Spectral and Temporal Cues in Cochlear Implant Speech Perception

Kaibao Nie, Amy Barco, and Fan-Gang Zeng

Objective: **Taking advantage of the flexibility in the number of stimulating electrodes and the stimulation rate in a modern cochlear implant, the present study evaluated relative contributions of spectral and temporal cues to cochlear implant speech perception.**

Design: **Four experiments were conducted by using a Research Interface Box in five MED-EL COMBI 40+ cochlear implant users. Experiment 1 varied the number of electrodes from four to twelve or the maximal number of available active electrodes while keeping a constant stimulation rate at 1000 Hz per electrode. Experiment 2 varied the stimulation rate from 1000 to 4000 Hz per electrode on four pairs of fixed electrodes. Experiment 3 covaried the number of stimulating electrodes and the stimulation rate to study the trade-off between spectral and temporal cues. Experiment 4 studied the effects of envelope extraction on speech perception and listening preference, including half-wave rectification, full-wave rectification, and the Hilbert transform. Vowels, consonants, and HINT sentences in quiet, as well as with a competing female voice served as test materials.**

Results: **Experiment 1 found significant improvement in all speech tests with a higher number of stimulating electrodes. Experiment 2 found a significant advantage of the high stimulation rate only on consonant recognition and sentence recognition in noise. Experiment 3 found an almost linear trade-off between the number of stimulation electrodes and the stimulation rate for consonant and sentence recognition in quiet, but not for vowel and sentence recognition in noise. Experiment 4 found significantly better performance with the Hilbert transform and the full-wave rectification than the half-wave rectification. In addition, envelope extraction with the Hilbert transform produced the highest rating on subjective judgment of sound quality.**

Conclusions: **Consistent with previous studies, the present result from the five MED-EL subjects showed that (1) the temporal envelope cues from a limited number of channels are sufficient to support high levels of phoneme and sentence recognition in quiet but not for speech recognition in a competing voice, (2) consonant recognition relies more on temporal cues while vowel recognition relies more on spectral cues, (3) spectral and temporal cues can be traded to some degree to produce similar performance in cochlear implant speech recognition, and (4) the Hilbert envelope improves both speech intelligibility and quality in cochlear implants.**

(Ear & Hearing 2006;27;208–217)

Hearing and Speech Research Laboratory, Departments of Biomedical Engineering and Otolaryngology – Head and Neck Surgery, University of California, Irvine (K.N., F.-G.Z.); and MED-EL Corporation, North America, Durham, North Carolina (A.B.).

Natural speech carries abundant acoustic cues in both spectral and temporal domains (Remez, Rubin, Pisoni, & Carrell, 1981; Rosen, 1992; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Smith, Delgutte, & Oxenham, 2002; Stevens, 1980). Modern cochlear implant devices can only crudely encode these spectral and temporal cues (Brill, Gstottner, Helms, von Ilberg, Baumgartner, Muller, et al., 1997; Seligman & McDermott, 1995; Vandali, Whitford, Plant, & Clark, 2000; Wilson, Finley, Lawson, Wolford, Eddington, & Rabinowitz, 1991). Typically, 6 to 22 spectral bands are used to encode a bandwidth of 5000 to 10,000 Hz, whereas temporal information is limited to band-specific envelopes up to several hundred hertz. In practice, the actual amount of information that can be transmitted to cochlear implant users is severely limited by a host of additional physical and physiological factors such as the electrode-nerve interface, nerve survival, and brain plasticity. These limitations result in 6 to 10 functional channels and poor pitch perception, compared with most normal-hearing listeners, in a typical cochlear implant user (Fishman, Shannon, & Slattery, 1997; Garnham, O'Driscoll, Ramsden, & Saeed, 2002; Kong, Cruz, Jones, & Zeng, 2004). The same signal processing and physical and physiological limitations also contribute to the problem facing most current cochlear implant users who can achieve a high level of speech recognition in quiet but suffer greatly in noise (Dorman, Loizou, Fitzke, & Tu, 1998; Fu & Shannon, 1999; Zeng & Galvin, 1999), particularly when the noise is temporally fluctuating, such as a competing voice (Nelson, Jin, Carney, & Nelson, 2003; Qin & Oxenham, 2003; Stickney, Zeng, Litovsky, & Assmann, 2004).

Recent research has focused on improving the

0196/0202/06/2702-0208/0 • Ear & Hearing • Copyright © 2006 by Lippincott Williams & Wilkins • Printed in the U.S.A.

representation of spectral and temporal cues in cochlear implants, for instance, using high-rate stimulation to improve representation of temporal envelope and fine structure (Rosen, 1992). To this date, mixed results have been obtained on the effect of high-rate stimulation on speech recognition and preference. Vandali et al. used three carrier stimulation rates (250, 807, and 1615 Hz) in five Nucleus-24 cochlear implant subjects with the ACE strategy and found no statistical differences between the 250- and 807-Hz carriers and actually slightly poorer performance for the 1615-Hz carrier in some subjects (Vandali, Whitford, Plant, & Clark, 2000). Holden et al. evaluated the effect of stimulation rate (720 versus 1800 Hz) on phoneme, word, and sentence recognition in 8 Nucleus-24 subjects with the ACE strategy and found a significant advantage of the higher rate with some tests in some subjects (Holden, Skinner, Holden, & Demorest, 2002). On the other hand, consistent improvement with high-rate stimulation (2000 versus 600 Hz) has been reported in MED-EL implant subjects with the CIS strategy that provided an overall stimulation rate up to 18,180 Hz (Kiefer, von Ilberg, Rupprecht, Hubner–Egner, & Knecht, 2000; Loizou, Poroy, & Dorman, 2000). The differences between previous studies were probably due to the trade-off and/or interactions between the number of electrodes and the stimulation rate. For example, the number of electrodes was held constant at 20 in both the Vandali et al. and Holden et al. studies but was at 6 to 8 in the Kiefer et al. and Loizou et al. studies. One of the goals in the present study was to address the high-rate stimulation issue by systematically varying the per-electrode carrier rate from 1000 to 4000 Hz, using four stimulating electrodes. Note that with the exception of Kiefer et al.'s study, the 4000-Hz per channel rate in the present study was twice the highest rate used in previous studies.

