Trends in Amplification
16(2) 67–82
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1084713812451159
http://tia.sagepub.com

Acoustic Cue Integration in Speech Intonation Recognition With Cochlear Implants

Shu-Chen Peng, PhD, CCC-A[1], Monita Chatterjee, PhD[2], and Nelson Lu, PhD[1]

Abstract

The present article reports on the perceptual weighting of prosodic cues in question-statement identification by adult cochlear implant (CI) listeners. Acoustic analyses of normal-hearing (NH) listeners' production of sentences spoken as questions or statements confirmed that in English the last bisyllabic word in a sentence carries the dominant cues (F0, duration, and intensity patterns) for the contrast. Furthermore, these analyses showed that the F0 contour is the primary cue for the question-statement contrast, with intensity and duration changes conveying important but less reliable information. On the basis of these acoustic findings, the authors examined adult CI listeners' performance in two question-statement identification tasks. In Task 1, 13 CI listeners' question-statement identification accuracy was measured using naturally uttered sentences matched for their syntactic structures. In Task 2, the same listeners' perceptual cue weighting in question-statement identification was assessed using resynthesized single-word stimuli, within which fundamental frequency (F0), intensity, and duration properties were systematically manipulated. Both tasks were also conducted with four NH listeners with full-spectrum and noise-band-vocoded stimuli. Perceptual cue weighting was assessed by comparing the estimated coefficients in logistic models fitted to the data. Of the 13 CI listeners, 7 achieved high performance levels in Task 1. The results of Task 2 indicated that multiple sources of acoustic cues for question-statement identification were utilized to different extents depending on the listening conditions (e.g., full spectrum vs. spectrally degraded) or the listeners' hearing and amplification status (e.g., CI vs. NH).

Keywords

cochlear implants, perception, intonation, prosody, cue weighting

Introduction

Speech perception involves multiple acoustic cues (Lisker, 1986; Stevens, 1980). Lisker, for example, identified as many as 16 patterns of acoustic properties that may contribute to NH listeners' identification of phonemic pairs in *rapid* versus *rabid*. It is known that listeners weight these available acoustic cues to different extents, depending on variables such as chronological age (e.g., Nittrouer, 2002, 2005), hearing status (e.g., Ferguson & Kewley-Port, 2002), and the presence of competing noise (e.g., Nittrouer, 2005).

In the present article, we focus on the use of such covarying acoustic cues in intonation recognition by listeners under conditions of spectral degradation, specifically the kind of degradation encountered by cochlear implant (CI) listeners in everyday life. While it is well established that the brain can learn to cope with the degraded/distorted speech signal transmitted via CIs, phonemes and words only constitute part of our everyday oral communication. Prosodic (or suprasegmental) components of speech convey information important for expressive functions of language in semantic, attitudinal, psychological, and social domains (Lehiste, 1970). Prosodic cues convey both the emotional state and the communicative intent of the listener. Critical elements of communication such as irony or sarcasm are often expressed in subtle changes in voice pitch, which are largely lost in CI information processing. Indeed, CI listeners have considerable difficulty in their correct recognition of the emotional content of natural utterances (Luo, Fu, Wei, & Cao, 2008). Perception of prosodic components of speech is important not only for its linguistic function (e.g., in lexical tones or intonation) but also for its role in facilitating spoken language development (Jusczyk et al., 1992; Soderstrom et al., 2003;

[1]From U.S. Food and Drug Administration, Silver Spring, MD, USA (SP, NL)
[2]University of Maryland, College Park, MD, USA (MC)

Corresponding Author:
Shu-Chen Peng, Division of Ophthalmic, Neurological, and Ear, Nose and Throat Devices, Office of Device Evaluation, U.S. Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993, USA
Email: shu-chen.peng@fda.hhs.gov

Thiessen, Hill, & Saffran, 2005). Today, CIs are indicated for use in children as young as 12 months of age in the United States, and there is a thrust toward early implantation. Given the important role of prosodic cues in early spoken language development, it is of considerable importance to understand the mechanisms underlying the processing of these cues via CIs.

