

Cochlear Implant Melody Recognition as a Function of Melody Frequency Range, Harmonicity, and Number of Electrodes

Sonya Singh,¹ Ying-Yee Kong,² and Fan-Gang Zeng³

Objective: The primary goal of the present study was to determine how cochlear implant melody recognition was affected by the frequency range of the melodies, the harmonicity of these melodies, and the number of activated electrodes. The secondary goal was to investigate whether melody recognition and speech recognition were differentially affected by the limitations imposed by cochlear implant processing.

Design: Four experiments were conducted. In the first experiment, 11 cochlear implant users used their clinical processors to recognize melodies of complex harmonic tones with their fundamental frequencies being in the low (104–262 Hz), middle (207–523 Hz), and high (414–1046 Hz) ranges. In the second experiment, melody recognition with pure tones was compared to melody recognition with complex harmonic tones in four subjects. In the third experiment, melody recognition was measured as a function of the number of electrodes in five subjects. In the fourth experiment, vowel and consonant recognition were measured as a function of the number of electrodes in the same five subjects who participated in the third experiment.

Results: Frequency range significantly affected cochlear implant melody recognition, with higher frequency ranges producing better performance. Pure tones produced significantly better performance than complex harmonic tones. Increasing the number of activated electrodes did not affect performance with low- and middle-frequency melodies but produced better performance with high-frequency melodies. Large individual variability was observed for melody recognition, but its source seemed to be different from the source of the large variability observed in speech recognition.

Conclusion: Contemporary cochlear implants do not adequately encode either temporal pitch or place pitch cues. Melody recognition and speech recognition require different signal processing strategies in future cochlear implants.

(*Ear & Hearing* 2009;30:160–168)

INTRODUCTION

Cochlear implants electrically stimulate the auditory nerve to restore hearing for those suffering from severe to profound hearing loss. Although great strides have been made in speech recognition for cochlear implant users, their music appreciation, particularly melody recognition, is still poor. With the exception of a recent study (Nimmons et al. 2008), which showed that an implant user (L4) scored 81% in a 12-item closed-set melody recognition task, all studies have found low-level performance (12–40% correct) in melody recognition tasks without rhythmic cues. As an extreme example, a “star” cochlear implant listener could hold a successful phone conversation but might not be able to tell apart two simple

nursery melodies (e.g., Gfeller et al. 2002; Kong et al. 2004; Spahr & Dorman 2004).

Both acoustic and perceptual differences between speech and music sounds contribute to this somewhat surprising dichotomy. In general, spectral and temporal envelope cues contribute to speech intelligibility (e.g., Remez 1981; Shannon et al. 1995), whereas spectral and temporal fine-structure cues contribute to melody recognition (e.g., Smith et al. 2002). Harmonicity is a particularly important fine structure that consists of multiples of the fundamental frequency. A harmonic sound usually produces salient pitch in normal-hearing listeners but seems to be inaccessible to cochlear-implant listeners.

Current cochlear implants all use similar signal processing: The algorithm first filters a sound into 12 to 24 narrowband signals and then extracts the temporal envelope from all or some of these narrowband signals with the highest levels; the algorithm compresses the temporal envelope, uses it to amplitude-modulate a fixed rate carrier, and stimulates the electrode corresponding to the narrowband signal’s center frequency (McDermott et al. 1992; Vandali et al. 2000; Wilson et al. 1991). The signal-processing algorithm in current cochlear implants provides a reasonably good representation of spectral and temporal envelope cues but cannot encode spectral and temporal fine-structure cues accurately. For instance, formants can be approximated by relative changes in current levels across the intracochlear electrode array, whereas harmonic fine structure is not explicitly extracted, and even if it is extracted, there is no guarantee that it will be encoded properly because of mismatched frequency-to-place maps in cochlear implants (e.g., Boex et al. 2006). Similarly, temporal envelopes can be accurately represented by amplitude modulation of a relatively high-rate carrier, but temporal fine structure is lost in cochlear implants (e.g., Wilson et al. 1997).

Overall, recent studies examining relative contributions of spectral and temporal cues have focused on speech stimuli with little or no attention paid to music perception (Geurts & Wouters 2001; Green et al. 2002, 2004; Laneau et al. 2004; Nie et al. 2006; Qin & Oxenham 2005; Xu et al. 2002; Zeng et al. 2005). These previous studies suggest that the current cochlear implants produce weak perception of voice pitch. Here, we evaluated relative contributions of spectral and temporal cues to melody recognition in cochlear implant listeners. First, we systemically measured recognition of melodies in three frequency ranges: low (104–262 Hz), middle (207–523 Hz), and high (414–1046 Hz). The low melody range was similar to the voice fundamental frequency range, whereas the high range was similar to the first formant frequency range (Hillenbrand et al. 1995). Because a 200 to 400 Hz low-pass cutoff frequency

¹School of Medicine, University of California; Departments of ²Cognitive Sciences, and ³Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology—Head and Neck Surgery, University of California, Irvine, California.

is typically used to extract temporal envelope, pitch information in the low and middle melody ranges would likely be conveyed by the temporal envelope cue, whereas pitch information in the high range would likely be conveyed by the electrode place cue in modern multi-electrode cochlear implants. Experiment 1 will explicitly test this hypothesis.

Second, we compared melody recognition between pure tones and complex harmonic tones over the same three frequency ranges. A pure tone is likely to stimulate one electrode or, at most, two electrodes if the frequency of a note happens to fall between two adjacent filters. Therefore, a pure tone should allow for a clearer pitch contour based on a single excitation site than a complex harmonic tone allows. Even if the complex harmonic tone can be resolved by cochlear implant filters, they will unlikely maintain the tonotopic relationship required for harmonic pitch perception because of the mismatched frequency-to-place map (e.g., Boex et al. 2006; Oxenham et al. 2004). Experiment 2 will test the hypothesis that melodies consisting of pure tones are easier to recognize than those consisting of complex harmonic tones, particularly in the low-frequency range where the cochlear implant filters' frequency resolution is high.