In addition to the high-rate carrier, extraction of the temporal envelope may affect cochlear implant speech performance. Three methods including the half-wave rectification, the full-wave rectification, and the Hilbert transform (Hilbert, 1912) have been used in cochlear implant speech processors, with the full-wave rectification being the most widely used in the Clarion and MED-EL devices. Only the most recent MED-EL behind-the-ear processor has implemented the Hilbert envelope in all spectral bands. The Nucleus Sprint processor uses quadrature rectification of the complex fast Fourier transform values to estimate the envelope, which yields identical envelopes to those obtained by using the Hilbert transform for low-frequency bands up to 1600 Hz (Vandali, personal communication). Theoretical analysis has suggested a possible advantage of the Hilbert envelope because the half-wave or full-wave rectification produces distorted frequency components (e.g., combination tones) in the modulation domain that are not physically present in the original signal, whereas the Hilbert transform provides a clear separation between a signal's temporal envelope and fine structure (Zeng, 2004). This possibility was partially supported by the observation that better speech performance was achieved with the MED-EL behind-the-ear processor using the Hilbert envelope than with its previous body-worn processor using the full-wave rectification (Anderson, Weichbold, & D'Haese, 2002; Helms, Muller, Schon, Moser, Arnold, Janssen et al., 1997; Helms, Muller, Schon, Winkler, Moser, Shehata–Dieler et al., 2001). Because additional factors might also contribute to the performance difference between the two processors, a second goal in the present study was to use the same MED-EL research interface to implement all three types of temporal envelope extraction and to evaluate their effects on cochlear implant speech performance.

A final motivation for the present study was to evaluate a potential trade-off between spectral and temporal cues, or more specifically the trade-off between the number of channels and the representation of temporal envelopes. By systematically varying the number of spectral bands and the cutoff frequency in the envelope extractor in normal-hearing listeners listening to cochlear implant simulations, Xu and his colleagues found a trade-off between spectral and temporal cues in Mandarin tone and phoneme recognition (Xu & Pfingst, 2003; Xu, Tsai, & Pfingst, 2002). Brill et al. also found evidence for the spectral-temporal trade-off in three MED-EL implant users (Brill, Gstottner, Helms, von Ilberg, Baumgartner, Muller et al., 1997). In at least one subject (FZ), they essentially found a linear trade-off between the number of stimulating electrodes and the carrier rate: the subject maintained 80 to 90% sentence recognition with essentially all combinations of the number of channels and the carrier rate (e.g., 2 channels with 9090-Hz rate or 10 channels with 1818-Hz rate).

The present study conducted four experiments to address this trade-off as well as other issues. Experiment 1 assessed the spectral contribution to cochlear implant performance by varying the number of stimulating electrodes from 4 to 12 while holding the stimulation rate at 1000 Hz per electrode. Experiment 2 addressed the temporal contribution by varying the stimulation rate from 1000 to 4000 Hz per electrode on the four stimulating electrodes. Experiment 3 addressed the trade-off between the number of stimulating electrodes and the stimulation rate by covarying them in a systematic manner.

TABLE 1. Biographical data of the five MED-EL cochlear implant subjects

Subject	Age	Gender	Cause of deafness	Age at onset	Month/year of implant	No. of electrodes used	Vowel score (%)	Consonant score (%)
S1	24	M	Unknown	0	5/1999	9	67	88
S2	40	F	Unknown	12	6/2000	12	31	48
S3	39	M	Infection	0	4/2000	12	71	64
S4	57	F	Medication	8	10/2002	10	38	60
S5	42	F	Hereditary	20	3/2001	8	63	60

Experiment 4 assessed the role of temporal envelope extraction (half-wave rectification, full-wave rectification, and the Hilbert transform) in cochlear implant performance.

METHODS

Subjects

Five adult subjects using the MED-EL cochlear implant participated in this study. Table 1 shows the biographical data from these MED-EL implant users. All subjects were native English speakers with at least 1 year of implant experience at the time of testing. Two subjects were prelingually deafened (S1 and S3) and the other three were postlingually deafened. Two subjects used all 12 available electrodes, but subjects S4, S1, and S5 used 10, 9, and 8 electrodes, respectively, due to side effects arising from stimulation on the omitted electrodes. In subjects S1 and S4, the 9- and 10-electrode condition was included in statistical analysis of the 12-electrode condition. On the other hand, subject S5 was not included in the analysis. As standard laboratory practice, all subjects were first tested with their clinically mapped speech processors on vowel and consonant recognition. The human subject protocol was approved by the local institutional review board.

Stimuli

The test materials included 12 vowels in /hVd/ context, 20 consonants in /aCa/ context, and 120 Hearing In Noise Test (HINT) sentences. The vowels included "had, haw'd, hayed, head, heed, herd, hid, hod, hoed, hood, hud, and who'd," produced by a female and a male talker (Hillenbrand, Getty, Clark, & Wheeler, 1995). The 20 consonants included "aba, acha, ada, afa, aga, aja, aka, ala, ama, ana, apa, ara, asa, asha, ata, atha, ava, awa, aya, and aza," also produced by a female and a male talker (Shannon, Jensvold, Padilla, Robert, & Wang, 1999). The 240 HINT sentences, produced by a male speaker, were divided into 24 lists consisting of 10 sentences each or a total of 50 key words (Nilsson, Soli, & Sullivan, 1994). Half of the sentences were presented in quiet, whereas the other half were presented with a competing voice from a female talker at 10-dB signal-to-noise ratio. The competing voice was always the sentence that had the longest duration among all stimuli (i.e., "they knocked on the window"). We chose to use a competing voice from a female talker because we found that compared to the traditionally used speech-spectrum-shaped noise, the competing voice produced the greatest difference in performance between the cochlear implant and normal-hearing subjects (Zeng, Nie, Stickney, Kong, Vongphoe, Bhargave et al., 2005). All stimuli were normalized to have the same root-mean-square level, which was presented to the subject at his or her most comfortable loudness level.