In a tonal language such as Mandarin Chinese, lexical meanings of syllables or words can be contrasted when lexical tones are varied. For example, when the syllable *ma* is produced with a high-level tone, it refers to *mother*, but it refers to *scold* when produced with a high-falling tone. In Mandarin Chinese, F0 variation serves as the major acoustic cue in lexical tone perception (Chao, 1968; Howie, 1976). However, additional acoustic properties, such as intensity and duration patterns, may also contribute to this perception (Whalen & Xu, 1992; Xu, Tsai, & Pfingst, 2002). In a nontonal language such as English, variation in speech prosodic components can also convey changes in linguistic functions. However, unlike in Mandarin Chinese where lexical tone contrasts occur at the level of syllables (or words), in English such contrasts may occur at various levels of linguistic units, such as words, phrases, or sentences. These prosodic contrasts are often referred to as "speech intonation" or "intonation" (Ladd, 1996; Lehiste, 1970, 1976). Similar to lexical tones, intonation is mainly conveyed via F0 variation, but this F0 variation often takes place in conjunction with variation in intensity and duration patterns (Cooper & Sorensen, 1981; Ladd, 1996; Lehiste, 1970, 1976). Listeners with NH are able to utilize F0, intensity, and duration cues to recognize speech intonation collectively (Fry, 1955, 1958; Lehiste, 1970, 1976).

While functional hearing may be restored with a CI in patients with a bilateral severe-profound sensorineural hearing loss, fundamental frequency (F0, or voice pitch) information is not adequately coded in contemporary CI devices, primarily due to the reduced spectral resolution of the speech information transmitted by CIs. The poor encoding of the harmonic structure of sounds affects CI listners' performance in voice gender recognition (e.g., Fu, Chinchilla, & Galvin, 2004; Fu, Chinchilla, Nogaki, & Galvin, 2005), music perception (Gfeller et al., 2005, 2002; Kong et al., 2004; Laneau, Wouters, & Moonen, 2006), recognition of prosodic aspects of speech, including intonation (Green, Faulkner, & Rosen, 2002, 2004; Peng, Tomblin, & Turner, 2008) and lexical tone recognition (Ciocca et al., 2002; Luo et al., 2008; Peng et al., 2004; Wei, Cao, & Zeng, 2004). Given this limited access to the F0 information, it is possible that CI patients' relative reliance on other acoustic cues, that is, intensity and duration cues, would increase compared to NH listeners with full access to the F0 information.

In NH listeners, voice pitch information can be perceived via temporal cues related to F0 (periodicity), as well as spectral cues, that is, resolved harmonic structures of voiced sounds (Rosen, 1992). With current CI devices, the slowly varying temporal envelope that conveys intensity and duration aspects of speech signals is relatively well transmitted to CI listeners. However, temporal coding is highly constrained in providing F0 information to CI listeners. In some coding strategies, such as the "spectral peak" (SPEAK), the carrier rates are low (about 250 to 300 pulses/sec per channel). The low carrier rate limits the maximum F0 that can be reliably transmitted via the envelope. More recent speech-coding strategies, for example, Advanced Combination Encoder (ACE), Continuous Interleaved Sampling (CIS), or High Resolution (HiRes), employ faster stimulation rates. Many CI listeners have difficulty in exploiting rapid temporal envelope fluctuations above 300 Hz (Chatterjee & Peng, 2008). In other words, with temporal cues transmitted via the envelope, CI listeners are limited in their access to periodicity cues of certain speakers, particularly children and women whose natural F0 often extends beyond 300 Hz.

