Original Article

Trends in Hearing
2014, Vol. 18: 1–14
© The Author(s) 2014
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216514553783
tia.sagepub.com

Speech Perception in Tones and Noise via Cochlear Implants Reveals Influence of Spectral Resolution on Temporal Processing

Andrew J. Oxenham[1,2], and Heather A. Kreft[2]

Abstract

Under normal conditions, human speech is remarkably robust to degradation by noise and other distortions. However, people with hearing loss, including those with cochlear implants, often experience great difficulty in understanding speech in noisy environments. Recent work with normal-hearing listeners has shown that the amplitude fluctuations inherent in noise contribute strongly to the masking of speech. In contrast, this study shows that speech perception via a cochlear implant is unaffected by the inherent temporal fluctuations of noise. This qualitative difference between acoustic and electric auditory perception does not seem to be due to differences in underlying temporal acuity but can instead be explained by the poorer spectral resolution of cochlear implants, relative to the normally functioning ear, which leads to an effective smoothing of the inherent temporal-envelope fluctuations of noise. The outcome suggests an unexpected trade-off between the detrimental effects of poorer spectral resolution and the beneficial effects of a smoother noise temporal envelope. This trade-off provides an explanation for the long-standing puzzle of why strong correlations between speech understanding and spectral resolution have remained elusive. The results also provide a potential explanation for why cochlear-implant users and hearing-impaired listeners exhibit reduced or absent masking release when large and relatively slow temporal fluctuations are introduced in noise maskers. The multitone maskers used here may provide an effective new diagnostic tool for assessing functional hearing loss and reduced spectral resolution.

Keywords

cochlear implants, speech perception, auditory perception, hearing, perceptual masking, temporal processing, spectral resolution

Introduction

Understanding speech in a background of noise is an important, and sometimes challenging, part of everyday human communication. This challenge is particularly acute for people with hearing loss. Despite the use of sophisticated signal processing techniques, neither hearing aids nor cochlear implants (CIs) are currently able to restore speech understanding in noise to normal (e.g., Humes, Wilson, Barlow, & Garner, 2002; Zeng, 2004). In particular, noise-reduction algorithms, such as spectral subtraction (Boll, 1979), improve the physical signal-to-noise ratio but generally produce little or no improvement in speech intelligibility (e.g., Jorgensen, Ewert, & Dau, 2013).

A new way to understand the failures of noise-reduction algorithms has been suggested by recent empirical and computational work, which emphasizes the role of the inherent temporal-envelope fluctuations in noise when masking speech (Dubbelboer & Houtgast, 2008; Jorgensen & Dau, 2011; Jorgensen et al., 2013; Stone, Fullgrabe, Mackinnon, & Moore, 2011; Stone, Fullgrabe, & Moore, 2012; Stone & Moore, 2014). The temporal envelope refers to the slowly varying changes in sound pressure over time, which are distinguished from the more rapid fluctuations of temporal fine structure (e.g., Rosen, 1992; Smith, Delgutte, & Oxenham, 2002). According to this approach, it is the modulation

[1]Department of Psychology, University of Minnesota, Minneapolis, MN, USA
[2]Department of Otolaryngology, University of Minnesota, Minneapolis, MN, USA

Corresponding author:
Andrew J. Oxenham, Department of Psychology, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, MN 55455, USA.
Email: oxenham@umn.edu

Creative Commons CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (http://www.creativecommons.org/licenses/by-nc/3.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (http://www.uk.sagepub.com/aboutus/openaccess.htm).

energy (i.e., the inherent envelope fluctuations) in noise that is the primary cause of masking, rather than the more traditional measure of overall noise energy (French & Steinberg, 1947; George, Festen, & Houtgast, 2008; Kryter, 1962). If confirmed, this new approach suggests that noise-reduction algorithms should aim at reducing the temporal fluctuations in noise, rather than simply reducing the overall noise energy. However, no studies have yet confirmed these effects in clinical populations—all studies so far have been carried out in normal-hearing listeners, using vocoder techniques to simulate certain aspects of CI processing (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

Here, we tested the hypothesis that the difficulties faced by CI users in understanding speech in noise are determined by the inherent temporal-envelope fluctuations present in the noise. We measured sentence intelligibility in backgrounds of noise (with inherent amplitude fluctuations), steady tones (with no inherent amplitude fluctuations), and modulated tones designed to produce similar amplitude fluctuations to those produced by the noise at the output of the CI. In dramatic contrast to the results from normal-hearing listeners, the CI users showed no benefit of eliminating the inherent fluctuations of the noise through the use of tone maskers. Follow-up experiments demonstrated that the CI users exhibited normal detection thresholds for coherent sinusoidal amplitude modulation, ruling out a lack of sensitivity to temporal-envelope modulation as the cause of the unexpected results. Instead, the lack of differentiation between tone and noise maskers in CI users may be due to the indirect effects of poor spectral resolution, resulting in an effective smoothing of the noise temporal envelopes.

Experiment 1: Speech Perception in Maskers With and Without Inherent Temporal Fluctuations

Methods

Listeners. A total of 12 CI users were tested. Individual details are provided in Table 1. In addition, four normal-hearing listeners (one female and three males, aged 21–38 years) were tested. Normal hearing was defined as having pure-tone audiometric thresholds less than 20 dB hearing level (HL) at all octave frequencies between 250 and 8000 Hz and reporting no history of hearing disorders. All experimental protocols were approved by the Institutional Review Board of the University of Minnesota, and all listeners provided informed written consent prior to participation.

Stimuli. Listeners were presented with sentences taken from the AzBio speech corpus (Spahr et al., 2012). The sentences were presented in three different types of masker: noise, tones, and noise-modulated tones (see

Table 1. Subject Information.

Subject code	Gender	Age (years)[a]	CI use (years)[b]	Etiology of deafness	Duration of hearing loss prior to implant (years)[c]	Speech processing strategy[d]
D02	F	63.8	12.0	Unknown	1	HiRes-P w/Fidelity120; NO ClearVoice
D10	F	59.5	10.9	Unknown	8	HiRes-S with Fidelity120; NO ClearVoice
D19	F	54.0	9.3	Unknown	11	HiRes-S with Fidelity120; NO ClearVoice
D27	F	61.8	4.3	Otoscerlosis	13	HiRes-S with Fidelity120; NO ClearVoice
D28	F	64.7	10.7	Familial progressive SNHL	7	HiRes-S with Fidelity120; NO ClearVoice
D30	F	53.7	7.1	Progressive SNHL; Mondinis	27	HiRes-S w/Fidelity120; ClearVoice LOW
D31	M	48.1	1.5	Meniere's	Unknown	HiRes-S with Fidelity120; NO ClearVoice
D34	F	72.7	1.3	Trauma; progressive	2	HiRes-P w/Fidelity120; NO ClearVoice
D35	F	54.2	2.7	High fever	Unknown	HiRes-S with Fidelity120; ClearVoice MED
D37	F	55.8	1.2	Unknown	<1	HiRes-P w/Fidelity120; NO ClearVoice
N13	M	75.8	23.4	Hereditary; progressive SNHL	4	SPEAK
N32	M	46.1	16.3	Maternal rubella	<1	SPEAK
	Average	59.2	8.4			

Note. M = male; F = female; CI = cochlear implant; SNHL = sensorineural hearing loss.
[a]Age of subject at time of experiment (in years).
[b]Duration of CI use in tested ear (in years).
[c]Duration of severe to profound hearing loss prior to implantation (in years).
[d]Speech processing strategy utilized during testing.

