Title
Speech and melody recognition in binaurally combined acoustic and electric hearing

Permalink
https://escholarship.org/uc/item/4dr5v9z3

Journal
Journal of the Acoustical Society of America, 117(3)

ISSN
0001-4966

Authors
Kong, Y Y
Stickney, G S
Zeng, F G

Publication Date
2005-03-01

Peer reviewed

Speech and melody recognition in binaurally combined acoustic and electric hearing

Ying-Yee Kong[a]
Hearing and Speech Research Laboratory, Department of Cognitive Sciences, University of California–Irvine, Irvine, California 92697

Ginger S. Stickney
Hearing and Speech Research Laboratory, Department of Otolaryngology—Head and Neck Surgery, University of California–Irvine, Irvine, California 92697

Fan-Gang Zeng[b]
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology—Head and Neck Surgery, University of California–Irvine, Irvine, California 92697

(Received 24 July 2003; revised 10 December 2004; accepted 10 December 2004)

Speech recognition in noise and music perception is especially challenging for current cochlear implant users. The present study utilizes the residual acoustic hearing in the nonimplanted ear in five cochlear implant users to elucidate the role of temporal fine structure at low frequencies in auditory perception and to test the hypothesis that combined acoustic and electric hearing produces better performance than either mode alone. The first experiment measured speech recognition in the presence of competing noise. It was found that, although the residual low-frequency (<1000 Hz) acoustic hearing produced essentially no recognition for speech recognition in noise, it significantly enhanced performance when combined with the electric hearing. The second experiment measured melody recognition in the same group of subjects and found that, contrary to the speech recognition result, the low-frequency acoustic hearing produced significantly better performance than the electric hearing. It is hypothesized that listeners with combined acoustic and electric hearing might use the correlation between the salient pitch in low-frequency acoustic hearing and the weak pitch in the envelope to enhance segregation between signal and noise. The present study suggests the importance and urgency of accurately encoding the fine-structure cue in cochlear implants. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1857526]

PACS numbers: 43.66.Ts, 43.71.Ky [PFA] Pages: 1351–1361

I. INTRODUCTION

Cochlear implants have successfully restored partial hearing in severely hearing-impaired individuals. Recent studies have reported that many implant users can recognize 70%–80% of sentences presented in quiet. However, understanding speech in noise and music appreciation still remains a challenge for most implant users, due to the limitations of the electrode design and the signal processing scheme employed in current cochlear implants.

The poor speech perception in noise and music appreciation in cochlear-implant listeners are mainly due to their inability to encode pitch. The limited spectral resolution, especially the inaccurate encoding of low-frequency information, is believed to be the main reason for their poor pitch perception performance. Low-frequency information is important for both musical and voice pitch perception. It has been shown that speech recognition in the presence of a competing talker can be achieved by segregating the components of each voice using the fundamental frequency ($F0$) as a cue. Brokx and Nooteboom (1982) showed that listeners could identify the keywords in sentences more accurately against a background of competing speech by increasing the difference in $F0$. The $F0$ cue was shown to be effective in segregating competing voices even at low signal-to-noise ratios when the target speech did not show distinct peaks in the spectrum (Summerfield and Culling, 1992).

A pitch percept in normal auditory system can be elicited by either the place mechanism with resolved low-numbered harmonics or by the temporal mechanism following the temporal fine structure of the input signal. However, both pitch encoding mechanisms fail in current cochlear implants. The low-frequency information is neither appropriately represented by the place of stimulation nor by the temporal fine structure of the neural firing pattern. First, the relatively shallow insertion depth of present electrode arrays severely limit the transfer of low-frequency spectral information. The average insertion depth for the Nucleus implant was estimated to be 20 mm (Ketten *et al.*, 1998), which corresponds to the acoustic frequency lower limit of about 1000 Hz (Greenwood, 1990). Even with the latest electrode designs (such as the Clarion HiFocus, Nucleus Contour, and Med-El Combi40+), which are intended to provide a deeper insertion of up to 30 mm, there is still no guarantee that

---

[a] Current address: MRC Cognition and Brain Sciences Unit, 15 Chaucer Rd., Cambridge, CB2 2EF, UK; electronic mail: ying-yee.kong@mrc-cbu.cam.ac.uk

[b] Electronic mail: fzeng@uci.edu

low-frequency neurons can be stimulated due to both reduced nerve survival in deafened individuals and nontonotopic distribution of low-frequency neurons in the cochlea (Nadol *et al.*, 1989; Linthicum *et al.*, 1991). Second, low-frequency temporal information is not appropriately encoded in current speech processing strategies. All current coding strategies, except for the analog-based strategies such as the Compressed Analog (CA, Eddington, 1980) and Simultaneous Analog Stimulation (SAS, Kessler, 1999), extract only the temporal envelope of incoming signals from 6 to 22 frequency bands using a low-pass filter with a cutoff frequency below 500 Hz and amplitude modulate it to a fixed-high-rate pulsatile carrier. In these strategies, the low-frequency temporal information, namely the slowly-varying envelope (<50 Hz) and the periodicity cues (50–500 Hz) (Rosen, 1992), can be preserved in the temporal envelope, but they are encoded in the ‘‘wrong places,’’ i.e., locations in the cochlea that are tuned to higher frequencies. Furthermore, fine structure information of the input signal, which is the phase information defined mathematically by the Hilbert transform (Hilbert, 1912), is discarded in such processing schemes due to the usage of a fixed-rate carrier. While the low-frequency information conveyed by the temporal envelope can support speech recognition in quiet (e.g., Shannon *et al.*, 1995), it is not sufficient to support speech recognition in noise with limited spectral cues (e.g., Fu *et al.*, 1998; Zeng and Galvin, 1999; Qin and Oxenham, 2003; Stickney *et al.*, 2004) and robust pitch perception (e.g., Burns and Viemeister, 1981; Green *et al.*, 2002; Kong *et al.*, 2004).

As the audiological criteria for implant candidacy have become less stringent, individuals with substantial residual low-frequency hearing have received cochlear implants. Recent development of short-electrode arrays allows the preservation of low-frequency acoustic hearing in these patients (von Ilberg *et al.*, 1999; Gantz and Turner, 2003). For those who are implanted with the conventional long-electrode arrays, low-frequency acoustic information is also available by combining electric hearing with acoustic hearing from the nonimplanted ear (Dooley *et al.*, 1993; Tyler *et al.*, 2002). Availability of these individuals allows a unique opportunity to study the role of fine-structure information at low frequencies in auditory perception, particularly in tasks that depend on pitch perception (i.e., music perception and speech recognition in the presence of a competing talker).