These limitations may also account for CI listeners' difficulty with question-statement identification based on voice pitch information (Chatterjee & Peng, 2008; Green et al., 2002, 2004). To date, the majority of the studies that assess CI listeners' speech recognition have adopted naturally uttered speech materials, including sentences, words, or phonemes (vowels and consonants). These materials permit estimations of CI listeners' capability of "real-world" speech perception. However, they also contain contextual or coarticulatory cues, which limits evaluations of listeners' utilization of multiple acoustic cues that may contribute to speech recognition. This limitation makes it difficult to study the processes and mechanisms underlying speech perception in CI listeners with these materials. Nevertheless, there have been a few notable studies on CI listeners' utilization of multiple sources of acoustic information in speech recognition, at the segmental level (i.e., consonants and vowels). For example, Xu et al. (2005) demonstrated that, under conditions of spectral degradation, temporal envelope cues are more involved in phoneme identification, particularly in the identification of consonants. Dorman et al. (1988) examined the relative salience of multiple cues (e.g., spectrum at signal onset and onset frequency and direction of change of formant transition) to stop consonant identification by one Symbion 4-channel CI user and 10 NH listeners. The results indicated that whereas NH listeners use both onset spectrum and formant transition in identifying stop consonants (/b/ vs. /d/), CI user rely more on onset spectrum than formant transition in identification. Similarly, Iverson et al. (Iverson, Smith, & Evans, 2006) manipulated the presence or absence of the formant movement and duration cues for vowel recognition. Their results indicated that while both formant movement and duration cues contribute to vowel recognition in CI listeners with full-spectrum stimuli and in NH listeners under acoustic CI simulations, NH listeners do not utilize these two cues in the same fashion with full-spectrum stimuli. More recently, Winn and colleagues (Winn, 2011; Winn, Chatterjee, & Idsardi, 2012) have found evidence for acoustic cue-trading in the perception of specific phonemic contrasts by both CI patients and NH listeners attending to degraded speech. Together, previous findings suggest a potential shift in perceptual cue weighting from normal hearing to electric hearing.

The studies cited above focused on CI listeners' recognition of phonemes (consonants and vowels). Relatively few studies have examined the processing of prosodic aspects of speech (e.g., lexical tones or intonation) under difficult listening conditions. Given the poor encoding of F0 via CIs, the extent to which intensity or duration cues, which naturally covary with F0 in questions and statements matched for their syntactic structure, might contribute to the identification of questions and statements by CI listeners is of considerable interest. In earlier studies, we reported on the processing of F0 cues by CI listeners and by NH listeners under various degrees of spectral degradation (Chatterjee & Peng, 2008; Peng, Lu, & Chatterjee, 2009) in an intonation recognition task. Specifically, Chatterjee and Peng examined exclusively on listeners' processing of F0-based information as well as its relationship to their psychophysical capability. Peng et al. focused on listeners' use of the F0 contour under a narrow set of specific conditions in which the secondary cue, intensity, was either conflicting or cooperating with the F0 cue. Neither of these studies examined listeners' overall relative use of the multiple acoustic cues (i.e., F0, intensity, and duration) that may be used by listeners to interpret speech intonation. In the present study, we present previously unpublished data on the three dominant acoustic features of naturally produced utterances in a question-statement contrast. Furthermore, we report on a full quantification of how listeners' weighting of these three different acoustic cues may vary under different listening conditions. The results presented here underscore the importance of the individual's listening strategy in auditory perception and have strong implications for CI patients' rehabilitation/training in clinical or other settings.

Acoustic Analyses of Naturally Uttered Sentences

Acoustic analyses of naturally uttered sentences were conducted as part of the first author's PhD dissertation (Peng, 2005). The set of speech stimuli used for experiments was also adopted by Chatterjee and Peng (2008), and was similar to those used by Green et al. (Green, Faulkner, Rosen, & Macherey, 2005), where sentences were used to assess the effectiveness of enhancing temporal periodicity cues in CI speech coding. Speech intonation can be used to convey linguistic functions, such as question versus statement contrasts. However, a yes-no question (e.g., "Is the girl on the playground?") can also be distinguished from its statement counterpart (e.g., "The girl is on the playground") on the basis of variation in syntactic structures. Thus, in order to focus on the perception of prosodic cues, speech stimuli were constructed so that they were only contrasted in prosodic patterns and were otherwise controlled for their syntactic structure. Based on this principle, the speech stimuli adopted in this study were comprised of 10 syntactically matched sentence pairs. These sentences were produced by each of six adult speakers (between 22 and 47 years of age; three per gender) as either a question or a statement (e.g., "The girl is on the playground?" vs. "The girl is on the playground"), by varying their prosodic patterns.