Figure 1). The noise was Gaussian noise, spectrally shaped to match the long-term spectrum of the AzBio speech corpus. The tones were selected to match the center frequencies of the CI channels for the individual CI users. The 16 center frequencies from the standard clinical map for Advanced Bionics CIs were used to generate the stimuli for the normal-hearing listeners. The center frequencies for the individual listeners, as well as the standard clinical map, are shown in Table 2. The amplitudes of the tones were set at the output levels of the respective vocoder channels, so that the overall level and spectral envelope matched that of the noise masker (and the speech corpus). The noise-modulated tones were generated using the same tones as in the tone masker, but each tone was modulated independently with the temporal envelope of a noise filtered into the vocoder subband with a center frequency corresponding to the tone's frequency.

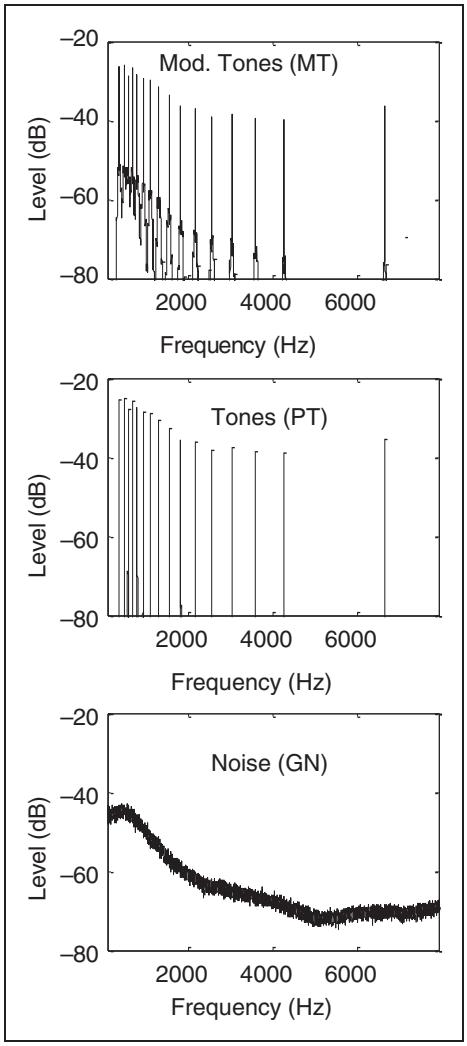

Figure 1. Representation of the three masker types used in Experiments 1 and 3. The three panels provide spectral representations of the modulated-tone (top), tone (middle), and noise (bottom) maskers. MT = modulated tones; PT = pure tones; GN = Gaussian noise.

The masker was gated on 1 s before the beginning of each sentence and was gated off 1 s after the end of each sentence. The masker in each trial was a sample of a longer 25-s sound file, cut randomly from within that longer waveform. The speech and masker were mixed before presentation, and the signal-to-masker ratios were selected in advance, based on pilot data, to span a range of performance between 0% and 100% word recognition.

The speech and the masker were mixed and were either presented unprocessed or were passed through a tone-excited envelope vocoder that simulates certain aspects of CI processing (Dorman, Loizou, Fitzke, & Tu, 1998; Whitmal, Poissant, Freyman, & Helfer, 2007). In most cases, the stimulus was divided into 16 frequency subbands, with cutoff frequencies and the center frequencies of each subband made equal to those in the clinical maps of the individual CI users; for the normal-hearing listeners, the standard clinical map used with Advanced Bionics CIs was used to set the subband frequencies (see Table 2). The temporal envelope from each subband was extracted using a Hilbert transform, and then the resulting envelope was lowpass filtered with a fourth-order Butterworth filter and a cutoff frequency of 50 Hz. This cutoff frequency was chosen to reduce possible voicing periodicity cues and to reduce the possibility that the vocoding produced spectrally resolved components via the amplitude modulation. Each temporal envelope was then used to modulate a pure tone at the center frequency of the respective subband. The speech was always presented at a level of 65 dB sound pressure level (SPL) measured at a position corresponding to the listener's head, and the level of the masker was adjusted to produce the desired signal-to-masker ratio.

Procedure. The stimuli were generated and processed using MATLAB (Mathworks, Natick, MA). The sounds were converted via a 24-bit digital-to-analog converter and were presented via an amplifier and a single loudspeaker, placed approximately 1 m from the listener at 0° azimuth and level with the listener's head. The listeners were seated individually in a double-walled sound-attenuating booth. Listeners responded to sentences by typing what they heard via a computer keyboard. They were encouraged to guess individual words, even if they had not heard or understood the entire sentence. Sentences were scored for words correct as a proportion of the total number of keywords presented. One sentence list (of 20 sentences) was completed for each masker type and masker level. The proportion correct scores were converted to rationalized arcsine units before statistical analysis (Studebaker, 1985). All reported analysis of variance (ANOVA) results include a Huynh-Feldt correction for lack of sphericity where applicable.

Table 2. Center frequencies (CFs) of each subject's clinical map, as well as those used for the normal-hearing listeners, that were utilized to customize the noises and implemented during vocoding in the experiment.

	Electrode number [apical (1) to basal (22)]																					
Subject code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
All NH	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665	NA	NA	NA	NA	NA	NA
D02	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665	NA	NA	NA	NA	NA	NA
D10	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665	NA	NA	NA	NA	NA	NA
D19	OFF	386	463	556	668	804	965	1160	1394	1674	2012	2417	2904	3490	4193	6638	NA	NA	NA	NA	NA	NA
D27	386	463	556	668	804	965	1160	1394	1674	2012	2417	2904	3490	4193	6638	OFF	NA	NA	NA	NA	NA	NA
D28	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665	NA	NA	NA	NA	NA	NA
D30	342	484	598	739	914	OFF	OFF	OFF	1129	1396	1725	2132	2635	3256	4025	6575	NA	NA	NA	NA	NA	NA
D31	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665	NA	NA	NA	NA	NA	NA
D34	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665	NA	NA	NA	NA	NA	NA
D35	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665	NA	NA	NA	NA	NA	NA
D37	333	455	540	642	762	906	1076	1278	1518	1803	2142	2544	3022	3590	4264	6665	NA	NA	NA	NA	NA	NA
N13	OFF	OFF	OFF	OFF	229	439	642	844	1045	1246	1447	1655	1895	2177	2500	2873	3316	3865	4529	5307	6217	OFF
N32	OFF	OFF	229	439	642	844	1045	1246	1447	1655	1895	2177	2500	2873	3316	3865	OFF	4529	5307	6217	OFF	OFF

Note. N32 and N13 are Cochlear Nucleus-22 devices. The electrode numbering system used in that device clinically goes from 22 (apical) to 1 (basal). Reported here is the opposite, so electrode 1 is the same as electrode 22 in the clinical numbering convention. This was done for ease of display. NH: normal-hearing listeners.

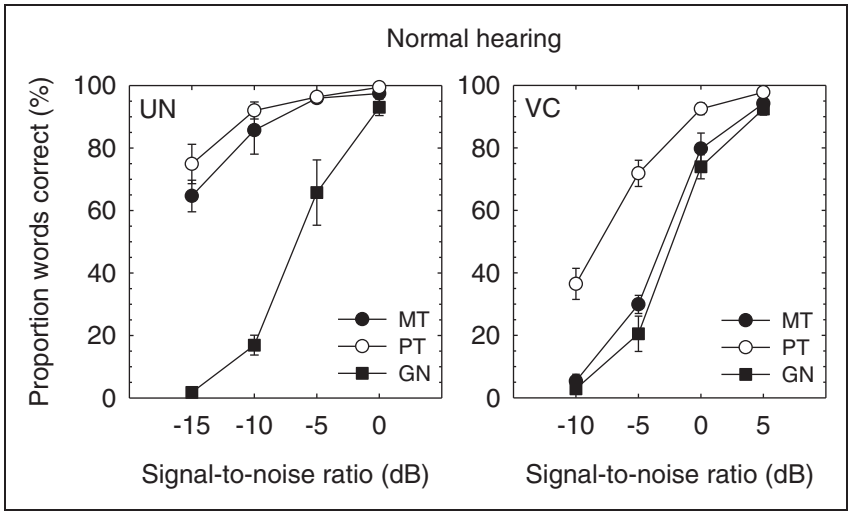

Figure 2. Results from normal-hearing listeners. The proportion of words reported correctly from sentences is plotted as a function of signal-to-masker ratio for maskers consisting of modulated tones (MT), pure tones (PT), and Gaussian noise (GN). Performance with unprocessed stimuli (UN) is shown in the left panel, and performance with the tone-vocoded stimuli (VC) is shown in the right panel. Error bars represent ±1 standard error of the mean between listeners.