Previous studies on speech perception with binaurally combined acoustic and electric hearing revealed mixed results (for adults, see Dooley *et al.*, 1993; Armstrong *et al.*, 1997; Tyler *et al.*, 2002; Ching *et al.*, 2004; for children, see Chmiel *et al.*, 1995; Ching *et al.*, 2001). Chmiel *et al.* (1995) reported significantly better speech performance in quiet with a combined use of hearing aids and cochlear implants in three of the six subjects. Similar results were also reported by Armstrong *et al.* (1997) and Ching *et al.* (2001 and 2004), showing better sentence and phoneme recognition performance with combined acoustic and electric hearing in both quiet and in multi-talker babble noise at a 10 dB signal-to-noise ratio. Anecdotally, some implant users reported that the additional low-frequency acoustic information improved both sound quality and sound localization (Armstrong *et al.*, 1997; Tyler *et al.*, 2002). Moreover, two of the three subjects tested in Tyler *et al.* (2002) reported that the acoustic and electric signals fused to form one integrated sound image. Potential incompatibility between acoustic and electric hearing has also been reported. For example, Tyler *et al.* (2002) reported that one of their subjects heard the acoustic and electric stimuli as separate sound sources. Blamey *et al.* (1996 and 2000) demonstrated a pitch mismatch and differences in the dynamic range and the shape of the iso-loudness curves between the acoustically and electrically stimulated ears. Dooley *et al.* (1993) also reported that some subjects discontinued using their hearing aids or cochlear implants after implantation. However, for those patients who adapted to both devices, the incompatibility between the two percepts did not seem to interfere with their speech recognition in both quiet and noise (Dooley *et al.*, 1993; Tyler *et al.*, 2002).

The first study by von Ilberg *et al.* (1999) on combined acoustic and electric hearing with specially designed short-electrodes showed better speech recognition in quiet with the additional low-frequency acoustic hearing compared to electric hearing alone. The range of improvement was 4 to 70 percentage points depending on the filtering configuration of the cochlear implant. A recent study by Turner and colleagues (2004) on short-electrodes also showed significant benefits of additional low-frequency acoustic hearing in speech recognition in noise. They compared the speech reception thresholds of spondee words in different noise backgrounds (steady-state noise versus competing sentences) in implant users with ‘‘short-electrodes’’ and traditional ‘‘long-electrodes.’’ They reported that speech reception thresholds improved by 15 dB in the competing talker background and 5 dB in the steady-state noise background in combined hearing recipients with the short-electrode implants compared to the traditional long-electrode users. They concluded that the better speech recognition performance in multi-talker babble noise with the additional low-frequency acoustic hearing was attributed to the ability of the listeners to take advantage of the voice differences between the target and the masker speech.

One of the reasons that current cochlear implant listeners have great difficulty in understanding speech in a fluctuating background of other talkers (Nelson *et al.*, 2003; Stickney *et al.*, 2004) is their impaired pitch perception ability. The goal of the present study was to investigate how residual low-frequency hearing from the nonimplanted ear provides information that is necessary for pitch perception, and in turn improves speech and music perception in cochlear implant listeners. Two experiments were conducted to reveal the role of low-frequency acoustic hearing in realistic listening situations that are exceptionally challenging for cochlear implant users. The first experiment was designed to evaluate speech recognition in the presence of another speech sound in three listening conditions: hearing aid (HA) alone, cochlear implant (CI) alone, and cochlear implant plus hearing aid (CI +HA). The second experiment was designed to evaluate melody recognition with primarily pitch cues in cochlear-implant users in the same three listening conditions. We hypothesized that the additional low-frequency acoustic infor-

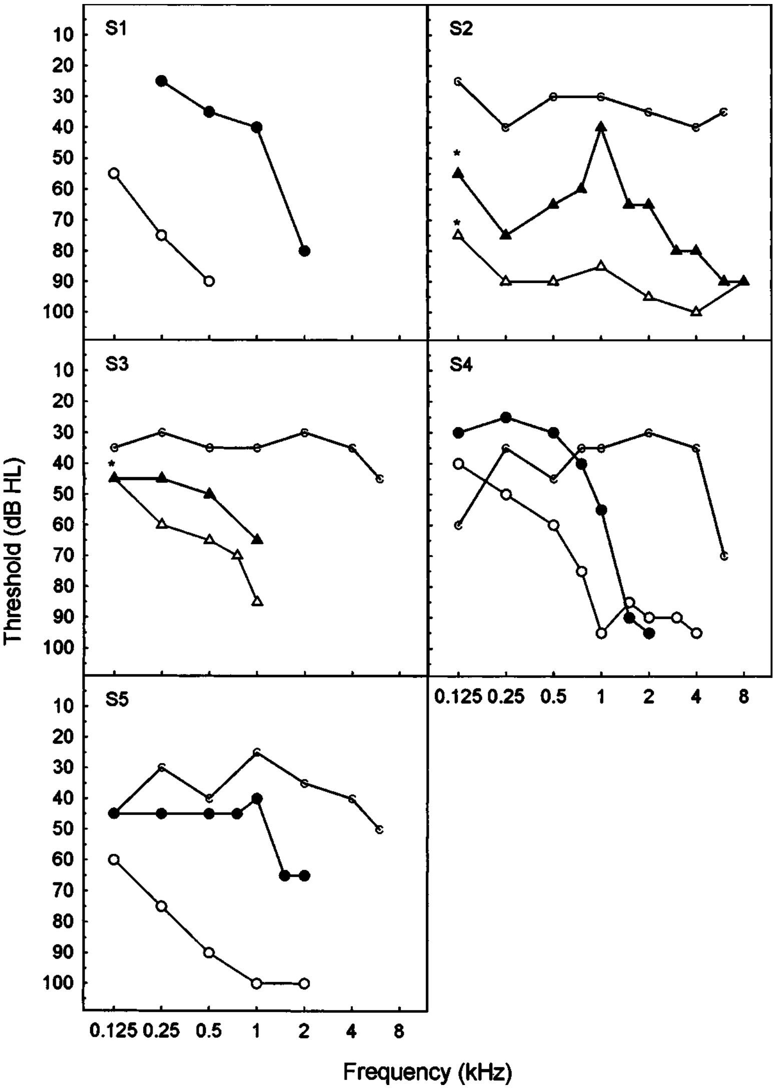

FIG. 1. Aided (closed symbols) and unaided (open symbols) thresholds in the nonimplanted ear (circles for the right ear; triangles for the left ear) and cochlear implant ear (indicated as "C"). Only thresholds at or below 100 dB HL were shown. The asterisk above the symbol indicates vibrotactile response. Implant thresholds in subject S1 were not obtained.

mation from the nonimplanted ear would provide more accurate pitch information to aid perceptual segregation of competing voices and to contribute significantly to musical pitch perception.

