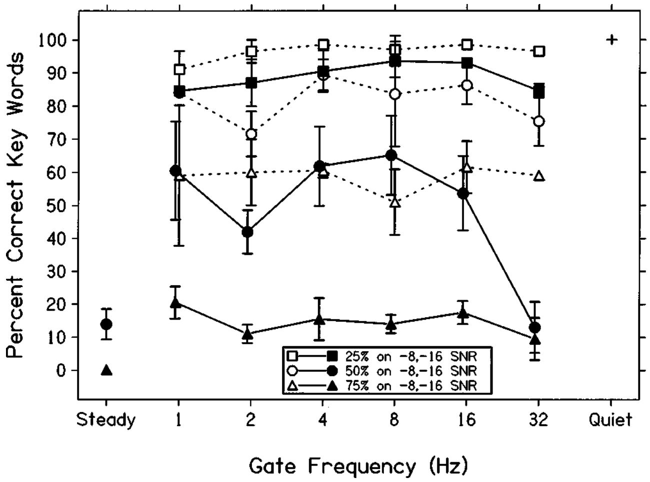

FIG. 1. Normal-hearing listeners' understanding of sentence key words in gated noise. Mean performance (±1 s.d.) is plotted in percent keywords correct as a function of gate frequency for noise presented at −16 dB (filled symbols) and −8-dB SNR. Gate frequencies range from 1 to 32 Hz. Performance in quiet is shown with a "+" symbol. Performance in steady noise is shown with a filled triangle at −16-dB SNR and with a filled circle at −8-dB SNR. A systematic effect of increasing duty cycle is seen.

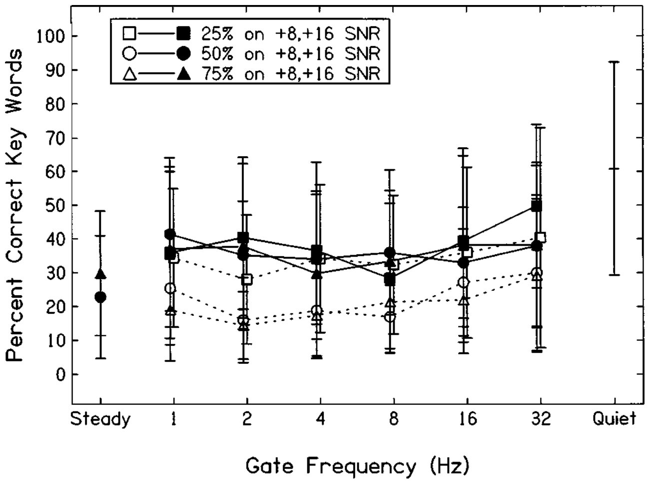

FIG. 2. Four-band simulation listeners' understanding of sentence keywords in gated noise. Mean performance (±1 s.d.) is plotted in percent keywords correct as a function of gate frequency for noise presented at +16 dB (filled symbols) and +8-dB SNR. Gate frequencies range from 1 to 32 Hz. Performance in quiet is shown with a "+" symbol. Performance in steady noise at +16-dB SNR is shown with a filled hourglass; performance in steady noise at +8-dB SNR is shown with an open hourglass. No systematic effect of masker duty cycle is noted.