Results

The mean results from the normal-hearing listeners are shown in Figure 2. The results from the unprocessed conditions are shown in the left panel, and the results from the tone-vocoded conditions are shown in the right panel. The different masker types are denoted by different symbols. As expected (Stone et al., 2011, 2012), the normal-hearing listeners were able to take advantage of the lack of modulation and the spectral sparsity of the tone maskers and were able to understand a substantial proportion of the sentences even at very low signal-to-masker ratios (Figure 2, left panel, open circles). Introducing the amplitude modulation to the tone maskers affected performance somewhat (filled circles), but performance remained much higher than in the presence

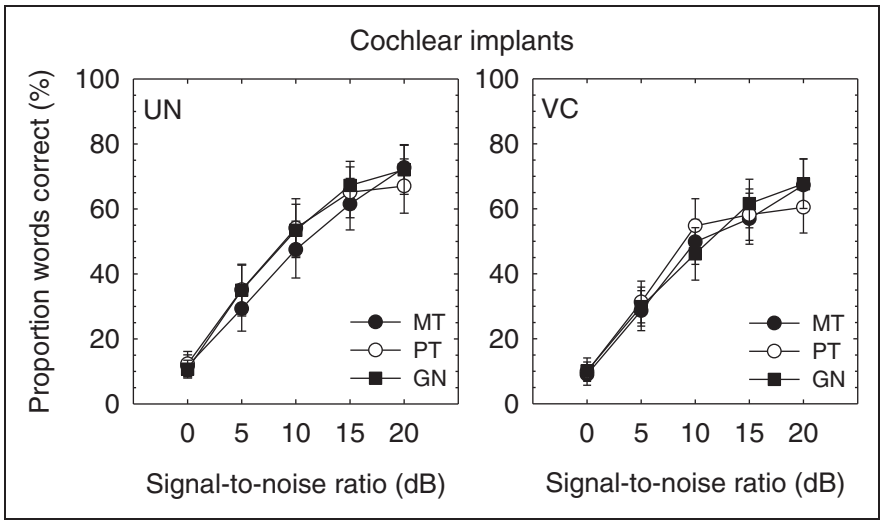

Figure 3. Results from cochlear implant (CI) users. The proportion of words reported correctly from sentences is plotted as a function of signal-to-masker ratio for maskers consisting of modulated tones (MT), pure tones (PT), and Gaussian noise (GN). Performance with unprocessed stimuli (UN) is shown in the left panel, and performance with the tone-vocoded stimuli (VC) is shown in the right panel. Error bars represent ±1 standard error of the mean between listeners.

of the spectrally dense noise masker (squares). A two-way within-subjects ANOVA confirmed that there were significant effects of both signal-to-masker ratio, $F(3, 9) = 95.6$, $p < .001$, partial $\eta^2 = .97$, and masker type, $F(1.6, 5) = 354.3$, $p < .001$, partial $\eta^2 = .992$, on the proportion of words reported correctly. Contrast analysis confirmed that there were significant differences between each of the three masker types ($p < .01$ in all cases). Passing the speech and maskers through a tone vocoder (Figure 2, right panel) led to poorer performance overall, as expected based on the loss of spectral resolution and the loss of the original temporal fine structure. Again, a within-subjects ANOVA revealed significant effects of both signal-to-masker ratio, $F(3, 9) = 359.1$, $p < .001$, partial $\eta^2 = .992$, and masker type, $F(1.9, 5.8) = 78.6$, $p < .001$, partial $\eta^2 = .963$. The modulated tone and noise maskers produced quite similar results, as predicted (contrast analysis: $p = .05$). This is because each tone in the masker was modulated with a temporal envelope designed to mimic the modulations inherent to the noise masker, so that the output from the vocoder was very similar in the noise and modulated-tone conditions. Most importantly, the tone masker produced substantially better performance than either the modulated tone or noise maskers (contrast analysis: $p = .005$ and $p = .001$, respectively), consistent with expectations for a masker with no inherent modulation (Stone et al., 2012).

The results from the CI users were quantitatively and qualitatively different from those of the normal-hearing listeners. There were individual differences in overall performance: The poorest performer reported 11% of all words correctly, averaged across all signal-to-masker ratios, whereas the best performer reported 74% of words correctly. Nevertheless, the pattern of results was quite similar across subjects, and so only the mean data are shown in Figure 3. In line with expectations, performance in noise was poorer than for normal-hearing listeners, with a higher signal-to-masker ratio required for equivalent performance. However, contrary to the results from normal-hearing listeners, the CI users appeared to gain little or no benefit from the tone maskers—the masker with no inherent fluctuations produced as much masking as the noise and the modulated-tone maskers (Figure 3, left panel).

The CI processors in the devices used by our listeners filter the incoming sound into frequency subbands and present the temporal envelope from each subband to a different electrode. Because the center frequencies of the tone vocoder subbands were selected to match the center frequencies of each of the CI subbands for each CI user individually, we expected little or no difference between the unprocessed and the tone-vocoded conditions: The vocoder should have been nearly perceptually transparent to the CI users, with the exception that some high-frequency envelope cues that may have been audible in the unprocessed condition were filtered out by the 50-Hz envelope lowpass filter in the vocoded condition. In line with expectations, results from the unprocessed and vocoded conditions were very similar (cf., left and right panels of Figure 3).

A three-way repeated-measures ANOVA with factors of vocoder status (vocoded or unprocessed), masking condition (tone, modulated tone, or noise), and signal-to-masker ratio (0–20 dB) confirmed no significant main effect of masking condition, $F(1.7, 19) = 1.8$, $p = .199$, or

vocoder status, $F(1, 11) = 2.4$, $p = .148$. As expected, there was a significant main effect of signal-to-masker ratio, $F(1.8, 19) = 106$, $p < .001$. The interaction between signal-to-masker ratio and masking condition was significant, $F(6.4, 71) = 4.1$, $p = .001$, presumably reflecting the apparent flatting of the curve in the pure-tone condition at high signal-to-masker ratios (Figure 3, open circles) that was not present in the other masking conditions. None of the other interaction terms were significant ($p > .1$ in all cases).

Discussion

Two related aspects of the results are noteworthy. The first aspect pertains to overall performance in the tone-masker conditions. In the unprocessed conditions, normal-hearing listeners were able to understand around 75% of words at a signal-to-masker ratio of −15 dB, whereas the CI users only approached that level of performance at a signal-to-masker ratio of +20 dB—a difference of about 35 dB between groups (cf., left panels, open circles in Figures 2 and 3). Even in the vocoded conditions (Figures 2 and 3, right panels, open circles), where the normal-hearing listeners had access to only 16 spectral channels, the signal-to-masker ratio required for equivalent performance was about 15 to 20 dB higher for the CI users.

