Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners

Leonid M. Litvak[a)]
Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California, 91342

Anthony J. Spahr
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287

Aniket A. Saoji
Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California, 91342

Gene Y. Fridman
Department of Biomedical Engineering, University of California Los Angeles, Los Angeles, California, 90095

(Received 11 December 2006; revised 16 May 2007; accepted 18 May 2007)

Spectral resolution has been reported to be closely related to vowel and consonant recognition in cochlear implant (CI) listeners. One measure of spectral resolution is spectral modulation threshold (SMT), which is defined as the smallest detectable spectral contrast in the spectral ripple stimulus. SMT may be determined by the activation pattern associated with electrical stimulation. In the present study, broad activation patterns were simulated using a multi-band vocoder to determine if similar impairments in speech understanding scores could be produced in normal-hearing listeners. Tokens were first decomposed into 15 logarithmically spaced bands and then re-synthesized by multiplying the envelope of each band by matched filtered noise. Various amounts of current spread were simulated by adjusting the drop-off of the noise spectrum away from the peak (40–5 dB/octave). The average SMT (0.25 and 0.5 cycles/octave) increased from 6.3 to 22.5 dB, while average vowel identification scores dropped from 86% to 19% and consonant identification scores dropped from 93% to 59%. In each condition, the impairments in speech understanding were generally similar to those found in CI listeners with similar SMTs, suggesting that variability in spread of neural activation largely accounts for the variability in speech perception of CI listeners. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749413]

PACS number(s): 43.64.Me, 43.71.An, 43.71.Es, 43.71.Ky [AJO] Pages: 982–991

I. INTRODUCTION

Although the average outcomes of cochlear implant (CI) procedures have improved over the past decade, large variability in the ability to understand speech by CI listeners is commonly reported (e.g., Firszt *et al.*, 2004). The variability in outcomes is not adequately explained by pre-operative factors such as age and duration of hearing loss. Although several hypotheses have been proposed in order to explain the variability among the patient population, one likely possibility is that differences in performance may be accounted for by differences in spatial selectivity in CI listeners. The evidence for this hypothesis is twofold. First, several psychophysical measures thought to be related to spatial selectivity have been found to correlate to speech perception in CI listeners (Eddington *et al.*, 1997a; Henry and Turner, 2003; Henry *et al.*, 2005; Saoji *et al.*, 2005). Second, normal-hearing (NH) subjects listening to simulations of reduced spectral resolution exhibit a range of performance that matches that observed in best CI patients (Shannon *et al.*, 1995; Friesen *et al.*, 2001). The goal of the present work is to determine whether the "spectral selectivity" hypothesis is sufficient to entirely account for the range of speech performance observed in CI listeners. In particular, the performances of NH listeners on a nonspeech spectral resolution task are matched with those of CI listeners, by introducing varying amounts of spectral smearing (Baer *et al.*, 1993; Baer and Moore, 1994) for the NH listeners. The performance of CI listeners on speech identification tasks is then compared to that of NH listeners, with respect to their performance on the nonspeech spectral resolution task.

Several metrics have been proposed as measures of spectral selectivity, including (1) forward masking patterns, which may be assessed either psychophysically (Boex *et al.*, 2003; Cohen *et al.*, 2003), or using evoked potentials (Abbas *et al.*, 2004), (2) simultaneous threshold interaction (Eddington *et al.*, 1997b), or (3) spectral shape perception (Henry and Turner, 2003; Henry *et al.*, 2005). All of the psychophysical measures have been reported to correlate in varying degrees to speech perception, with spectral shape perception measures producing the best correlations, possibly because those measures assess spectral resolution across the whole cochlea. In the spectral shape task, subjects are asked to discriminate between locations of spectral peaks. For example, Henry *et al.* (2005) reported on a task where the subjects are

[a)]Author to whom correspondence should be addressed. Electronic mail: leonidl@advancedBionics.com

required to discriminate between two spectra, each of which resembles a full-wave rectified sinusoid in the frequency domain. Each subject's spectral resolution was quantified by the highest ripple frequency (in cycles/octave) with 30 dB contrast that the subject could identify as different from one where the spectral peaks versus valleys have been reversed in frequency location. More recently, Saoji *et al.* (2005) reported on spectral modulation transfer functions of cochlear implant recipients using methods that are similar to those previously utilized in normal-hearing listeners (Bernstein and Green, 1987). In that study, the authors quantified the necessary peak-to-valley contrast to differentiate between a spectral ripple and white noise as a function of spectral ripple frequency. They showed that spectral modulation thresholds (SMTs) corresponding to the lowest spectral ripple frequencies (0.25 and 0.5 cycles/octave) best related to speech recognition. Because ability to understand speech is of greatest relevance to the present study, SMT at these ripple frequencies was chosen as the measure of spectral resolution.

Studies of acoustic hearing have demonstrated that speech recognition of NH listeners is decreased with spectral smearing (Baer *et al.*, 1993; Baer and Moore, 1994; Shannon *et al.*, 1995; Dorman *et al.*, 1997; Friesen *et al.*, 2001). Several studies [e.g. (Shannon *et al.*, 1995; Dorman *et al.*, 1997; Friesen *et al.*, 2001)] simulated performance of CI patients by exposing NH listeners to "noise vocoders," which reduced spectral information into a limited number of "channels." Friesen *et al.* (2001) demonstrated that for 7–8 channels, NH listeners using an equivalent number of channels performed better than the average CI performer, but comparably to the best CI performers. However, it is unclear from Friesen *et al.* whether performance of poorer listeners relates specifically to loss of spectral resolution in CI listeners. Fu and Nogaki (2005) showed that performance of NH listeners can be reduced further by spectrally smearing the output of the vocoder bands by using overlapping noise carriers. However, unlike in the present study, no attempt was made to match the degree of smearing to specific CI subjects.

II. METHODS

A. Subjects

Ten normal hearing subjects ranging in age from 22 to 26 years participated in these experiments. Each subject had normal hearing based upon pure-tone thresholds (<20 dB Hearing Level (HL) from 250 to 8000 Hz, ANSI, 1996) and screening tympanograms (*Y*, 226 Hz). All studies have been approved by a private Institutional Review Board (IRB), as well as by the institutional IRB at the Arizona State University.