Signal Processing

The present study used a Diagnostic Interface Box (DIB) (Reference Note 1) and a Research Interface Box (RIB) (Reference Note 2) provided by the University of Innsbruck to work with the MED-EL cochlear implant. Both interfaces were controlled using a personal computer by a serial communication port (RS 232). To set up a subject's map, the DIB was first used to measure the threshold level (THR) and the maximum comfortable loudness level (MCL) on all electrodes in the map. The computer then processed a sound file (.wav) offline and generated a data file that contained all parameters describing an electric stimulus, including electrode number, current amplitude, pulse duration, interpulse interval, and stimulation rate. During the test, the RIB would download the data file, generate its corresponding signal and send the signal to the internal receiver through a radio frequency link coil.

All stimuli were sampled at 22,025 Hz and normalized to have the same RMS level. A sound was pre-emphasized (first-order Butterworth high-pass filter with 1200-Hz cutoff), then divided into 4, 8, and 12 bands, using a bank of sixth-order Butterworth band-pass filters with equal bandwidths on a logarithmic scale from 300 to 5500 Hz. For instance, cutoff frequencies for an eight-band system were

300, 432, 621, 893, 1285, 1848, 2658, 3824, and 5500 Hz. Within each band, the temporal envelope was extracted by a half-wave or full-wave rectifier followed by a second-order Butterworth filter (experiments 1 through 3) or by the Hilbert transform (experiment 4). To match between the large acoustic dynamic range and the narrow electric dynamic range, the temporal envelope was further compressed in amplitude by a logarithmic function: where c = 500.

$$y = \log(1 + cx)/\log(1 + c)$$

where x is the acoustic amplitude and y is the electric amplitude.

Experiment 1 emulated three speech processors with 4, 8, or 12 stimulating electrodes, each of which was stimulated by a constant carrier rate at 1000 Hz. Electrodes 1, 4, 8, and 12 were activated in the four-electrode processor, whereas electrodes 1, 3, 4, 6, 7, 9, 10, and 12 were activated in the eight-electrode processor. Experiment 2 fixed the number of electrodes at four but varied the per-electrode stimulation rate from 1000 to 2000 and 4000 Hz. To avoid aliasing, the cutoff frequency of the low-pass filter used in the rectification methods was set to be half of the stimulation rate (i.e., fc = 500 Hz for the 1000-Hz stimulation rate, fc = 1000 Hz for the 2000-Hz stimulation rate, and fc = 2000 Hz for the 4000-Hz stimulation rate). Note in practice that good temporal envelope representation requires the carrier rate be at least four times the highest modulation frequency in the envelopes (McKay, McDermott, & Clark, 1994; Wilson, Finley, Lawson, & Zerbi, 1997). Experiment 3 covaried the number of electrodes and the stimulation rate to maintain a total stimulation rate at 16,000 Hz, including 4 electrodes with 4000 Hz, 8 electrodes with 2000 Hz, and 12 electrodes with 1333 Hz. Although experiments 1 through 3 used the full-wave rectification method to extract the temporal envelope, experiment 4 compared performance among three temporal envelope extraction methods including half-wave rectification, full-wave rectification, and the Hilbert transform. A second-order Butterworth low-pass filter with 500-Hz cutoff frequency was used to smooth the temporal envelope and to avoid aliasing in all methods. To facilitate comparison, the number of electrodes was varied from 4 to 8 and to 12 while the stimulation rate was fixed at 1000-Hz per electrode. Monopolar configuration was used in all experiments.

Procedures

Four sequential test sessions were used in this study, starting with consonant recognition, then vowel recognition, sentence recognition in quiet and finally sentence recognition in noise. A 10-minute practice session was given to each subject for familiarization with the experimental processor before each test session. Within each test session, the order in which the experimental conditions were evaluated was randomized. In phoneme tests, a customized graphic user interface (GUI) was used with 12 vowels or 20 consonants displayed on the computer screen. The subject had to make a choice in response to the presented stimulus. Feedback regarding the correct response was given after each presentation. In sentence tests, the subject was presented with a sentence and then had to type in as many key words as possible presented in the sentence. The keywords were defined as all words except for "a," "an," and "the" in the sentence. The order of typed keywords did not affect the scoring. No feedback was given in the sentence test.

After the recognition tests, a subjective sound quality rating was also conducted by using the three envelope extraction methods used in experiment 4. The subject heard the same HINT sentences as used in the recognition tests and then had to rate the sound quality with a score from 1 to 10, with 1 representing the poorest sound quality and 10 representing the best sound quality.

RESULTS

Experiment 1: Number of Stimulating Electrodes

Figure 1 shows percent correct scores as a function of the number of stimulating electrodes for consonant recognition (top panel), vowel recognition (middle panel), and sentence recognition in quiet and in noise (bottom panel). In Figure 1, as well as the remaining figures, the data were averaged across five subjects, except for the 12-electrode data, which did not include data from S5. With the exception of sentence recognition in noise, performance increased monotonically with the number of stimulating electrodes for all tests. The average improvement in percentage points from 4 to 12 electrodes was 13 for consonant recognition, 30 for vowel recognition, 21 for sentence recognition in quiet, and 18 for sentence recognition in noise. One-way analysis of variance (ANOVA) with repeated measures confirmed this observation: the number of electrodes was a significant factor for consonant recognition [$F(2,6) = 6.9$, $p < 0.05$], vowel recognition [$F(2,6) = 5.7$, $p < 0.05$], and sentence recognition in quiet [$F(2,4) = 8.8$, $p < 0.05$], but not for sentence recognition in noise [$F(2,4) = 3.8$, $p > 0.05$].