The second noteworthy aspect of the data pertains to the effect of inherent masker fluctuations on speech reception. In the normal-hearing listeners, consistent with other recent studies (Stone et al., 2011, 2012; Stone & Moore, 2014), inherent fluctuations in the masker played an important role in limiting speech intelligibility. In the vocoded conditions (Figure 2, right panel), the tonal masker with no temporal-envelope fluctuations produced significantly less masking than the maskers with inherent fluctuations, resulting in an improvement in speech reception threshold (the signal-to-masker ratio at which 50% of words are correctly reported) of about 5 dB. In the unprocessed conditions (Figure 2, left panel), the difference in performance between the tone- and noise-masker conditions was much greater, presumably because the normal-hearing listeners were also able to make use of the spectral gaps between adjacent tones and the differences in temporal fine structure between the masker and speech. However, even here, adding temporal-envelope fluctuations to the tone masker resulted in a significant decrement in performance (cf., open and filled circles in left panel of Figure 2). The novel aspect of the data is that the CI users showed no significant effect of masker fluctuations on speech intelligibility. In clear contrast to the results from the normal-hearing listeners in either the unprocessed or the vocoded conditions, the CI users gained no benefit from eliminating the inherent fluctuations from the masker. Thus, in contrast to conclusions drawn from the data of normal-hearing listeners (Stone & Moore, 2014), it seems that the overall masker energy, *not* the modulation energy, determines speech intelligibility in CI users.

What explains the qualitative difference in results between the normal-hearing listeners and the CI users? Three possible explanations are considered here. The first possibility is that the CI users may be less sensitive to the inherent fluctuations in the masker, to the extent that the fluctuations no longer interfere with speech intelligibility. Most studies of amplitude-modulation detection in CI users have presented the stimuli using direct stimulation, rather than acoustically via the subjects' clinical speech processor, and have generally found good sensitivity to amplitude modulation and a similar dependence of threshold on modulation rate (e.g., Chatterjee & Oberzut, 2011; Shannon, 1992). At least two studies used broadband noise as a carrier and presented the stimuli via loudspeaker to the CI users. The first study (Won, Drennan, Nie, Jameyson, & Rubinstein, 2011) found somewhat poorer modulation sensitivity in CI users when compared with data from normal-hearing listeners from the literature (Bacon & Viemeister, 1985), particularly at higher modulation frequencies. They also found a correlation between modulation sensitivity and speech perception, in line with earlier findings by Fu (2002) using direct stimulation. However, some of the differences between normal-hearing listeners and CI users at high modulation frequencies may have been due to the mode of presentation: The normal-hearing listeners were presented with sounds over headphones, whereas the CI users were presented with sounds via a loudspeaker in a sound booth. It is possible that the acoustics of the booth may have effectively smoothed the temporal-envelope modulations at higher modulation frequencies. The second study measured amplitude-modulation detection thresholds only at 8 Hz (Gnansia et al., 2014). That study reported a wide range of detection thresholds, from −1 to −22 dB relative to 100% modulation. Average thresholds were around −10 dB, which was worse than that found by Won et al. (2011). Gnansia et al. (2014) provide several reasons for why their thresholds may have been poorer. More importantly, they reported a significant correlation between modulation detection thresholds and consonant and vowel recognition in quiet, but not in noise (although a trend toward correlation was present in all their comparisons). Overall, there is little reason to believe that the lack of sensitivity to inherent noise fluctuations in CI users is due to poor sensitivity to amplitude modulation.

The second possibility is that the tones used in the tonal masker interacted with one another within the analysis filters of the CI and generated temporal beating

patterns that in turn produced modulation masking of the speech. This possibility seems unlikely, given that the lowest beating frequency of about 122 Hz (the frequency difference between the two lowest tonal carriers) is much higher than the frequencies typically associated with speech perception, which tend to be between 4 and 16 Hz (Drullman, Festen, & Plomp, 1994a, 1994b). The large difference in frequency between the beating modulation and the important speech modulation frequencies makes it unlikely that the beating frequencies masked the modulation frequencies associated with speech perception (Bacon & Grantham, 1989; Dau, Kollmeier, & Kohlrausch, 1997; Houtgast, 1989). In some cases, higher modulation frequencies, associated with the F0 of the speaker, can also assist in speech intelligibility (e.g., Stone, Fullgrabe, & Moore, 2008). In our case, such cues were filtered out from the vocoder simulations and, as vocoding had no significant effect on the performance of CI users, those cues do not seem to play a large role in the current experiments.

The third possible explanation for the difference in the pattern of results between the normal-hearing listeners and the CI users relates to the different signal-to-masker ratios at which the two groups were tested. It is possible that at the high signal-to-masker ratios necessary to test CI users, the inherent fluctuations of noise become unimportant, even in normal acoustic hearing. This explanation is rendered unlikely by the fact that performance was measured in both CI users and normal-hearing listeners at a signal-to-masker ratio of 0 dB. At this signal-to-masker ratio, the normal-hearing listeners showed a clear effect of masker fluctuations when listening through a vocoder (Figure 2, right panel), whereas the CI users did not (Figure 3). Thus, the difference in the pattern of the results between normal-hearing listeners and CI users is unlikely to be due solely to the different signal-to-masker ratios tested.

Experiment 2: Amplitude-Modulation Detection in Cochlear-Implant Users and Normal-Hearing Listeners

Rationale

As discussed in the previous section, it seems unlikely that insensitivity to amplitude modulation on the part of CI users, or modulation masking through beats generated by the tonal carriers, can explain the lack of effect of inherent noise fluctuations on speech intelligibility in CI users. Nevertheless, this hypothesis was tested here by measuring detection thresholds for sinusoidal amplitude modulation at 8 Hz—a frequency in the middle of the range thought to be most important for speech perception (e.g., Drullman et al., 1994a, 1994b). Detection thresholds were measured for a single carrier and for multiple carriers, with frequencies corresponding to all the center frequencies listed in Table 2. If the CI users were insensitive to amplitude modulation, then their detection thresholds should be higher (worse) than those of normal-hearing listeners in both the single-carrier and multi-carrier conditions. If beating between carriers had the effect of masking amplitude modulation in the speech range, then CI users should have poorer thresholds specifically in the multicarrier condition, which included potential beating between carriers.

Methods

Subjects. The same 12 CI users who participated in Experiment 1 also took part in this experiment. Six new normal-hearing listeners (four females and two males, aged 19–34 years), as defined in Experiment 1, were tested.

Stimuli and Procedure. Each stimulus was 500 ms long, including 10-ms raised-cosine onset and offset ramps, and was presented at an average level of 65 dB SPL in the soundfield, as described in Experiment 1. The stimulus level was roved by ±3 dB on each trial with uniform distribution to reduce potential cues based on loudness. Two conditions were tested. In the first condition (single carrier), a single tone was presented with a frequency corresponding to the center frequency of an electrode in the middle of the array (usually electrode 8, or 1278 Hz for the normal-hearing listeners); in the second condition (multiple carriers), all carrier frequencies were presented simultaneously. In each trial, the tones were presented three times, separated by silent interstimulus intervals of 500 ms. In one of the three presentations, chosen at random on each trial, the tones were modulated in amplitude by an 8-Hz sinusoid. The listeners' task was to identify which of the three tones in each trial was modulated. Feedback was provided after each trial. At the beginning of a run, the modulation depth was 100% (modulation index, $m = 1$ or 0 dB). After two consecutive correct responses, the modulation index was decreased; after a single incorrect response, the modulation index was increased. This two-down, one-up adaptive procedure tracks the 70.7% correct point on the psychometric function (Levitt, 1971). Initially, the modulation index was changed in steps of 3 dB; after two reversals in the direction of the threshold tracking procedure, the step size was reduced to 2 dB; and after a further two reversals, the step size was reduced to its final size of 1 dB. Each run terminated after six reversals at the final step size, and the mean of the modulation index (in dB) at the final six reversal points was taken as

the threshold estimate for the run. Each of the two conditions was repeated five times in each listener in an interleaved random order, and the mean of the five measurements were taken to be the threshold for that condition and listener.