The speech perception and SMT data for CI subjects reported here are taken from Saoji *et al.* (2005) and will be published in a separate article. Twenty five Advanced Bionics CII or HR/90k cochlear implants (ranging in age from 38 to 65 years) listeners with varying speech perception abilities participated in that experiment. All subjects had at least one year of experience with their cochlear implants.

B. Stimuli

All stimuli were generated using MATLAB software (Mathworks, 2006). The stimuli were generated in the frequency domain assuming a sampling rate of 44,100 Hz. First, the desired spectral shape was generated using the equation

$$|F(f)| = \begin{cases} 10^{\frac{C}{2}\sin(2\pi(\log_2(f/350))\cdot f_c+\theta_0)/20} & 350 < f < 5600 \\ 0 & otherwise \end{cases}, \quad (1)$$

where $F(f)$ is the amplitude of a bin with center frequency f Hz, f_c is the spectral modulation frequency (in cycles/octave), and θ_0 is the starting phase. The desired noise band was synthesized by adding a random phase to each bin, and taking an inverse Fourier transform. The flat noise stimuli were generated using a similar technique, except that spectral contrast C was set to 0. The amplitude of each stimulus was adjusted to an overall level of 60 dB sound pressure level (SPL). Independent noise stimuli were presented on each observation interval. The stimulus duration was 400 ms.

Speech understanding was assessed using a vowel and consonant identification task. Vowel stimuli consisted of 13 vowels created with the use of KLATT software (Klatt, 1980) in /bVt/ format ("bait, bart, bat, beet, bert, bet, bit, bite, boat, boot, bought, bout, but"). The vowels were brief (90 ms) and of equal duration so that vowel length would not be a cue to identity (Dorman *et al.*, 1989). The stimuli for the tests of consonant identification were 16 male-voice consonants in the /aCa/ context, originally taken from the Iowa laser video disk (Tyler *et al.*, 1986).

C. Spectral modulation thresholds

Normal-hearing listeners were tested in a double-walled sound treated room using Sennheiser HD 25-SP1 circumaural headphones and all stimuli were presented at 60 dB SPL.

As reported by Saoji *et al.* (2005), the CI listeners used their everyday program in all test conditions. The stimuli were output from a standard PC to an Audiophile soundcard. The output of the soundcard was fed to the body worn Platinum Series Processor through the Advanced Bionics Direct Connect® system. Sound card output was attenuated using an inline attenuator such that for all stimuli, the electric input to the DirectConnect® system was equivalent to a 60 dB SPL acoustic input to the microphone of the speech processor. A cued two interval, two-alternative, forced choice procedure was used to collect data. Prior to data collection subjects received a few sample trials for any new condition to familiarize them with the stimuli. On each trial, the three observation intervals were separated by 400 ms silent intervals. In the first interval the standard stimulus was always presented. This cuing or reminder interval is helpful in cases where listeners can hear a difference between the signal and the standard stimulus but cannot identify which is which. The standard stimulus had a flat spectrum with bandwidth extending from 350 to 5600 Hz. The signal and the second standard were randomly presented in the other two intervals. Thresholds were estimated using an adaptive psychophysical

procedure employing 60 trials. The signal contrast level was reduced after three consecutive correct responses and increased after a single incorrect response. Initially the contrast was varied in a step size of 2 dB, which was reduced to 0.5 dB after three reversals in the adaptive track (Levitt, 1971). Threshold for the run was computed as the average modulation depth corresponding to the last even number of reversals, excluding the first three. The equilibrium point of such a procedure is 79.4% correct. Using the above procedure, modulation detection thresholds were obtained for the modulation frequencies of 0.25 and 0.5 cycles/octave. A threshold was defined as the average of three 60 trial runs. The average of the thresholds for the two modulation frequencies was used as the measure of spectral resolution, as this measure was found to best correlate to consonant and vowel recognition (Saoji *et al.*, 2005).

D. Vowel and consonant recognition

During both the vowel and consonant identification tasks, listeners completed a practice session, in which they heard each vowel presented twice while the text representation of the token was visually displayed on the computer screen. Subjects then completed two repetitions of the test procedure, with feedback, as a final practice condition. In the test condition, vowel identification performance was measured in two blocks of 78 trials in which each of the 13 vowels was presented six times in random order. Thus, vowel identification and the resulting confusion matrix for each subject were based on 12 presentations of each vowel stimulus. Likewise, consonant identification performance was measured in two blocks of 80 trials in which each of the 16 consonants was presented five times in random order. Consonant identification and the resulting confusion matrix for each subject were based on a total of ten presentations per consonant stimulus. The vowel and consonants were presented at an overall level of 60 dB SPL and the overall level was randomized by 3 dB (±1.5 dB) in 0.5 dB steps. The randomization would discourage the listeners from using loudness cues while performing the vowel and consonant recognition task.

As reported by Saoji *et al.* (2005), due to time constraints, consonant identification was not measured in one CI subject. In addition, a loss of data occurred on the experimental computer, whereby the confusion matrices were not stored for three CI subjects on the vowel task, and for four CI subjects on the consonant task.

E. Vocoder simulations

Vocoder simulations utilized in this study were designed to model both the processing typically performed in a cochlear implant, and spread of excitation that may occur in electrically stimulated cochlea. Each token was digitally sampled at 17,400 Hz. Short-time Fourier transform was computed with resolution of 256 bins, and temporal overlap of 192 samples (Oppenheim and Schafer, 1975). Next, individual bins were grouped into 15 nonoverlapping, logarithmically spaced analysis channels. The envelope of each channel was computed on a frame-by-frame basis by computing the square root of the total energy in the channel. The energy computation implies an implicit envelope detector with a low-pass filter whose cutoff equals the bandwidth of each bin, or 68 Hz. The output of each channel was used to modulate a noise band. The noise band was similarly synthesized in the frequency domain. The center frequency of the noise band was identical to the center frequency of the corresponding analysis channel. The rate of the drop-off of the noise spectrum away from the center frequency was varied from 5 to 40 dB/octave, to simulate various amounts of spread of excitation that may occur in an electrically stimulated cochlea. The desired time-frequency output pattern for each channel was computed by multiplying its instantaneous energy by the corresponding spectral envelope of the noise band. The time-spectral patterns corresponding to each channel were then added to compute the total time-spectral pattern. The overall bandwidth of the signal was limited to 8700 Hz. Finally, temporal output wave form was computed by first adding random phase to each bin, and then computing an inverse short-time Fourier transform (Oppenheim and Schafer, 1975).