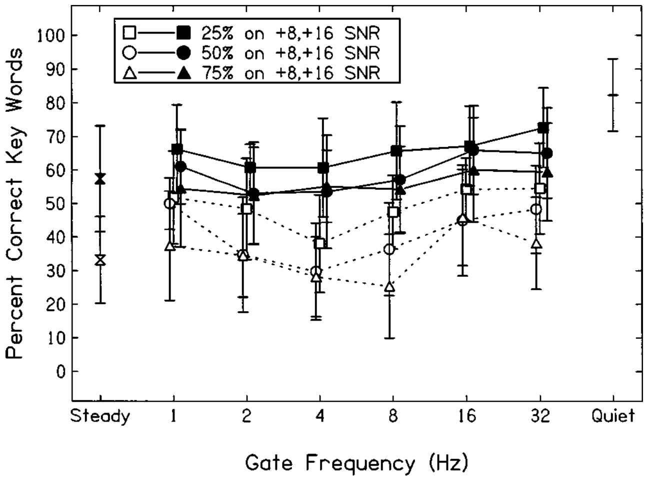

FIG. 3. Implant listeners' understanding of sentence keywords in gated noise. Mean performance (±1 s.d.) is plotted in percent keywords correct as a function of gate frequency for noise presented at +16 dB (filled symbols) and +8-dB SNR. Gate frequencies range from 1 to 32 Hz. Performance in quiet is shown with a "+" symbol. Performance in steady noise at +16-dB SNR is shown with a filled hourglass; performance in steady noise at +8-dB SNR is shown with an open hourglass. No systematic effect of masker duty cycle is noted.

key words correct as a function of gate frequency for noise presented at −16 dB (filled symbols) and −8-dB SNR (open symbols). Gate frequencies range from 1 to 32 Hz. For both SNRs, a predictable relationship between duty cycle and percent correct can be seen, demonstrating the energetic masking that occurs as the noise duty cycle increases from 25% onto 75% similar to the results from Punch (1978).

Normal-hearing listeners identify approximately 10% of keywords at 4 and 8 Hz gate frequencies at −16-dB SNR when the noise duty cycle is 75%. When the duty cycle is changed to 50%, they identify approximately 60% correct at the same gate frequencies. For the 25% duty cycle condition, they identify nearly 90% of the keywords correctly. The same trend is seen for the +8-dB SNR conditions, but the observed difference between 50% and 25% duty cycles is restricted presumably because performance approaches 100%. A repeated measure analysis of variance (ANOVA) indicated a significant difference in performance between each duty cycle for both SNRs ($F[2,10]=168.8, p < 0.0001$). All Bonferroni corrected paired t tests were significant (at $p < 0.01$). There is a clear systematic relationship between the energy of the masker and listeners' performance, suggesting that the energetic masking has a significant effect on performance in gated noise. Normal-hearing listeners' performance was best in noise with 25% duty cycle, moderate in noise with 50% duty cycle, and poorest in noise with 75% duty cycle. At the fastest modulation rates for −16-dB SNR, performance in the 50% duty cycle noise decreased compared to the slower rates. We presume this is due to the effects of forward masking, as noted in Nelson *et al.* (2003). At these fast rates and the highest noise levels, the gated noise may perceptually fill the silent interval, causing a sharp reduction in recognition.

In contrast, Figs. 2 and 3 show simulation and implant listeners' understanding of sentence keywords in gated noise, respectively. Performance is plotted in percent keywords correct as a function of gate frequency for noise presented at +16-dB (filled symbols) and +8-dB SNR (open symbols). Gate frequencies again range from 1 to 32 Hz. There is no apparent systematic effect of the noise duty cycle for implant or simulation listeners for either +8- or +16-dB SNR. Performance for the 25% duty cycle condition is slightly better than performance at 50% or 75% for the simulation listeners at +8-dB SNR ($F[2,22]=34.3, p < 0.01$). Bonferroni-corrected paired t tests ($t[11]=0.04, p > 0.05$) indicated that there is no significant difference in performance for 50% and 75% duty cycle noise. The implant listeners' performance is also slightly better for the 25% than for the 50% and 75% duty cycles. Bonferroni-corrected paired t tests ($t[17]=2.1, p > 0.05$) indicated that there is no significant difference in performance for 50% and 75% duty cycle noise. Performance only differs by approximately 10% when the noise duty cycle varies from 25% "on" to 75% "on."

The data are summarized more clearly by collapsing across gate frequencies from 4 to 16 Hz, as shown in Figs. 4(a) (normal-hearing listeners) and 4(b) (simulation and implant listeners). Because there was no significant effect of gate frequency ($F[2,16]=3.78, p > 0.05$), the data could be collapsed across these gate frequencies to more clearly illustrate the effect of the noise duty cycle.