The F0-related (F0 height and F0 contour), intensity, and duration characteristics of the set of sentences described above were acoustically analyzed in order to derive the resynthesized stimuli adopted in Task 2 (see below for details). The two F0-related parameters, F0 height and F0 contour, were measured using the autocorrelation pitch extraction algorithm in Praat (version 4.3; Boersma & Weenink, 2004). The F0 height values were obtained by averaging the F0 values at the vocalic portions of the nonutterance-final words (see Peng et al., 2008, for a description of these sentences in detail). Words at the nonutterance-final position were chosen because of questions, as the F0 height values at this position were less affected by those values at the utterance-final position. While the average F0 height of utterances by female participants ($M = 234.54$ Hz, $SD = 60.83$ Hz) was significantly higher than that of their male counterparts ($M = 120.99$ Hz, $SD = 28.60$ Hz), $t(59) = 12.37$, $p < .001$, male and female utterances shared similar patterns in terms of F0 contour, intensity, and duration characteristics. Acoustic data were thus collapsed across gender for the remaining acoustic parameters, which are described as follows:

F0 contour. Although F0 fluctuates steadily over the entire utterance, it was consistently observed to be different between questions and statements at the utterance-final bisyllabic word (i.e., rising and falling for questions and statements, respectively). This observation was consistent with the findings in the literature (for details, see Cruttenden, 1986). The acoustic findings of this parameter are thus limited to bisyllabic tokens at the utterance-final position. Using the above-mentioned autocorrelation pitch extraction algorithm, the absolute F0 values were recorded at the onset and offset of the vocalic portion of each bisyllabic word. Figure 1(a) shows a rising F0 contour for questions (i.e., F0 at offset minus F0 at onset > 0) and a falling contour for statements (i.e., F0 at offset minus F0 at onset < 0). The average amount of F0 variation for questions at the utterance-final position was significantly greater than that for statements, $t(59) = 21.65$, $p < .001$.

Intensity. The intensity characteristic reported here focuses on the ratio of the peak intensity at the 2nd syllable of each utterance-final word to its 1st syllable. Again, this was because the intensity ratio was found to be consistently different at the utterance-final position of each sentence (i.e., greater for questions than statements). Figure 1(b) displays the distribution of the intensity characteristics for utterances