Results and Discussion

The mean results from both the CI users and the normal-hearing listeners are shown in Figure 4. The mean modulation detection thresholds for both groups were around −17.5 dB (relative to 100% modulation) for the single-carrier and around −23 dB for the multicarrier condition. The thresholds for the normal-hearing listeners were poorer than those from earlier studies (e.g., Kohlrausch, Fassel, & Dau, 2000), where thresholds of −25 to −30 dB have been reported. This difference may be due to less practice on the part of our subjects, as well as the free-field presentation method used here, compared with headphone presentation in the earlier studies. There was no significant difference in mean amplitude-modulation detection thresholds between the two groups, $F(1, 15) = 0.53$, $p = .821$, and no significant interaction between subject group and the number of carriers, $F(1, 15) = 0.126$, $p = .727$; however, there was a main significant effect for the number of carriers, $F(1, 15) = 13.3$, $p = .002$, partial $\eta^2 = .470$.

With the single carrier, the lack of a significant difference in modulation detection thresholds between the CI users and the normal-hearing listeners suggests that the results from Experiment 1 cannot be explained in terms of poorer sensitivity to speech-relevant amplitude modulation on the part of the CI users. The fact that there was also no difference between the two groups with the multiple carriers suggests that any beating between carriers (if at all audible) did not act to mask speech-relevant modulation energy.

In summary, therefore, the fact that CI users were insensitive to inherent masker fluctuations in Experiment 1 cannot be explained in terms of reduced sensitivity to modulation per se, either because of higher absolute modulation thresholds or because of modulation masking produced by interacting carriers.

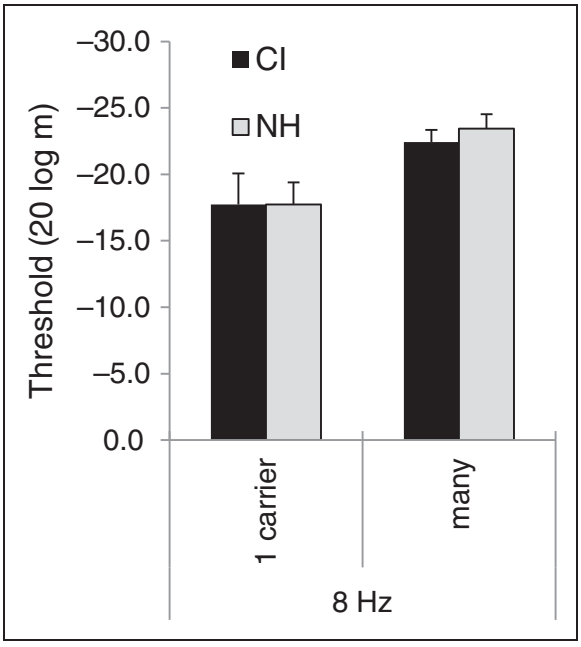

Figure 4. Detection thresholds for sinusoidal amplitude modulation at a rate of 8 Hz in normal-hearing listeners (NH) and cochlear-implant users (CI). Thresholds are shown for the single-carrier condition (left) and the multicarrier condition (right), in which all carriers (up to 16) were present and modulated coherently.

Experiment 3: Simulating the Effects of Current Spread on Modulation Perception in Normal-Hearing Listeners

Rationale

In this final experiment, we tested the hypothesis that reduced spectral resolution could account for the surprising finding in Experiment 1 that CI users' speech was not affected by inherent fluctuations in the masker. The spectral resolution of CIs is limited in part by the spread of current from each electrode to remote locations within the cochlea. In this way, the signals from neighboring electrodes can interfere and sum with each other. The temporal envelope of Gaussian noise has a Rayleigh distribution, and the sum of independent Rayleigh-distributed variables has reduced variance (i.e., reduced modulation energy), relative to the mean (i.e., the overall energy; Hu & Beaulieu, 2005). The biological interface between the CI electrodes and the spiral ganglion of the auditory nerve may therefore result in an overlap and *smoothing* of the temporal-envelope fluctuations inherent in noise, to the extent that the fluctuations no longer interfere with speech perception. This hypothesis was tested by presenting normal-hearing listeners with an acoustic simulation of the effects of current spread using a variant of the tone-excited envelope vocoder from Experiment 1.

Methods

The same six normal-hearing listeners who took part in Experiment 2 were tested in this experiment. As in Experiment 1, listeners were presented with sentences taken from the AzBio speech corpus (Spahr et al., 2012) in the same three types of masker: noise, tones,

and noise-modulated tones. The methods of presentation and analysis were essentially identical to those of Experiment 1.

The novel aspect of this experiment was that the speech and the masker were mixed and then passed through a tone vocoder that simulated the spectral smearing associated with current spread from monopolar stimulation. The new vocoder also split the incoming signal into 16 frequency subbands and extracted the temporal envelope from each subband as before. However, the output of each subband was no longer just the envelope from the corresponding input subband but was instead a weighted sum of the intensity envelopes from all the input subbands (Figure 5). This procedure is similar to one used in an earlier study (Crew, Galvin, & Fu, 2012), with the exception that the earlier study summed amplitudes, rather than intensities. The following equation was used to define the temporal envelope, e_i, at the output of subband i

$$e_i(t) = \sqrt{\sum_{n=1}^{16} \left(w_{i,n} e_n(t)\right)^2} \quad (1)$$

where n is the input subband number, e_n is the temporal envelope after filtering through subband n, and $w_{i,n}$ is the weight applied to e_n when summing for the output envelope i. The weight was simply an attenuation of 8 dB/octave on either side of the subband center frequency, in line with estimates of spatial tuning curve slopes from CI users with monopolar stimulation (Nelson, Kreft, Anderson, & Donaldson, 2011).

Results and Discussion

Limiting the spectral resolution of normal-hearing listeners (through simulated current spread) had a dramatic effect on speech intelligibility (Figure 6). First, performance in the presence of the noise masker was severely reduced, relative to the vocoder simulations without

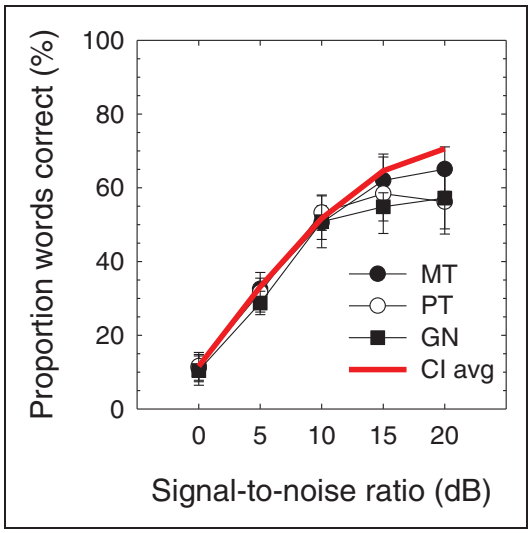

Figure 6. Proportion of words correctly reported from sentences as a function of signal-to-masker ratio for normal-hearing listeners, listening through a tone vocoder with the simulated effects of reduced spectral resolution. The three masker types are modulated tones (MT), pure tones (PT), and Gaussian noise (GN). Error bars represent ±1 standard error of the mean between listeners. The mean performance of the cochlear-implant users in Experiment 1 (Figure 3) is shown as a red curve for comparison.