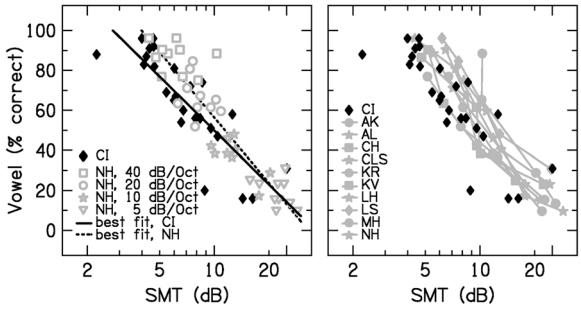

FIG. 1. (Color online) Vowel speech performance of the CI and NH listeners plotted against their respective SMT scores. Panels A and B contain the same data. In each panel, the data from CI listeners are indicated with black diamonds. Panel A shows the performance of the NH listeners with open symbols, where each symbol corresponds to the noise slope of the vocoder used in the simulation. Panel B shows that the individual performance for each NH listener follows the same trend as the average data. The lines in panel A are the best-fit lines for the CI patients and the NH listeners.

Ten normal-hearing (NH) listeners participated in this study. Each NH listener listened to four different vocoder simulations, which differed only in the slopes of the noise bands. Based on a small pilot study, slopes of 5, 10, 20, and 40 dB/octave were chosen. For each simulation condition, the SMT at 0.25 and 0.5 cycles/octave, as well as performance on vowel and consonant recognition, was measured.

III. RESULTS

A. Vowel recognition

Figure 1 panels show the results of vowel recognition as a function of SMT for the CI patients and NH listeners. Panels A and B demonstrate different aspects of the same data. The data collected from the CI listeners are shown in black diamonds in all two panels.

Large CI subject variability was observed in both vowel recognition and SMT tasks. As in previously reported studies (Henry *et al.*, 2005; Saoji *et al.*, 2005), the SMT is strongly correlated with vowel recognition ($r=-0.84$). For the NH

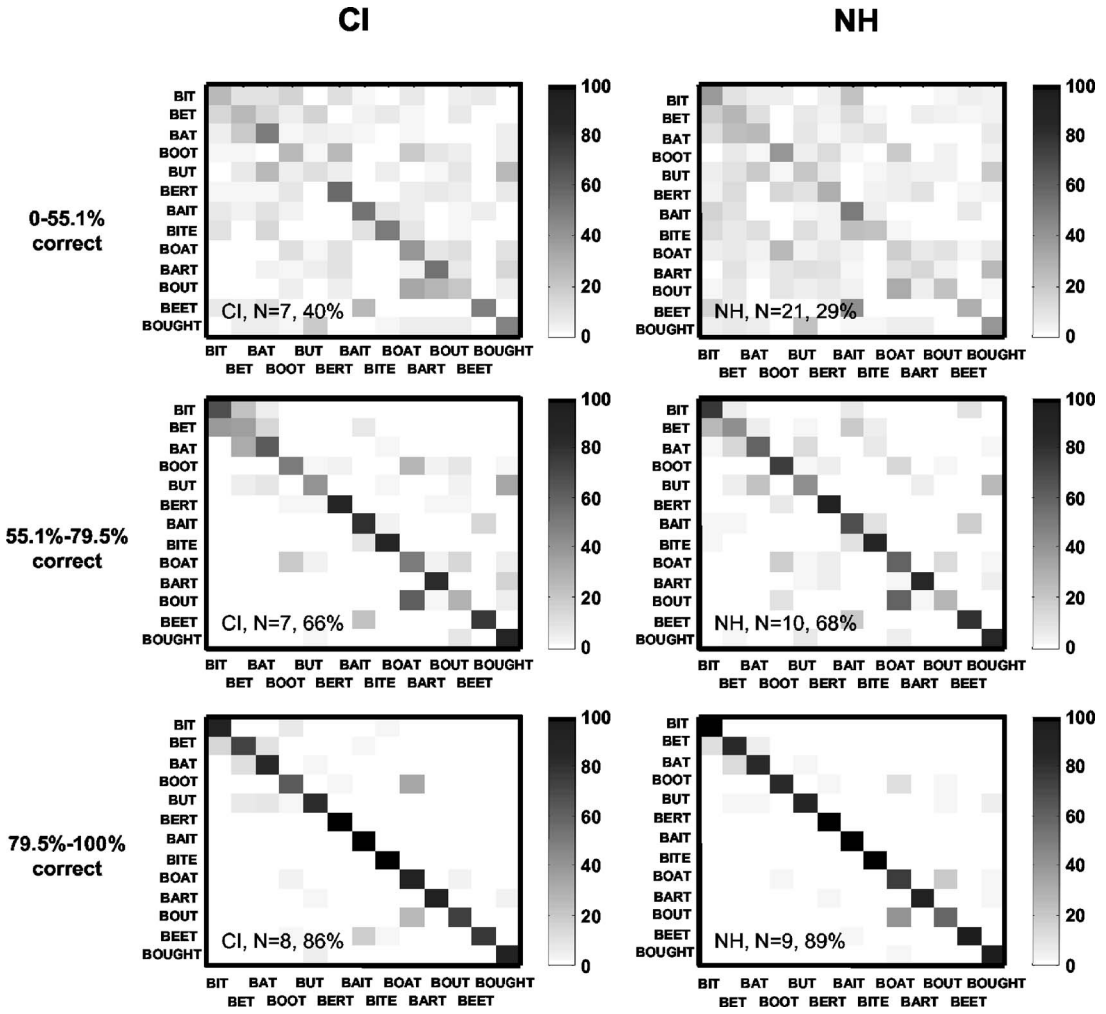

FIG. 2. Patterns of errors which occurred during the vowel speech tests. Confusion matrices for the vowel speech scores for the CI patients are presented in the left panels and those for the NH listeners in the right panels. The subjects were separated based on their vowel test scores (indicated in percent correct on the left of the matrices), so that each group contained the data from approximately the same number of CI subjects. The *y* axis corresponds to the stimulus, and the *x* axis corresponds to the response.