Figure 4(a) (top panel) shows the mean performance of the normal-hearing listeners as a function of duty cycle (25% on, 50% on, 75% on, and steady) at two SNRs. Error bars represent one standard deviation. The clear trend for reduced performance with increasing duty cycle can be seen. Simply, the greater the amount of time the noise is on, the more the masking that occurs, suggesting that the effect of the gated noise is primarily energetic masking.

Figure 4(b) (lower panel) shows the mean performance of implant (black bars) and simulation (gray bars) listeners as a function of duty cycle (25% on, 50% on, 75% on, and steady) at two SNRs. Error bars represent one standard de-

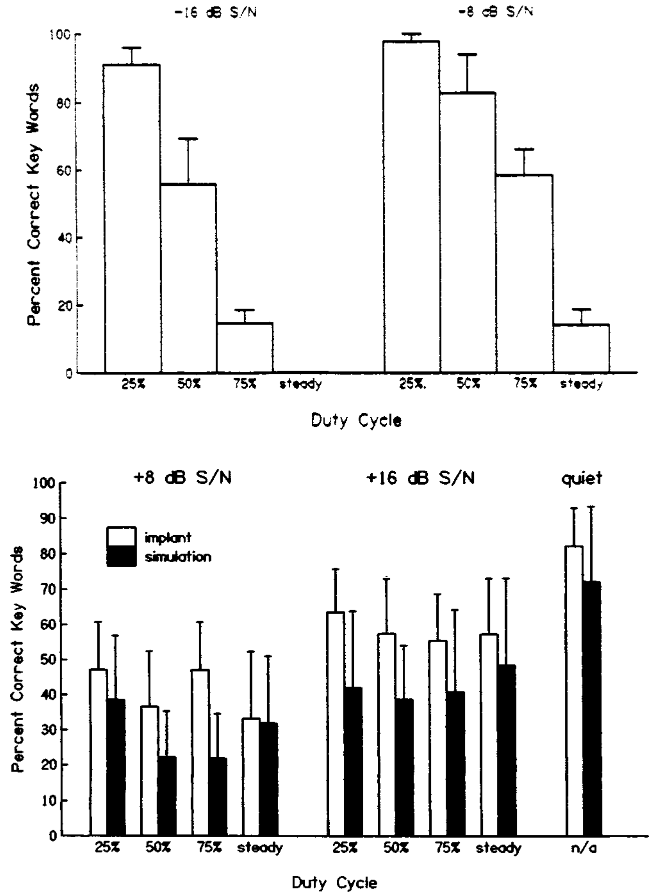

FIG. 4. In Panel (a) (top), mean performance of normal-hearing listeners is shown as percent keywords correct as a function of duty cycle (25% on, 50% on, 75% on, and steady) at two SNRs. Error bars represent one standard deviation. A clear trend for increased masking with increasing duty cycle can be seen. In (b) (bottom), mean performance of four-band simulation (black bars) and implant (gray bars) listeners is shown as percent keywords correct as a function of duty cycle (25% on, 50% on, 75% on, and steady) at two SNRs. Error bars represent one standard deviation. No clear trend for increased masking with increasing duty cycle can be seen.

viation. Implant listeners performed better on the task than did the four-band simulation listeners. However, no clear trend for increased masking with increasing duty cycle can be seen for either group. Instead, there is little change in performance for implant and simulation listeners from the condition of steady noise to any of the masker duty cycle conditions, especially for the +16-dB SNR condition. The fact that the SNR was quite favorable for the implant and simulation listeners (+8 and +16 dB), and the lack of change with duty cycle suggest that the effect of the noise may not be so much energetic masking of the signal, but informational masking, possibly caused by the fluctuations in the noise masker being misinterpreted as part of the signal. In contrast, the normal-hearing listeners showed clear energetic masking at −8- and −16-dB SNR.

The results of Experiment 1a demonstrate that the interference caused by gated noise maskers for implant (and simulation) listeners is not only related to energetic masking. For these stimuli and maskers, there is no systematic relationship between the percentage of time the masker is present and keyword recognition. Instead, the gated noise masker used in Experiment 1 seems to disrupt the flow of information for simulation listeners and implant users.