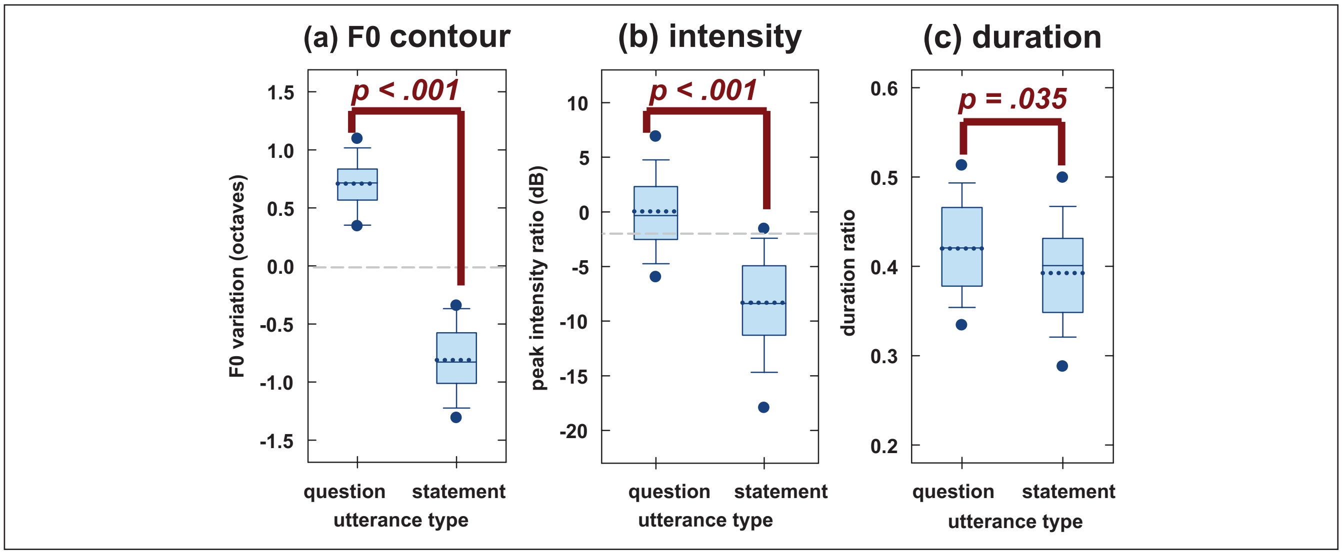

Figure 1. Distributions of the acoustic characteristics for multitalker, naturally uttered bisyllabic words, produced as questions or statements. Panels (a) through (c) show the amounts of F0 variation, peak intensity ratio, and duration ratio, respectively. The x-axis displays the utterance types (i.e., "question" vs. "statement"); the y-axis displays the overall amount of changes in each acoustic dimension. The mean and median are displayed by the dotted and solid lines across each box, respectively. The upper and lower bounds of each box represent the quartiles, the whisker away from the box bounds showed the ± 1.25 *SD* of the mean, and the filled circles represent the 5th and 95th percentiles bounds, if they are outside of the end of whisker. The *p* value is indicated on each panel.

produced as questions or statements. The intensity ratio ranged from –8.02 to 12.30 dB for questions, but was less than 0 dB for statements. The mean peak intensity ratio (between the 2nd syllable and the 1st syllable of each bisyllabic word) for questions was significantly greater than that for statements, $t(59) = 12.66$, $p < .001$.

Duration. Unlike F0 contour and peak intensity ratio, duration characteristics did not appear to differ consistently between questions and statements. Among several duration parameters, including (a) the duration of the entire utterance, (b) the duration of the bisyllabic word, (c) the ratio between the duration of the utterance-final (bisyllabic) word and that of the entire utterance, and (d) the duration ratio between the 1st and 2nd syllables of the bisyllabic word, only (a) and (c) were found to be statistically significant (p values < .05). Since the primary purpose of the present acoustic analyses was to explore whether adult speakers' utterances (in this case, sentences) showed consistent prosodic differences between questions and statements, the duration parameter reported here focuses on (c), as this parameter takes different speaking rates among speakers and different sentence lengths into consideration. Figure 1(c) shows that the range of the mean duration ratio (between the duration of the utterance-final [bisyllabic] word and that of the entire utterance) was slightly, but significantly, greater for questions than for statements, $t(59) = 2.16$, $p = .035$.

Resynthesized Tokens

The resynthesized tokens were identical to those adopted in Chatterjee and Peng (2008) and Peng et al. (2009). As illustrated in Figures 1(a) through 1(c), the acoustic analysis revealed very different patterns of F0 contour between questions and statements (i.e., rising for questions and falling for statements) but overlapping amounts of changes in intensity and duration properties (duration, in particular) between questions and statements. Nonetheless, all three dimensions, F0 contour, intensity, and duration properties, were observed to be different between questions and statements. Based on the above-mentioned acoustic findings, one bisyllabic word ("popcorn") was acoustically manipulated in a systematic manner using the Praat software (version 4.3; Boersma & Weenink, 2004). Bisyllabic tokens were adopted for acoustic resynthesis, as the 1st syllable could be used as the reference for the parametric variation in intensity and duration properties of the 2nd syllable. Reynthesized tokens were generated by varying four acoustic parameters orthogonally; see (a) through (d) below for details. This factorial design permitted evaluations of main effects of each of the acoustic parameters as well as their interactions. The entire stimulus set contained 360 tokens (1 bisyllabic word × 2 steps of F0 height × 9 steps of F0 contour × 5 steps of peak intensity ratio × 4 steps of duration ratio). Figure 2 displays examples of the target via parametric manipulations