listeners using vocoder simulations, the vowel scores decreased as the noise-band slopes were broadened. The average as well as individual SMTs also tended to deteriorate from 6.3 dB for 40 dB/octave (narrow noise band) condition to 22.5 dB for 5 dB/octave (broad noise band) condition. Thus, use of progressively shallower noise-band slopes had an effect of decreasing spectral resolution abilities of NH listeners. The range of SMTs observed in NH subjects listening to simulations was similar to the range observed in CI listeners (2.25–25 dB). In addition, as shown in panel B, the individual scores on the vowel recognition task decreased with shallower noise-band slopes. Comparing the NH data to the corresponding data for CI recipients, a good quantitative agreement is reached in that the NH data overlays the CI data. For quantitative comparison, panel A shows the data from all of the simulation conditions, pooled together so that vowel identification abilities of CI patients can be directly compared to that of NH subjects under condition of similar spectral resolution. The pooling of the data is justified because the relationship between the SMT and vowel recognition within each group (different open symbols in Panel A) appears to be similar to the relationship across groups. The lines indicate the best linear fits of the CI and NH data. Although one can observe slight differences in the fits, both the differences between the two lines in the range from 4 to 30 dB are not statistically different [$p=0.57$, permutation test (Good, 1994)]. Thus, both statistical analysis and visual inspection suggest that a modified vocoder that models spectral resolution of cochlear implant listeners also models vowel identification performance.

Figure 2 compares the pattern of errors made by NH subjects listening in various simulation conditions and matched CI subjects. Each of the rows of panels shows the average confusion matrices for CI subjects (left) and NH listeners (right) with overall performance of 0–55.1% correct (top), 55.1–79.5% correct (middle), and 79.5–100% correct (bottom). These performance ranges were chosen such that there is an approximately equal number of CI subjects in each range. The data for NH group were combined across simulation conditions. The similarity in scores was chosen as the basis for assigning subjects to a group so that the results in this section can be readily compared to those for consonant recognition task, where the equivalent performance is not achieved at the same spectral resolution between the two

TABLE I. Correlation coefficients between the confusion matrices for the vowel data computed either with or without the main diagonal. The probability that the two populations have the same confusion patterns is indicated by the p level, and was computed using a nonparametric method described in the text.

	Number		Correlation between NH and CI data	
Category	NH group	CI group	With diagonal	Without diagonal
0.0–55.1	(N=21)	(N=7)	0.772(P=26.8%)	0.695(P=22.0%)
55.1–79.5	(N=10)	(N=7)	0.971(P=57.9%)	0.858(P=45.6%)
79.5–100.0	(N=9)	(N=8)	0.985(P=12.9%)	0.716(P=4.4%)

groups. The results for vowels would be similar if spectral resolution would be chosen as the basis for the grouping.

Visual inspection of the two confusion patterns in each row of Fig. 2 indicates many similarities, but also some differences, between the two plots. Table I shows the correlations between the confusion matrices for NH and CI listeners. The correlation coefficient between the two matrices (computed by treating each matrix as a set of ordered numbers) is higher than 0.75 for all groups, and is especially high for the two higher-performing groups (0.97 and 0.99, respectively). The correlations remain reasonably high (close to or above 0.7) even if the entries on the main diagonal of both the NH and CI matrices are set to zero prior to computing the correlation. The latter manipulation is of interest because only the errors contribute to the off diagonal entries; hence correlation without the off diagonal allows comparison of the error pattern almost independently of the overall score.

A version of a significance test based on the bootstrap method (Efron and Tibshirani, 1993) was undertaken to determine whether the differences between the corresponding matrices are due to the variability seen between subjects and between tests, or whether the variability reflects true differences in the errors made by CI and NH listeners. The null hypothesis was that both sets of confusion matrices (i.e., those from CI and NH listeners) correspond to the same population. Under the null hypothesis, expected range of correlations between the average NH and CI matrices could be computed by mixing up the individual data, and re-dividing the subjects arbitrarily into the two groups corresponding to the original NH and CI groups, and computing the correlation between the average matrices for the two new groups. The "re-drawing" procedure was repeated 1000 times, and the p values were computed as proportions of the time that the correlations obtained from the bootstrap procedure were higher than the observed correlation between NH and CI matrices. Note that if the observed correlation is significantly lower than the distributions under the null hypothesis, then the null hypothesis is invalidated, which means that the two sets of matrices come from significantly different populations. For all of the three performance groups, the correlation coefficients observed in the original data were not significantly different from those observed under the null hypothesis ($P>4.4\%$), suggesting that the two populations are not significantly different.

B. Consonant recognition

The panels of Fig. 3 plot consonant recognition as a function of SMT. As in Fig. 1, the panels show different aspects of the same data. As in Fig. 1, the data collected from the CI listeners are shown in black diamonds in all three panels.

In panel A, the data for NH listeners are shown with open symbols. The symbol shapes correspond to the particular noise-band slopes applied in the vocoder simulations. Panel B shows performance of individual NH subjects. As for vowels, SMT was strongly correlated with consonant recognition of both CI subjects and NH subjects listening to simulations ($r=-0.82$ and -0.88, respectively). However, as compared to the vowel scores shown in Fig. 1, the consonant scores dropped off less with decreased spectral resolution (16% per doubling versus 26% per doubling). The lower dependence of consonant scores on spectral resolution is consistent with the observation that consonant recognition is less dependent on spectral cues as compared to vowel recognition. As with the vowel identification task, the highest scores and the lowest SMTs correspond to the condition with the largest noise-band slope (40 dB/octave). Panel B shows similar performance trends of individual NH subjects for each condition.

