

Research Article

Masking Release and Modulation Interference in Cochlear Implant and Simulation Listeners

Su-Hyun Jin,^a Yingjiu Nie,^b and Peggy Nelson^b

Purpose: To examine the effects of temporal and spectral interference of masking noise on sentence recognition for listeners with cochlear implants (CI) and normal-hearing persons listening to vocoded signals that simulate signals processed through a CI (NH-Sim).

Method: NH-Sim and CI listeners participated in the experiments using speech and noise that were processed by bandpass filters. Depending on the experimental condition, the spectra of the maskers relative to that of speech were set to be completely embedded with, partially overlapping, or completely separate from, the speech. The maskers were either steady or amplitude modulated and were presented at +10 dB signal-to-noise ratio.

Results: NH-Sim listeners experienced progressively more masking as the masker became more spectrally overlapping with speech, whereas CI listeners experienced

masking even when the masker was spectrally remote from the speech signal. Both the NH-Sim and CI listeners experienced significant modulation interference when noise was modulated at a syllabic rate (4 Hz), suggesting that listeners may experience both modulation interference and masking release. Thus, modulated noise has mixed and counteracting effects on speech perception.

Conclusion: When the NH-Sim and CI listeners with poor spectral resolution were tested using syllabic-like rates of modulated noise, they tended to integrate or confuse the noise with the speech, causing an increase in speech errors. Optional training programs might be useful for CI listeners who show more difficulty understanding speech in noise.

Key Words: cochlear implants, hearing loss, speech perception

Typical environmental noises such as background conversations are temporally varying in frequency and amplitude. Listeners with normal hearing (NH) can take advantage of gaps in these fluctuating maskers. They are able to “listen in the dips” of temporally varying noise to extract information about the speech signal, thereby experiencing improvement in speech recognition (e.g., Bernstein & Grant, 2009; Festen & Plomp, 1990; Jin & Nelson, 2006). Such performance improvement in the presence of fluctuating compared to steady-state noise conditions is known as *masking release*. Previous studies have reported that NH listeners’ speech recognition scores could improve by as much as 80 percentage points when noise was modulated versus steady (Jin & Nelson, 2006).

However, significant masking release reduction or no masking release has been found in cochlear implant (CI) users or in NH listeners identifying vocoded speech that

simulates speech processed by a CI device (NH-Sim; Fu & Nogaki, 2004; Kwon, Perry, Wilhelm, & Healy 2012; Nelson & Jin, 2004; Nelson, Jin, Carney, & Nelson, 2003; Qin & Oxenham, 2003; Stickney, Zeng, Litovsky, & Assmann, 2004). For example, Nelson and colleagues (Nelson & Jin, 2004; Nelson et al., 2003) compared the performance of three listener groups (NH, CI, and NH-Sim) for sentence recognition in the presence of different masking noises, including steady-state noise and gated noise modulated at different frequencies. They found that the NH listeners showed significant masking release even at high levels of noise (−8 and −16 dB signal-to-noise ratios [SNRs]), whereas the NH-Sim and CI listeners showed limited or no masking release at more favorable SNRs (+8 and +16 dB SNRs).

Typically, robust and redundant speech cues (both spectral and temporal) help NH listeners to segregate relevant speech signals out of a noise mixture and to integrate the signals into a continuous perceptual stream. However, when spectral cues are reduced, via CI processor, for example, a broadband modulated noise might seem perceptually quite similar to the speech: Its frequency components are overlapped with those of speech, and the rate of fluctuation of noise and speech are similar. Without fine-grained spectral information in speech, listeners might rely more on temporal cues to process consonants and vowels. In the presence of modulated noise, such temporal information in speech is obscured by the noise modulations,

^aThe University of Texas at Austin

^bUniversity of Minnesota, Minneapolis

Correspondence to Su-Hyun Jin: shjin@utexas.edu

Editor: Larry Humes

Received August 21, 2012

Revision received January 13, 2013

Accepted January 17, 2013

DOI: 10.1044/1059-0889(2013)12-0049

and NH-Sim and CI listeners may experience modulation interference rather than (or in addition to) masking release (Kwon & Turner, 2001). As a result, little to no masking release has been observed from these groups of listeners.

One thing in common found from the previous studies of speech perception in noise was the positive SNRs used for CI or NH-Sim listeners (e.g., Fu & Nogaki, 2004; Kwon & Turner, 2001; Nelson et al., 2003; Qin & Oxenham, 2003; Stickney et al., 2004). Because these listeners were listening to impoverished spectral speech cues, they were tested at favorable (positive) SNRs in which the signal and noise are more similar in level in order to avoid a floor effect (Nelson et al., 2003). At those positive SNRs, CI or NH-Sim listeners may confuse the modulated noise with the speech and will not be able to take advantage of “dip listening” (Ihlefled, Deeks, Axon, & Carlyon, 2010).

When the characteristics of a noise are quite different from those of speech, the negative effect of masking interference on CI or NH-Sim listeners might be negligible. In studies of spatial masking, when speech and noise were presented from different locations, NH listeners as well as CI and NH-Sim listeners showed spatial release from masking (Ihlefled et al. 2010; Loizou et al., 2009). Furthermore, Kwon et al. (2012) showed that when the envelopes of modulated noise were less temporally overlapped with those of speech signals, ~50% of CI participants were able to benefit from the dips of modulated noise, resulting in significant masking release. These results suggest that if speech and noise are perceptually segregated, CI listeners would be able to experience masking release.

In the current study, we investigated differences in the masking of speech experienced by NH-Sim and CI listeners for maskers that overlap with speech in audio frequency and modulation rate. The purpose of the current study was to examine the conditions in which temporal and spectral interference affect sentence recognition for NH-Sim and CI listeners, and to determine whether masking release might be seen in CI listeners when the spectrum of speech are distinct from that of noise.

We asked two research questions. First, do NH-Sim and CI listeners experience similar systematic increases in masking as the frequencies contained in a broadband masker overlap more closely with the speech signal? To answer this question, both speech and noise were processed by series of bandpass filters. Depending on the experimental condition, the spectra of the maskers relative to that of speech were set to be completely embedded, partially overlapping more (PO-more), partially overlapping less (PO-less), or completely separate from the speech (remote). We hypothesized that NH-Sim listeners would show a systematic reduction in masking of a speech signal as the masker overlapped less with the speech, but that CI listeners may not show such a clear relationship.

Second, under what conditions do NH-Sim and CI listeners experience masking release or modulation interference? Previous investigations indicated that the modulation index of speech is ~3 Hz to 4 Hz, which corresponds to the syllable rate in speech (Drullman, Festen, & Plomp, 1994;

Houtgast & Steenken, 1985). We hypothesized that, if NH-Sim and CI listeners rely more on the temporal envelope cues of speech, then masking interference may be strongest when the noise is amplitude modulated at these rates. Although studies on both NH-Sim and CI listeners (Ihlefled et al., 2010; Kwon & Turner, 2001; Nelson et al., 2003; Qin & Oxenham, 2003) examined the effect of modulation rate on speech perception, no clear answers were reported. We hypothesized that NH-Sim and CI listeners would show no masking release but might show masking interference when the noise is amplitude modulated at a rate approaching the syllable rate. Furthermore, the masking interference associated with the rate effect may occur even when the speech is spectrally remote from the noise because CI listeners are not able to segregate speech from noise even when they are spectrally remote (Stickney et al., 2004).