II. SPEECH RECOGNITION IN NOISE

A. Methods

1. Subjects

Four cochlear-implant subjects with significant residual acoustic hearing in the nonimplanted ear were recruited to participate in this study. They were two females and two males, with ages ranging from 49 to 79. Figure 1 shows their aided and unaided thresholds in the nonimplanted ear and the thresholds in the implanted ear. Their unaided thresholds showed moderate to profound loss at frequencies from 125 to 8000 Hz, but their aided thresholds showed only mild to severe loss at frequencies at or below 1000 Hz. The aided threshold averaged over 125, 250, and 500 Hz was 30, 70, 48, and 28 dB HL for subjects S1, S2, S3, and S4, respectively. While all other subjects had better aided thresholds below 1000 Hz, subject S2 had the lowest threshold (40 dB HL) at 1000 Hz and poorer thresholds below 1000 Hz. Thresholds from the cochlear implant for subjects S2, S3, and S4 were between 25 and 70 dB HL from 125 to 6000 Hz. The implant thresholds for S1 were not tested. Three out of the four subjects (S1, S3, and S4) continued to use their hearing aids on a daily basis, whereas S2 discontinued using his hearing aid after implantation in spite of residual hearing in the nonimplanted ear. Subject S2 did not use his hearing aid because of poor speech recognition rather than any perceived incompatibility between his hearing aid and cochlear implant.

All subjects were postlingually deafened and had at least one year of implant usage at the time of the test. They were native speakers of American English. Table I shows additional information regarding hearing history and implant type. Two subjects had the Clarion device, with S1 having the Clarion precurved electrode and S2 having the Clarion Hi Focus II with the positioner. The remaining two subjects (S3 and S4) used the Nucleus 24 device. Subject S1 used two different speech processing strategies: simultaneous analog stimulation (SAS) and multiple pulsatile sample (MPS), depending on the listening situations. Subject S2 and S4 used the continuous interleaved sampling (CIS) strategy, and S3 used the advanced combination encoder (ACE) strategy.

TABLE I. Biographical information on five cochlear-implant subjects.

Subject	Age[a]	Mus.[b]	Age onset[c]	Etiology	Yrs. exp.[d]	Device[e]	Strategy[f]	HA use[g]	Consonant[h] (%)	Vowel[i] (%)
S1	49	>20	4	Unknown	4	Clarion precurved	SAS/MPS	Y	46/54	51/40
S2	50	0	25	Unknown	2	Clarion HiFocus II	CIS	N	86	68
S3	69	0	37	Unknown	3	Nucleus 24	ACE	Y	54	51
S4	79	0	36	Unknown	1	Nucleus 24	CIS	Y	58	45
S5	19	0	3	Unknown	3	Clarion precurved	SAS	Y	41	33

[a] Age of the subject at the time of the experiments.
[b] Years of formal musical training.
[c] Age at the onset of hearing loss.
[d] Years of experience with the implant.
[e] Implant type.
[f] Processing strategy in the speech processor used during the experiments.
[g] Consistent use of hearing aid in the nonimplanted ear after implantation.
[h] Score (% correct) on consonant recognition in quiet in /aCa/ context.
[i] Score (% correct) on vowel recognition in quiet in /hVd/ context.

2. Stimuli

A subset of IEEE sentences (1969) recorded by Hawley *et al.* (1999) was used in this experiment. Each list consisted of ten sentences with five keywords per sentence. The target sentence was spoken by a male voice. Another sentence (competing sentence) spoken by a different male talker or by a female talker was used as a masker. The same competing sentence was used throughout testing (“Port is a strong wine with a smoky taste”). The target sentence was either presented alone or in the presence of the masker. The target and masker had the same onset, but the masker’s duration was always longer than the target sentence. The target sentence was presented at approximately 65 dBA whereas the level of the masker varied from 45 dB to 65 dBA to produce five signal-to-noise ratios (SNR): +20, +15, +10, +5, and 0 dB.

3. Procedure

Subjects were evaluated under three listening conditions: hearing aid (HA) alone, cochlear implant (CI) alone, and cochlear implant with hearing aid (CI+HA). The subject’s cochlear implant was turned off in the HA alone condition, and their hearing aid was turned off and the nonimplanted ear was plugged in the CI alone condition. These three listening conditions were evaluated in random order for each subject.

All tests were performed in a double-walled sound-treated booth. Both the target and masker sentences were presented via a loud speaker directly in front of the subject. Subjects used their own hearing aid and cochlear implant volume and sensitivity settings during the entire test session. All subjects were tested with the male masker, with subjects S2, S3, and S4 also being tested in a second test session with the additional female masker. S1 was not available for the second test session with the female masker. Prior to the test session, subjects were presented with two practice sessions of ten sentences each binaurally. In the first practice session, subjects were presented with sentences in quiet. The second practice session was used to familiarize listeners listening to the target sentence in the presence of the masker. In this practice session, two sentences were presented for each of the five SNR conditions used in the actual experiment. In the test session, each subject was presented with all five SNR conditions in a random order. There were 10 randomized sentences (5 keywords each), for a total of 50 keywords per SNR and 50 sentences for the test session. The subjects typed their responses at the keyboard and were encouraged to guess if unsure. Responses were collected and scored in terms of the number of words correctly identified using MATLAB software.

B. Results

Figure 2 shows percent correct scores as a function of SNR for sentence recognition in the presence of the male masker for both individual and average data (bottom right panel). Panels S1a and S1b represent results from subject S1 using the SAS and the MPS strategy, respectively. Results from the three listening conditions, hearing aid (HA) alone, cochlear implant (CI) alone, and the combined devices (CI +HA), are represented by closed squares, closed circles, and open triangles, respectively.

Both the individual and average data show the same trend: the hearing aid alone produced essentially zero speech recognition across different SNRs [$F(4,16)=1.88$, $p>0.05$], while the cochlear implant alone and the combined devices produced monotonically increasing performance as a function of SNR [CI alone: $F(4,16)=10.98$, $p<0.01$; CI +HA: $F(4,16)=23.09$, $p<0.001$]. The most interesting finding is that the combined hearing produced significantly better performance than the CI alone particularly in the higher SNR conditions by an average of 8 percentage points at a 15 dB SNR [$F(1,4)=10.78$, $p<0.05$] and 20 percentage points at a 20 dB SNR [$F(1,4)=27.13$, $p<0.01$].