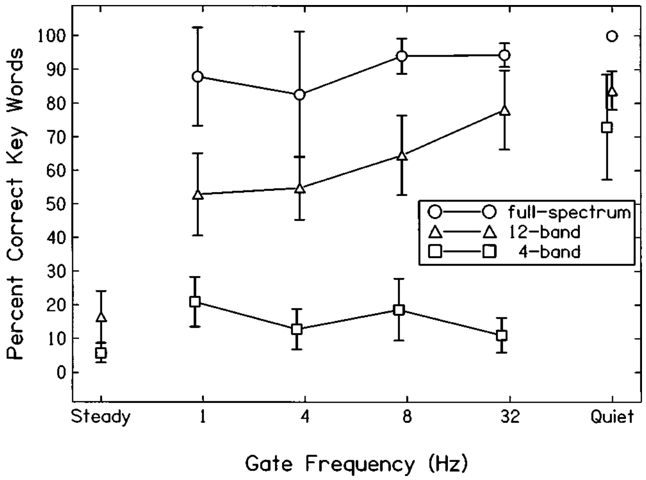

FIG. 5. Performance of normal-hearing listeners is shown for their understanding of sentences in gated noise when the stimuli were full-spectrum (open circles), 12-band simulations (open triangles), and 4-band simulations (open squares). Mean percent correct keywords (±1 s.d.) is plotted as a function of noise gate frequency. All stimuli were presented at −4-dB SNR.

B. Experiment 1b: Understanding speech in gated noise with increasing spectral channels

In Experiment 1b, normal-hearing listeners identified keywords from sentences in gated noise when the stimuli were full-spectrum, 12-band simulations, or 4-band simulations. Participants were eight additional normal-hearing college students who had not previously participated in any experiments. The duty cycle of the noise remained constant at 50%. It was hypothesized that the differences in spectral information present in the three sets of stimuli would have significant effects on listener performance. The results are shown in Fig. 5.

Figure 5 demonstrates the performance of normal-hearing listeners for their understanding of sentences in gated noise when the stimuli were full-spectrum (open circles), 12-band simulations (open triangles), and 4-band simulations (open squares). Mean percent correct keywords is shown as a function of noise gate frequency while the SNR remained constant at −4 dB. All three types of stimuli were highly intelligible in quiet (72% correct for 4 bands, 83% correct for 12 bands), but were almost completely unintelligible (5% and 12%, respectively) in steady noise.

The current listeners showed no release from masking in the four-band simulation condition, as was demonstrated previously in Nelson *et al.* (2003). In this condition, the identification of keywords in gated noise was not significantly different from identification in steady noise. The results from the 12-band simulations are quite different, however. Listeners showed significant release from masking when the stimuli had 12 bands of spectral information. A two-factor repeated measures ANOVA indicated a significant difference among spectral conditions (full-spectrum, 12-band and 4-band: $F[2,16]=409.5$, $p<0.0001$). There was also a significant difference among gate frequencies ($F[3,24]=14.6, p<0.0001$) and a significant gate frequency by spectrum interaction ($F[6,48]=9.3, p<0.01$). For the 12-band simulation stimuli, for which the results were not limited by ceiling or floor effects, the faster the gate frequency, the greater the release from masking. Performance for 12-band gated noise conditions improved (over steady noise) by approximately 40% for slow gate frequencies, and by as much as 60% for the fastest gate frequency (32 Hz). These results suggest that sentence stimuli with greater spectral detail are more easily segregated from gated background noise than are stimuli with poor spectral representation.

The results of Experiment 1 suggest that when implant listeners attempt to understand speech in gated noise, the fluctuations in the noise appear disruptive and confusing, more than energetically masking. Performance of implant and simulation listeners was not significantly better for the 50% duty cycle condition compared to the 75% condition, which would be expected if the disruption were due to the energy of the noise. In addition, it is presumed that energetic masking at +16-dB SNR should be minimal. We hypothesize that the envelope disruptions reduce masking release at least partially because there is not a clear segregation of the speech signal from the noise. As Qin and Oxenham (2003) hypothesize, the weak representation of pitch may contribute to the lack of segregation. As a result, implant and simulation listeners are not able to listen in the dips of the noise, not because the target speech in the dips is inaudible when the speech is 65 dBA and the noise 49 dBA, but because listeners do not group consecutive glimpses of the quiet signal. Masking release is facilitated by improving the spectral representation of the speech signal, presumably because the greater spectral representation of the signal allows more successful grouping of the speech signals and segregation of speech from noise.