Experiment 1: Sentence Recognition of NH-Sim Listeners

Method

Participants

Ten undergraduate or graduate students (5 male and 5 female) participated in the experiment. They were 19 to 32 years of age and were Native American English speakers. Their hearing was ≤ 20 dB HL at audiometric frequencies of 250, 500, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz (American National Standards Institute, 2010).

Stimuli

Both Institute of Electrical and Electronics Engineers sentences (IEEE, 1969) spoken by a female speaker and white noise were filtered through 16 filter bands whose cutoff frequencies were set based on the work of Fu and Nogaki (2004). Bandpass filters (4th-order Butterworth) were used. The output of each bandpass-filtered stimulus was numbered in order from 1 to 16: Band #1 contained the lowest frequency components, ranging from 200 Hz to 272 Hz, and Band #16 contained the highest frequency components, ranging from 5768 Hz to 7000 Hz. Table 1 shows the low and high cutoff frequencies of each of the 16 bandpass filters.

The speech temporal envelopes were extracted from the 10 lower individual speech bands, 1–10, by low-passing the bands of speech through a 4th-order Butterworth filter at a cutoff frequency of 160 Hz. Our previous study (Nie & Nelson, 2009) showed that low-pass filtered natural IEEE sentences at a cutoff frequency of 2000 Hz were equally intelligible as those sentences with a full spectrum to NH listeners and listeners with hearing impairment. In the pilot study (Nie & Nelson, 2009), the 10-band CI simulation satisfied our criterion of a $\geq 50\%$ key word identification in NH listeners. The speech envelopes from the 10 bands were half-wave rectified and imposed on the white noise filtered through the same band. Each band carrying speech temporal envelopes was refiltered through the same bandpass filter

Table 1. The low and high cutoff frequencies for each bandpass filter.

Cutoff frequency	Bandpass filter															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Low	200	272	359	464	591	745	931	1155	1426	1753	2149	2627	3205	3904	4748	5768
High	272	359	464	591	745	931	1155	1426	1753	2149	2627	3205	3904	4748	5768	7000

to eliminate the spectral spread resulting from imposing envelopes on noise. The 10 bands of noise carrying a speech envelope were added to formulate the vocoded-speech stimuli.

White noise was bandpassed and then superimposed with sinusoidal amplitude modulation (SAM) to be presented as background noise. The spectra of the background noise were set to completely embedded, partially overlapping (PO), adjacent and not overlapping, or remote, with that of the speech. Table 2 shows spectral components of the noise bands for each noise condition. For the completely embedded condition, the individual noise bands 5 through 10 were combined into one noise band. There were two partially overlapping conditions: PO-more and PO-less, and their noise band components were 7–12 and 9–14, respectively. The noise band components of the adjacent condition included Bands 11–16. For the remote condition, the individual noise bands 13–16 were combined. Each SAM noise condition was modulated at five modulation rates: 0 Hz (unmodulated noise), 4 Hz, 16 Hz, 32 Hz, and 64 Hz. Total numbers of conditions per listener were 26: 5 noise-band conditions \times 5 modulation rates plus 1 quiet condition.

Procedure

Both speech and noise stimuli were processed by a customized MATLAB code with a sampling rate of 22050 Hz and were presented via a SoundMAX card controlled by a Pentium 4 computer. The speech stimuli were presented at 70 dB A for the low-pass spectrum monaurally to the right ear when no background noise was present; when background noise was present, the speech stimuli were presented at 70 dB A and the noise was at an SNR of 10 dB. The SNR was determined by the difference between

Table 2. The low and high cutoff frequencies of the vocoded-speech and each noise condition presented to the implant simulation (NH-Sim) listeners. The name of each noise condition represents the spectral distance between speech and noise: embedded (EM), partially overlapping (PO-more and PO-less), adjacent (AD) to, and remote (RM) from speech bands.

Noise condition	Low cutoff frequency	High cutoff frequency
EM	591	2149
PO-more	931	3205
PO-less	1426	4748
AD	2149	7000
RM	3205	7000
Speech	200	2149

the level of the speech pass-band (i.e., 70 dB A) and the level of the pass-band of the background noise.

In each test condition (in quiet and in noise with different noise spectrum at different modulation rates), two randomly chosen lists of 10 sentences were presented to the participants. After each sentence was presented, participants wrote down on the response sheet what they heard and then pressed the “Enter” key on the computer keyboard to listen to the next sentence. Each participant’s response was scored for the percentage of correct key words identified. The total number of key words from two IEEE lists was 100. If participants responded in homophones of key words, they were scored correct for these corresponding key words. No correct feedback was provided.

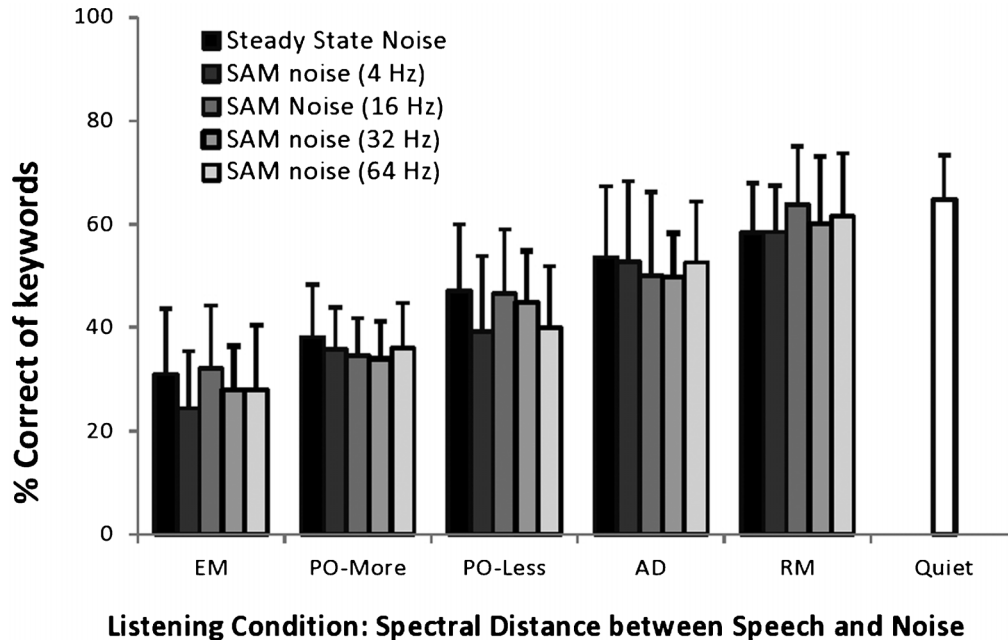
The order of experimental conditions was randomized across the 10 participants. In addition, the order of sentence presentation within each list was randomized. Participants took 4–6 hr in total to complete the speech experiment, which was broken into 1- to 2-hr listening sessions. Participants undertook the sessions on different days. Practice listening was given at the beginning of the first session of the speech experiment. For the practice session, two randomly chosen IEEE lists were presented to the participants. One of the two practice lists was presented without background noise (quiet condition); the other one was presented with background noise of Bands 13 through 16 (remote condition) modulated at 64 Hz.