As with the vowel scores, the consonant scores of NH listeners decreased with shallower noise-band slopes. However, whereas the vowel data for CI listeners matched closely to the NH listeners with the same SMT, this was not the case for the consonant data. In particular, at the same spectral resolution, performance on the consonant task was generally higher for the NH vocoder listeners. The difference between the two is quantified in panel A where the data from NH listeners are pooled across different simulation conditions, and the best-fit lines were fit to both sets of data. Permutation test indicated significant ($p=0.006$) differences between the two line fits in the region of SMT from 3.5 to 30 dB (Good, 1994). The slope of the best-fit line was not different in the

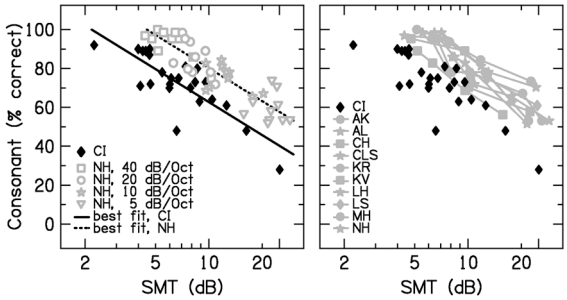

FIG. 3. (Color online) Consonant speech performance of the CI and NH listeners plotted against their respective SMT scores. The format of the plots is identical to that in Fig. 1.

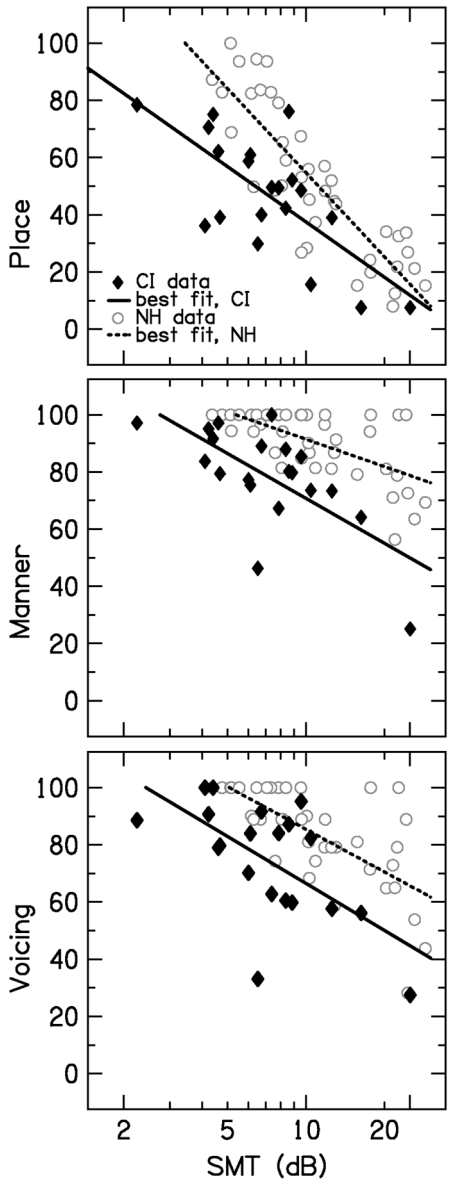

FIG. 4. (Color online) Identification abilities of place, manner, and voicing cues versus the SMT for CI subjects and NH listeners listening through vocoder simulations. The data for NH listeners is pooled across simulation conditions. The best-fit lines show the similarity in the data trends for each group.

two conditions ($p=0.65$), indicating that the degree of degradation of the consonant scores of CI and NH listeners with equivalent decrease of SMT was similar. As shown in Fig. 4, similar difference was observed in all classical production-based features of voicing, manner, and place of articulation (Miller and Nicely, 1955). Although NH listeners tended to perform better on all features, the differences across individual features were not significant ($p>0.09$).

Figure 5 and Table II repeat the error pattern analysis for the consonant recognition. The subjects were split into three groups based on the consonant test performance. The statistical analysis for the correlations between the matrices obtained for the CI listeners and those obtained for the NH listeners is described in the previous section. The correlations between the NH and CI confusion matrices were comparable to those obtained on vowel matrix comparison. However, the p-value analysis of the confidence of the correlation measures revealed that these correlation measures were not highly significant. This statistical analysis suggests that, despite the high degree of similarity, there were also significant differences between the confusion matrices of the NH compared to the CI listeners.

IV. DISCUSSION AND CONCLUSIONS

A. Vowels versus consonants

When matched in spectral resolution abilities, performance of NH subjects listening through vocoder simulations was similar to that of CI patients. The match between the two populations was greatest for the vowel recognition task, where the average performance, variability in performance, and confusion patterns were all similar between the two groups. In addition, the decrease in the consonant scores observed in NH listeners as a function of decreased spectral resolution was similar to that of the CI patients as reported by Saoji *et al.* (2005). However, even at equivalent spectral resolution abilities, somewhat lower overall scores were observed for the CI patients in the consonant recognition task. While many studies have reported that CI listeners perform more poorly than NH subjects on spectral tasks, these results suggest that CI patients have also deficits in perception of nonspectral (amplitude/temporal) cues as compared to NH listeners. Such deficits should only affect the consonant scores because the "synthetic" vowel recognition task relies almost exclusively on spectral cues. For example, Fu, 2002 found the highest correlation between consonants and temporal modulation and a smaller correlation between vowels and temporal modulation. Alternatively, these temporal deficits may be partially explained by the age differences between the NH and CI groups (Ohde and Abou-Khalil, 2001).

While this observation would suggest that CI patients have deficits in the temporal/amplitude domain as compared with NH listeners, this assertion is not fully supported by the data in Fig. 4. As this figure indicates, roughly equivalent differences between the NH and CI listeners are observed for features with primarily spectral cues (such as place), and primarily temporal/amplitude cues (such as voicing or manner). More data will be required to ascertain whether deviations in the features which are primarily based on the temporal and amplitude cues are more significant than deviations of the features with the spectral cues.