Third, we tested melody recognition while varying the number of activated electrodes. Previous studies found that speech recognition improved with an increased number of electrodes but reached a plateau unexpectedly with 7 to 10 electrodes (Fishman et al. 1997; Friesen et al. 2001; Garnham et al. 2002). Because pitch perception of complex harmonic tones and speech recognition rely on different acoustic cues, one might predict a different pattern of results between melody recognition and speech recognition as a function of the number of electrodes. To explicitly test this hypothesis, experiment 3 tested melody recognition, whereas experiment 4 tested speech recognition as a function of the number of electrodes as a comparison.

EXPERIMENT 1: MELODY RECOGNITION AS A FUNCTION OF FREQUENCY RANGE

Previous studies on cochlear implant melody performance used a similar 200 to 600 Hz middle-frequency frequency range and found generally poor performance in different subject populations with different devices (Gfeller et al. 2002;

Gfeller & Lansing 1991; Kong et al. 2004; McDermott 2004; Nimmons et al. 2008; Spahr & Dorman 2004). To determine the contribution of the stimulus to melody recognition independent of the contribution of the device and subject, we measured within-subject performance across a wide range of melody frequencies in a sample size of 11 cochlear implant users.

Methods

Subjects • Eleven cochlear implant users (C1–C11) were recruited based on availability for this experiment. They all grew up in the United States and were familiar with melodies used in the present study. None of them had received formal music training before or as a part of this study. Some of them had participated in earlier research in our laboratory. They were all native English speakers and had relatively good phoneme recognition in quiet (48–91% correct for consonants and 31–82% for vowels). Table 1 describes age at the time of testing, etiology, age of onset, duration of deafness, duration of implant use, implant device, speech-processing strategy, number of electrodes, and other relevant information. The average age of the participants was 56 yr at the time of the experiment. The average duration of implant use was 3 yr. Depending on the device, as well as availability, the number of activated electrodes could be 12, 14, 16, or 20.

Stimuli • Twelve isochronous melodies were used, including: “Old MacDonald Had a Farm”; “Twinkle, Twinkle, Little Star”; “London Bridge is Falling Down”; “Mary Had a Little Lamb”; “This Old Man”; “Yankee Doodle”; “She’ll be Coming ’Round the Mountain”; “Happy Birthday”; “Lullaby, and Goodnight”; “Take Me Out to the Ball Game”; “Auld Lang Syne”; and “Star Spangled Banner”. The fundamental frequency of these melodies spanned three different frequency ranges: low (104–262 Hz), middle (207–523 Hz), and high (414–1046 Hz). The low-range melodies were one octave lower, whereas the high-range melodies were one octave higher than the middle-range melodies, which had been used in a previous study conducted in the same laboratory (Kong et al. 2005). These melodies were created using a software synthesizer (ReBirth RB-338, version 2.0.1). Each melody had 12 to 14 notes and contained no rhythmic cues. Each note had the same duration (350 MS) with a silent period of 150 msec

TABLE 1. Background information for the 11 cochlear implant subjects involved in the study

Subject	Age	Etiology	Onset Age	Onset Type	Duration of Deafness	Duration of CI	Device-Strategy	# Electrodes	Cons	Vowel
C1	70	Tumor and virus	63	Sudden	1	7	N24-ACE	20	84	41
C2	72	Unknown	56	Progressive	11	4	N24-ACE	16	54	51
C3	79	Unknown	65	Progressive	12	2	N24-CIS	12	58	49
C4	42	Otosclerosis	36	Progressive	3	3	N24-SPEAK	20	78	64
C5	51	Unknown	45	Progressive	2	4	CII-CIS	16	86	68
C6	69	Nerve damage	50	Sudden	16	3	CII-CIS	14	74	70
C7	26	Genetic	Birth	Congenital	19	7	Medel-CIS	12	88	67
C8	41	High fever	6	Sudden	31	4	Medel-CIS	12	48	31
C9	29	Genetic	Birth	Congenital	27	2	CII-CIS	16	NT	NT
C10	61	Genetic	40	Progressive	7	1	N24-ACE	20	91	82
C11	77	Blood clot	38	Sudden	38	1	N24-ACE	20	76	64

Age, onset age, duration of deafness, and duration of cochlear implant (CI) use variables all expressed in the unit of years.

Cons represents percent correct scores for consonant recognition in quiet. Vowel represents percent scores for vowel recognition in quiet. NT = Not Tested.

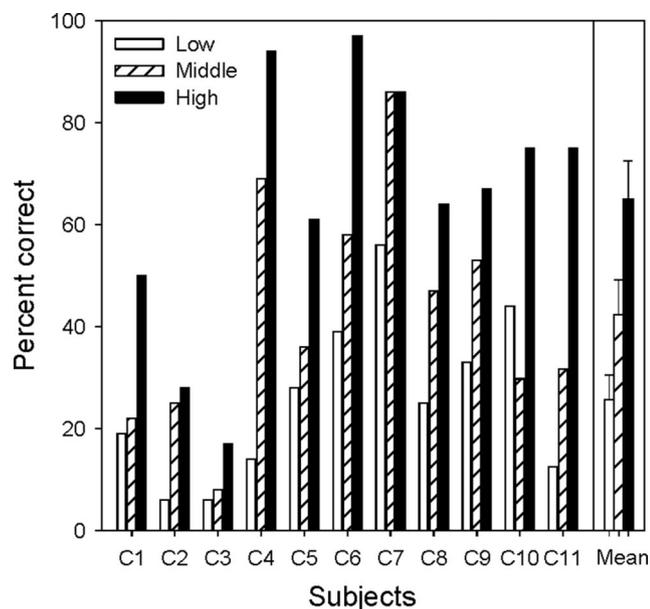


Fig. 1. Melody recognition scores in the low-, middle-, and high-frequency ranges for 11 cochlear implant users, as well as their mean scores. Error bars represent one standard error from the mean.

between notes. The harmonic structure contained no apparent resonance frequencies and had a roughly -20 dB per octave slope from 200 to 4000 Hz.