Two additional points are also noted in the data. First, consonant recognition seems to be less dependent on the number of electrodes than vowel recognition: consonant recognition reached a plateau at 8 electrodes ($p > 0.05$), whereas vowel recognition still

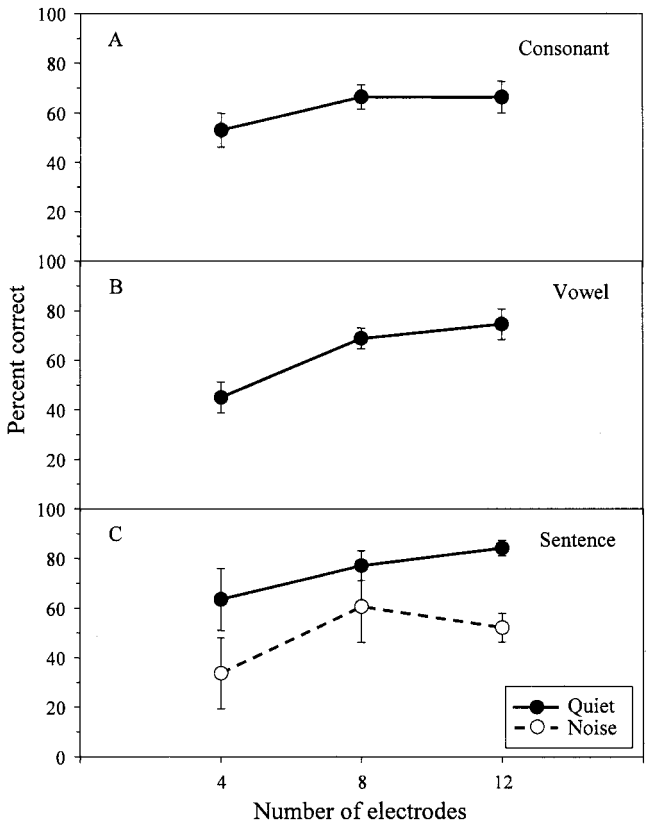

Fig. 1. Consonant (A), vowel (B), and sentence (C) recognition scores as a function of the number of electrodes from 4, 8, to 12, using a stimulation rate of 1000 Hz. The 12-electrode data in this figure, as well as in Figure 3, Figure 4, and Figure 5, were from four subjects only (excluding S5). In C, filled circles connected by the solid line represent sentence recognition data in quiet, whereas open circles connected by the dashed line represent sentence recognition data in a competing talker. Error bars represent standard errors.

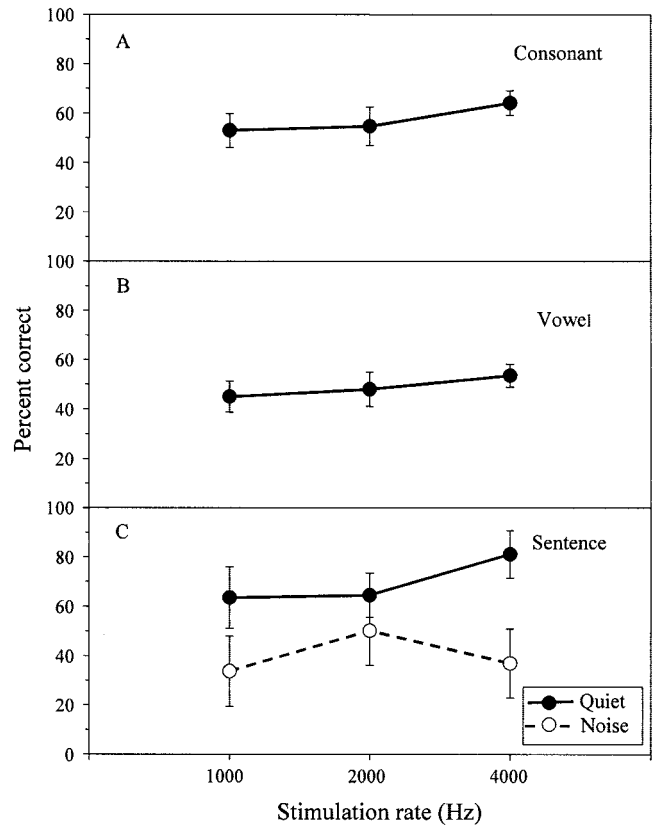

Fig. 2. Consonant (A), vowel (B), and sentence (C) recognition scores as a function of the stimulation rate from 1000, 2000, to 4000 Hz, using a 4-electrode condition. In C, filled circles connected by the solid line represent sentence recognition data in quiet, whereas open circles connected by the dashed line represent sentence recognition data in a competing talker. Error bars represent standard errors.

increased by 7 percentage points from 8 to 12 electrodes ($p < 0.05$). Second, high levels of performance were achieved for sentence recognition in quiet: 63% correct score with 4 electrodes and 84% with 12 electrodes. Compared with performance in quiet, the noise decreased the sentence recognition score by 30 percentage points with 4 electrodes and 32 percentage points with 12 electrodes ($p < 0.05$).

Experiment 2: Effect of Stimulation Rate

Figure 2 shows percent scores as function of stimulation rate for consonant recognition (top panel), vowel recognition (middle panel), and sentence recognition in quiet and in noise (bottom panel). Similar to changes in the number of electrodes, increasing stimulation rate also generally improved cochlear implant performance. However, compared with the change in the number of electrodes from 4 to 12, the average improvement was much smaller with the stimulation rate being increased from 1000 to 4000 Hz: 11 percentage points for consonant, 9 for vowel, 17 for sentence in quiet, and 4 for sentence recognition in noise. Correspondingly, only consonant recognition was found to be significantly dependent on stimulation rate [$F(2,8) = 5.1$, $p < 0.05$].