Results

Figure 1 shows the percentage correct key word identification for the NH-Sim listeners. The scores varied greatly from one listener to another for the same conditions. For instance, the NH-Sim listeners correctly identified 53%–80% of the key words, with an average of 64.7%, in the quiet condition and 10%–54% of the key words, with an average of 20%, when the noise spectrum was completely embedded in the speech spectrum.

Repeated measurement analysis of variance (ANOVA) on conditions with background noise revealed significant differences in the spectral (band) condition; that is, as the spectrum of the background noise approached or overlapped more with the spectrum of speech, listeners identified fewer key words, $F(4, 36) = 100.851, p < 0.001$. However, different amplitude modulated (AM) rates of the modulated bandpass noise did not result in significant improvement in speech understanding, $F(4, 36) = 1.707, p = 0.170$. Speech perception with modulated noise did not improve from that with steady-state noise, indicating no masking release. The

Figure 1. Average performance of the implant simulation (NH-Sim) listeners for understanding Institute of Electrical and Electronics Engineers (IEEE, 1969) sentences. Percentage correct key word identification is shown as a function of the spectral distance between speech and noise. The error bars represent +1 standard error. The name of each noise condition represents the spectral distance between speech and sinusoidal amplitude modulating (SAM) noise: embedded (EM), partially overlapping (PO-more and PO-less), adjacent (AD) to, and remote (RM) from speech bands.



interaction between the spectral separation (i.e., between speech and noise) and modulation rate was not significant, $F(16, 144) = 0.809, p = 0.673$.

Further pair-wise comparisons with Bonferroni adjustment were performed on the factor of spectral separation between noise and speech. The key word identification score of each noise–speech spectral separation was significantly higher than those scores obtained in the smaller noise–speech spectral separations (all $p < 0.05$). That is, with the five noise–speech spectral separations, from small to large—embedded, PO-more, PO-less, adjacent, and remote—the speech understanding score significantly improved progressively.

As no difference in key word identification was found between AM-rate separations, the scores from different AM-rate separations in the same noise–speech spectral (band) separation were averaged to yield a single score for each noise–speech spectral separation. Pair-wise comparisons from the repeated measurement of ANOVAs conducted based on these averaged scores and the scores in the quiet condition showed that the performance in the quiet condition was significantly better than the performance in the conditions where the noise was present, except the remote condition.

Overall, the NH-Sim listeners showed two findings. First, they experienced a systematic increase in masking (decrease in speech recognition) as the masker was more spectrally overlapping with the speech signal. No masking was seen from the spectrally remote masker. Second, they showed neither masking release nor modulation interference

from the modulated maskers in any noise–speech spectral separation condition at any modulation rate, suggesting that the large majority of the masking experienced was based on the spectral content of the speech and noise signals, with minimum effect of the envelopes of the maskers.

Experiment 2: Sentence Recognition of CI Listeners

We conducted a similar experiment with the CI listeners to confirm whether the findings of the NH-Sim listeners would carry over to actual users of CIs. We hypothesized that because CI listeners are more reliant on temporal envelopes, they may experience more modulation interference than the simulation listeners. Further, because of current spread from the CI devices, we hypothesized that CI listeners may experience masking from noises that were presumed to be spectrally remote from the speech signal.

Method

Participants

Seven postlingually deafened adult listeners with CI participated in this experiment. A detailed description of the CI listeners is shown in Table 3. Their mean age was 48 years, ranging from 21 to 73 years, and their average length of deafness before implantation was 26 years, ranging from

Table 3. Demographic and device information for the cochlear implant (CI) listeners. Their devices were set at the participant's everyday listening setting.

Participant	Sex	Age (years)	Duration of deafness (years)	Duration of device use (years)	Brand of CI	Processor	Device use	Ear tested	Etiology
CI 1	Male	55	25	9	Cochlear	Nucleus Freedom	Bilateral	Right	Unknown
CI 2	Female	57	20	7	AB	Harmony Fidelity	Bilateral	Right	Unknown
CI 3	Female	21	18	2	Cochlear	Nucleus5 CP810	Bilateral	Left	Meningitis
CI 4	Male	73	39	2	Cochlear	Nucleus Freedom	Bilateral	Right	Noise/Age
CI 5	Male	58	42	13	AB	Harmony Fidelity	Bilateral	Left	Unknown
CI 6	Female	42	19	9	AB	Harmony HiRes	Bimodal	Right	Genetic
CI 7	Female	32	25	2	Cochlear	Nucleus Freedom	Bilateral	Right	Unknown

18 to 42 years. All listeners had worn their CIs for >2 years ($M = 6$ years, range = 13 years).

Stimuli

In order to create different degrees of spectral interference for the CI listeners, both speech and noise were spectrally filtered and were used to construct experimental conditions with various spectral distances between the speech signals and noise. In order to have the frequency components of the speech signals for CI listeners similar to those of the vocoded-speech for NH-Sim listeners (see Table 1), IEEE sentences were low-pass filtered at 2149 Hz, which is the high cutoff frequency of bandpass noise #10 used for the vocoded-speech. However, if the CI listeners did not achieve speech recognition scores of at least 50% in quiet, the low-pass cutoff frequency was changed to 3205 Hz, which is the high cutoff frequency of Band #11. This allowed those CI listeners to have more spectral information of the speech available. Two out of the seven CI listeners (CI 3 and 5) required the wider speech spectrum to achieve 50% correct in quiet. Because there was no significant modulation effect on speech recognition in the NH-Sim listeners, and previous studies by Nelson and colleagues (Nelson & Jin, 2004; Nelson et al., 2003) using speech materials with a full spectrum showed no significant performance difference between 16 Hz and higher modulation rates for CI listeners, three noise modulation rates—0, 4, and 16 Hz—were used to shorten the length of experiment time for the CI listeners.

Similar to the noise conditions used for the NH-Sim listeners, a series of bandpass-filtered noises were created to design four noise conditions: embedded (noise completely embedded within the speech band), PO-more (partially overlapping more), PO-less (partially overlapping less), and remote (not overlapping) relative to the speech spectrum. The low and high cutoff frequencies for each noise condition are shown in Table 4.