Figure 3 shows sentence recognition scores with the female masker from subjects S2, S3, and S4. Similar to the male masker condition, qualitative trends were observed with the female masker: (1) the HA alone produced essentially zero speech recognition, (2) both the CI alone and the combined devices produced monotonically increasing perfor-

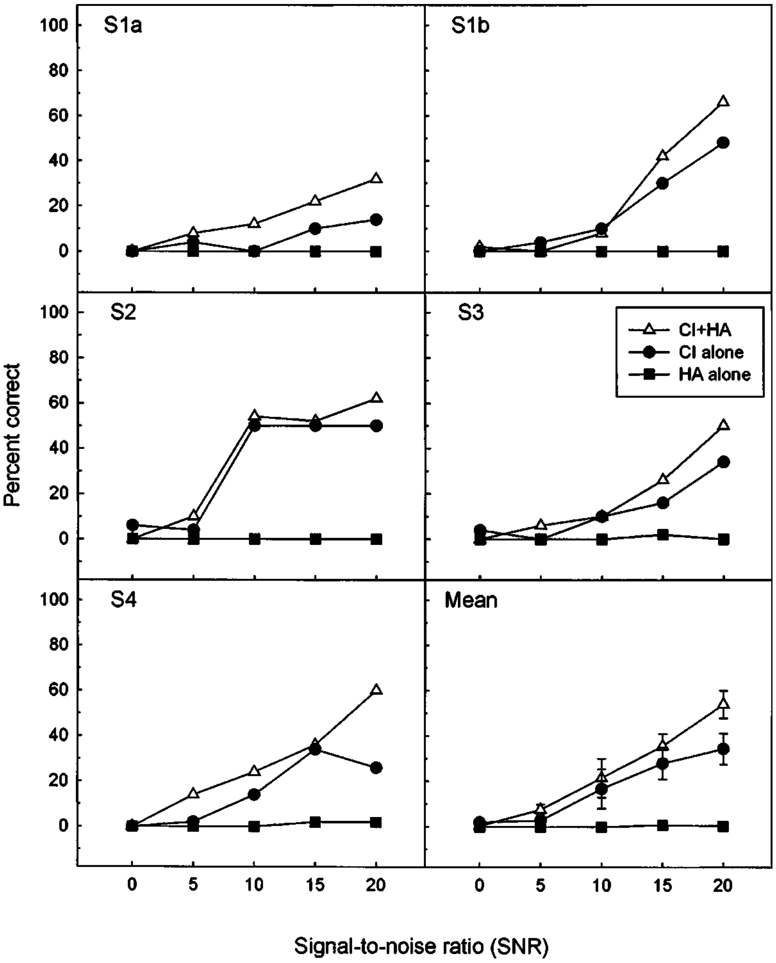

FIG. 2. Individual and mean sentence recognition scores (% correct) as a function of signal-to-noise ratio (SNR) with the male masker. The three functions in each panel represent the CI+HA (open triangles), CI alone (closed circles), and HA alone (closed squares) conditions. The vertical error bars in the mean graph represent the standard error of the mean.

mance as a function of SNR, and (3) the combined hearing produced better performance than either mode alone. The benefits of combined hearing compared to CI alone were observed in all subjects, with average improvement ranging from 8 to 25 percentage points from 0 to 20 dB SNR. However, due to the limited number of subjects and the large

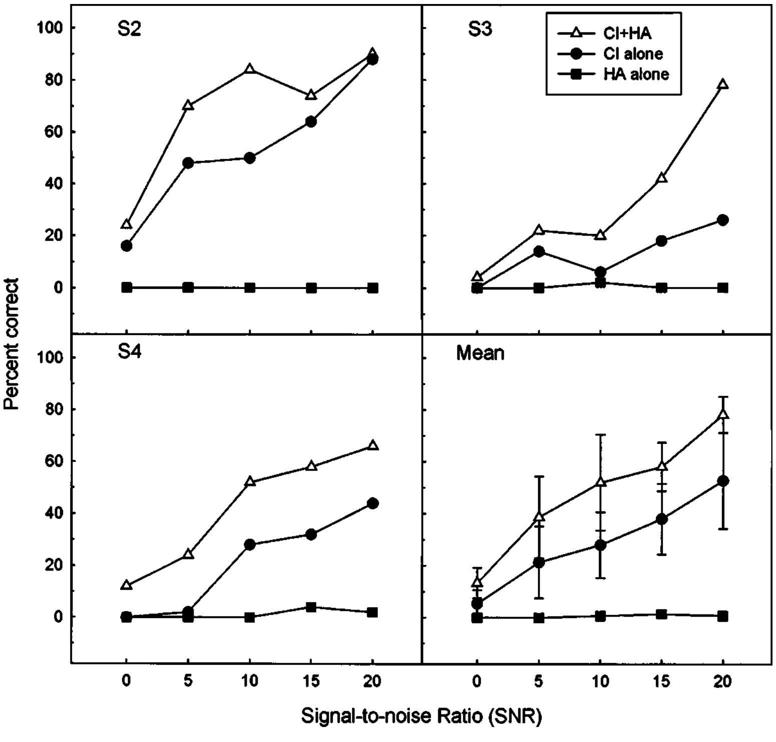

FIG. 3. Sentence recognition scores (% correct) as a function of SNR for subjects S2, S3, and S4 with the female masker.

intersubject variability of performance, the superior performance of combined hearing over CI alone was only found to be significant at 15 [$F(1,2)=19.10$, $p<0.05$] and 10 dB SNRs [$F(1,2)=18.75$, $p<0.05$]. Large differences between the male and the female maskers were observed in the combined hearing condition at all SNRs with an average of 21 percentage points better with the female masker than with the male masker. Significant difference between the two maskers in combined hearing was found in more challenging SNRs at 15 [$F(1,2)=100.00$, $p<0.01$] and 10 dB SNRs [$F(1,2)=19.24$, $p<0.05$]. In contrast, with the CI alone, the improvement of speech recognition with the female masker compared to the male masker was relatively small (average 9 percentage points) and no significant difference between maskers was found at any SNR [$F(1,2)=2.07$, $p>0.1$]. The improvement of the combined hearing over the CI alone was much greater for the female masker (average 19 percentage points) than the male masker (average 7 percentage points). For example, subject S2 improved by about 5 percentage points with the male masker, but the improvement with the female masker was 26 and 34 percentage points at 5 and 10 dB SNR, respectively. Similarly, subject S3 improved by 15 or less percentage points with the male masker, but the improvement with the female masker was 24 and 52 percentage points at 15 and 20 dB SNR, respectively.

III. MELODY RECOGNITION

A. Methods

1. Subjects

Five cochlear-implant subjects, including the same four cochlear-implant subjects from experiment I and an additional subject S5 participated in the melody recognition experiment. Only S1 had extensive musical training, while the rest had very limited music experience. Subject S5 was a non-native speaker of English, but he attended kindergarten in the United States at age 6. He reported learning all the melodies, except one, used in the experiment at a young age and was able to hum the tunes of these melodies. He was implanted with the Clarion precurved electrode and was using the SAS processing strategy. Like most of the subjects in this experiment, S5 continued to use his hearing aid on a daily basis. His average aided threshold for 125, 250, and 500 Hz was 45 dB HL.

2. Stimuli

Three sets of 12 familiar melodies, played by single notes, were generated using a software synthesizer (ReBirth RB-338, version 2.0.1). For each melody, rhythmic information was removed by using notes of the same duration (quarter notes with 350 ms in duration) with a silent period of 150 ms between notes. Therefore, pitch was the only available cue for melody recognition. Each melody consisted of 12–14 notes of its initial phrase. Three sets of the twelve melodies were generated in low-, mid-, and high-frequency ranges. In the low-frequency melody condition, all melodies were within a frequency range from 104 (G#2) to 261 Hz (C4), whereas the mid- (208 to 523 Hz) and high-range (414 to 1046 Hz) melodies were one and two octaves above the low-range melodies, respectively. The largest semitone difference between the highest and the lowest notes of the melody was 16 and the smallest difference was 7. Table II shows the titles of the melodies used in this experiment and the frequency components of each melody (for detailed information, see Kong *et al.*, 2004).