Procedure • A 12-item closed-set, forced-choice procedure was used. All subjects used their clinical maps in this experiment. The melodies were presented via direct connection from the speech processor to the computer. The volume was adjusted to a comfortable level as judged by each individual subject. Melodies with the same frequency range were treated as a block, and the presentation order of the blocks was randomized. Within each block, each melody was played three times in random order, producing a block of 36 melodies. The subjects had a short practice session (10–15 min) before taking the formal test to become familiar with both the graphic-user interface and the stimuli. The subjects received feedback on accuracy during training but no feedback during testing.

Results

Figure 1 shows percent correct scores for melody recognition as a function of melody frequency range (low range = unfilled bars; middle range = hatched bars; high range = filled bars) by 11 cochlear implant users. The mean melody recognition score at each frequency range plus one standard error from the mean is depicted at the rightmost position on the x axis. Note the monotonic increase in performance as the melody frequency range is increased. A repeated-measures analysis of variance (ANOVA) showed that the melody frequency range was a significant factor influencing cochlear implant melody recognition ($F_{(2,20)} = 43.0$; $p < 0.001$). On average, the percent score increased significantly from 25.7% for low-frequency melodies to 42.3% for middle-frequency melodies and to 64.9% for high-frequency melodies (post hoc, two-tailed, paired t -test; $p < 0.025$).

Also note the large individual variability in Figure 1. The individual percent scores ranged from chance performance at 6

(C2 and C3) to 56% (C7) for the low-range melodies, from 8 (C3) to 86% (C7) for the middle-range melodies, and from 17% (C3) to almost perfect performance at 97% (C6) for the high-range melodies.

EXPERIMENT 2: MELODY RECOGNITION WITH PURE-TONE STIMULI

Because experiment 1 used complex harmonic tones, their pitch can be theoretically conveyed by multiple mechanisms, including the salient pitch of the resolved lower harmonics and the less salient pitch of the unresolved harmonics (Oxenham et al. 2004; Ritsma 1967; Schouten et al. 1962). The variable, but generally poor, performance in implant melody recognition in experiment 1 could be a result of poor temporal representation, poor place representation, or both. Experiment 2 used pure tones as the stimuli in an attempt to improve melody recognition. The working hypothesis was that a pure tone would accurately represent the temporal pitch at least up to the roughly 400 Hz envelope extraction frequency (Kiefer et al. 2000; Loizou et al. 2003; Vandali 2001). In addition, the pure tone would represent the place pitch more accurately than the complex harmonic tone because the former activates one or two electrodes, whereas the latter activates more electrodes.

Four of the 11 cochlear implant subjects participated in this experiment (C1, C2, C8, and C9). The same three sets of 12 melodies were used in this experiment. In contrast to complex harmonic tones in experiment 1, the notes in the melodies were presented using pure tones. However, the same procedure was used in this experiment as in experiment 1.

Results

Figure 2 compares the percent of melodies identified correctly for complex harmonic tones (unfilled bars, Fig. 1) and pure tones (filled bars) in low-, middle-, and high-frequency ranges (low = top panel; middle = middle panel; high = bottom panel). Individual data are shown on the x axis, whereas the rightmost position of each panel shows the mean score and standard error from the mean. Overall, pure-tone melodies produced significantly better performance than complex harmonic tones ($F_{(1,3)} = 84.2$; $p < 0.05$). The frequency range was also a significant factor in determining the amount of improvement ($F_{(2,6)} = 15.5$; $p < 0.05$). The improvement was 20.3 percentage points in the low-frequency range, 24.5 percentage points in the middle-frequency range, and 8.0 percentage points in the high-frequency range. These results are consistent with the working hypothesis that better representation of both the temporal and place pitches produced better melody recognition with pure tones than complex harmonic tones in cochlear implants, particularly for low- and middle-range melodies. The present results also indicate that current cochlear implants cannot process complex pitch effectively.

EXPERIMENT 3: MELODY RECOGNITION AS A FUNCTION OF NUMBER OF ELECTRODES

Another means of evaluating relative contributions of temporal and spectral cues to cochlear implant performance is to reduce the number of activating electrodes systematically. This method has been used effectively to study speech recognition (Fishman et al. 1997; Friesen et al. 2001; Garnham et al. 2002)

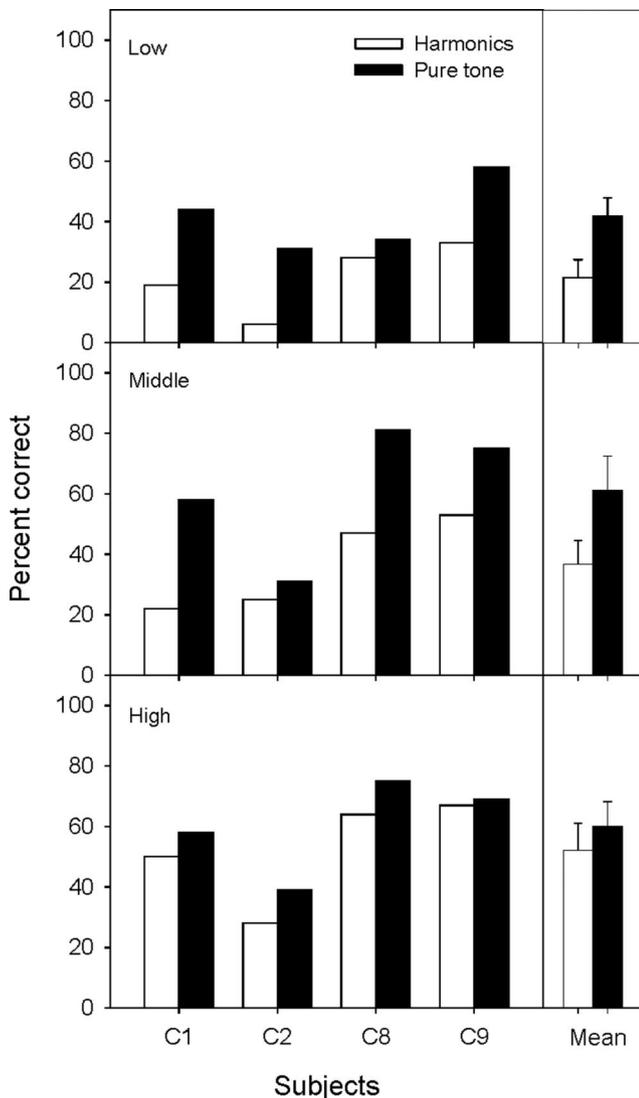


Fig. 2. Melody recognition scores for four cochlear implant users in the low-, middle-, and high-frequency ranges with melodies presented as either complex harmonic tones or pure tones. The rightmost part of the figure has mean melody recognition scores for complex harmonic tones and pure tones in each respective melody range. Error bars represent one standard error from the mean.