Procedure

Six out of the seven CI listeners wore implants bilaterally. Throughout the experiment, they were asked to use only one device that works better and that he or she relies on more in everyday life. The other device was turned off. One CI listener, CI 6, used a single device without a hearing

aid on the other side. Thresholds through each listener's CI device in the sound field were measured for filtered noises with identical spectra of speech and each masking noise (i.e., embedded, PO-more, PO-less, and remote) in order to confirm that speech and noise were fully audible to the CI listeners. Noises were presented monaurally via a loudspeaker located at 45° azimuth and 0° elevation, and in 1-meter distance to the ear with the CI device.

For the speech recognition test, the loudspeaker was set at the same azimuth and distance to the listeners' CI device as those set in the masking-noise threshold test. The level of speech was 70 dB A. When background noise was present, the noise was at a +10 dB SNR. In each condition, two randomly chosen lists of 10 sentences were presented to the participants. Participants were seated in a sound booth and were asked to repeat back what they heard after each sentence was presented. An examiner outside the booth was able to hear the verbal responses via a talk-back system and marked the key words that were repeated correctly on an answer sheet. After each response, the examiner pressed the "Enter" key on the computer keyboard to present the next sentence. Each participant was scored for the percentage of correct key words identified. The same scoring method used for the NH-Sim listeners was applied to calculate identification scores for each condition. No correct feedback was provided to the participants.

Conditions were randomized across the participants, as was the order of the 10 sentences within each list. Practice

Table 4. The low and high cutoff frequencies of speech and each noise condition presented to the CI listeners. Spectral allocation A refers to the speech and noise spectrum used for CI 1, 2, 4, 6 and 7, and spectral allocation B refers to the speech and noise spectrum used for CI 3 and 5, who needed to have more spectral information of the speech available.

Noise condition	Allocation A		Allocation B	
	Low cutoff frequency	High cutoff frequency	Low cutoff frequency	High cutoff frequency
RM	3205	7000	3904	7000
PO-less	1426	4748	2149	4748
PO-more	931	3205	1753	5768
EM	591	2149	591	3205
Speech	200	2149	200	3205

Table 5. Detection thresholds (in dB SPL) of individual CI listeners for each broadband noise whose spectrum was similar to the speech or noise conditions.

Listening condition	CI 1	CI 2	CI3	CI4	CI5	CI6	CI7
Speech	25.42	22.27	25.5	38.45	26.75	22.25	25.60
RM	19.17	23.67	26.17	38.33	25.00	20.58	25.70
PO-less	22.83	19.00	25.08	35.25	23.00	21.44	27.30
PO-more	24.25	19.25	26.80	37.83	27.00	21.58	26.80
EM	22.75	21.17	25.13	36.42	26.56	21.75	24.30

listening was given at the beginning of the first session of the speech experiment. First, CI listeners were presented with full-spectrum speech (unprocessed speech) in quiet to make sure their best speech recognition score was at least 70%. Second, they listened to low pass filtered speech at 2149 Hz in quiet. If the CI listeners were able to identify the key words at least 50%, this high cutoff frequency was employed to process speech stimuli. If CI listeners could not reach 50% key word identification, the cutoff frequency was increased to 3205 Hz, as described earlier. Once the appropriate cutoff frequency was set for each CI listener, another practice run was played in the remote noise condition with noise modulation of 16 Hz.

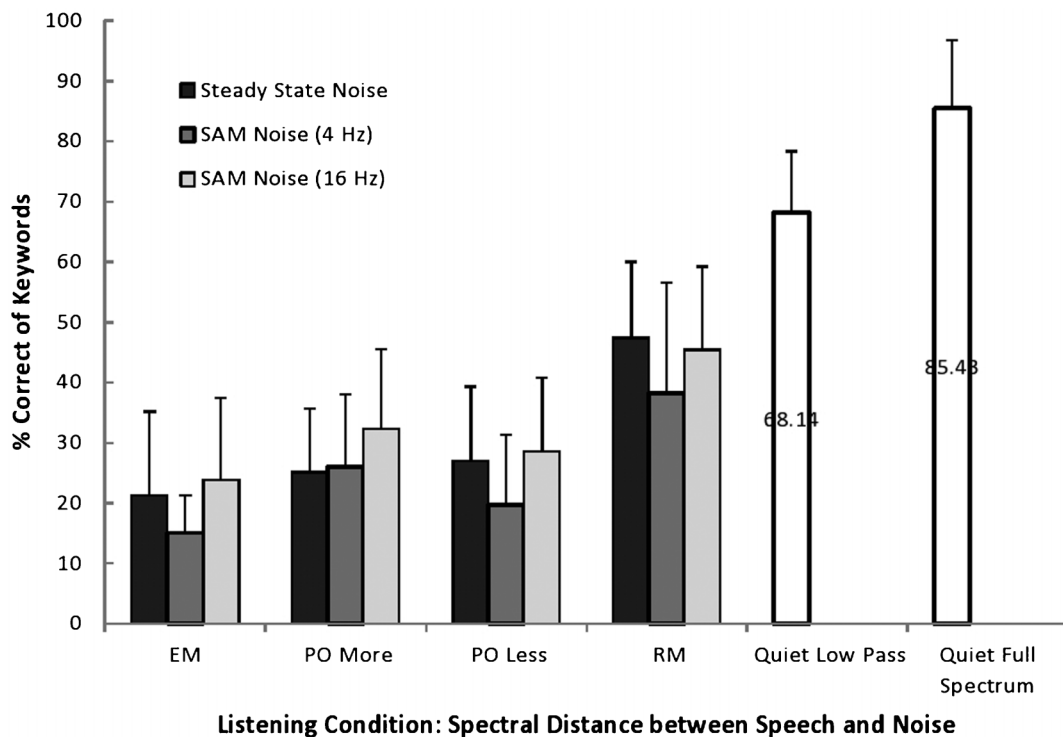
Results

Table 5 shows detection thresholds for individual CI listeners in each band of noise whose spectrum was similar to

the speech or noise conditions. Detection thresholds for speech and each noise condition were measured in order to ensure that both speech and noise were audible to the CI listeners. Thresholds ranged from 19 dB SPL to 38 dB SPL, which were below the presentation levels for speech and noise.

Figure 2 shows the average scores of percentage correct key word identification for the CI listeners as a function of different listening conditions. In quiet, CI listeners understood 85% of full-spectrum speech on average, ranging from 72% to 100%. When the stimuli were low-pass filtered at 2149 Hz (for CI 1, 2, 4, and 7) or at 3205 Hz (for CI 3 and 5), the average performance in quiet was 68%, ranging from 54% to 86%. To compare speech recognition scores in quiet with those in noise, a one-way ANOVA was carried out with five levels of spectral listening conditions: embedded, PO-more, PO-less, remote, and quiet. Because there was no AM modulation applied to the quiet conditions, only speech recognition scores in steady-state noise (0 Hz

Figure 2. Average performance of the cochlear implant (CI) listeners for understanding IEEE (1969) sentences.



modulation) at each spectral distance condition were analyzed. A significant overall spectral separation effect was revealed, $F(4, 30) = 19.017$, $p < 0.001$. Bonferroni post hoc tests showed that CI listeners' performance in the remote condition was significantly worse than in quiet ($p = 0.023$) and significantly better than in the rest of the noise conditions (all p values < 0.026). The tests also indicated better performance in quiet than in all other spectral separation conditions (all p values ≤ 0.023). The results suggest that CI listeners' speech understanding is adversely affected by the presence of noise, even when the noise is spectrally remote from speech.