a. *F0 height*. As illustrated in Figure 2(a), flat F0 contours with 120- and 200-Hz initial F0 heights were generated to simulate male and female speakers' typical F0 heights.
b. *F0 contour*. As illustrated in Figure 2(b), each of the F0 height-adjusted tokens was further manipulated to vary the F0 contour (linear glides), from

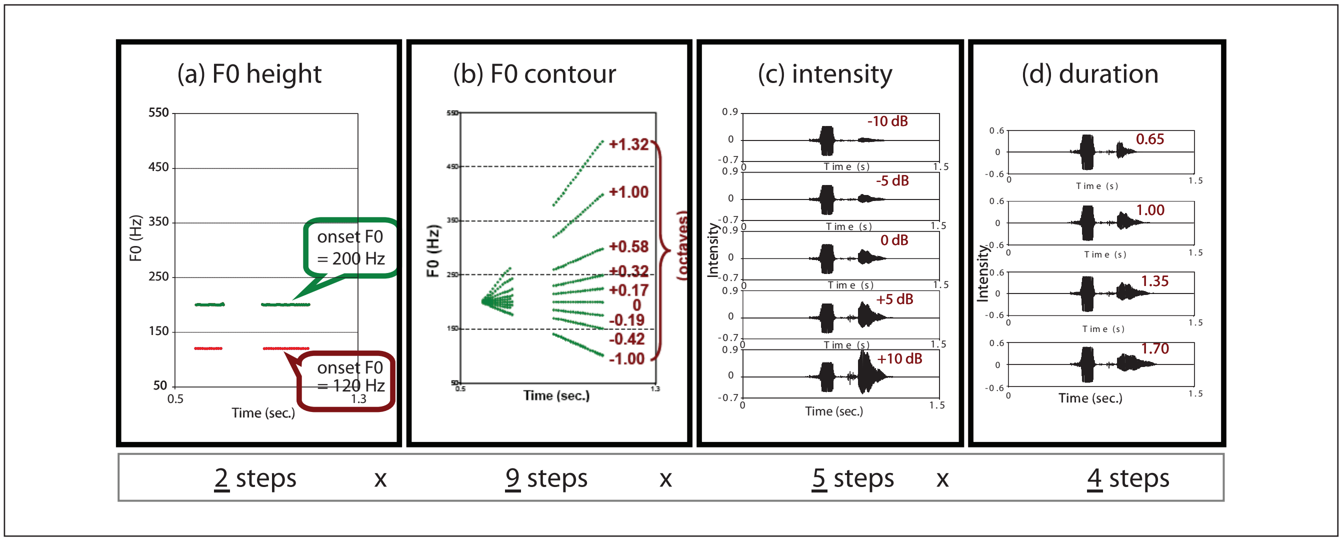

Figure 2. Illustrations of acoustic dimensions that were manipulated orthogonally, resulting in 360 tokens. Panels (a) through (d) display the steps specific to each acoustic dimension, that is, F0 height, F0 contour, peak intensity ratio, and duration ratio, respectively. The number of steps for each dimension is also indicated below each panel.
Source: This figure was reproduced from Peng, Lu, & Chatterjee (2009), with permission from the publisher.

the onset to offset of the bisyllabic token in nine steps (–1.00, –0.42, –0.19, 0, 0.17, 0.32, 0.58, 1.00, and 1.32 octaves). Target F0s at the offset were 60, 90, 105, 120, 135, 150, 180, 240, and 300 Hz for stimuli with the 120-Hz F0 height; for the tokens with the F0 height of 200-Hz, the target F0s at the offset were 100, 120, 150, 175, 200, 225, 300, 400, and 500 Hz. As shown in the acoustic analyses, statements were associated with a falling F0 contour (parameter values < 0), while questions were associated with a rising contour (parameter values > 0).