Experiment 3: Trade-off Between Number of Electrodes and Stimulation Rate

Figure 3 shows percent correct scores as a function of increasing number of electrodes coupled with decreasing stimulation rate for consonant recognition (top panel), vowel recognition (middle panel), and sentence recognition in quiet and in noise (bottom panel). If there were a trade-off between the number of electrodes and the stimulation rate, then no change in performance would be observed for these combinations of the number of electrodes and the stimulation rate, namely, a horizontal line would be displayed in the graph. Indeed, no signifi-

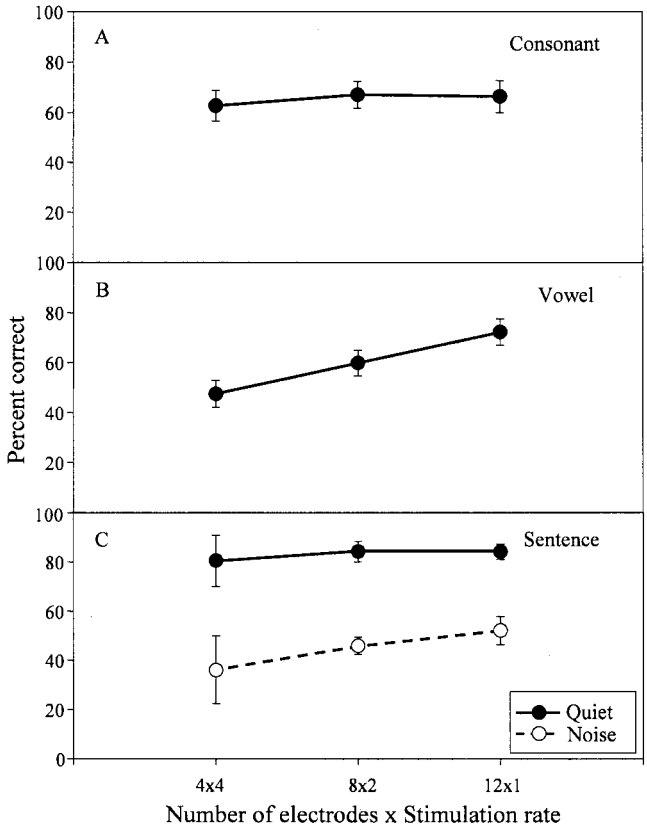

Fig. 3. Consonant (A), vowel (B), and sentence (C) recognition scores as a function of combinations of the number of electrodes and the stimulation rate. In C, filled circles connected by the solid line represent sentence recognition data in quiet, whereas open circles connected by the dashed line represent sentence recognition data in a competing talker. Error bars represent standard errors.

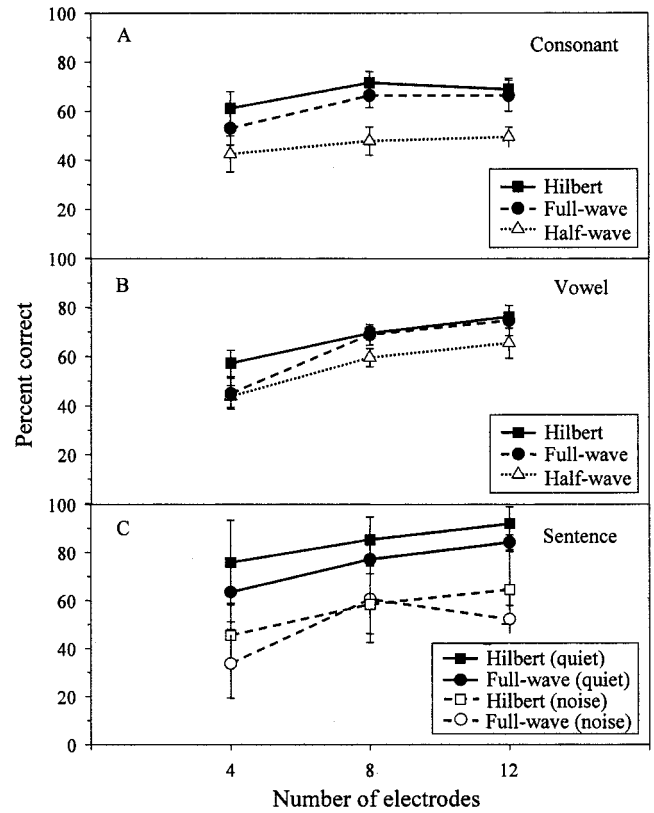

Fig. 4. Consonant (A), vowel (B), and sentence (C) recognition scores as a function of number of electrodes. Three envelope-extraction methods are represented by squares (the Hilbert transform), circles (full-wave rectification), and triangles (half-wave rectification). Data with full-wave rectification are the same as in Figure 1. In C, the filled symbols connected by the solid line represent sentence recognition data in quiet, whereas the open symbols connected by the dashed line represent sentence recognition data in noise. Error bars represent standard errors.

cant difference was found between these combinations of number of electrodes and stimulation rate for consonant recognition [$F(2,6) = 0.776$, $p > 0.05$] and sentence recognition in quiet [$F(2,4) = 1.15$, $p > 0.05$], suggesting a linear trade-off between number of electrodes and stimulation for these tasks. However, a significant monotonic improvement was observed for vowel recognition from 51% with the 4-electrode, 4-kHz stimulation rate condition to 74% with the 12-electrode, 1.3-kHz rate condition [$F(2,6) = 5.1$, $p < 0.05$]. This improvement in performance was clearly due to the increased number of electrodes, suggesting a dominant role of spectral resolution in vowel recognition. A significant difference was also observed for sentence in noise scores [34% with 4 × 4 kHz to 52% with 12 × 1 kHz, $F(2,4) = 6.9$, $p < 0.05$]. However, it is unclear, given the results of experiments 1 and 2, whether this improvement is the result of increased numbers of electrodes or the use of rates less than 4000 Hz.