Two-way repeated-measures ANOVAs (Spectral Separation \times AM Rate) with sentence recognition scores as the dependent variables showed significant effects of both spectral, $F(3, 18) = 22.48$, $p < 0.001$, and AM-rate differences, $F(2, 12) = 19.86$, $p < 0.001$. However, there was no significant interaction between the two factors, $F(6, 36) = 1.39$, $p > 0.05$. Bonferroni post hoc tests for the effect of spectral separation between speech and noise revealed that the performance of the CI listeners in the remote noise condition was significantly better than those in three other noise conditions, embedded, PO-more, and PO-less ($p < 0.001$). Interestingly, speech understanding scores were not significantly different ($p > 0.05$) when the spectrum of noise was either completely embedded with or partially overlapping (PO-more and PO-less) the speech bandwidth. This suggests that the internal spectral masking was greater than what would be expected by the physical spectral characteristics of the noise (the distance from the speech signals), which might be due to the fact that CI listeners experienced significant spectral interaction between the masker and speech.

Examining the effect of AM rate, Bonferroni post hoc tests showed that the CI listeners' sentence recognition was significantly poorer in noise modulated at 4 Hz than in steady noise (0 Hz) or in noise modulated at 16 Hz regardless of spectral separation. This is illustrated in Figure 2, showing the lower performance across noise spectra at 4 Hz compared to the performance at 0 Hz and 16 Hz.

In general, the CI listeners showed evidence of modulation interference when the masker was modulated at 4 Hz, which is a rate similar to the syllabic rate of the speech stimuli. Further, the CI listeners showed significant masking even in the presence of spectrally remote or slightly overlapping noises, indicating significant internal current spread of speech and noise spectra.

Comparison Between NH-Sim and CI Listeners

To examine performance differences between the two listener groups, we compared the results of Experiments 1 and 2. First, a one-way (listener group) ANOVA with speech recognition scores in quiet as the dependent variable showed that there was no significant listener group difference in quiet, $F(1, 15) = 0.565$, $p > 0.05$, which indicates that the NH-Sim and CI listeners were able to understand the low-pass filtered speech equally well in quiet.

To examine performance differences between the NH-Sim and CI listeners in noise, a three-way repeated-measures ANOVA was carried out with two within-subject factors (Spectral Separation \times AM Rate) and one between-subject factor (listener group). Only the noise conditions used for both listener groups were included in the analysis (four spectral separations: embedded, PO-more, PO-less, and remote, and three AM rates: 4 Hz, 16 Hz, and steady state). Figure 3 shows the performance of both the NH-Sim and CI listeners as a function of noise conditions at each AM rate (the top panel for 0 Hz, the middle for 4 Hz, and the bottom for 16 Hz).

Overall, the CI listeners performed significantly worse than the NH-Sim listeners in the presence of noise (between-subject factor) regardless of the modulation rate, $F(1, 15) = 8.312$, $p < 0.05$. In addition, significant effects of spectral separation, $F(3, 45) = 86.079$, $p < 0.001$, and AM rate, $F(2, 30) = 17.723$, $p < 0.001$, were found. A Bonferroni post hoc test showed that when noise was modulated at 4 Hz, listeners performed worse compared to their performance in steady and 16-Hz modulated noise (all p values < 0.001). Listeners' performance in steady noise was not significantly different from that in the modulated noise at 16 Hz ($p > 0.813$). There were no significant two- or three-way interactions among the three factors ($p > 0.05$). This result suggests that noise modulated at a syllabic rate (4 Hz) caused both the NH-Sim and CI listeners greater difficulty understanding speech than a steady noise or a noise with faster modulations.

The nonsignificant interaction between the AM rate and listener group was somewhat unexpected. Based on the previous separate analysis for the NH-Sim group and for the CI group, the CI listeners were significantly affected by the AM rate difference (with 4-Hz modulated noise showing a negative effect on performance), whereas the NH-Sim listeners were not. This ambiguous finding might be due to the fact that the NH-Sim listeners were originally tested with more AM-rate conditions. To test this assumption, we conducted a two-way repeated-measures ANOVA on the performance of the NH-Sim group, including only the conditions that were the same as those for the CI group. The results showed that there was a significant effect of modulation rate, $F(2, 18) = 5.23$, $p = 0.016$, and spectral separation, $F(3, 27) = 80.07$, $p < 0.01$. A Bonferroni post hoc test revealed a significant difference in the NH-Sim listeners' performance between 4 Hz and 16 Hz ($p = 0.025$), but not between steady and 4 Hz or 16 Hz (all p values > 0.05). Thus, we conclude that both the CI and the NH-Sim listeners were somewhat negatively affected by the 4-Hz AM rate, resulting in modulation interference, although the interference was stronger for the CI group than for the NH-Sim group. On the other hand, there was no performance difference between the 16-Hz modulated noise and the steady noise conditions for both listener groups, which suggests neither masking release nor modulation interference for the faster AM rate.

Modulation of the noise seemed to affect CI users in a manner somewhat different from the simulation listeners. Figure 4 shows the average performance difference between modulated noise (either 4 Hz or 16 Hz) and steady-state

Figure 3. Average performances of both the NH-Sim and CI listeners for understanding IEEE sentences are shown as a function of noise conditions at each amplitude modulated (AM) rate (the top panel is for 0 Hz, the middle for 4 Hz, and the bottom for 16 Hz). The percentage correct scores from Figures 1 and 2 are replotted here to compare the performances of the two listener groups at each noise condition.