C. Experiment 2a: Understanding gated sentences

In Experiment 2, listeners identified sentences that were themselves gated, or interrupted by periods of silence. Identifying gated speech was considered to be a direct measure of auditory fusion of the speech signal. The more that listeners are able to fuse the interrupted speech signal into a coherent stream, the more successful they may be at identifying key words in gated sentences.

Figure 6 shows the performance of normal-hearing listeners (open squares), implant listeners (open circles) and four-band simulation listeners (open triangles) for understanding gated sentences in quiet. Mean percent keywords correct is shown as a function of the gate frequency. As a reference point, recall from Figs. 2 and 3 that the quiet performance of the implant listeners was approximately 80% correct, and for the simulation listeners' quiet performance was approximately 60% correct.

Normal-hearing listeners showed improved performance with increasing gate frequencies above about 4 Hz, with significantly better performance at 16 and 32 Hz than at 2 and 4 Hz. Their identification of gated sentences at 16 and 32 Hz gate frequency was very good, approximately 80% correct. Thus, original full-spectrum speech stimuli that were interrupted with brief periods of silence were easily fused into an intelligible sentence.

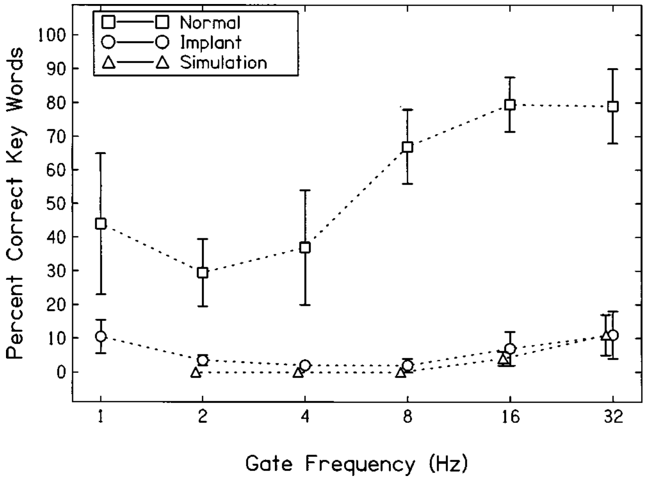

FIG. 6. Performance of normal-hearing listeners (open squares), implant listeners (open circles), and four-band simulation listeners (open triangles) for understanding gated sentences in quiet. Mean percent keywords correct (±1 s.d.) is shown as a function of the gate frequency.

In contrast, implant and four-band simulation listeners performed surprisingly poorly on this task at all gate frequencies. There was no significant difference between implant and the four-band simulation listeners on this task. They rarely scored above 10% correct for any condition, suggesting that they did not tolerate interruptions of any kind in the continuous speech signal. They showed little evidence of fusing across the brief silent periods. Presumably, the lack of fusing might be due to the sparse spectral information available to the implant and four-band simulation listeners. In addition, it may be because the periods of silence disrupted the temporal envelope of the speech signal.

D. Experiment 2b: Understanding gated sentences with increasing spectral channels

Experiment 2b tested the hypothesis that improving the spectral representation of speech would improve the simulation listeners' ability to fuse (and consequently understand) interrupted speech. Figure 7 shows the performance of eight additional normal-hearing listeners for their understanding of gated sentences when the stimuli were full-spectrum (open circles), 12-band simulations (open triangles), and four-band simulations (open squares). Mean percent correct keywords is shown as a function of sentence gate frequency.