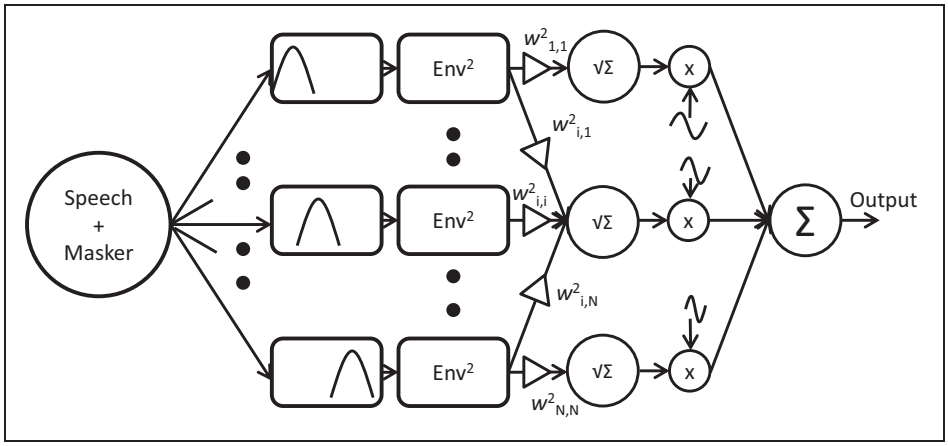

Figure 5. Schematic diagram of the algorithm used to simulate the effects of reduced spectral resolution in cochlear-implant users due to current spread. The temporal envelope is extracted from each frequency subband, and the squared weighted contributions from each subband are summed and then square-rooted to produce the final subband envelopes, which are then used to modulate the sinusoidal carriers before the output signal is created by summing the modulated carriers.

the reduced spectral resolution (cf., Figure 2, right panel). Second, the large benefit previously found for tone maskers was no longer present; in fact, there was no longer any significant difference between the three conditions, with no main effect of masker type, $F(2, 10) = 0.981$, $p = .408$, or interaction between masker type and signal-to-masker ratio, $F(4.4, 22) = 0.555$, $p = .712$. Again, as expected, the main effect of signal-to-masker ratio was highly significant, $F(4, 20) = 45.1$, $p < .001$, partial $\eta^2 = .9$. As shown by the red curve on the graph, the results in the three conditions were very similar to the performance of CI users in Experiment 1, averaged across the three conditions. Thus, a simulated reduction of spectral resolution, similar to that experienced by CI users, is sufficient to account fully for the large differences in performance between the normal-hearing listeners and the CI users and for the lack of effect of inherent noise fluctuations found for the CI users in the original experiment. An analysis of the modulation spectrum of the output of the vocoder simulations of spectral smearing and current spread shows how the relative modulation energy in the stimulus is reduced by the spectral smearing. This is illustrated in Figure 7, which shows the modulation spectrum of the vocoder in response to speech-shaped noise with and without spectral smearing, both for the individual 16 channels and for the sum of the 16 channels.

General Discussion

Overview of Results

Our results reveal dramatic qualitative differences between acoustic and electric hearing in understanding speech in different backgrounds: The speech perception ability of normal-hearing listeners was severely disrupted by the amplitude fluctuations inherent in noise, whereas the CI users' speech perception was not affected. The lack of effect of inherent fluctuations in CI users is not due to a lack of sensitivity to amplitude modulation and instead can be explained by the effects of reduced spectral resolution, presumably caused by interactions within the cochlea of the electrical signals from neighboring electrodes in the CI. Contrary to our current understanding (based on acoustic hearing) that speech perception in noise is limited primarily by the inherent fluctuations in the noise (Stone et al., 2012; Stone & Moore, 2014), speech perception in electric hearing seems to be limited primarily by the overall energy of the masker in each CI channel, independent of fluctuations.

Tone Versus Noise Carriers in Vocoders

There has been some debate about how best to simulate aspects of CI processing in normal-hearing listeners, with the most popular simulations involving noise- or tone-excited vocoders (e.g., Whitmal et al., 2007). Spectral spread has been simulated using noise-excited vocoders with variable slopes of the reconstruction filters to vary the spread of excitation from each channel (Bingabr, Espinoza-Varas, & Loizou, 2008; Fu & Nogaki, 2005). As illustrated by the present results, a disadvantage of using noise carriers is that the inherent temporal-envelope fluctuations of the vocoder noise bands may influence the results; this may explain why noise-excited vocoders generally produce poorer intelligibility than tone-excited vocoders under similar conditions (Whitmal et al., 2007). As shown here, and in an earlier study (Crew et al., 2012), it is also possible to simulate aspects of spectral (or current) spread using a tone-excited vocoder, by modulating each tone carrier with a weighted sum of temporal envelopes from adjacent channels. The similarity of the results from the CI users and the vocoded normal-hearing listeners (Figure 6) suggests that the vocoder was successful in simulating at least some aspects of CI processing and perception. Of course, discrete tones do not simulate the more continuous spread of excitation produced by actual electrodes in a CI. Nevertheless, they may still produce a *functionally* relevant simulation by ensuring that the independence of the information from each channel is limited by the summation of the envelope information prior to modulation of each carrier.

Speech in Noise: A Trade-Off Between Spectral Resolution and Noise-Envelope Fluctuations?

Although it is widely acknowledged that spectral resolution affects speech perception, especially in noise (Friesen, Shannon, Baskent, & Wang, 2001; Won, Drennan, & Rubinstein, 2007), not all studies have found strong correlations between measures of spectral resolution and speech perception in noise across individual CI users (Anderson, Oxenham, Nelson, & Nelson, 2012). The present results provide a new interpretation for the lack of a clear relationship: As spectral resolution decreases, speech perception becomes more difficult, but at the same time the detrimental effect of inherent noise fluctuations diminishes, due to the effects of across-channel temporal envelope summation. Thus, a form of trade-off occurs between spectral resolution and inherent temoral-envelope fluctuation depth. This insight suggests that noise maskers may not provide the best test of spectral resolution in clinical populations, and that tonal maskers may be more suitable for clinical tests because they maximize the effects of spectral resolution without incorporating any potentially confounding effects of temporal masker fluctuations. The present results also highlight the importance of ongoing efforts to improve the spectral resolution of CIs, through the use of