c. *Peak intensity ratio*. As illustrated in Figure 2(c), each of the F0 height- and variation-adjusted tokens was manipulated to have specific peak intensity ratios of the 2nd syllable relative to the 1st syllable of a reference token (i.e., –10, –5, 0, 5, 10 dB). The reference token ("0 dB") is shown in Figure 2(c). This reference token was a neutral token that was naturally produced as part of a question. Note that negative and positive ratios roughly correspond to statements and questions, respectively.
d. *Duration ratio*. As illustrated in Figure 2(d), each token was further processed to vary the duration of the 2nd syllable, again relative to a reference token (a neutral token that was naturally produced as part of a question, see the panel indicating "1.00" in Figure 2[d]). The duration of the 2nd syllable was multiplied by a factor of 0.65, 1.00, 1.35, or 1.70. A greater ratio (i.e., longer duration) generally corresponds to questions; as observed in the acoustic analyses of naturally produced sentences, however, this parameter is not as reliably used by speakers to mark question-statement contrasts.

Intonation Recognition

Each participant was asked to perform two tasks: The purpose of Task 1 was to examine CI and NH listeners' performance in recognizing questions and statements with naturally produced sentences, and the purpose of Task 2 was to examine these listeners' perceptual weighting of multiple acoustic cues using resynthesized bisyllabic stimuli.

Study Participants

The 13 adult CI users who participated in Peng et al. (2009) served as the participants in this study. As mentioned previously, Peng et al. focused on results obtained with these participants with a subset of the stimulus set, whereas the present study reports on the overall patterns obtained with the entire stimulus set. The CI users ranged from 23 to 70 years of age (M = 59 years of age) and had at least 1 year of device experience at test time. The majority of the participants, except for CI-5 and CI-13, were postlingually deaf. They used either a Cochlear Nucleus device (Cochlear Americas, Denver, CO), or a Clarion device (Advanced Bionics, Sylmar, CA). Table 1 provides a list of each CI participant's background information.

In addition, data are also reported here that were obtained with the identical small group of adult NH listeners (N = 4) in Peng et al. (2009). As with the CI patients, results reported in the present study were obtained with the entire set of stimuli rather than the smaller set described by Peng et al. All of the NH participants passed the hearing screening at 20 dB HL from 250 to 8000 Hz at octave intervals, bilaterally. The inclusion of this NH group was to provide reference to NH listeners' performance under different listening conditions

Table 1. Background Information of CI Participants

Participant	Etiology	Onset of Deafness[a]	Age at Testing	Device Experinence (Years)	Gender	Device	Processing Strategy
CI-1	Unknown	Post	60	3	Male	Nucleus 24	ACE
CI-2	Unknown	Post	52	7	Female	Clarion S	MPS
CI-3	Genetic	Post	63	3	Female	Nucleus 24	ACE
CI-4	Possibly genetic	Post	56	9	Female	Nucleus 24	ACE
CI-5	Unknown	Pre	52	4	Female	Clarion CII	HiRes
CI-6	Unknown	Post	65	6	Male	Clarion S	MPS
CI-7	Possibly genetic	Post	64	1	Female	Nucleus Freedom	ACE
CI-8	Unknown	Post	83	5	Male	Nucleus 24	ACE
CI-9	Trauma	Post	49	14	Male	Nucleus 22	SPEAK
CI-10	Genetic	Post	70	13	Female	Nucleus 22	SPEAK
CI-11	Unknown	Post	65	14	Male	Nucleus 22	SPEAK
CI-12	German measles/ ototoxicity	Post	64	1	Female	Nucleus Freedom	ACE
CI-13	Unknown	Pre	23	7	Male	Nucleus 24	ACE

Note: CI = cochlea implant; NH = normal hearing.
[a]Prelingually deaf—became deaf before age 3; postlingually deaf—became deaf at or after age 3.
Source: Modified from Table 1 in Peng, Lu, and Chatterjee (2009) with permission from the publisher.