A two-factor repeated measures ANOVA indicated a significant difference among spectral conditions (full-spectrum, 12-band and 4-band: $F[2,16]=452.4$, $p<0.0001$). There was also a significant difference among gate frequencies ($F[3,24]=181.5, p<0.0001$) and a significant gate frequency by spectrum interaction ($F[6,48]=76.6, p<0.0001$). At faster gate frequencies, interrupted speech recognition improved for the 12-band and full-spectrum stimuli.

As in Experiment 2a, the listeners in this experiment showed improved keyword recognition with increasing gate frequency. The performance of listeners identifying full-spectrum stimuli interrupted by 32 Hz silent intervals was not significantly different from their identification of continuous sentences. Natural speech that was interrupted by brief silent gaps (16 ms) was highly intelligible, although

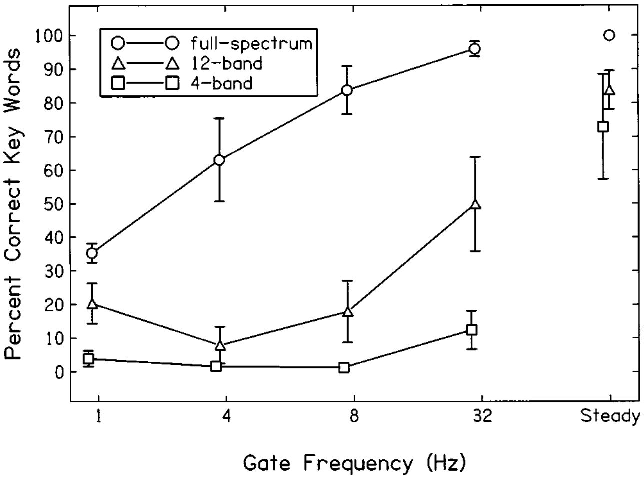

FIG. 7. Performance of normal-hearing listeners is shown for gated sentences when the stimuli were full-spectrum (open circles), 12-band simulations (open triangles), and 4-band simulations (open squares). Mean percent correct keywords (±1 s.d.) is shown as a function of sentence gate frequency.

listeners reported that it sounded "rough." In contrast, speech that was interrupted by longer silent gaps had missing phonemes and syllables that caused a reduction in intelligibility. Performance for the four-band stimuli in this study was very poor and was equivalent to that reported in Experiment 2a above. Simulation listeners with only four bands of modulated noise were unable to understand more than about 10% of keywords in interrupted sentences at any gate frequency. Listener performance for the 12-band simulation condition was significantly better than for the 4-band conditions, and was significantly poorer than the full-spectrum condition. This result implies that spectral information is important for fusing interrupted sentences. As the amount of spectral information increases, listeners improve in their ability to integrate across silent intervals in gated speech.

IV. GENERAL DISCUSSION

Results from this investigation and from the previous related work (Nelson *et al.*, 2003) suggest that cochlear implant users are significantly and negatively affected by fluctuating signals and by fluctuating background noise. Even when temporal gaps in noise are as long as 500 ms, listeners do not show significant release from masking when compared to their performance in steady noise. Background noises as low as +16-dB SNR cause significant disruption in speech perception. The amount of masking caused by fluctuating noise seems to be only weakly related to the modulation frequency or the duty cycle of the noise. The fluctuating nature of the noise appears to serve as a disruption and an informational masker, rather than as an energetic masker.

Implant users not only show limited improvement from temporal gaps in fluctuating maskers, they also do not demonstrate successful auditory fusion of interrupted speech signals. Normal-hearing listeners presented with natural speech show improving word recognition performance with increasing rate of fluctuation. Implant and simulation listeners did not. It has been proposed by McKay and Carlyon (1999) and others (e.g., Guerts and Wouters, 2001; Xu, Tsai, and Pfingst, 2002; Green, Faulkner and Rosen, 2002) that implants have a relatively weak temporal pitch that does not convey voice fundamental frequency in a robust manner. The current results are consistent with the idea that implant listeners' perception of voice pitch is not strong enough to aid in segregating speech from noise, or in fusing interrupted speech. Implant and simulation listeners do not appear to successfully segregate speech signals from gated background noise, and do not fuse interrupted speech signals into a coherent speech stream.