but has not been applied to music perception. Because recognition of speech and melody relies on different cues (Smith et al. 2002), it would be interesting to see whether melody recognition produces different patterns of results than speech recognition with a reduced number of electrodes.

Methods

Subjects • Only Nucleus users participated in the next two experiments because of limited subject availability and access to the Nucleus device-mapping interface. These Nucleus subjects were C1, C2, C3, C10, and C11.

Stimuli • The same 12 melodies consisting of complex harmonic tones were used in this experiment. Again, the melodies were presented in the low-, middle-, and high-frequency ranges.

Procedure • This experiment followed the same protocol as used in previous cochlear implant speech studies (e.g., Fishman

et al. 1997). The intent of this experimental protocol was to study the effects of processing parameters rather than the effect of learning on melody recognition. Experimental continuous interleaved sampling strategies with different numbers of electrodes were implemented using a laboratory-owned speech processor. Six conditions were created to produce experimental maps with either 1, 2, 4, 7, 10, or 20 activated electrodes for subjects C1, C10, and C11. Because C2 had only 16 functioning electrodes, this subject's six maps had either 1, 2, 4, 7, 10, or 16 electrodes. Accordingly, C3 had only four maps with either 1, 4, 7, or 12 electrodes. The test order of these experimental maps was randomized for each subject.

Table 2 shows frequency ranges mapped to the activated electrodes for the six maps in subject C1. Subject C10 and C11 had similar maps. Because subjects C2 and C3 had only 16 and 12 usable electrodes, respectively in their original map, the exact number of their activated electrodes was different from C1, C10, and C11. These electrodes were chosen to be evenly spaced and as far apart as possible. For example, subject C2 had electrodes 11 and 19 activated for the two-electrode map condition. All subjects used MP1 + 2-stimulation mode. Threshold (T) and comfortable loudness (C) levels were determined individually for each electrode. Because the per-channel stimulation rate increased with the number of activated electrodes being reduced when making a new map, the T and C levels were reassessed to ensure that the new map produced a sensation between threshold and comfortable loudness. All subjects practiced with a body-worn loaner processor until they felt comfortable to take the melody test.

Results

Figure 3 shows percent correct scores in melody recognition as a function of the number of electrodes in the low- (top panel), middle- (middle panel), and high- (bottom panel) frequency ranges. The open symbols represent individual data, whereas the thick line represents the mean data. To perform statistical analysis, subject C2's data with 16 electrodes were used as if 20 electrodes had been activated, but subject C3's data were not used because of two missing data points.

Overall, the number of activated electrodes was a significant factor ($F_{(5,15)} = 15.2$; $p < 0.05$), but the frequency range of the melodies was not significant ($F_{(2,6)} = 4.3$; $p > 0.05$). Because there was a significant interaction between the two factors ($F_{(10,30)} = 24.8$; $p < 0.05$), the effect of the number of electrodes was analyzed for each frequency range. The results showed that, on average, the implant subjects performed equally poorly (20–30% correct) regardless of the number of electrodes for both the low-frequency range ($F_{(5,15)} = 0.2$; $p > 0.05$) and the middle-frequency range ($F_{(5,15)} = 2.9$; $p > 0.05$). Note the relatively large individual variability, particularly for C1, who scored 60% correct for the low-frequency range with only 1 electrode. In contrast, the melody recognition scores increased from around 20% with 1 to 4 electrodes to 40 to 65% with 7 to 20 electrodes for the high-frequency range ($F_{(5,15)} = 34.0$; $p < 0.05$). These results suggest that current cochlear implant users only have access to the pitch cues under limited conditions. They have limited access to the temporal pitch cue

TABLE 2. Maps for subject C1 showing frequencies presented to the activated electrodes when 20, 10, 7, 4, 2, or 1 electrode(s) are activated

Electrode	20	10	7	4	2	1
Freq in Hz	187-312	187-312	187-312	187-312	187-312	187-312
(20)	312-437	312-437	312-437	312-437	312-437	312-437
Electrode	10	7	4	2	1	
Freq in Hz	187-437	187-437	187-437	187-437	187-437	
(10)	437-687	437-687	437-687	437-687	437-687	
Electrode	7	4	2	1		
Freq in Hz	187-562	187-562	187-562	187-562		
(7)	562-1062	562-1062	562-1062	562-1062		
Electrode	4	2	1			
Freq in Hz	187-1062	187-1062	187-1062			
(4)	1062-2062	1062-2062	1062-2062			
Electrode	2	1				
Freq in Hz	187-1062	187-1062				
(2)	1062-7937	1062-7937				
Electrode	1					
Freq in Hz	187-7937					
(1)						