TABLE II. The 12 familiar melodies and their frequency ranges.

Melody[a]	Low range	Mid range	High range	Largest interval[b]	Int extent[c]
1	131–220 (C3-A3)	261–440 (C4-A4)	522–880 (C5-A5)	5th	9
2	122–184 (B2-F#3)	245–369 (B3-F#4)	490–738 (B4-F#5)	m3rd	7
3	110–174 (A2-F3)	220–348 (A3-F4)	440–696 (A4-F5)	4th	8
4	110–184 (A2-F#3)	220–369 (A3-F#4)	440–738 (A4-F#5)	6th	9
5	110–174 (A2-F3)	220–348 (A3-F4)	440–696 (A4-F5)	4th	7
6	131–196 (C3-G3)	261–392 (C4-G4)	522–784 (C5-G5)	m3rd	7
7	110–220 (A2-A3)	220–440 (A3-A4)	440–880 (A4-A5)	Octave	12
8	146–220 (D3-A3)	292–440 (D4-A4)	584–880 (D5-A5)	4th	7
9	131–196 (C3-G3)	261–392 (C4-G4)	522–784 (C5-G5)	5th	7
10	103–261 (G#2-C4)	207–523 (G#3-C5)	414–1046 (G#4-C6)	m6th	16
11	104–233 (G#2-A#3)	207–466 (G#3-A#4)	414–932 (G#4-A#5)	4th	14
12	110–184 (A2-F#3)	220–369 (A3-F#4)	440–738 (A4-F#5)	5th	9

[a] 1=Twinkle, Twinkle, Little Star; 2=This Old Man; 3=She'll be Coming Round the Mountain; 4=Old MacDonald Had a Farm; 5=Lullaby, and Good Night; 6=Mary Had a Little Lamb; 7=Take Me Out to the Ball Game; 8=London Bridge is Falling Down; 9=Happy Birthday; 10=Star Spangled Banner; 11=Auld Lang Syne; 12=Yankee Doodle.

[b] Largest interval in melody (m=minor).

[c] Range in semitones between the highest and the lowest notes in the melodies.

3. Procedure

All subjects were tested in three listening conditions (HA alone, CI alone, and CI+HA) and three melody conditions (low, mid, and high) for a total of 9 conditions. For the HA alone and CI alone conditions, stimuli were presented at the subject's most comfortable level while they wore their hearing aid or cochlear implant at their usual settings. For the combined CI+HA condition, the presentation level was set the same as in the HA alone condition while the speech processor volume was adjusted to achieve the most comfortable loudness. The presentation level ranged from 70 to 85 dB SPL.

The titles of the 12 melodies were displayed on a computer screen and the subject was asked to choose the melody that was presented. A practice session with feedback was given before the actual test. For each experimental condition, melodies were presented three times in random order. Repetition of the stimulus was not allowed and visual feedback regarding the correct response was given immediately after

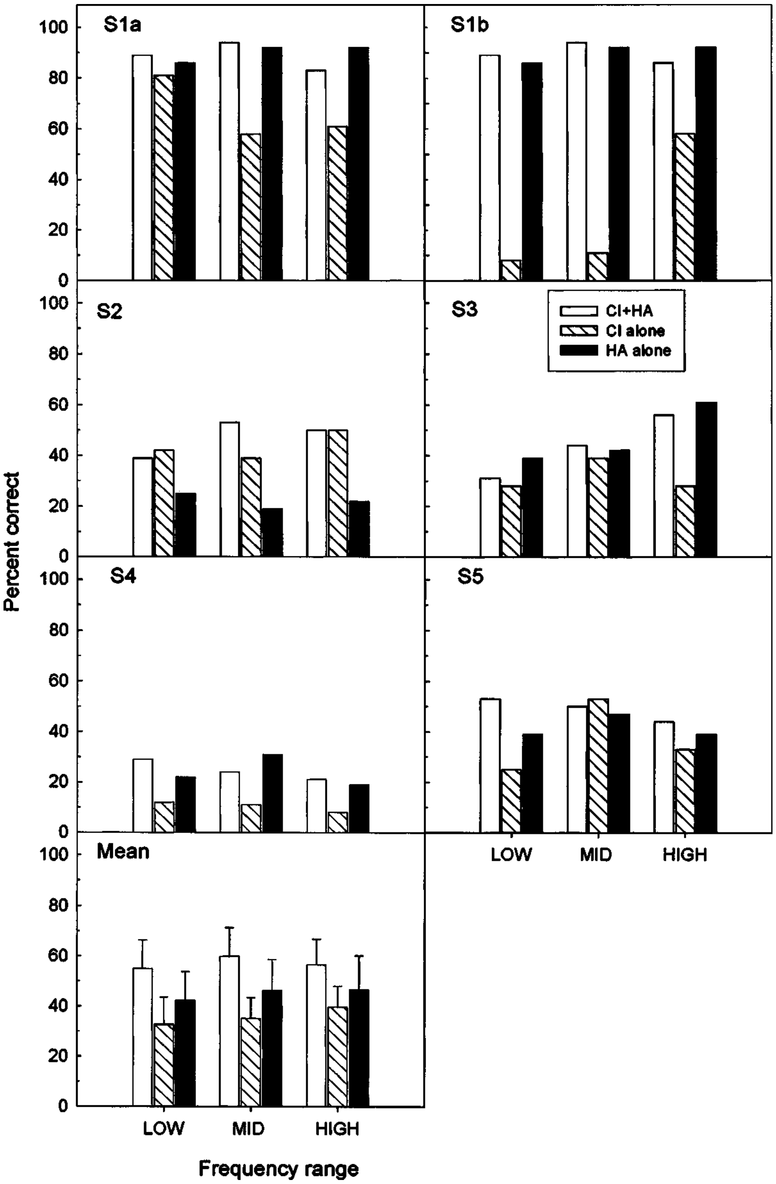

FIG. 4. Individual and mean melody recognition scores (% correct) for the three listening (CI alone, HA alone, and CI+HA) and three melody conditions (Low, Mid, and High). Vertical error bars represent the standard error of the mean.

the subject's response. As in experiment I, all three melody and all three listening conditions were presented in random order.