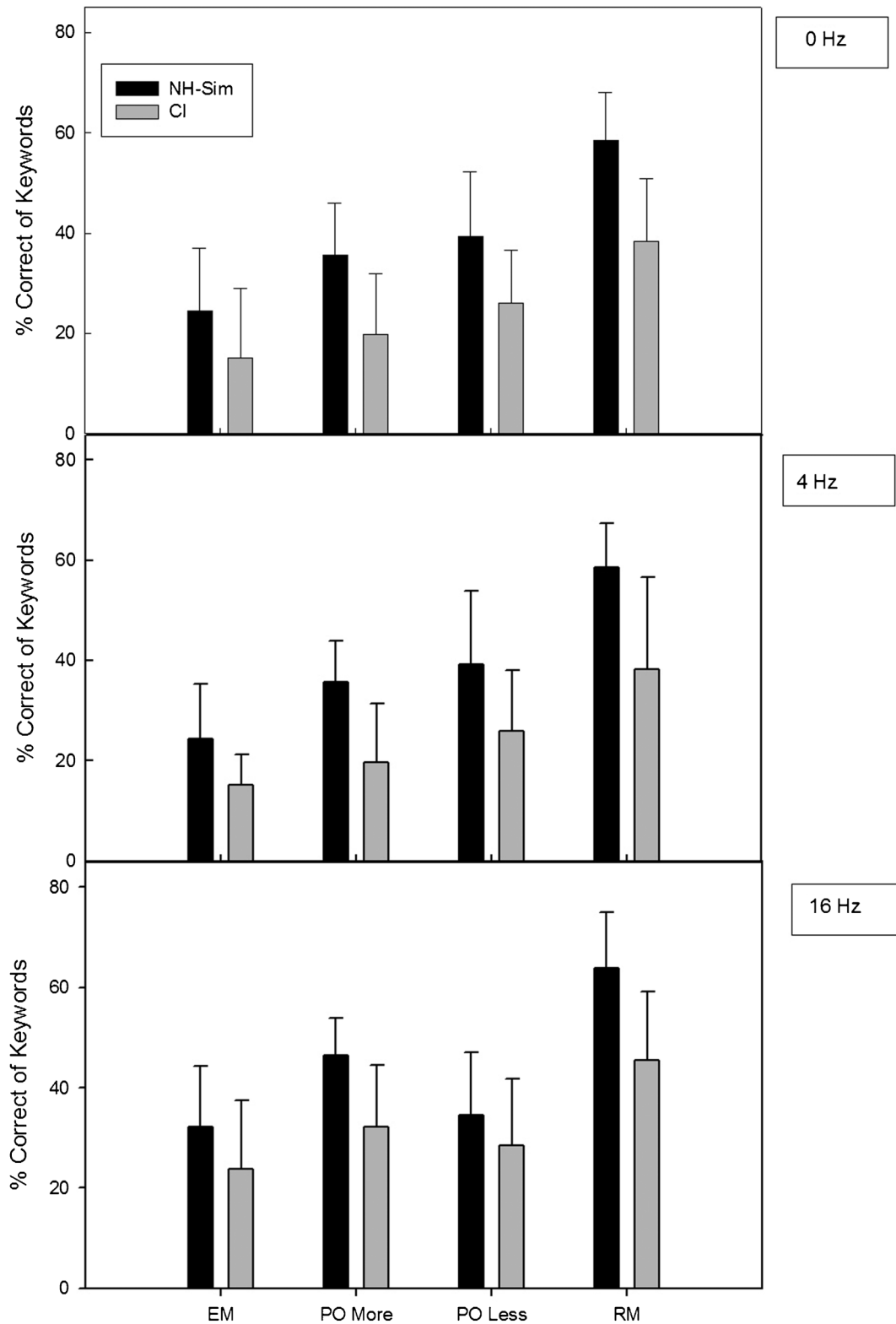
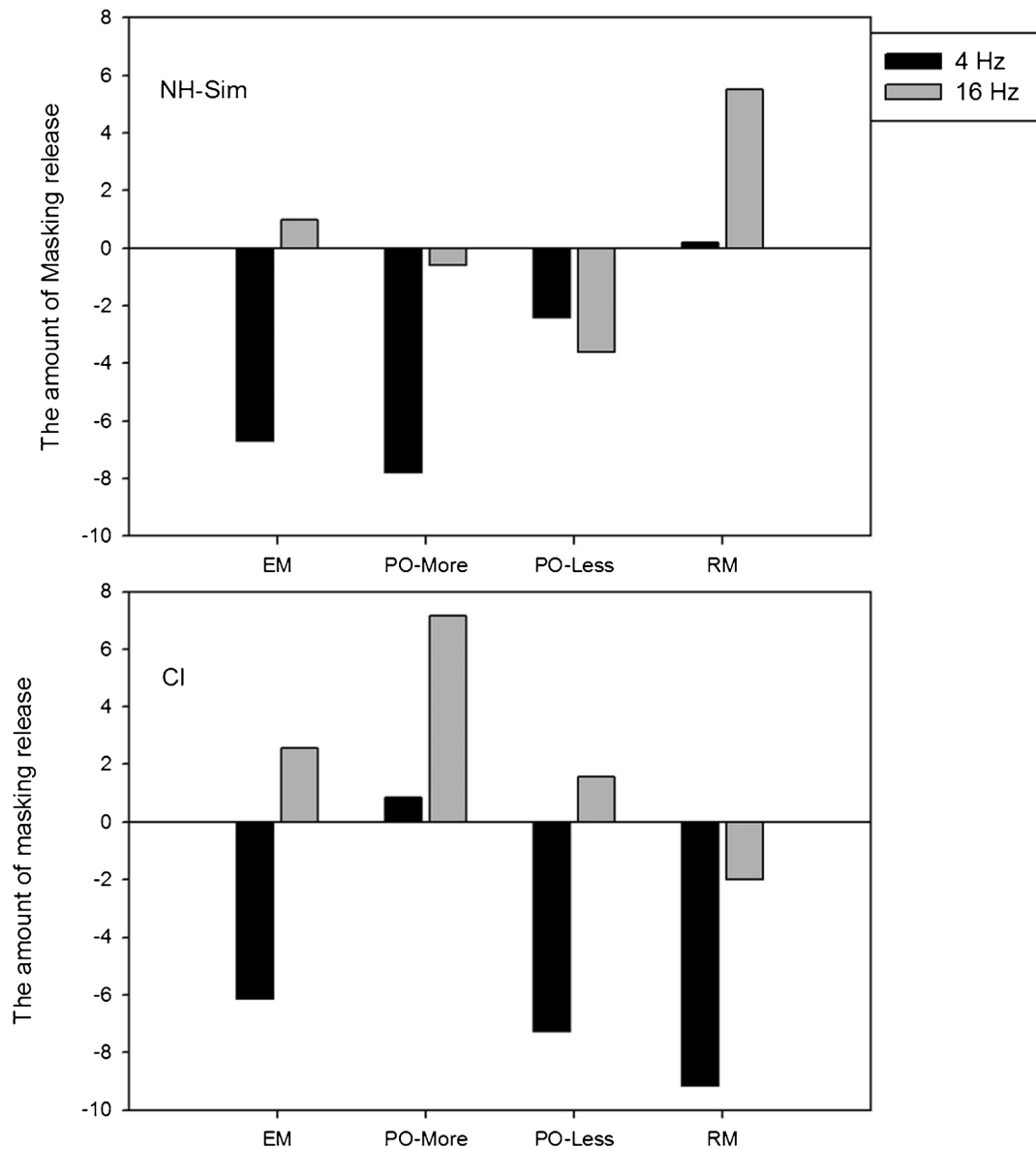


Figure 4. The average performance differences between modulated noise (at 4 Hz and 16 Hz) and steady noise are shown for the NH-Sim (upper panel) and CI listeners (lower panel). The differences >0 indicate masking release; those <0 indicate modulation interference.



noise (0 Hz) for the NH-Sim (upper panel) and the CI listeners (lower panel) as a function of the noise conditions. The difference >0 indicates masking release (i.e., the score in modulated noise exceeds the score in steady noise), whereas the difference <0 suggests modulation interference, with the performance in modulated noise poorer than that in steady noise. For NH-Sim listeners, the performance differences in most noise conditions except remote at 4 Hz were <0, showing masking interference. At 16 Hz, the difference for NH-Sim listeners was >0, indicating some masking release. For CI users, the performance differences in 4-Hz modulated noise were mostly <0 and were always poorer than those in

16-Hz modulated noise. This suggests that at the 4-Hz modulation rate, CI listeners seem to be experiencing modulation interference, especially when compared to their performance at 16 Hz.