The nature of the impairment seems to be that implant listeners receive a signal with reduced spectral cues, which then results in greater perceptual importance of the amplitude envelope of incoming signals for word and sentence identification (as in Hedrick and Carney, 1997). The envelopes of the signals, however, are affected by the fluctuations introduced by gating the speech or the background noise, and thus they are unreliable cues. Improving the spectral representation of speech, as shown in Experiments 1b and 2b and in the work of Qin and Oxenham (2003), seems to improve simulation listeners' understanding of sentences, and therefore improves their resilience to fluctuations in background noise or speech. Increasing the number of simulation channels does not, however, restore performance to that obtained with natural speech. If improved spectral resolution is important for implant listeners, then the "best" implant listeners would presumably be those who have better spectral (spatial) resolution, better word recognition in quiet, and better masking release and auditory fusion of interrupted sentences. This issue warrants further investigation. It seems logical that those listeners with the best spectral resolution would have the highest word identification scores (as in Henry and Turner, 2003), but it is unknown whether those better listeners have the greatest masking release and/or best identification of gated sentences. To test the hypothesis at a first approximation, correlations between the keyword identification score (percent correct in quiet) and amount of masking release (expressed as the difference between performance in gated and continuous noise in percent correct) were calculated. Similarly, we obtained correlations between percent correct words in quiet and auditory fusion (calculated by the average percent correct gated sentence score). We observed a moderate correlation ($r = 0.5$) between keywords correct in quiet and masking release for these nine cochlear implant users, and no apparent correlation ($r = 0.1$) between keywords correct in quiet and gated sentence performance. It may be, as Qin and Oxenham (2003) hypothesized, that improved spectral representation of speech signals is not sufficient for full masking release and/or auditory fusion of interrupted speech. Instead, they hypothesize that a stronger representation of voice pitch may be necessary for implant and simulation listeners to approach normal performance on such tasks. An additional investigation of the role of spectral resolution versus voice pitch strength is needed.

Zeng (2002) has proposed future cochlear implant-algorithms that might improve implant listeners' pitch resolution and range. The work of Qin and Oxenham (2003) supports the premise that an improved representation of F0 may help implant users significantly. Those proposed potential solutions include: increasing the physical number of electrodes, introducing virtual channels between electrodes (e.g., McDermott and McKay, 1994), and restoring stochastic electric responses (e.g., Zeng, Fu, and Morse, 2000). The current findings support the premise that improvements in implant processing could improve implant listeners' masking release and auditory fusion, provided that the underlying (neural) spatial resolution is adequate.

V. SUMMARY AND CONCLUSIONS

Cochlear implant users are significantly affected by fluctuating signals and by fluctuating background noise. Based on the results of Experiment 1, we conclude that implant and simulation listeners show limited improvement from temporal gaps in fluctuating maskers, and do not demonstrate successful auditory fusion of interrupted speech signals. Implant listeners apparently receive a signal with reduced spectral information that results in listeners' reliance on fluctuations in the amplitude of incoming signals. However, the amplitude of the signal is affected by the fluctuations introduced by gating the speech or the background noise. Based on Experiment 2, we conclude that implant and simulation listeners do not appear to successfully fuse interrupted speech signals into a coherent speech stream. Improving the spectral representation of speech seems to improve listeners' stream segregation and resilience to fluctuations in background noise or speech. Speech processing strategies that convey voice pitch more successfully may be better for providing masking release and fusion of interrupted speech.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. P50-DC00110. The collegial help of Arlene Carney and Gail Donaldson, and the work of research assistants Brian Urban and Laura Price are gratefully appreciated. The helpful comments of two anonymous reviewers provided significant improvements to the manuscript.