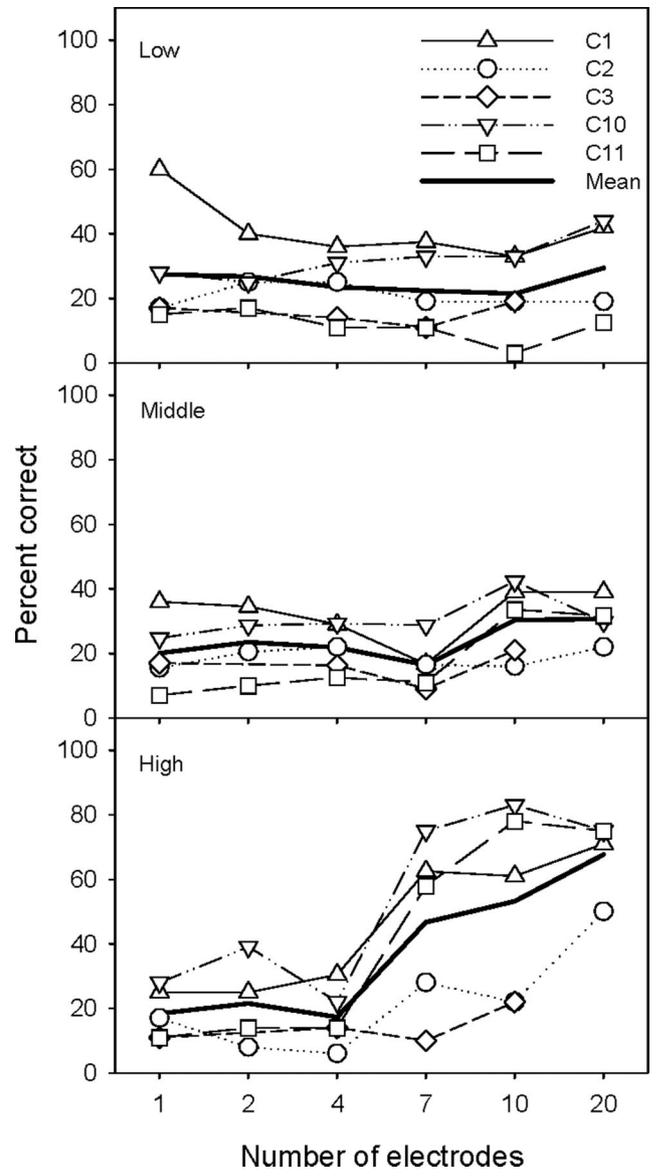


Fig. 3. Melody recognition for five cochlear implant users in the low-, middle-, and high-frequency ranges with 1, 2, 4, 7, 10, or 20 activated electrodes. The thick lines represent the mean scores.

when the pitch range is relatively low and to the place pitch cue when the melody frequency range is high (400 to 1000 Hz).

EXPERIMENT 4: COMPARISON WITH SPEECH RECOGNITION

To directly compare the difference between speech and music perception, experiment 4 measured phoneme recognition and compared it to melody recognition in the same cochlear implant subjects tested in experiment 3.

Methods

Subjects • The same five cochlear implant users who participated in experiment 3 participated in this experiment (C1, C2, C3, C10, and C11).

Stimuli • Vowel recognition was tested using 12 /hvd/ syllables, including heed, hod, hud, hid, hawed, heard, head, hood,

hayed, had, who'd, and hoed (Hillenbrand et al. 1995). Ten different talkers were used, including five males and five females.

Consonant recognition was tested using 20 /aCa/ syllables, including aba, aga, ama, asa, ava, acha, aja, ana, asha, awa, ada, aka, apa, ata, aya, afa, ala, ara, atha, and aza (Shannon et al. 1999). Eight talkers were used, including five males and three females.

Procedure • Stimuli were presented via direct connection from the sound card of a computer to a loaner speech processor. The loaner speech processor, just as in experiment 3, had both the clinical map and the experimental maps with different numbers of activated electrodes.

The subjects practiced as long as needed to feel comfortable with the interface and stimulus before taking each of the vowel and consonant tests. The vowel and consonant tests were given in a sound attenuated room. The volume for each of the tests was adjusted to a comfortable level deemed by the subject by changing both the volume on the computer and the volume on the loaner speech processor. Each token from each talker was presented three times in random order. Overall percent correct scores were reported as the outcome measure.

Results

Figure 4 shows vowel (top panel) and consonant (bottom panel) recognition scores as a function of the number of activated electrodes. Open symbols represent individual data, and the thick line represents the mean data. The chance

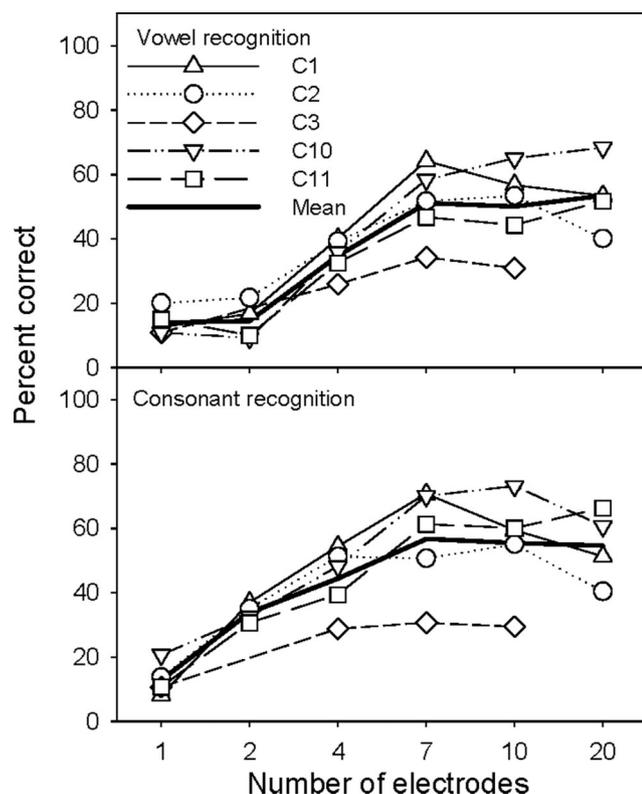


Fig. 4. Vowel and consonant recognition scores for five cochlear implant users with 1, 2, 4, 7, 10, or 20 activated electrodes. The thick lines represent the mean scores.

performance was 8% for vowel recognition and 5% for consonant recognition. Although the same overall trend was observed for both the vowel and consonant data, there was a significant difference between the sets of data ($F_{(1,3)} = 18.4$; $p < 0.05$). A post hoc test showed a significant difference for the two-electrode condition (vowel = 14% and consonant = 34%) and the four-electrode condition (vowel = 35% and consonant = 44%). In contrast to the flat pattern observed with low- and middle-range melody recognition, both vowel and consonant data showed a monotonic function with the number of electrodes ($F_{(5,15)} = 52.7$; $p < 0.05$). Similar to previous results (e.g., Fishman et al. 1997), both vowel and consonant recognition reached an asymptotic level of performance (approximately 50%) at seven electrodes.