Experiment 4: Effect of Temporal Envelope Extraction

Figure 4 shows percent correct scores as a function of the number of electrodes for consonant recognition (top panel) and vowel recognition (middle panel) and sentence recognition in quiet and in noise (bottom panel). The three envelope-extraction methods are the Hilbert transform (squares), the full-wave rectification (circles), and the half-wave rectifier (triangles; this method was not tested in sentence recognition). A repeated measures, two-way ANOVA was used to test whether the number of electrodes and/or the envelope extraction technique had a significant effect on speech recognition. For consonant recognition, there was a significant effect of envelope extraction [$F(2,4) = 29.2$, $p < 0.01$] but no significant effect of the number of electrodes [$F(2,4) = 3.45$, $p > 0.05$]. Post hoc tests revealed a significant difference of 21 percentage points be-

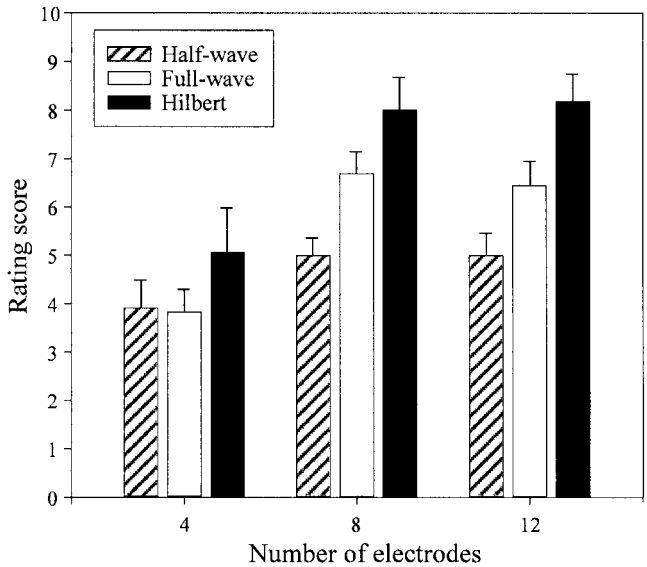

Fig. 5. Subject rating scores for the temporal envelope extracted by the half-wave rectification (shaded bars), the full-wave rectification (white bars), and the Hilbert transform (black bars) method with 4, 8, and 12 stimulating electrodes. Error bars represent standard errors.

tween the Hilbert transform and the half-wave rectification (paired t-test, $p < 0.01$), 15 percentage points between the full-wave and half-wave rectification ($p < 0.01$), and 7 percentage points between the Hilbert transform and the full-wave rectification ($p < 0.05$). For vowel recognition, not only was there a significant effect of envelope extraction [$F(2,4) = 11.1$, $p < 0.05$] but also a significant effect of the number of electrodes [$F(2,4) = 21.4$, $p < 0.01$]. However, post hoc tests revealed only a significant difference of about 11 percentage points between the Hilbert transform and the half-wave rectification ($p < 0.01$). For sentence recognition, the small improvement by the Hilbert transform over the full-wave rectification was not significant in quiet [$F(1,2) = 0.85, p > 0.05$] or in noise [$F(1,2) = 0.77, p > 0.05$]. This nonsignificant effect of envelope extraction might be due to a lack of statistical power (or data points), given that the half-wave rectification condition was not included in the sentence test. On the other hand, the number of electrodes was a significant factor affecting both sentence recognition in quiet [$F(1,2) = 22.6, p < 0.01$] and in noise [$F(1,2) = 8.1, p < 0.05$].

Figure 5 shows subjective rating of sound quality as a function of number of electrodes with three methods of temporal envelope extraction: the half-wave rectification (hatched bars), the full-wave rectification (open bars), and the Hilbert transform (filled bars). Two-way ANOVA revealed a significant effect of both the number of electrodes [$F(2,6) = 5.6$, $p < 0.05$] and the envelope extraction [$F(2,6) = 7.4$, $p < 0.05$] on sound quality rating. The lowest rating was obtained with the half-wave and full-wave rectification using 4 electrodes (3.8) compared with 5.1 rating with the Hilbert transform using the same 4 electrodes. The highest rating was obtained with the Hilbert transform using either 8 or 12 electrodes (8.0). Subjectively, subjects commented that the half-wave rectified speech sounded "indistinct," "muffled," "robotic," or "nasal"; the full-wave rectified speech sounded "nasal" but "clear"; whereas the Hilbert transform speech sounded "clear" or "distinguishable."

DISCUSSION

Spectral and Temporal Cues

The present results shed light onto relative contributions of spectral and temporal cues to speech perception. First, the relative contributions can be observed by the apparent distinctive pattern of results between consonant and vowel recognition. Similar to previous findings (Kiefer, von Ilberg, Rupprecht, Hubner–Egner, & Knecht, 2000; Loizou, Poroy, & Dorman, 2000), the present result showed that consonant recognition is improved by high-rate stimulation but is essentially independent of the number of electrodes between 4 and 12, whereas vowel recognition is hardly affected by the stimulation rate but is critically dependent on the number of electrodes. Together, these results suggest that consonant recognition relies more on temporal cues, whereas vowel recognition relies more on spectral cues.

Second, the relative contributions can be observed by the significant difference in sentence recognition between quiet and noise conditions. Similar to previous studies (Dorman, Loizou, & Fitzke, 1998; Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001; Garnham, O'Driscoll, Ramsden, & Saeed, 2002), the present study found high levels of phoneme and sentence recognition in quiet with as few as 4 electrodes but a significant drop in performance with noise. These results support the argument that while temporal cues from a limited number of spectral bands are adequate for speech recognition in quiet (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), high spectral resolution is critically needed to support speech recognition in noise (Dorman, Loizou, Fitzke, & Tu, 1998; Fu & Shannon, 1999; Nelson, Jin, Carney, & Nelson, 2003; Qin & Oxenham, 2003; Stickney, Zeng, Litovsky, & Assmann, 2004; Zeng & Galvin, 1999).