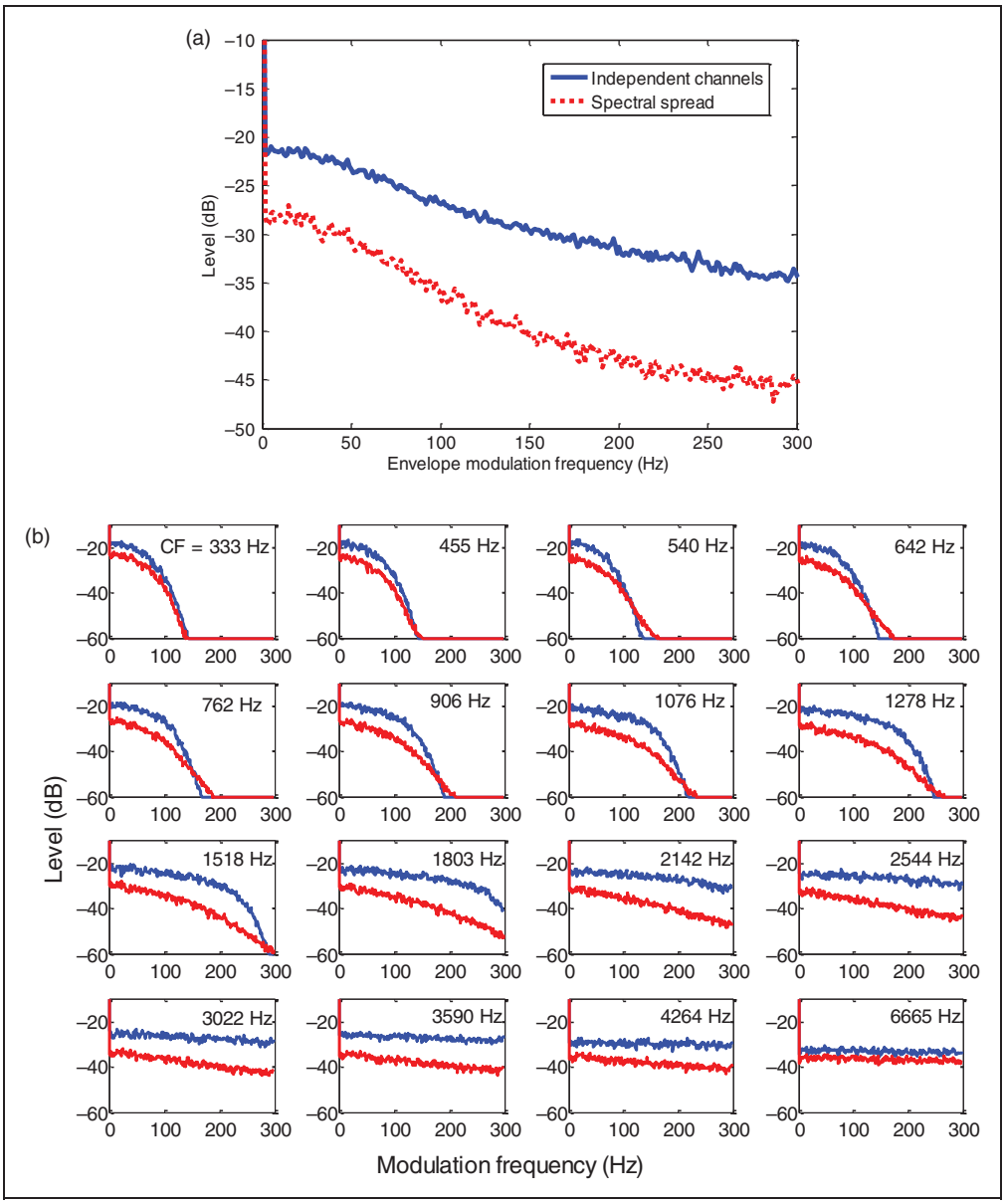

Figure 7. Modulation spectra for the Gaussian noise maskers, showing how modulation energy is reduced by summing to simulate poorer spectral resolution. (a) Modulation spectrum for the Gaussian noise, averaged across the 16 frequency subbands used in the experiments for the independent channels (blue solid line) and after simulation of reduced spectral resolution due to current spread (red dotted line). (b) Modulation spectra at the outputs of the individual frequency subbands, showing that the reduction in modulation energy is observed at all frequencies.

techniques such as focused stimulation (e.g., Bierer, 2007; Zhu, Tang, Zeng, Guan, & Ye, 2012).

Tone Maskers as a Diagnostic Tool

As mentioned in the discussion of Experiment 1, the use of pure-tone maskers accentuates the difference in performance between CI users and normal-hearing listeners, to the extent that equivalent performance can require differences in signal-to-masker ratio of around 30 dB (cf., open circles in the left panels of Figures 2 and 3). The large differences are probably due to a combination of the spectrally sparsity and flat temporal envelopes of the tone maskers, both of which can be exploited by normal-hearing listeners but not by CI users. This type of masker may therefore prove to be a useful diagnostic tool in the clinical assessment of CI users and in the evaluation of new algorithms and techniques designed to improve the spectral resolution of CIs. A diagnostic tool could use pure-tone maskers to probe spectral

resolution without the potentially confounding effects of inherent noise fluctuations.

Implications for Masking Release With Cochlear-Implant Users and Hearing-Impaired Listeners

Inserting temporal gaps or imposing slow amplitude fluctuations on an otherwise stationary noise masker can lead to dramatic improvements in speech intelligibility, known as *masking release* (Miller & Licklider, 1950). A long-standing puzzle has been why people with hearing loss or CIs typically show much less masking release than people with normal hearing (Bacon, Opie, & Montoya, 1998; Eisenberg, Dirks, & Bell, 1995; Festen & Plomp, 1990; Gregan, Nelson, & Oxenham, 2013; Nelson & Jin, 2004; Nelson, Jin, Carney, & Nelson, 2003; Peters, Moore, & Baer, 1998; Stickney, Zeng, Litovsky, & Assmann, 2004). Various hypotheses to explain the loss of masking release have been proposed and tested over the past two decades, including loss of audibility and dynamic range (Desloge, Reed, Braida, Perez, & Delhorne, 2010), poorer coding of temporal fine structure (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Qin & Oxenham, 2003), loss of cochlear amplification and compression (Gregan et al., 2013), and baseline signal-to-noise ratio at testing (Bernstein & Brungart, 2011). However, none of these proposals have fully explained the patterns of deficits exhibited by hearing-impaired listeners and CI users (Bacon et al., 1998; Freyman, Griffin, & Oxenham, 2012; Gregan et al., 2013; Oxenham & Simonson, 2009). Our findings provide a new perspective and a potential solution to the problem. It has already been shown that normal-hearing listeners do not exhibit masking release when the original masker has no inherent fluctuations (Stone et al., 2012). The explanation for this is that masking is normally produced primarily by the inherent fluctuations in the noise, so that when the inherent fluctuations are eliminated, masking is greatly reduced; therefore, when additional fluctuations are introduced through amplitude modulation, rather than produce masking release, the additional fluctuations actually act to mask the speech (Stone et al., 2011, 2012; Stone & Moore, 2014). Our new results suggest that CI users' perception of the fluctuations inherent in noise may be severely reduced (due to the loss of spectral resolution and the resulting smoothing of the temporal envelopes). If the noise is perceived as an effectively *steady* masker, then no masking release would be expected because the additional fluctuations will worsen, rather than improve, performance (Kwon & Turner, 2001; Stone & Moore, 2014).

Similar reasoning may explain why people with cochlear hearing loss also often exhibit reduced masking release. Cochlear hearing loss often results in reduced spectral resolution or frequency selectivity (Moore, 2007). The mechanisms of reduced spectral resolution are different in hearing-impaired listeners than in CI users because spectral smearing in acoustic hearing occurs before (rather than after) the extraction of the temporal envelope in the inner hair cells. Nevertheless, the perceptual effects may be similar. Reduced frequency selectivity, through a widening of the cochlear filter bandwidths, results in a *flattening* of the modulation spectrum and hence a reduction in the relative contribution of the lower modulation frequencies that are most responsible for masking speech (Jorgensen & Dau, 2011). This reduction in the relative energy of lower modulation frequencies may in turn result in reduced speech masking and hence reduced masking release when additional modulation is imposed.

Summary

Speech intelligibility via electric hearing is not affected by the inherent fluctuations in noise, in contrast to the results from normal acoustic hearing. This surprising outcome can be understood in terms of the reduced spectral resolution provided by current CIs. The results provide new insights into the long-standing puzzle of why CI users and other hearing-impaired individuals exhibit a reduced release from masking when additional fluctuations are imposed on a noise masker. The results also suggest that the use of tones to mask speech provides a highly sensitive measure of spectral resolution that could be of diagnostic benefit in the treatment of hearing loss and in the evaluation of new algorithms and devices.

Acknowledgments

We thank Edward Carney for programming assistance. We also thank the editor, Monita Chatterjee, and an anonymous reviewer for helpful comments on an earlier version of this paper.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grant R01 DC 012262 from the National Institutes of Health.