Figure 5 contrasts mean phoneme recognition scores (open symbols) with mean melody scores (filled symbols) obtained from the same subjects. There were clear similarities and differences between melody and speech data. High-frequency melody recognition resembled the monotonically increasing function for vowel and consonant recognition, whereas low- and middle-frequency melody recognition functions were flat. ANOVA showed no statistically significant difference between high-frequency melody recognition and either consonant recognition ($F_{(1,3)} = 1.1$; $p > 0.05$) or vowel recognition ($F_{(1,3)} = 0.2$; $p > 0.05$).

However, ANOVA revealed a statistically significant difference between low-frequency melody recognition and consonant recognition ($F_{(1,3)} = 11.6$; $p < 0.05$), as well as between middle-frequency melody recognition and consonant recognition ($F_{(1,3)} = 64.1$; $p < 0.05$). There was a statistically significant difference between middle-frequency melody recognition and vowel recognition ($F_{(1,3)} = 30.6$; $p < 0.05$), but there was no statistically significant difference between low-frequency melody recognition and vowel recognition ($F_{(1,3)} = 5.6$; $p = 0.09$). If only 1-, 4-, 7-, and 10-electrode conditions were considered, which would allow for C3's data to be included, then a significant difference would be observed between low-frequency melody recognition and vowel recognition ($F_{(1,4)} = 11.3$; $p < 0.05$).

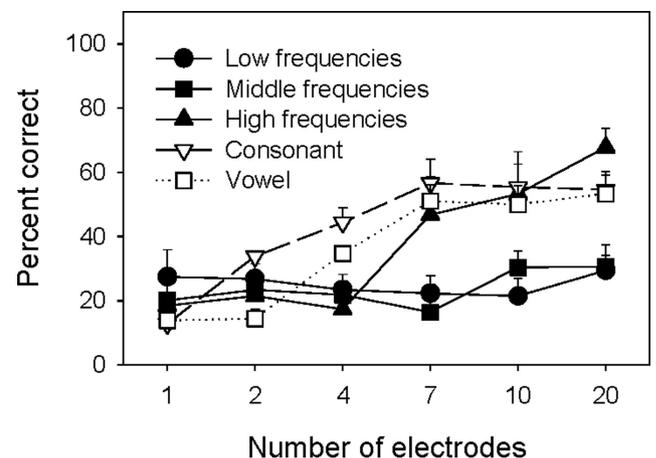


Fig. 5. Mean melody recognition scores for cochlear implant users in the low-, middle-, and high-frequency ranges using various numbers of electrodes (filled symbols). Mean vowel and consonant scores for cochlear implant users are also depicted as open symbols. Error bars represent one standard error from the mean.

DISCUSSION

Melody Recognition and Speech Recognition

To further examine the relationship between melody and speech recognition, a two-tailed, bi-variant, Pearson correlation was performed between three-range melody recognition (Fig. 1) and both consonant and vowel recognition scores (the last two columns in Table 1). Not surprisingly, there was significant correlation between low-range and middle-range melody recognition ($r = 0.64$; $p < 0.05$), low-range and high-range melody recognition ($r = 0.62$; $p < 0.05$), and middle-range and high-range melody recognition ($r = 0.80$; $p < 0.01$). Similarly, there was significant correlation between consonant and vowel recognition ($r = 0.72$; $p < 0.05$). However, there was no significant correlation between any range of melody recognition and phoneme recognition ($p > 0.05$). This lack of correlation supports the hypothesis that melody recognition and speech recognition involve two different processes (e.g., Smith et al. 2002; Zatorre et al. 2002).

Individual Variability

Although the overall performance was low, relatively large individual variability was still observed in cochlear implant melody recognition (Fig. 1). It is possible that some of this large variability was caused by device limitation. For example, subject C4, the user of the SPEAK strategy, which typically has a pulse rate of 250 Hz and, therefore, is not likely to convey useful temporal pitch cues for even the low-frequency range melodies, shows the largest difference in performance across frequency ranges of any of the subjects. To further probe possible causes and predictors of melody recognition, correlational analysis was performed between subject variables (Table 1) and melody recognition (Fig. 1). Except for a significant negative correlation ($r = -0.79$; $p < 0.01$) between age and middle-range melody recognition, no significant correlation was found among any of the three subject variables (age, duration of deafness, and duration of implant use) or any of the three ranges of melody recognition scores. The present study has a relatively small sample size, but its results suggest that, at present, we do not understand the causes of the large individual variability, nor can we predict it.

Melody Recognition and Frequency Resolution

Because psychophysical data have shown that cochlear implant users can discriminate temporal pitch up to at least 300 Hz (e.g., Pijl 1995; Zeng 2002), one might be surprised to find that the low-range melody recognition actually produced the poorest result. The reason for this discrepancy is that the psychophysical data are obtained by varying stimulation rates on a single electrode, whereas the melody recognition data are collected through the implant user's speech processor (continuous interleaved sampling, spectral peak or advanced combination encoder) that does not explicitly extract and encode the fundamental frequency. This discrepancy clearly reveals the limitation of current cochlear implants in extracting and encoding low-frequency temporal information (e.g., Geurts & Wouters 2001; Green et al. 2002, 2004; Qin & Oxenham 2005; Xu et al. 2002; Zeng et al. 2005).