---

[1] The quiet speech signals were processed and later mixed with noise. We considered mixing the speech and noise prior to processing during the early stages of our experiments, but found that the number of stimuli (60 lists of 10 talkers) and conditions (five SNRs, three sets of simulation channels, three duty cycles) was unwieldy for preprocessing and storage of processed stimuli. In addition, the fact that the noise would be ''frozen'' for each stimulus was undesirable. As a comparison, we mixed a few sentences from several talkers with gated noise at +8-dB SNR and processed the mixed stimuli. Those stimuli were compared visually and via listening with those that were used (the processed sentences plus noise.) We could not detect a difference between the stimuli, and the results of pilot listening showed very similar recognition. In addition, the results obtained from the experiments are very similar to other reports from the literature regarding implant simulations. Results from the four-band simulations are quite similar to those from our actual implant listeners, as has also been noted by others. As such, we decided to process all the sentences prior to mixing them with noise in real time. We recognize that this is not exactly what implant listeners would experience, but believe that the results are revealing about the nature of implant listening in noise.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). ''The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds,'' J. Speech Lang. Hear. Res. **41**, 549–563.

Bashford, J. A., and Warren, R. M. (**1987**). ''Multiple phonemic restoriations follow the rules for auditory induction,'' Percept. Psychophys. **42**, 114–121.

Chatterjee, M., and Galvin III, J. J. (**2002**). ''Auditory streaming in cochlear implant listeners,'' J. Acoust. Soc. Am. **111**, 2429.

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., and Shunn-Cunningham, B. G. (**2003**). ''Note on informational masking,'' J. Acoust. Soc. Am. **113**, 2984–2987.

Green, T., Faulkner, A., and Rosen, S. (**2002**). ''Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants,'' J. Acoust. Soc. Am. **112**, 2155–2164.

Grimault, N., Bacon, S., and Micheyl, C. (**2002**). ''Auditory stream segregation on the basis of amplitude-modulation rate,'' J. Acoust. Soc. Am. **111**, 1340–1348.

Guerts, L., and Wouters, J. (**2001**). ''Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants,'' J. Acoust. Soc. Am. **109**, 713–726.

Gustafsson, H. A., and Arlinger, S. D. (**1994**). ''Masking of speech by amplitude-modulated noise,'' J. Acoust. Soc. Am. **95**, 518–529.

Hedrick, M. S., and Carney, A. E. (**1997**). ''Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants,'' J. Speech Lang. Hear. Res. **40**, 1445–1457.

Hedrick, M. S., and Jesteadt, W. (**1996**). ''Effect of relative amplitude, presentation level, and vowel duration on perception of voiceless stop consonants by normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **100**, 3398–3407.

Henry, B. A., and Turner, C. W. (**2003**). ''The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners,'' J. Acoust. Soc. Am. **113**, 2861–2873.

IEEE (**1969**). ''IEEE recommended practice for speech quality measurements,'' IEEE Trans. Audio Electroacoust. **17**, 225–246.

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.

McDermott, H. J., and McKay, C. M. (**1994**). ''Pitch ranking with nonsimultaneous dual-electrode stimulation of the cochlea,'' J. Acoust. Soc. Am. **101**, 1622–1631.

McKay, C. M., and Carlyon, R. P. (**1999**). ''Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,'' J. Acoust. Soc. Am. **105**, 347–357.

Miller, G. A., and Licklider, J. C. R. (**1950**). ''The intelligibility of interrupted speech,'' J. Acoust. Soc. Am. **19**, 167–173.

Nelson, P., 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.

Punch, J. L. (**1978**). ''Masking of spondees by interrupted noise in hearing-impaired listeners,'' J. Am. Audiol. Soc. **3**, 245–252.

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.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech perception without spectral cues,'' Science **270**, 303–304.

Warren, R. W. (**1999**). *Auditory Perception: A New Analysis and Synthesis* (Cambridge University Press, New York).

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). ''Children's detection of pure-tone signals: Informational masking with contralateral maskers,'' J. Acoust. Soc. Am. **113**, 3297–3305.

Xu, L., Tsai, Y., and Pfingst, B. (**2002**). ''Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses,'' J. Acoust. Soc. Am. **112**, 247–258.

Zeng, F. G. (**2002**). ''Temporal pitch in electric hearing,'' Hear. Res. **174**, 101–106.

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