References

Anderson, E. S., Oxenham, A. J., Nelson, P. B., & Nelson, D. A. (2012). Assessing the role of spectral and intensity cues in spectral ripple detection and discrimination in cochlear-implant users. *The Journal of the Acoustical Society of America*, *132*(6), 3925–3934.

Bacon, S. P., & Grantham, D. W. (1989). Modulation masking: Effects of modulation frequency, depth, and phase. *The Journal of the Acoustical Society of America*, *85*(6), 2575–2580.

Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech Language and Hearing Research*, *41*(3), 549–563.

Bacon, S. P., & Viemeister, N. F. (1985). Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects. *Audiology*, *24*, 117–134.

Bernstein, J. G., & Brungart, D. S. (2011). Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio. *The Journal of the Acoustical Society of America*, *130*(1), 473.

Bierer, J. A. (2007). Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration. *The Journal of the Acoustical Society of America*, *121*(3), 1642–1653.

Bingabr, M., Espinoza-Varas, B., & Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants. *Hearing Research*, *241*(1–2), 73–79.

Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics Speech and Signal Processing*, *27*(2), 113–120.

Chatterjee, M., & Oberzut, C. (2011). Detection and rate discrimination of amplitude modulation in electrical hearing. *The Journal of the Acoustical Society of America*, *130*(3), 1567–1580.

Crew, J. D., Galvin, J. J. 3rd, & Fu, Q. J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *The Journal of the Acoustical Society of America*, *132*(5), EL429–EL435.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers. *The Journal of the Acoustical Society of America*, *102*(5), 2892–2905.

Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., & Delhorne, L. A. (2010). Speech reception by listeners with real and simulated hearing impairment: Effects of continuous and interrupted noise. *The Journal of the Acoustical Society of America*, *128*(1), 342–359.

Dorman, M. F., Loizou, P. C., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *The Journal of the Acoustical Society of America*, *104*(6), 3583–3585.

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *Journal of the Acoustical Society of America*, *95*(5), 2670–2680.

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, *95*(2), 1053–1064.

Dubbelboer, F., & Houtgast, T. (2008). The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. *The Journal of the Acoustical Society of America*, *124*(6), 3937–3946.

Eisenberg, L. S., Dirks, D. D., & Bell, T. S. (1995). Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing. *Journal of Speech and Hearing Research*, *38*, 222–233.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736.

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America*, *19*, 90–119.

Freyman, R. L., Griffin, A. M., & Oxenham, A. J. (2012). Intelligibility of whispered speech in stationary and modulated noise maskers. *The Journal of the Acoustical Society of America*, *132*(4), 2514–2523.

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. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163.

Fu, Q. J. (2002). Temporal processing and speech recognition in cochlear implant users. *Neuroreport*, *13*(13), 1635–1639.

Fu, Q. J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, *6*(1), 19–27.

George, E. L., Festen, J. M., & Houtgast, T. (2008). The combined effects of reverberation and nonstationary noise on sentence intelligibility. *The Journal of the Acoustical Society of America*, *124*(2), 1269–1277.

Gnansia, D., Lazard, D. S., Leger, A. C., Fugain, C., Lancelin, D., Meyer, B., . . . , Lorenzi, C. (2014). Role of slow temporal modulations in speech identification for cochlear implant users. *International Journal of Audiology*, *53*(1), 48–54.

Gregan, M. J., Nelson, P. B., & Oxenham, A. J. (2013). Behavioral measures of cochlear compression and temporal resolution as predictors of speech masking release in hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *134*(4), 2895–2912.

Houtgast, T. (1989). Frequency selectivity in amplitude-modulation detection. *The Journal of the Acoustical Society of America*, *85*(4), 1676–1680.

Hu, J., & Beaulieu, N. C. (2005). Accurate simple closed-form approximations to Rayleigh sum distributions and densities. *IEEE Communications Letters*, *9*, 109–111.

Humes, L. E., Wilson, D. L., Barlow, N. N., & Garner, C. (2002). Changes in hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *Journal of Speech Language and Hearing Research*, *45*(4), 772–782.

Jorgensen, S., & Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, *130*(3), 1475–1487.

Jorgensen, S., Ewert, S. D., & Dau, T. (2013). A multi-resolution envelope-power based model for speech intelligibility. *The Journal of the Acoustical Society of America*, *134*(1), 436–446.

Kohlrausch, A., Fassel, R., & Dau, T. (2000). The influence of carrier level and frequency on modulation and

beat-detection thresholds for sinusoidal carriers. *The Journal of the Acoustical Society of America*, *108*(2), 723–734.

Kryter, K. D. (1962). Methods for the calculation and use of the Articulation Index. *The Journal of the Acoustical Society of America*, *34*, 467–477.

Kwon, B. J., & Turner, C. W. (2001). Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference? *The Journal of the Acoustical Society of America*, *110*(2), 1130–1140.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2), 467–477.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences U S A*, *103*(49), 18866–18869.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, *22*, 167–173.

Moore, B. C. J. (2007). *Cochlear hearing loss: Physiological, psychological and technical issues*. Chichester, England: Wiley.

Nelson, D. A., Kreft, H. A., Anderson, E. S., & Donaldson, G. S. (2011). Spatial tuning curves from apical, middle, and basal electrodes in cochlear implant users. *The Journal of the Acoustical Society of America*, *129*(6), 3916–3933.

Nelson, P. B., & Jin, S. H. (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, *115*(5 Pt. 1), 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. *The Journal of the Acoustical Society of America*, *113*(2), 961–968.

Oxenham, A. J., & Simonson, A. M. (2009). Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference. *The Journal of the Acoustical Society of America*, *125*(1), 457–468.

Peters, R. W., Moore, B. C. J., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, *103*(1), 577–587.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, *114*(1), 446–454.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *336*(1278), 367–373.

Shannon, R. V. (1992). Temporal modulation transfer functions in patients with cochlear implants. *Journal of the Acoustical Society of America*, *91*(4 Pt 1), 2156–2164.

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

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*, 87–90.

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C.,. . . ,Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear Hear*, *33*(1), 112–117.

Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, *116*(2), 1081–1091.

Stone, M. A., Fullgrabe, C., Mackinnon, R. C., & Moore, B. C. J. (2011). The importance for speech intelligibility of random fluctuations in "steady" background noise. *The Journal of the Acoustical Society of America*, *130*(5), 2874–2881.

Stone, M. A., Fullgrabe, C., & Moore, B. C. J. (2008). Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region. *The Journal of the Acoustical Society of America*, *124*(4), 2272–2282.

Stone, M. A., Fullgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, *132*(1), 317–326.

Stone, M. A., & Moore, B. C. J. (2014). On the near non-existence of "pure" energetic masking release for speech. *The Journal of the Acoustical Society of America*, *135*(4), 1967–1977.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*(3), 455–462.

Whitmal, N. A., Poissant, S. F., Freyman, R. L., & Helfer, K. S. (2007). Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *The Journal of the Acoustical Society of America*, *122*(4), 2376–2388.

Won, J. H., Drennan, W. R., Nie, K., Jameyson, E. M., & Rubinstein, J. T. (2011). Acoustic temporal modulation detection and speech perception in cochlear implant listeners. *The Journal of the Acoustical Society of America*, *130*(1), 376–388.

Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, *8*(3), 384–392.

Zeng, F. G. (2004). Trends in cochlear implants. *Trends in Amplification*, *8*(1), 1–34.

Zhu, Z., Tang, Q., Zeng, F. G., Guan, T., & Ye, D. (2012). Cochlear-implant spatial selectivity with monopolar, bipolar and tripolar stimulation. *Hearing Research*, *283*(1–2), 45–58.