Better performance with the high frequency range, as shown in the present study, is likely a result of increased frequency resolution in the frequency-to-electrode map. Current implants

typically use the Greenwood-like map, which provides proportionally more channels in the middle- to high-frequency range than the low-frequency range, at least on the logarithmic scale (Greenwood 1990). For example, to divide a broad bandwidth from 100 to 8000 Hz into 20 narrow frequency bands, the Greenwood map would allocate two channels in the 200 to 400 Hz frequency range but four to five channels in the 500 to 1000 Hz range (Table 2). Indeed, psychophysical evidence has shown lower pitch difference limens by allocating more filters in the F0 range (Geurts & Wouters 2001) or better melody recognition with semitone-spaced filters for midfrequency range melodies (Kasturi & Loizou 2007).

However, caution needs to be taken when attempting to generalize the relationship between increased frequency resolution and functional improvement (e.g., Carroll & Zeng 2007). Assuming roughly an equal frequency range so that frequency resolution is proportional to the number of electrodes, one would expect that users of the 22-electrode Nucleus device might do better than users of the 16-electrode Clarion or the 12-electrode Med-El devices. The sample size was too small in the present study to make statistically significant comparison, but no such trend was observed (Fig. 1). A much larger sample size between 15 Clarion and 15 Nucleus users found no difference in melody recognition either (Spahr & Dorman 2004). Channel interactions may be a factor limiting the extent to which spectral resolution can be improved by increasing the number of stimulating electrodes.

Pure Tones Versus Complex Harmonic Tones

The present study showed significantly better melody recognition with pure tones than with complex harmonic tones. This difference in performance demonstrates a lack of coding of complex pitch in current implants. When listening to pure tones, typically one electrode is being stimulated per note. However, when listening to complex harmonic tones, several electrodes are likely to be stimulated per note. If the lower harmonics were resolved and appropriately mapped to the right place in the cochlea, one would not expect such a difference in performance between pure tones and harmonic stimuli (Dai 2000; Oxenham et al. 2004; Ritsma 1967). Existing evidence suggests that we have not restored a pure-tone sensation in terms of frequency resolution and frequency selectivity, let alone the harmonic sensation in electric hearing (Zeng 2004).

Encoding Melody in Cochlear Implants

Although music is seldom made up of pure tones in real-life listening situations, better performance with pure tones than harmonic tones suggests that as an intermediate step, pure-tone-like stimuli can be used to improve encoding of melody in cochlear implants. The reason for taking this intermediate step is twofold. First, in the foreseeable future, it is going to be difficult to restore harmonic sensation, particularly the salient resolved harmonic pitch, in cochlear implants. Second, the pure-tone encoding approach can be relatively easily implemented to address the shortcoming of current commercial cochlear implants, which only implicitly encode the pitch information in the temporal envelope domain.

Explicit coding of pitch can be implemented by varying the carrier rate on one or more electrodes (e.g., Fearn & Wolfe 2000; Pijl 1995; Zeng 2002) or the modulation frequency of a fixed carrier (e.g., Geurts & Wouters 2001; Laneau et al. 2006; McKay et al. 1994). Melody recognition or melody-related

performance can be improved using such a strategy but often at the expense of speech recognition (e.g., Green et al. 2005).

An alternative approach is to use place cues to improve melody or voice pitch recognition. Wouters et al. have suggested that a greater number of electrodes be used than are currently allocated in the fundamental frequency range (Geurts & Wouters 2004; Laneau et al. 2004). Given the limited total available number of electrodes in the current devices, a trade-off is required in channel densities between low and high frequencies. Other options of increasing the number of electrodes will require new electrode design and technology (Hillman et al. 2003) or the application of virtual channels (Donaldson et al. 2005; Townshend et al. 1987). It is likely that the extent to which the actual cochlear implant users may benefit from these enriched signals or alternative processing schemes will depend on their residual capacities.

Summary

Cochlear implant melody recognition is best when listening to high-frequency melodies (414–1046 Hz) as opposed to low- or middle-frequency melodies (104–262 and 207–523 Hz) and when listening to pure tones instead of complex harmonic tones. The number of electrodes had no effect on low- and middle-frequency melody recognition but affected high-frequency melody recognition in a way similar to phoneme recognition. Similar to speech recognition, cochlear implant melody recognition exhibited large individual variability. The source of the large variability in melody recognition seems to be different from the source of the variability in speech recognition.

ACKNOWLEDGMENTS

The authors thank all of the cochlear implant users who devoted their time to this study, Rachel Cruz for her help in programming the cochlear implant user's map, and three anonymous reviewers for their helpful comments. Sonya Singh received a research fellowship from the UCI Undergraduate Research Opportunity Program and is currently a medical student at UCI. Ying-Yee Kong is currently affiliated with the Department of Speech-Language Pathology and Audiology at Northeastern University, Boston, MA.

This work was supported by a grant from the National Institutes of Health (2RO1 DC002267).

Address for correspondence: Fan-Gang Zeng, 364 Med Surge II, University of California, Irvine, Irvine, CA 92697. E-mail: fzeg@uci.edu.

Received January 26, 2007; accepted September 14, 2008.