B. Factors affecting speech perception abilities of CI patients

The ability to recognize speech varies substantially across CI users. Several hypotheses have been advanced as to the underlying causes of this variability. First, it has been suggested that variability may be due to variability in electrode placement relative to the natural frequency-to-place alignment in the cochlea. While such frequency-to-place misalignment might exist in some subjects, and while acute changes to the frequency-to-place alignment in the vocoder simulations and normal-hearing listeners can lead to large decreases in performance (for a recent review see Baskent

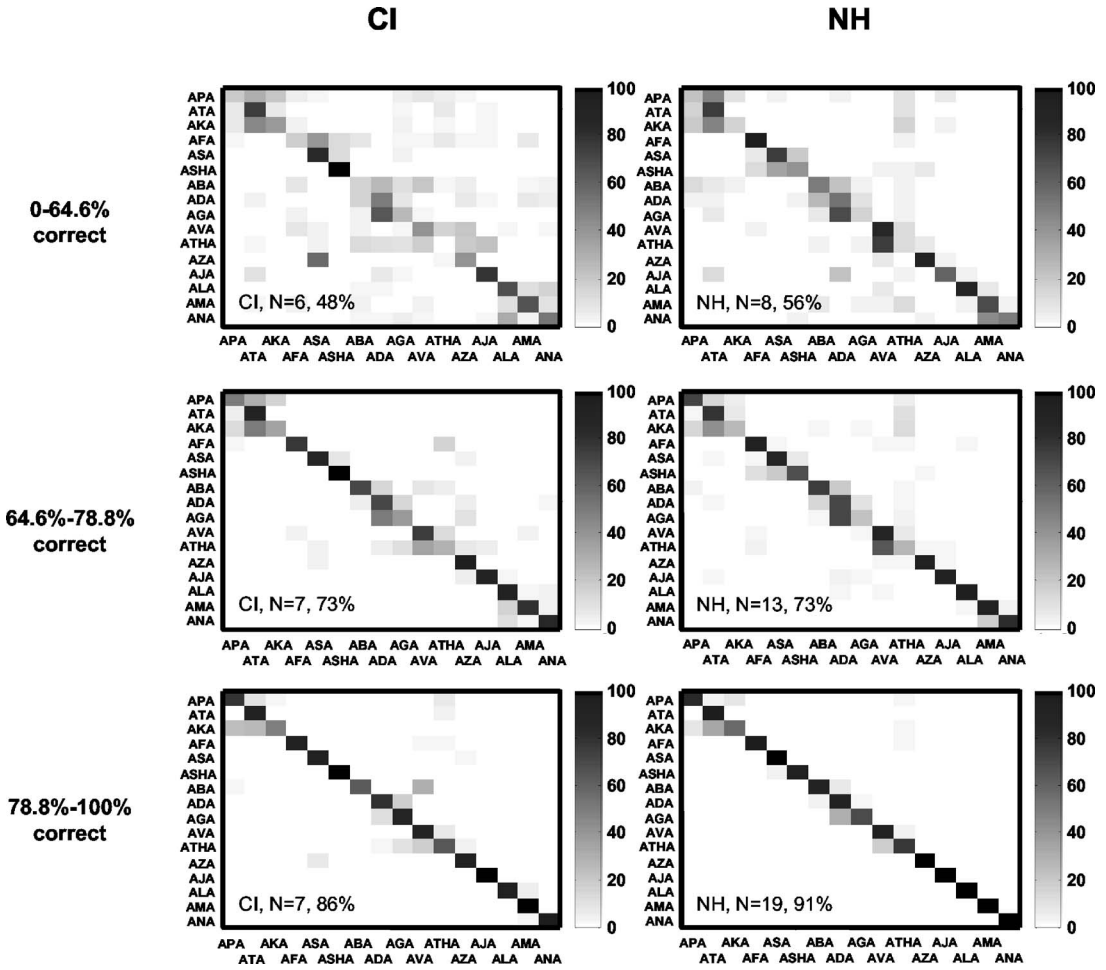

FIG. 5. Patterns of errors which occurred during the consonant speech tests. Confusion matrices for the consonant speech scores for the CI patients are presented in the left panels and those for the NH listeners in the right panels. The subjects were separated based on their consonant test scores (indicated in percent correct on the left of the matrices), so that each group contained the data from approximately the same number of CI subjects.

and Shannon, 2003), several reports suggest that both CI as well as NH listeners are able to partially adapt to such misalignment over time (Rosen *et al.*, 1999; Harnsberger *et al.*, 2001; Fu *et al.*, 2002). For example, Harnsberger *et al.* (2001) examined perceptual "vowel spaces" of synthetic vowels which differed only by first two formants. They found that with one exception, there were no systematic shifts observed in the perceptual vowel spaces of CI listeners relative to those established for NH subjects. Since the subjects that participated in the present study had substantial experience with their CI (on average more than 2 years), it is unlikely that variability observed in speech scores can be accounted by such frequency misalignments.

CI subjects also differ in their ability to perceive temporal modulations that are essential for proper perception of speech. Perceptions of temporal modulations (above 400 Hz) have been shown to not be correlated to speech scores (Cazals *et al.*, 1994). However, several investigators reported strong correlation between measures of low frequency ($<$400 Hz) temporal modulation and speech recognition tasks (Cazals *et al.*, 1994; Fu, 2002), suggesting a causal relationship between the two measures. Low frequency information provides periodicity and envelope cues, indicative of voicing and manner, respectively (Rosen, 1992). The relationship between low frequencies and manner and voicing is consistent with the results obtained in (Fu, 2002) that indicates a stronger correlation between temporal modulation thresholds and consonants as opposed to vowels.

However, a causal relationship between temporal modu-

TABLE II. Comparison of the confusion matrices for the consonant data. The format is similar to that of Table I.