Discussion

Effect of Noise on NH-Sim and CI Listeners

The purpose of the current study was to investigate the possible contribution of temporal and spectral interference

to sentence recognition in modulated noise for NH-Sim and CI listeners. Both speech and noise were processed via various sets of bandpass or low-pass filters (Fu & Nogaki, 2004) to create different spectral distances between the target speech and noise. The amplitudes of noise masker were also modulated at different rates to examine whether a certain modulating rate might cause more interference than others. Overall, the results suggest that even when the NH-Sim and CI listeners are equated for their speech understanding in quiet, there are important differences between the two groups in noise conditions. For example, sentence recognition of the CI listeners was significantly poorer than that of the NH-Sim listeners in all types of noise, even though their performances were equivalent in quiet, suggesting that there is considerable spectral spread for “real” CI listeners that is not captured in NH-Sim listeners. Syllabic rates of modulation had negative modulation interference effects on both listener groups, with more noticeable interference for the CI listeners.

When speech and masker noise were spectrally separated, the NH-Sim listeners were not adversely affected by the noise, which resulted in similar performance in quiet and in spectrally remote noise. When the noise and target speech were spectrally overlapped, greater spectral overlapping of vocoded speech and masker resulted in a larger and systematic decrease in speech understanding. The NH-Sim listeners did not show significant masking release when the noise masker was amplitude modulated (seen in Figure 3). Furthermore, the various AM rates on masker did not result in much increased difficulty in understanding vocoded speech, suggesting that the amplitude modulation itself might not substantially interfere with the temporal envelope cue for vocoded speech. However, it should be noted that there was large intersubject variability for listeners’ understanding vocoded speech in noise, similar to that seen by others (Fu & Nogaki, 2004; Nelson et al., 2003).

The CI listeners were significantly affected by noise even when the spectrum of noise was remote from that of speech, which was different from the NH-Sim listeners, whose performance did not drop in the remote condition. The sentence recognition scores of the CI listeners dropped even further when the noise spectrum was overlapped with the speech spectrum. For the three overlapping noise conditions—embedded, PO-more, and PO-less—the CI listeners’ performance remained poor and approximately the same. It should be noted, however, that a potential limitation of the current study might have affected generalization of the findings to the CI population.

In order to examine the effect of spectral interference of noise on speech perception when spectral distances between noise and speech signal were varied, we decided to use speech stimuli that were low-pass filtered at 2149 Hz or 3205 Hz instead of full-spectrum speech, which is a more natural and realistic listening condition. Overall for CI users, the AM rate of the modulated noise had an effect, stronger than their NH-Sim counterparts. Their 4-Hz modulated noise performance was significantly lower than both steady ($p < .001$) and 16 Hz ($p = .005$) noise, with no difference

between steady and 16-Hz modulated noise. Overall, no masking release was observed. In fact, within a given spectral condition (remote, overlapping), performance in steady noise was often better than that in modulated noise, indicating some modulation interference. In particular, performance in 4-Hz AM noise (shown in filled bars with zero or negative masking release) was the poorest.

Clinical Implications for CI Listeners

One of the common findings from the studies of CI users is individual variability in their speech recognition scores regardless of speech materials or noise conditions (e.g., Fu & Galvin, 2008; Gifford, Shallop, & Peterson, 2008; Kwon et al., 2012), which was also found in the current study. Table 6 shows the performance of individual CI listeners in several conditions: quiet with full-spectrum speech and low-pass filtered speech, remote noise for the three AM conditions, and embedded noise for the same AM conditions. Note that CI 2 and CI 6, who performed best in quiet conditions (both full-spectrum and low-pass filtered speech), understood sentences better in most of the noise conditions than the rest of the CI listeners. CI 7 performed almost as well in quiet but scored poorer in noise. Among the remaining CI listeners, CI 3 and 4 had the worst scores in quiet; however, their recognition scores in the noise conditions were not consistently lower than the others. Overall, the current findings of individual variability suggest that a CI listener’s performance in quiet does not always predict how much he or she is affected by noise in real life (as suggested by Smoorenberg, 1992).

Unlike the NH-Sim listeners, the CI users were significantly affected by the noise even when its spectrum was distant from the speech spectrum, as in the remote condition (see Figure 2). As suggested before, this might be due to channel interactions between speech and noise, which is a peripheral effect of the remote masker. It could also be possible that such a spectrally distinct noise masker may interfere with the central processing of a speech signal. This

Table 6. Percentage correct key word identification scores of individual CI listeners in quiet and in noise conditions.

Listening condition	CI 1	CI 2	CI 3	CI 4	CI 5	CI 6	CI 7
Quiet full	78	100	72	81	76	98	93
Quiet low pass	72	73	54	61	67	86	64
RM 0 Hz	42	61	47	44	30	67	41
RM 4 Hz	32	66	38	28	24	61	19
RM 16 Hz	40	62	40	32	40	68	36
PO-more 0 Hz	28	35	14	33	12	37	17
PO-more 4 Hz	31	29	11	38	16	42	15
PO-more 16 Hz	31	38	20	44	20	53	20
PO-less 0 Hz	25	46	23	34	19	34	8
PO-less 4 Hz	7	38	13	23	14	32	11
PO-less 16 Hz	19	42	19	26	27	49	18
EM-0 Hz	6	38	19	13	24	41	8
EM-4 Hz	9	20	9	20	14	24	10
EM-16 Hz	13	30	21	25	13	51	14

central masking could result from the listener's reduced ability to separate the target and masker due to the high degree of target-masker similarity, sometimes also referred to as *informational masking* (Arbogast, Mason, & Kidd, 2002; Brungart, 2001).