B. Results

Figure 4 shows individual and mean melody recognition results for the three melody (low, mid, and high) and listening conditions (HA alone=closed bars, CI alone=slanted bars, and CI+HA=open bars). Panels S1a and S1b represent results from subject S1 using the SAS and MPS strategies, respectively. Melody recognition performance varied remarkably from subject to subject in all listening conditions. Performance ranged from an average of 19% for S4 to 90% for S1 in the HA alone condition, from 8% for S4 to 81% for S1 in the CI alone condition, and from 21% for S4 to 92% for S1 in the CI+HA condition. Consistent with Kong *et al.* (2004), a difference in processing strategies was observed, with the SAS strategy producing better melody recognition than CIS-type strategy. For subject S1, her SAS strategy produced 73, 47, and 3 percentage points better performance than her MPS strategy for the low-, mid-, and high-range melodies, respectively. However, inconsistent with earlier studies on melody recognition, the melody recognition performance with cochlear implants alone in some of the subjects was considerably better than the chance performance level (e.g., Gfeller *et al.*, 2002; Kong *et al.*, 2004). It should be noted that the melody recognition performance with the mid-range melodies reported in Kong *et al.* (2004) was primarily obtained from cochlear implant users with the older devices (Clarion precurved and Nucleus-22) and with the envelope extraction processing strategies, namely MPS, CIS, and SPEAK. Preliminary data from our laboratory on a small group of users implanted with Nucleus 24, Clarion HiFocus II, and Med-El devices showed a different level of performance, with some performing similarly to the older device users at chance level and others in the range of 40%–80% correct. The reasons for this remarkable difference in performance between the newer and older devices will need further investigation, but it is not in the scope of discussion in this study.

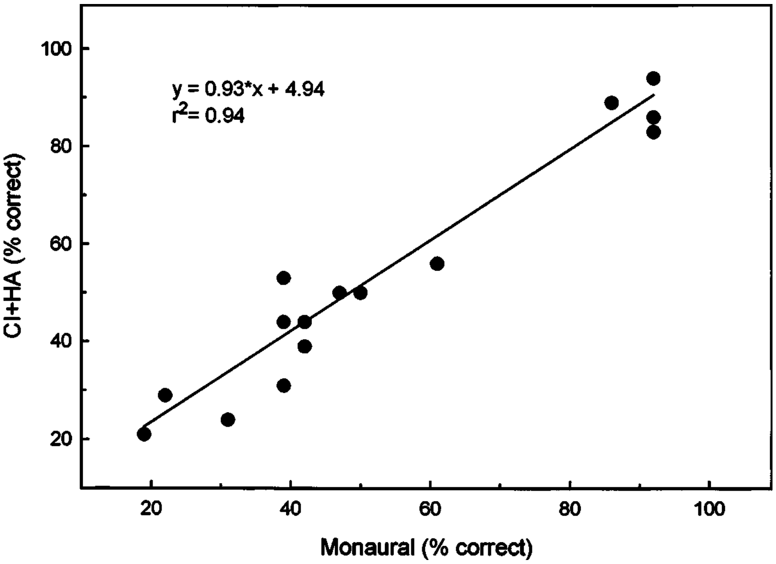

FIG. 5. Correlation of melody recognition (% correct) with binaural (CI+HA) and monaural (HA alone or CI alone) stimulation. Monaural data from S1a, S1b, S3, S4, and S5 are from the HA alone condition, but from the CI alone condition for S2. The slope of the regression line is 0.93 and the intercept is 4.94.

With the HA alone, the average melody recognition performance across all subjects and conditions was 45% correct. This is in direct contrast with the score of 0% obtained for the speech recognition in noise task from the first experiment. On average, HA alone produced an average of 17 percentage points better melody recognition than the averaged CI alone performance, but showed similar performance to the combined hearing condition. These patterns of results were observed in four out of the five subjects. The only exception was subject S2, who discontinued the regular use of his hearing aid after implantation and had very unusually poor aided thresholds in the frequency range (<1000 Hz) that was tested in this experiment. Due to the large intersubject differences in the melody recognition scores, a repeated measures ANOVA did not show significant difference between the HA alone and CI alone performance. Nevertheless, the trend of better performance with the HA alone than the CI alone was found in 14 out of the 18 cases, the probability of obtaining this result by chance is only 1.2%.

IV. DISCUSSION

A. Comparison between speech and melody recognition

Cochlear implant speech recognition performance in quiet has improved with advances in technology, but speech recognition in competing backgrounds and music perception remains challenging for implant users. One of the reasons for their poor performance in speech recognition in noise and music perception is their impaired pitch perception ability caused by both the limitations of the electrode design and the signal processing scheme employed in current cochlear implants. We hypothesize that providing the additional fine-structure information at low frequencies via the nonimplanted ear may allow for better encoding of pitch, which in turn can improve music appreciation and enhance speech recognition in competing backgrounds.

The present study showed that speech recognition in noise improved with combined acoustic and electric hearing compared to electric hearing alone, consistent with findings in earlier combined hearing studies (Armstrong *et al.*, 1997; Ching *et al.*, 2001 and 2004; Tyler *et al.*, 2002) and the results reported regarding "short-electrode" cochlear implants (Turner *et al.*, 2004). Turner *et al.* (2004) showed significantly lower spondee word reception thresholds in the presence of two simultaneously presented sentences than in steady-state white noise in "short-electrode" users, but not in the traditional long-electrode users. In contrast to speech recognition, the advantage of combined hearing was not observed in the melody recognition task. Instead, the performance with combined hearing was determined by the better ear (i.e., acoustic ear in S1, S3, S4, and S5, implant ear in S2), as indicated by Fig. 5 showing a highly significant correlation between the binaural and the best monaural conditions [$r^2 = 0.94$, $p < 0.001$] and close to the unit slope (0.93) of the linear regression function.

The differential speech and music results reinforce the recently reported dichotomies in auditory perception, i.e., the fine-structure cue at low frequencies dominates pitch perception while the envelope cue dominates speech recognition (Smith *et al.*, 2002). The present results demonstrate this dichotomy with opposite patterns of results between hearing aid and cochlear implant performance in speech and melody recognition: the hearing aid (containing the fine-structure cue at low frequencies) produced no speech recognition but significant melody recognition, while the cochlear implant (containing the temporal envelope cue) produced significant speech recognition but relatively poor melody recognition. The inability to recognize speech with only low-frequency information is consistent with classic articulation index studies, where low-pass filtered speech ($\lesssim 800$ Hz) was relatively unintelligible (e.g., French and Steinberg, 1947; Pavlovic *et al.*, 1986). However, this additional low-frequency acoustic information, when combined with the cochlear implant, produced significantly better speech recognition in noise than the implant alone condition.