Category	Number		Correlation between NH and CI data	
	NH group	CI group	With diagonal	Without diagonal
0.0–64.6	(N=8)	(N=6)	0.752 (P=0.0%)	0.545 (P=0.0%)
64.6–78.8	(N=13)	(N=7)	0.964 (P=9.0%)	0.820 (P=2.1%)
78.8–100.0	(N=19)	(N=7)	0.985 (P=4.5%)	0.627 (P=0.2%)

lation perception and speech recognition is not consistent with results from vocoder simulations, which demonstrate that speech performance is only slightly affected by severe degradations in temporal information. Xu *et al.* (2005) examined reductions in temporal information in a vocoder simulation by low-pass filtering the envelope down to as much as 1 Hz. Since Xu *et al.* (2005) utilized the second order Butterworth low-pass filter, which decreases the response by 12 dB for every doubling in frequency beyond the low-pass cutoff, such filtering would translate to an 80 dB attenuation of temporal modulations in the 100 Hz range (Cazals *et al.*, 1994; Fu, 2002). The effect of reduction in temporal information on vowel perception was approximately 20%. Deficits in processing of temporal modulations alone are therefore not sufficient to account for almost an 80% difference in scores observed between the best and the worst performers on the vowel identification task.

Across-subject differences in stimulation selectivity of electrically stimulated cochlea have been observed using a variety of techniques (Boex *et al.*, 2003; Cohen *et al.*, 2003, 2004, 2005). Recently, several measures of spatial selectivity that simultaneously assess selectivity across the whole array have been found to moderately correlate to speech perception (Henry and Turner, 2003; Henry *et al.*, 2005). These correlations have been improved using techniques invoked in the present and previous manuscripts which rely on measuring spectral perception specifically at the lowest modulation frequencies which are apparently most important for speech perception of CI listeners (Saoji *et al.*, 2005). The relatively high correlations between spectral resolution and speech perception are consistent with a causal relationship between speech perception and spatial selectivity. The present study with NH subjects listening through the modified vocoder provides further support for the causal relationship, because it shows that performance of NH subjects with similar reductions in spectral selectivity is similar to that observed in CI patients. The variability in performance observed in NH listeners under these conditions is similar to variability observed in CI patients. In the case of consonant recognition, the data suggest a uniform temporal deficit in CI listeners.

C. Is the SMT a measure of peripheral spatial selectivity?

SMT can be interpreted as a measure of spectral selectivity (Dubno and Dorman, 1987; Horst, 1987). Poor peripheral spatial selectivity can smear the spectral contrast and thereby increase the SMT. Alternatively, differences in SMTs across patients may also reflect differences in sensitivity to internal contrast. This "efficiency" hypothesis would suggest that across-subject differences in SMT should be roughly independent of the spectral modulation frequency. Saoji *et al.* (2005) reported on SMT to spectral modulation frequencies of 0.25, 0.5, 1 and 2 cycles/octave. Subjects that were sensitive to spectral modulations of 0.25 and 0.5 cycles/octave tended to have thresholds that were least dependent on spectral modulation frequency, while subjects with the elevated SMTs at 0.25 and 0.5 cycles/octave tended to have greater increases in SMT for the frequencies of 1 and 2 cycles/octave. Thus, Saoji *et al.* (2005) supports the notion that the SMTs at 0.25 and 0.5 cycles/octave reflect spectral selectivity of CI listeners.

If SMT of CI listeners is partially determined by spatial selectivity at the periphery, then these results would suggest that more selective arrays or stimulation paradigms will lead to an improvement in speech recognition. Although the simpler of the approaches, a "peripheral selectivity" hypothesis is not the only possibility. Because central processes may play a role in recognition of spectral patterns, it is also possible that loss of spectral resolution may result without peripheral loss of selectivity. For example, with long duration of deafness, the structure of the auditory cortex changes profoundly. Such changes may negatively affect the ability of the listeners to discriminate spectral patterns (Irvine *et al.*, 2000).

It has been well documented that stimulation strategies employed in CIs may lead to activation patterns in the cochlea which may be very different from normal. For example, electrical stimulation may not properly preserve the crucial phase relationships observed in the firing patterns of healthy auditory nerves to realistic sounds (Loeb *et al.*, 1983; Carney, 1994; Loeb, 2005). In addition, electrical stimulation may lead to activation patterns that are unnaturally synchronized to the individual electrical pulses (Dynes and Delgutte, 1992; Litvak *et al.*, 2001; Ferguson *et al.*, 2003). The differences in individual performance on speech perception as well as nonspeech spectral tasks may therefore reflect the varying ability of subjects to adapt to such unnatural activation patterns.

Finally, it is possible that a deficit in more central processes that are unrelated to specifics of electrical stimulation patterns may be responsible for effective loss of spectral resolution. One argument in favor of the last possibility is that similar spectral and temporal deficits are encountered in hearing impaired (HI) listeners, although to a lesser degree than in the CI patients (Bacon and Viemeister, 1985; Glasberg and Moore, 1986; Summers and Leek, 1994; Henry *et al.*, 2005).

D. Comparison with other vocoder simulations

The results of the present study as well as those of Fu and Nogaki (2005) and Shannon *et al.* (1998) suggest that speech perception abilities of normal hearing listeners listening to CI simulations can be degraded by varying the overlap of the noise carriers. In contrast, several earlier reports accomplished the same task by varying the simulated number of analysis channels (Shannon *et al.*, 1995; Friesen *et al.*, 2001; Xu *et al.*, 2005). At face value, the "overlap" approach has greater resemblance to the clinical processors of most CI patients, who are almost always fit with strategies that utilize a greater number of analysis channels and electrodes than equivalent "effective channels" (Shannon *et al.*, 1995). It is possible therefore that the overlap approach may be responsible for the close match between the vowel confusion patterns between CI subjects and NH listeners listening to simulations. If similar correspondence cannot be achieved with the "effective channel" simulations, then the "overlap" simu-

lations, possibly extended to include some deficits in temporal processing, may be a more effective tool for modeling performance of CI listeners.