A caveat for this argument is that the number of channels may be as high as 34 in terms of producing better performance in noise in acoustic simulations

of cochlear implants but rarely exceeds 7 to 8 electrodes in actual cochlear implant users. As a matter of fact, some drop in performance was observed from 8 to 12 numbers of electrodes for the sentence in noise recognition as observed in the present study. A similar pattern can also be found in the Friesen et al. (2001) and Garnham et al. (2002) studies, suggesting that a disassociation between the effective number of channels and the number of electrodes in actual cochlear implant users. The reason for this disassociation may be mismatched frequency-to-electrode maps and electrode interactions.

Finally, the relative contributions can be observed by the trade-off between spectral and temporal cues. The present result showed a trade-off between number of electrodes and stimulation rate for consonant and sentence recognition in quiet but not for vowel and sentence recognition in noise. Xu and Pfingst (2003) found a similar trade-off between the number of spectral bands and the cutoff frequency of the envelope filter for both consonant and vowel recognition. The underlying physiological mechanisms for such a trade-off are not clear but may reflect the difference in brain processing between temporal cues (left hemisphere) and spectral cues (right hemisphere) (Zatorre & Belin, 2001).

High-Rate Stimulation

Although cochlear implant manufacturers have advocated high-rate stimulation in their current devices, little or no scientific evidence exists to support the high-rate stimulation advantage. Increasing the per-electrode carrier rate from a few hundred hertz to 1000 to 2000 Hz seemed to improve speech recognition under certain conditions, but increasing the rate further to 4000 Hz may actually degrade cochlear implant performance (Holden, Skinner, Holden, & Demorest, 2002; Kiefer, von Ilberg, Rupprecht, Hubner-Egner, & Knecht, 2000; Loizou, Poroy, & Dorman, 2000; Vandali, Whitford, Plant, & Clark, 2000). Using a competing voice as the noise masker (10 dB signal-to-noise ratio), the present result showed a significant drop in sentence recognition with this high stimulation rate (43 percentage point drop with 4000 Hz versus 30 with 1000 Hz, relative to quiet performance, as shown in Figure 2C). It is possible that high-rate stimulation produces greater electrode interaction, effectively reducing the number of functional channels (de Balthasar, Boex, Cosendai, Valentini, Sigrist, & Pelizzone, 2003; Middlebrooks, 2004). Although high-rate stimulation can theoretically produce improved temporal representation and naturalistic stochastic responses (Rubinstein, Wilson, Finley, & Abbas, 1999), the trade-off between high-rate stimulation and electrode interaction may pose a practical upper limit for high-rate stimulation in cochlear implant performance.

Temporal Envelope Extraction

The present result shows that the half-wave rectification produced the worst performance compared with the full-wave rectification and the Hilbert transform. In all test conditions, the Hilbert transform produced better performance than the full-wave rectification, but this difference did not reach the significance level except for the consonant recognition and the sound quality rating experiment. The advantage of the Hilbert transform over the full-wave rectification was reported in speech recognition in noise in a larger MED-EL cochlear implant population (Anderson, Weichbold, & D'Haese, 2002). Together, these two studies suggest that the Hilbert transform or similar methods, such as the quadrature rectification of the complex fast Fourier transform values used in the Nucleus Sprint processor, be used to extract the temporal envelope in future cochlear implants or acoustic simulations of the cochlear implant.

Training Effect

The present experimental design only provided 10 minutes for the implant subjects to become familiar with the experimental processor. The limited practice was a compromise of both the limited availability of the subject's test time and the large number of test conditions in the experimental design. Although previous studies used similarly short practice time (Fishman, Shannon, & Slattery, 1997), we note that a significant training effect has been observed with novel processors after familiarization with the novel processor from a few weeks to 3 months (Fu, Shannon, & Galvin, 2002; Skinner, Holden, Whitford, Plant, Psarros, & Holden, 2002). Therefore, it is possible that different outcomes to those observed in the present study may be obtained, had the present study used the long-duration training session in the experimental design.

CONCLUSION

Using the MED-EL cochlear implant capable of providing a total stimulation rate of 18,180 Hz, the present study evaluated the effect of number of electrodes, stimulation rate, and temporal envelope extraction on speech recognition in quiet and in noise. The present result from five MED-EL subjects, together with previous relevant studies, supported the following conclusions:

(1) Increasing the number of stimulating elec-

trodes from 4 to 12 generally increases speech performance, particularly for closed-set vowel recognition and sentence recognition in quiet.

(2) Increasing the per-electrode carrier rate from 1000 to 4000 Hz increases consonant recognition and sentence recognition in quiet but does not affect vowel recognition and even degrades sentence recognition in the presence of a competing voice.

(3) A linear trade-off exists between the number of electrodes (from 4 to 12) and the stimulation rate (from 4000 to 1333 Hz) for consonant recognition and sentence recognition in quiet but not for vowel recognition and sentence recognition in the presence of a competing voice.

(4) The temporal envelope extracted by the Hilbert transform produces the best performance in both intelligibility and sound quality compared with the half-wave and full-wave rectification methods.

(5) Consonant recognition relies more on temporal cues, whereas vowel recognition relies more on spectral cues.

(6) High-rate stimulation up to 2000 Hz is beneficial to speech recognition, but further increase to 4000 Hz may degrade performance due to possibly excessive electrode interaction at this high rate.

ACKNOLEDGMENTS

We thank the cochlear implant subjects for their participation in this study. Abby Copeland, Jason Edwards, Rosalie Uchanski, Andrew Vandali, and an anonymous reviewer provided helpful comments on the manuscript. Kaibao Nie is currently with the Virginia Merrill Bloedel Hearing Research Center in University of Washington, Seattle. This research was funded by National Institutes of Health (RO1-DC002267), with additional support from MED-EL Corporation, North America, which covered the research interface box cost, as well as the implant subjects' testing and travel costs.

Address for correspondence: Fan-Gang Zeng, 364 Med Surge II, Irvine, CA 92697–1275. E-mail: fzeng@uci.edu

Received December 3, 2004; accepted September 20, 2005.