It has been reported that CI users may have an effective number of four to eight channels, even though more channels are built in the speech processor (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001; Garnham, O'Driscoll, Ramsden, & Saeed, 2002). Thus, the temporal envelopes of the speech represented at the central auditory system may be perceived as more similar to those of the masker for CI users than for the simulation listeners. The central/informational masking was demonstrated in the Stickney et al. (2004) study, where CI listeners did not show better speech recognition when the talker of single-talker maskers was different from the signal talker. If this is true, then peripheral manipulations on the current commercial CI devices such as developments in signal processing algorithms or mapping strategies might not be enough to reduce the negative effect of background noise on speech perception of CI listeners. Perhaps, specific auditory training such as speech-in-noise training or targeted speech training in modulated maskers could be recommended based on listeners' individual needs (Fu & Galvin, 2008).

Previous studies reported that auditory training was helpful for listeners with hearing impairment to detect/recognize simple stimuli such as musical tones (Galvin et al., 2007; Gfeller et al., 2002) and to understand speech recognition in difficult listening conditions (Fu & Galvin, 2008). Furthermore, behavioral changes due to auditory training seem to be correlated with training-related neurophysiological changes (Tremblay et al., 1998). Further studies need to be conducted to develop and verify an effective training method for CI users listening in modulated noise, in order to improve their segregation of speech from modulated background noise. In the meanwhile, clinicians may recommend commercially available computer-based training such as Computer Assisted Speech Training (Fu, Galvin, Wang, & Nogaki, 2005), Listening and Auditory Communication Enhancement (Sweetow & Henderson-Sabes, 2004), or a word-based auditory training program (Humes, Burk, Strauser, & Kinney, 2009), to CI users who have more difficulty understanding speech in noise. These programs have been shown to alleviate their stress levels in unfavorable listening conditions and to improve their understanding of speech in general.

Conclusion

CI users show more spectrally based masking than do their NH-Sim counterparts. Noise that was thought to be spectrally remote from the speech signal nevertheless interfered with CI listeners' perception of bandpass speech signals. Even when noise was designed to be spectrally remote from the speech signal, the presence of the noise (modulated or unmodulated) produced masking in CI listeners but not in simulation listeners. This can be presumed to result from the current spread of the speech and noise

signals, and could be one reason why even the presumably remote masker showed some modulation interference for the CI users. In the simulation listeners, performance in the remote masker was equivalent to that in the quiet low-pass speech, and no modulation interference is seen. Presumably, then, the simulation listeners could still segregate the speech from the remote noise, and the envelope of the modulated remote noise did not cause interference or confusion with the use of the speech envelope.

Both NH-Sim and CI listeners show some evidence of modulation interference when listening to speech in modulated noise, with modulation interference showing up more strongly in CI users than in their NH-Sim counterparts. This interference happened whether the noise was spectrally embedded, overlapping, or remote from the speech signal. This could be an important factor in understanding the lack of masking release in these listeners, and could add to our understanding of their poor performance in background noise. When listeners have poor spectral resolution and they are tested using syllabic-like rates of modulated noise even at favorable SNRs, they apparently tend to be disrupted by the modulated noise, resulting in increased speech processing errors.

Acknowledgment

Portions of this project were supported by a Texas Speech-Language Hearing Foundation grant to the first author, the Doctoral Dissertation Fellowship at the University of Minnesota to the second author, and the National Institute on Deafness and other Communication Disorders (DC008306) to the third author.

References

- Arbogast, T., Mason, C., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *Journal of the Acoustical Society of America*, *112*, 2086–2098.
- Bernstein, J. G., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *125*, 3358–3372.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, *109*, 1101–1109.
- Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *Journal of the Acoustical Society of America*, *95*, 2670–2680.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, *88*, 1725–1736.
- Fishman, K. E., Shannon, R. V., & Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *Journal of Speech, Language, and Hearing Research*, *40*, 1201–1215.
- Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America*, *110*, 1150–1163.

- Fu, Q. J., & Galvin, J. J. III.** (2008). Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hearing Research, 242*, 198–208.
- Fu, Q.-J., Galvin, J. J. III, Wang, X., & Nogaki, G.** (2005). Moderate auditory training can improve speech performance of adult cochlear implant users. *Acoustic Research Letters Online*, 106–111.
- Fu, Q. J., & Nogaki, G.** (2004). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology, 6*, 19–27.
- Galvin, J. J. III, Fu, Q. J., & Nogacki, G.** (2007). Melodic contour identification in cochlear implants. *Ear and Hearing, 28*, 302–319.
- Garnham, C., O'Driscoll, M., Ramsden, R., & Saeed, S.** (2002). Speech understanding in noise with a Med-El 40 cochlear implant using reduced channel sets. *Ear and Hearing, 23*, 540–552.
- Gfeller, K., Wiit, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., & Ringgenberg, S.** (2002). Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of American Academy of Audiology, 13*, 132–145.
- Gifford, R. H., Shallop, J. K., & Peterson, A. M.** (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology & Neurotology, 13*, 193–205.
- Houtgast, T., & Steenken, H. J. M.** (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America, 77*, 1069–1077.
- Humes, L. E., Burk, M. H., Strauser, L. E., & Kinney, D. L.** (2009). Development and efficacy of a frequent-word auditory training protocol for older adults with impaired hearing. *Ear and Hearing, 30*, 613–627.
- Ihlefeld, A., Deeks, J., Axon, X., & Carlyon, R.** (2010). Simulations of cochlear-implant speech perception in modulated and unmodulated noise. *Journal of the Acoustical Society of America, 128*, 870–880.
- Institute of Electrical and Electronics Engineers.** (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio Electroacoustic, 17*, 225–246.
- Jin, S. H., & Nelson, P. B.** (2006). Speech perception in gated noise: The effects of temporal resolution. *Journal of the Acoustical Society of America, 119*, 3097–3108.
- Kwon, B. J., Perry, T. T., Wilhelm, C. L., & Healy, E. W.** (2012). Sentence recognition in noise promoting or suppressing masking release by normal-hearing and cochlear-implant listeners. *Journal of the Acoustical Society of America, 131*, 3111–3119.
- Kwon, B. J., & Turner, C. W.** (2001). Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference? *Journal of the Acoustical Society of America, 110*, 1130–1140.
- Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., & Roland, P.** (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *Journal of the Acoustical Society of America, 125*, 372–383.
- Nelson, P. B., & Jin, S. H.** (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America, 115*, 2286–2294.
- Nelson, P. B., Jin, S. H., Carney, A. E., & Nelson, D. A.** (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America, 113*, 961–968.
- Qin, M. K., & Oxenham, A. J.** (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America, 114*, 446–454.
- Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P.** (2004). Cochlear implant speech recognition with speech maskers. *Journal of the Acoustical Society of America, 116*, 1081–1091.
- Smootenberg, G. F.** (1992). Speech reception in quiet and in noise conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *Journal of the Acoustical Society of America, 91*, 421–437.
- Sweetow, R., & Henderson-Sabes, J.** (2004). The case for LACE: Listening and auditory communication enhancement training. *Hearing Journal, 75*, 32–40.
- Trembly, K., Kraus, N., & McGee, T.** (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *Neuroreport, 9*, 3556–3560.

Copyright of American Journal of Audiology is the property of American Speech-Language-Hearing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.