The "overlap" approach advocated in this paper provides a rich space in which to explore effects of spatial selectivity on performance. For example, if tools can be established that effectively measure spatial selectivity in separate regions of the cochlea, then "overlap" simulations can be modified accordingly to include various overlaps in various bands. Another productive direction may be to manipulate the "speech processor" part of the simulation. If strategies can be proposed that attempt to overcome poor spectral resolution of some CI listeners, then these strategies can be incorporated into the "speech processor" part of the simulation, and thus evaluated in NH subjects in parallel with CI patients. Comparison of effects between the two groups may lead to better understanding of limitations of CI performance.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Robert Shannon and Dr. Michael Dorman for their encouragement throughout this research, and useful feedback on an earlier version of the manuscript. This research was sponsored by Advanced Bionics Corporation.

Abbas, P. J., Hughes, M. L., Brown, C. J., Miller, C. A., and South, H. (**2004**). "Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential," Audiol. Neuro-Otol. **9**, 203–213.

American National Standards Institute (ANSI) (**1996**). "American standard specification for audiometers" (American National Standards Institute, New York).

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

Baer, T., and Moore, B. C. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Baer, T., Moore, B. C., and Gatehouse, S. (**1993**). "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times," J. Rehabil. Res. Dev. **30**, 49–72.

Baskent, D., and Shannon, R. V. (**2003**). "Speech recognition under conditions of frequency-place compression and expansion," J. Acoust. Soc. Am. **113**, 2064–2076.

Bernstein, L. R., and Green, D. M. (**1987**). "Detection of simple and complex changes of spectral shape," J. Acoust. Soc. Am. **82**, 1587–1592.

Boex, C., Kos, M. I., and Pelizzone, M. (**2003**). "Forward masking in different cochlear implant systems," J. Acoust. Soc. Am. **114**, 2058–2065.

Carney, L. H. (**1994**). "Spatio-temporal encoding of sound level: Models for normal encoding and recruitment of loudness," Hear. Res. **76**, 31–44.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," J. Acoust. Soc. Am. **96**, 2048–2054.

Cohen, L. T., Lenarz, T., Battmer, R. D., Bender von Saebelkampf, C., Busby, P. A., and Cowan, R. S. (**2005**). "A psychophysical forward masking comparison of longitudinal spread of neural excitation in the contour and straight nucleus electrode arrays," Int. J. Audiol. **44**, 559–566.

Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. S. (**2003**). "Spatial spread of neural excitation in cochlear implant recipients: Comparison of improved ECAP method and psychophysical forward masking," Hear. Res. **179**, 72–87.

Cohen, L. T., Saunders, E., and Richardson, L. M. (**2004**). "Spatial spread of neural excitation: Comparison of compound action potential and forward-masking data in cochlear implant recipients," Int. J. Audiol. **43**, 346–355.

Dorman, M. F., Dankowski, K., McCandless, G., and Smith, L. (**1989**). "Identification of synthetic vowels by patients using the Symbion multichannel cochlear implant," Ear Hear. **10**, 40–43.

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

Dubno, J. R., and Dorman, M. F. (**1987**). "Effects of spectral flattening on vowel identification," J. Acoust. Soc. Am. **82**, 1503–1511.

Dynes, S. B., and Delgutte, B. (**1992**). "Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea," Hear. Res. **58**, 79–90.

Eddington, D., Tierney, J., and Long, C. (**1997a**). *Cochlear Implants* (RLE, Cambridge, MA).

Eddington, D. K., Rabinowitz, W. R., Tierney, J., Noel, V., and Whearty, M. (**1997b**). "Speech processors for auditory prostheses," 8th Quarterly Progress Report, NIH Contract No. N01–DC-6–2100."

Efron, B., and Tibshirani, R. (**1993**). *An Introduction to the Bootstrap* (Chapman and Hall, New York).

Ferguson, W. D., Collins, L. M., and Smith, D. W. (**2003**). "Psychophysical threshold variability in cochlear implant subjects," Hear. Res. **180**, 101–113.

Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., Runge-Samuelson, C. L., and Wackym, P. A. (**2004**). "Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems," Hear. Res. **25**, 375–387.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

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

Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fu, Q. J., Shannon, R. V., and Galvin, J. J., III. (**2002**). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," J. Acoust. Soc. Am. **112**, 1664–1674.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Good, P. I. (**1994**). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (Springer-Verlag, New York).

Harnsberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., and Meyer, T. A. (**2001**). "Perceptual 'vowel spaces' of cochlear implant users: Implications for the study of auditory adaptation to spectral shift," J. Acoust. Soc. Am. **109**, 2135–2145.

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

Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.

Horst, J. W. (**1987**). "Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects," J. Acoust. Soc. Am. **82**, 874–885.

Irvine, D. R. F., Rajan, R., and McDermott, H. J. (**2000**). "Injury-induced reorganization in adult auditory cortex and its perceptual consequences," Hear. Res. **147**, 188–199.

Klatt, D. H. (**1980**). "Software for cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), 467–477.

Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2001**). "Auditory nerve fiber responses to electric stimulation: Modulated and unmodulated pulse trains," J. Acoust. Soc. Am. **110**, 368–379.

Loeb, G. E. (**2005**). "Are cochlear implant patients suffering from perceptual dissonance?," Ear Hear. **26**, 435–450.

Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). "Spatial cross-correlation. A proposed mechanism for acoustic pitch perception," Biol. Cybern. **47**, 149–163.

Miller, G., and Nicely, P. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Ohde, R. N., and Abou-Khalil, R. (**2001**). "Age differences for stop conso-

nant and vowel perception in adults," J. Acoust. Soc. Am. **110**, 2156–2166.
Oppenheim, A. V., and Schafer, R. W. (**1975**). *Digital Signal Processing* (Prentice–Hall, Englewood Cliffs, N.J.).
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.
Saoji, A., Litvak, L., Emadi, G., and Spahr, A. (**2005**). "Spectral modulation transfer function in cochlear implant listeners," in *Conference on Implantable Auditory Prostheses*, Asilomar, California.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shannon, R. V., Zeng, F. G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467–2476.
Summers, V., and Leek, M. R. (**1994**). "The internal representation of spectral contrast in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3518–3528.
Tyler, R. S., Preece, J. P., Lansing, C. R., Otto, S. R., and Gantz, B. J. (**1986**). "Previous experience as a confounding factor in comparing cochlear-implant processing schemes," J. Speech Hear. Res. **29**, 282–287.
Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.