REFERENCES

- Boex, C., Baud, L., Cosendai, G., et al. (2006). Acoustic to electric pitch comparisons in cochlear implant subjects with residual hearing. *J Assoc Res Otolaryngol*, *7*, 110–124.
- Carroll, J., & Zeng, F. G. (2007). Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations. *Hear Res*, *231*, 42–53.
- Dai, H. (2000). On the relative influence of individual harmonics on pitch judgment. *J Acoust Soc Am*, *107*, 953–959.
- Donaldson, G. S., Kreft, H. A., & Litvak, L. (2005). Place-pitch discrimination of single- versus dual-electrode stimuli by cochlear implant users (L). *J Acoust Soc Am*, *118*, 623–626.
- Fearn, R., & Wolfe, J. (2000). Relative importance of rate and place: experiments using pitch scaling techniques with cochlear implants recipients. *Ann Otol Rhinol Laryngol Suppl*, *185*, 51–53.
- 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. *J Speech Lang Hear Res*, *40*, 1201–1215.
- Friesen, L. M., Shannon, R. V., Baskent, D., et al. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*, *110*, 1150–1163.
- Garnham, C., O'Driscoll, M., Ramsden, R., et al. (2002). Speech understanding in noise with a Med-El COMBI 40+ cochlear implant using reduced channel sets. *Ear Hear*, *23*, 540–552.
- Geurts, L., & Wouters, J. (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am*, *109*, 713–726.
- Geurts, L., & Wouters, J. (2004). Better place-coding of the fundamental frequency in cochlear implants. *J Acoust Soc Am*, *115*, 844–852.
- Gfeller, K., & Lansing, C. R. (1991). Melodic, rhythmic, and timbral perception of adult cochlear implant users. *J Speech Hear Res*, *34*, 916–920.
- Gfeller, K., Turner, C., Mehr, M., et al. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants Int*, *3*, 29–53.
- Green, T., Faulkner, A., & 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.
- Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am*, *116*, 2298–2310.
- Green, T., Faulkner, A., Rosen, S., et al. (2005). Enhancement of temporal periodicity cues in cochlear implants: effects on prosodic perception and vowel identification. *J Acoust Soc Am*, *118*, 375–385.
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am*, *87*, 2592–2605.
- Hillenbrand, J., Getty, L. A., Clark, M. J., et al. (1995). Acoustic characteristics of American English vowels. *J Acoust Soc Am*, *97*, 3099–3111.
- Hillman, T., Badi, A. N., Normann, R. A., et al. (2003). Cochlear nerve stimulation with a 3-dimensional penetrating electrode array. *Otol Neurotol*, *24*, 764–768.
- Kasturi, K., & Loizou, P. C. (2007). Effect of filter spacing on melody recognition: acoustic and electric hearing. *J Acoust Soc Am*, *122*, EL29–EL34.
- Kiefer, J., von Ilberg, C., Rupprecht, V., et al. (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. *Ann Otol Rhinol Laryngol*, *109*, 1009–1020.
- Kong, Y. Y., Cruz, R., Jones, J. A., et al. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear Hear*, *25*, 173–185.
- Kong, Y. Y., Stickney, G. S., & Zeng, F. G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *J Acoust Soc Am*, *117*, 1351–1361.
- Laneau, J., Wouters, J., & Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *J Acoust Soc Am*, *116*, 3606–3619.
- Laneau, J., Wouters, J., & Moonen, M. (2006). Improved music perception with explicit pitch coding in cochlear implants. *Audiol Neurootol*, *11*, 38–52.
- Loizou, P. C., Stickney, G., Mishra, L., et al. (2003). Comparison of speech processing strategies used in the Clarion implant processor. *Ear Hear*, *24*, 12–19.
- McDermott, H. J. (2004). Music perception with cochlear implants: a review. *Trends Amplif*, *8*, 49–82.
- McDermott, H. J., McKay, C. M., & Vandali, A. E. (1992). A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *J Acoust Soc Am*, *91*, 3367–3371.
- McKay, C. M., McDermott, H. J., & Clark, G. M. (1994). Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. *J Acoust Soc Am*, *96*, 2664–2673.
- Nie, K., Barco, A., & Zeng, F.-G. (2006). Spectral and temporal cues in cochlear implant speech perception. *Ear Hear*, *27*, 208–217.
- Nimmons, G. L., Kang, R. S., Drennan, W. R., et al. (2008). Clinical assessment of music perception in cochlear implant listeners. *Otol Neurotol*, *29*, 149–155.
- Oxenham, A. J., Bernstein, J. G., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proc Natl Acad Sci USA*, *101*, 1421–1425.

- Pijl, S. (1995). Musical pitch perception with pulsatile stimulation of single electrodes in patients implanted with the Nucleus cochlear implant. *Ann Otol Rhinol Laryngol Suppl*, 166, 224–227.
- Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear*, 26, 451–460.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., et al. (1981). Speech perception without traditional speech cues. *Science*, 212, 947–949.
- Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *J Acoust Soc Am*, 42, 191–198.
- Schouten, J. F., Ritsma, R. J., & Cardozo, B. L. (1962). Pitch of the residue. *J Acoust Soc of Am*, 34, 1418–1424.
- Shannon, R. V., Jensvold, A., Padilla, M., et al. (1999). Consonant recordings for speech testing. *J Acoust Soc Am*, 106, L71–L74.
- Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.
- Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416, 87–90.
- Spahr, A. J., & Dorman, M. F. (2004). Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices. *Arch Otolaryngol Head Neck Surg*, 130, 624–628.
- Townshend, B., Cotter, N., Van Compernelle, D., et al. (1987). Pitch perception by cochlear implant subjects. *J Acoust Soc Am*, 82, 106–115.
- Vandali, A. E. (2001). Emphasis of short-duration acoustic speech cues for cochlear implant users. *J Acoust Soc Am*, 109, 2049–2061.
- Vandali, A. E., Whitford, L. A., Plant, K. L., et al. (2000). Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear Hear*, 21, 608–624.
- Wilson, B. S., Finley, C. C., Lawson, D. T., et al. (1991). Better speech recognition with cochlear implants. *Nature*, 352, 236–238.
- Wilson, B. S., Finley, C. C., Lawson, D. T., et al. (1997). Temporal representations with cochlear implants. *Am J Otol*, 18(6 Suppl), S30–S34.
- Xu, L., Tsai, Y., & Pfingst, B. E. (2002). Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *J Acoust Soc Am*, 112, 247–258.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn Sci*, 6, 37–46.
- Zeng, F. G. (2002). Temporal pitch in electric hearing. *Hear Res*, 174, 101–106.
- Zeng, F. G. (2004). Auditory Prostheses: Past, Present, and Future. In F. G. Zeng, A. N. Popper, & R. R. Fay (Eds.). *Cochlear Implants: Auditory Prostheses and Electric Hearing* (Vol. 20, pp. 1–13). New York: Springer-Verlag.
- Zeng, F. G., Nie, K., Stickney, G. S., et al. (2005). Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci USA*, 102, 2293–2298.