B. Auditory segregation and grouping

The superior speech recognition performance in binaurally combined hearing over the CI alone may arise from the benefits of (1) binaural processing including the binaural squelch effect (Carhart, 1965; Colburn, 1977) and/or diotic summation, a small benefit arising from listening with two ears compared to one ear with the identical signal and noise (Day *et al.*, 1988), (2) a monaurally based grouping and segregation mechanism, or (3) a combination of both. However, we argue that the presently observed improved speech recognition in noise with binaurally combined acoustic and electric hearing cannot be due to the binaural advantage. First, there are apparently no preserved level and phase differences between the acoustic and electric hearing, as required by the traditional binaural squelch. Both the target speech and masker were presented directly in front of the subjects in our study. Second, the advantage from diotic summation is small (Cox *et al.*, 1981) and it results mainly in better speech recognition in quiet (Kaplan and Pickett, 1981). This cannot account for the considerably large improvement of speech recognition in noise (averaged 19 percentage points with the female masker) with the combined hearing in our study. Third, similar improvement was obtained with combined acoustic and electric hearing on the same side with the short-electrode implant, providing evidence strongly against the binaural advantage hypothesis. Fourth, speech recognition in noise was improved more with the female masker than with the male masker, suggesting a monaurally based grouping and segregation mechanism.

Previous studies in which speech recognition in noise improved with the separation of the fundamental frequency have demonstrated the importance of voice pitch cues for segregating speech from competing backgrounds (e.g., Brokx and Nooteboom, 1982; Gardner *et al.*, 1989; Assmann and Summerfield, 1990). During voicing, the pulsing of the vocal folds gives rise to a consistent pattern of periodicity in the time wave form and corresponding harmonicity in the spectrum. Different from acoustic hearing, low harmonics cannot be resolved in current cochlear implants. The only pitch information available in the implants is from the reduced salience pitch cue provided by the temporal envelope (Burns and Viemeister, 1981; Faulkner *et al.*, 2000; Green *et al.*, 2002). The nonsalience of temporal envelope pitch can be demonstrated by the much poorer discriminability of modulation frequency and electric pulse rate than pure-tone frequency discrimination (Formby, 1985; Grant, 1998; Zeng, 2002). Thus, we hypothesize that the pitch difference in the temporal envelope is not robust enough to reliably separate the target and masker, particularly when both are dynamic speech sounds (e.g., Green *et al.*, 2002). We further hypothesize that the fine-structure information at low frequencies in the combined acoustic and electric hearing provides better $F0$ information that allows the cochlear-implant users to segregate the target from the masker. In the present study, the average fundamental frequency was 108 Hz for the target, 136 Hz for the male masker, and 219 Hz for the female masker (measured by the STRAIGHT program, courtesy of Kawahara, 1997). The significantly better speech recognition performance with the female masker compared to the male masker supported the idea that the availability of the fundamental frequency cue in combined acoustic and electric hearing was critical for separating the target speech from the masker speech.

The encoding of voice pitch in normal-hearing listeners can be achieved by the place coding or temporal coding mechanism, or both. A number of models have been proposed to investigate the auditory and perceptual processes by which normal-hearing listeners utilize the $F0$ difference when identifying the constituents of double vowels. Assmann and Summerfield (1990) tested different models to predict performance in normal-hearing listeners in identifying concurrent vowels with different fundamental frequencies. They reported that the place-time models, which estimated voice pitch using a periodicity analysis of the wave forms in each channel, were superior to the place models in the context of vowel identification. A purely temporal model by Meddis and Hewitt (1992) could even predict the improvement of segregation of simultaneous vowels as a function of $F0$ difference based on the pooled periodicity information which were summed across channels. These models suggested that temporal information, namely the periodicity cues, are critical for the segregation of competing sound sources.

The underlying mechanism for segregating competing sounds with combined acoustic and electric hearing is unclear. We propose that the segregation of target speech from the masker is based on the temporal periodicity cues in both the acoustic and electric signals. While the periodicity cue carried in the envelope alone does not provide sufficient $F0$ sensitivity to perceptually segregate target speech from the masker (Faulkner *et al.*, 2000; Green *et al.*, 2002), the presence of the additional salient temporal fine-structure cue at low frequencies in acoustic hearing, which is correlated with the periodicity cue in the temporal envelope in electric hearing, increases perceptual segregation between the signal and noise as well as improves grouping of the signal and that of the noise. This hypothesis is consistent with the recently reported poor (23% correct) speaker identification performance (Vongphoe and Zeng, 2004) and the absence of talker effect for speech recognition in the presence of a competing talker (Stickney *et al.*, 2004) in cochlear implant users. Even though there is no direct evidence to support this hypothesis at this stage, several predictions can be made to test its validity in the future. For example, should fundamental frequency be the main cue used by low-frequency acoustic hearing to improve electric hearing, we would predict minimal improvement for voiceless speech segments. Additionally, any mismatch between the fundamental frequency provided by acoustic hearing and the temporal envelope provided by electric hearing would result in a reduced benefit in combined hearing.

V. CONCLUSIONS

The present study implicates a dichotomy between the envelope and fine-structure cues at low frequencies in speech and melody recognition. The temporal envelope cue is sufficient for speech recognition, but not for melody recognition. On the other hand, the fine-structure cue at low frequencies is sufficient for pitch perception, but not for speech recognition. However, when the fine-structure cue in acoustic hearing is combined with the envelope cue in electric hearing, significant improvement can be observed in speech recognition in a competing background. The greatest improvement was observed when the target and the masker had the largest difference in fundamental frequency, suggesting a monaurally based grouping mechanism rather than a binaurally based mechanism for the observed advantage with the combined acoustic and electric hearing.

The present study suggests the importance of appropriately encoding the fine-structure cue in cochlear implants. Although this fine-structure cue at low frequencies produces negligible intelligibility for speech recognition in quiet, it is critical for music perception, speech recognition in noise, and other listening situations including speaker identification and sound source segregation.

ACKNOWLEDGMENTS

We thank our cochlear-implant subjects for their time and dedication. We also thank Kaibao Nie for programming the user interface, J. Ackland Jones for generating the music stimuli and assisting in data collection and manuscript preparation, and Peter Assmann and Chris Turner for helpful suggestions. This work was supported in part by the National Institutes of Health, Department of Health and Human Services (A research supplement award to Y.-Y.K., F32 DC005900-01 to G.S.S., and 2 RO1-DC-02267 to F.G.Z.).

Armstrong, M., Pegg, P., James, C., and Blamey, P. (**1997**). ‘‘Speech perception in noise with implant and hearing aid,’’ Am. J. Otolaryngol. **18**, S140–S141.

Assmann, P. F., and Summerfield, Q. (**1990**). ‘‘Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,’’ J. Acoust. Soc. Am. **88**, 680–696.

Blamey, P. J., Dooley, G. J., James, C. J., and Parisi, E. S. (**2000**). ‘‘Monaural and binaural loudness measures in cochlear implant users with contralateral residual hearing,’’ Ear Hear. **21**, 6–17.

Blamey, P. J., Dooley, G. J., Parisi, E. S., and Clark, G. M. (**1996**). ‘‘Pitch comparsions of acoustically and electrically evoked auditory sensations,’’ Hear. Res. **99**, 139–150.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ‘‘Intonation and the perception separation of simultaneous voices,’’ J. Phonetics **10**, 23–36.

Burns, E. M., and Viemeister, N. F. (**1981**). ‘‘Played again SAM: Further observations on the pitch of amplitude-modulated noise,’’ J. Acoust. Soc. Am. **70**, 1655–1660.