REFERENCES

Anderson, I., Weichbold, V., & D'Haese, P. (2002). Recent results with the MED-EL COMBI 40+ cochlear implant and TEMPO+ behind-the-ear processor. *Ear, Nose, & Throat Journal*, *81*, 229–233.

Brill, S. M., Gstottner, W., Helms, J., von Ilberg, C., Baumgartner, W., & Muller, J., et al. (1997). Optimization of channel number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI 40+. *American Journal of Otolaryngology*, *18 (6 Suppl)*, S104–S106.

de Balthasar, C., Boex, C., Cosendai, G., Valentini, G., Sigrist, A., & Pelizzone, M. (2003). Channel interactions with high-rate biphasic electrical stimulation in cochlear implant subjects. *Hearing Research*, *182*, 77–87.

Dorman, M. F., Loizou, P. C., & Fitzke, J. (1998). The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors. *Ear and Hearing*, *19*, 481–484.

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. *Journal of the Acoustical Society of America*, *104*, 3583–3585.

Fishman, K. E., Shannon, R. V., & Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *Journal of Speech, Language, and Hearing Research*, *40*, 1201–1215.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America*, *110*, 1150–1163.

Fu, Q. J., & Shannon, R. V. (1999). Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping. *Journal of the Acoustical Society of America*, *106*, L18–L23.

Fu, Q. J., Shannon, R. V., & Galvin, 3rd J. J. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *Journal of Acoustical Society of America*, *112*, 1664–1674.

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

Helms, J., Muller, J., Schon, F., Moser, L., Arnold, W., & Janssen, T., et al. (1997). Evaluation of performance with the COMBI40 cochlear implant in adults: a multicentric clinical study. *ORL J Otorhinolaryngol Relat Spec*, *59*, 23–35.

Helms, J., Muller, J., Schon, F., Winkler, F., Moser, L., Shehata-Dieler, W, et al. (2001). Comparison of the TEMPO+ ear-level speech processor and the cis pro+ body-worn processor in adult MED-EL cochlear implant users. *ORL Journal for Oto-rhino-laryngology and its Related Specialties*, *63*, 31–40.

Hilbert, D. (1912). Grundzuge einer allgemeient Theorie der linearen integralgleichungen (Teubner, Leipzig)

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97 (5 Pt 1)*, 3099–3111.

Holden, L. K., Skinner, M. W., Holden, T. A., & Demorest, M. E. (2002). Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy. *Ear and Hearing*, *23*, 463–476.

Kiefer, J., von Ilberg, C., Rupprecht, V., Hubner-Egner, J., & Knecht, R. (2000). Optimized speech understanding with the continuous interleaved sampling speech coding strategy in patients with cochlear implants: effect of variations in stimulation rate and number of channels. *Annals of Otology, Rhinology, and Laryngology*, *109*, 1009–1020.

Kong, Y. Y., Cruz, R., Jones, J. A., & Zeng, F. G. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing*, *25*, 173–185.

Loizou, P. C., Poroy, O., & Dorman, M. (2000). The effect of parametric variations of cochlear implant processors on speech understanding. *Journal of the Acoustical Society of America*, *108*, 790–802.

McKay, C. M., McDermott, H. J., & Clark, G. M. (1994). Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. *Journal of the Acoustical Society of America*, *96 (5 Pt 1)*, 2664–2673.

Middlebrooks, J. C. (2004). Effects of cochlear-implant pulse rate and inter-channel timing on channel interactions and thresholds. *Journal of the Acoustical Society of America*, *116*, 452–468.

Nelson, P. B., Jin, S. H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: cochlear

implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*, *113*, 961–968.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, *95*, 1085–1099.

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

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech-Perception without Traditional Speech Cues. *Science*, *212*, 947–950.

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*, 367–373.

Rubinstein, J. T., Wilson, B. S., Finley, C. C., & Abbas, P. J. (1999). Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation. *Hearing Research*, *127*, 108–118.

Seligman, P., & McDermott, H. (1995). Architecture of the Spectra 22 speech processor. *Annals of Otology, Rhinology, and Laryngology Supplement*, *166*, 139–141.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., & Wang, X. (1999). Consonant recordings for speech testing. *Journal of the Acoustical Society of America*, *106*, L71–L74.

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

Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., & Holden, T. A. (2002). Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear and Hearing*, *23*, 207–223.

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

Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America*, *68*, 836–842.

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

Vandali, A. E., Whitford, L. A., Plant, K. L., & Clark, G. M. (2000). Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear and Hearing*, *21*, 608–624.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). Better Speech Recognition with Cochlear Implants. *Nature*, *352*, 236–238.

Wilson, B. S., Finley, C. C., Lawson, D. T., & Zerbi, M. (1997). Temporal representations with cochlear implants. *American Journal of Otology*, *18 (6 Suppl)*, S30–S34.

Xu, L., & Pfingst, B. E. (2003). Relative contributions of spectral and temporal cues for phoneme perception as revealed by acoustic simulations of cochlear implants. *Abstract of the 26th annual research meeting, 26.*

Xu, L., Tsai, Y., & Pfingst, B. E. (2002). Features of stimulation affecting tonal-speech perception: implications for cochlear prostheses. *Journal of the Acoustical Society of America*, *112*, 247–258.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*, 946–953.

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

Zeng, F. G., & Galvin, J. J. (1999). Amplitude mapping and phoneme recognition in cochlear implant listeners. *Ear and Hearing*, *20*, 60–74.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., & Bhargave, A., et al. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Science of the United States of America*, *102*, 2293–2298.

REFERENCE NOTES

1 Diagnostic Interface Box (DIB), the University of Innsbruck, 2001.

2 RIB Research Interface Box System: Manual V1.0, the University of Innsbruck, 2001.

Erratum

Gates, G. (2006) Letter to the Editor. *Ear and Hearing, 27,* 91.

In the letter, the author was listed as Geifrgi A. Gates. The author's correct name is George A. Gates.