Carhart, R. (**1965**). ‘‘Monaural and binaural discrimination against competing sentences,’’ Int. J. Audiol. **4**, 5–10.

Ching, T. Y. C., Incerti, P., and Hill, M. (**2004**). ‘‘Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears,’’ Ear Hear. **25**, 9–21.

Ching, T. Y. C., Psarros, C., Hill, M., Dillon, H., and Incerti, P. (**2001**). ‘‘Should children who use cochlear implants wear hearing aids in the opposite ear?,’’ Ear Hear. **22**, 365–380.

Chmiel, R., Clark, J., Jerger, J., Jenkins, H., and Freeman, R. (**1995**). ‘‘Speech perception and production in children wearing a cochlear implant in one ear and a hearing aid in the opposite ear,’’ Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 314–316.

Colburn, H. S. (**1977**). ‘‘Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise,’’ J. Acoust. Soc. Am. **61**, 525–533.

Cox, R., DeChicchis, A. R., and Wark, D. (**1981**). ‘‘Demonstration of binaural advantage in audiometric test rooms,’’ Ear Hear. **2**, 194–201.

Day, G., Browning, G., and Gatehouse, S. (**1988**). ‘‘Benefit from binaural hearing aids in individuals with a severe hearing impairment,’’ Br. J. Audiol. **22**, 273–277.

Dooley, G. J., Blamey, P. J., Seligman, P. M., Alcantara, J. I., Clark, G. M., Shallop, J. K., Arndt, P., Heller, J. W., and Menapace, C. M. (**1993**). ‘‘Combined electrical and acoustical stimulation using a bimodal prosthesis,’’ Arch. Otolaryngol. Head Neck Surg. **119**, 55–60.

Eddington, D. K. (**1980**). ‘‘Speech discrimination in deaf subjects with cochlear implants,’’ J. Acoust. Soc. Am. **68**, 885–891.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). ‘‘Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: implications for cochlear implants,’’ J. Acoust. Soc. Am. **108**, 1877–1887.

Formby, C. (**1985**). ‘‘Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners,’’ J. Acoust. Soc. Am. **78**, 70–77.

French, N. R., and Steinberg, J. C. (**1947**). ‘‘Factors governing the intelligibility of speech sounds,’’ J. Acoust. Soc. Am. **19**, 90–119.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). ‘‘Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,’’ J. Acoust. Soc. Am. **104**, 3586–3596.

Gantz, B. J., and Turner, C. W. (**2003**). ‘‘Combining acoustic and electrical hearing,’’ Laryngoscope **113**, 1726–1730.

Gardner, R. B., Gaskill, S. A., and Darwin, C. J. (**1989**). ‘‘Perceptual grouping of formants with static and dynamic differences in fundamental frequency,’’ J. Acoust. Soc. Am. **85**, 1329–1337.

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., Witt, S., and Stordahl, J. (**2002**). ‘‘Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults,’’ Cochlear Implants Int. **3**, 29–53.

Grant, K. W., Summers, V., and Leek, M. R. (**1998**). ‘‘Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners,’’ J. Acoust. Soc. Am. **104**, 1051–1060.

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

Greenwood, D. D. (**1990**). ‘‘A cochlear frequency-position function for several species—29 years later,’’ J. Acoust. Soc. Am. **87**, 2592–2605.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). ‘‘Speech intelligibility and localization in multi-source environment,’’ J. Acoust. Soc. Am. **105**, 3436–3448.

Hilbert, D. (**1912**). *Grundzuge einer Allgemeinen Theorie der linearen Integralgleichungen* (Teubner, Leipzig).

IEEE (**1969**). ‘‘IEEE recommended practice for speech quality measurement,’’ IEEE Trans. Audio Electroacoust. **17**, 225–246.

Kaplan, H., and Pickett, J. (**1981**). ‘‘Effects of dichotic/diotic versus monotic presentation on speech understanding in noise in elderly hearing-impaired listeners,’’ Ear Hear. **2**, 202–207.

Kawahara, H. (**1997**). ‘‘Speech representation and transformation using adaptive interpolation of weighted spectrum: VOCODER revisited,’’ Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp. 1303–1306.

Kessler, D. K. (**1999**). ‘‘The Clarion mulit-strategy cochlear implant,’’ Ann. Otol. Rhinol. Laryngol. **108**, 8–16.

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., and Neely, J. G. (**1998**). ‘‘In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays,’’ Ann. Otol. Rhinol. Laryngol. Suppl. **175**, 1–16.

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

Linthicum, F. H., Fayad, J., Otto, S. R., Galey, F. R., and House, W. F. (**1991**). ‘‘Cochlear implant histopathology,’’ Am. J. Otolaryngol. **12**, 245–311.

Meddis, R., and Hewitt, M. J. (**1992**). ‘‘Modeling the identification of concurrent vowels with different fundamental frequencies,’’ J. Acoust. Soc. Am. **91**, 233–245.

Nadol, J. B., Young, Y.-S., and Glynn, R. J. (**1989**). ‘‘Survival of spiral ganglion cells in profound sensorineural hearing loss: Implications for cochlear implantation,’’ Ann. Otol. Rhinol. Laryngol. **98**, 411–416.

Nelson, P., Jin, S.-H., Carney, A., and Nelson, D. (**2003**). ‘‘Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners,’’ J. Acoust. Soc. Am. **113**, 961–968.

Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). ‘‘An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals,’’ J. Acoust. Soc. Am. **80**, 50–57.

Qin, M. K., and Oxenham, A. J. (**2003**). ''Effects of simulated cochlear-implant processing on speech reception in fluctuating masker,'' J. Acoust. Soc. Am. **114**, 446–454.

Rosen, S. (**1992**). ''Temporal information in speech and its relevance for cochlear implants,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

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

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

Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). ''Cochlear implant speech recognition with speech masker,'' J. Acoust. Soc. Am. **116**, 1081–1091.

Summerfield, Q., and Culling, J. F. (**1992**). ''Auditory segregation of competing voices: Absence of effects of FM or AM coherence,'' Philos. Trans. R. Soc. London, Ser. B **336**, 357–366.

Turner, C. W., Gantz, B. J., Vidal, C., and Behrens, A. (**2004**). ''Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing,'' J. Acoust. Soc. Am. **115**, 1729–1735.

Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (**2002**). ''Patients utilizing a hearing aid and a cochlear implant: speech perception and localization,'' Ear Hear. **23**, 98–105.

Von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Sturzebecher, E., and Klinke, R. (**1999**). ''Electric-acoustic stimulation of the auditory system,'' ORL J. Otorhinolaryngol. Relat. Spec. **61**, 334–340.

Vongphoe, M., and Zeng, F.-G. (**2004**). ''Speaker recognition with temporal cues in acoustic and electric hearing,'' J. Acoust. Soc. Am. (submitted).

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

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