(spectrally degraded vs. full spectrum) as well as to provide a reference against which the results obtained with the CI listeners could be compared. Although such comparisons are routine in the CI literature, interpretation of such comparisons should be made with caution, owing to the differences in age and hearing status between the two groups. All participants gave written informed consent approved by the University of Maryland, College Park Institutional Review Board (IRB) prior to the task; all participants were paid for participation.

Acoustic CI Simulations

The signal processing technique that was used to create CI simulations has been described in previous studies (Chatterjee & Peng, 2008; Peng et al., 2009) in detail. All sentence stimuli (Task 1) and resynthesized tokens (Task 2) were subjected to noise-band vocoding (Shannon et al., 1995). Briefly, noise vocoders were implemented as follows: The speech stimuli were bandpass filtered into logarithmically spaced frequency bands (16, 8, 4, and 1 frequency bands for Task 1 and 8 and 4 frequency bands for Task 2). The input frequencies ranged from 200 to 7000 Hz. The corner frequencies for each analysis band followed Greenwood's equation for a 35-mm cochlear length (Fu & Shannon, 1999; Greenwood, 1990; also see Table 2 of Fu & Nogaki, 2004). Temporal envelope extraction from each frequency band was performed by half-wave rectification and lowpass filtering (cutoff at 400 Hz, –24 dB/octave). The resulting envelope was used to modulate amplitude white noise. The modulated noise was bandpass filtered using the same filters that were used for frequency analysis. The outputs of all channels were finally summed. The long-term RMS amplitude was matched to that of the full-spectrum, original signals.

Test Procedure

The test equipment and environment were identical to those in Peng et al. (2009): A Matlab-based user interface was used to control the tasks. The stimuli were presented via a single loudspeaker (Tannoy Reveal) in soundfield, at about 65 dB SPL (A-weighting) in a double-walled sound booth. The CI listeners were tested using their own speech processors, with volume and sensitivity settings at their everyday levels. The NH listeners were presented with both the full-spectrum stimuli and acoustic CI simulations; the CI listeners only listened to the full-spectrum stimuli. A single-interval, 2-alternative forced-choice (2AFC) paradigm was used to measure intonation recognition in both tasks: The stimulus was presented once, and the participant was asked to indicate whether it sounded like a "statement" or a "question" by clicking on the appropriate box on the display screen. The entire stimulus set (120 naturally uttered sentences or 360 resynthesized bisyllabic tokens) was presented in one run, with the order of presentation fully randomized. The mean of two full runs of each condition was calculated to obtain the final data. Learning effects were evaluated in a preliminary study, where stable performance and comparable results were observed across 10 runs of the same block presented to one adult CI participant.

All participants listened to the sentences, followed by the resynthesized stimuli. For the sentences, each NH participant was tested with the full-spectrum stimuli and to 16-, 8-, 4- and 1-channel acoustic CI simulations. For the resynthesized stimuli, each NH participant was tested with the

Table 2. Accuracy in Task 1 and Estimated Coefficients in the Logistic Models[a] Fitted to Individual Data (for CI Listeners) and Group Data (for Both CI and NH Listeners) in Task 2

Listener/Group/Condition	Accuracy in Task 1 (% +*SD*)	F0 Height	F0 Contour	Peak Intensity Ratio[b]	Duration Ratio
Full spectrum					
CI-1	78.33	0.0632	2.2657**	0.0610*	0.3557
CI-2	94.17	0.2506	6.1637**	0.1251**	1.5800**
CI-3	98.33	0.0306	6.1041**	–0.0565**	–0.2227
CI-4	98.33	1.9140**	4.5714**	0.1580**	–2.6335**
CI-5	63.33	0.2162	1.7730**	0.1786**	0.8243**
CI-6	95.83	0.6709**	3.2154**	0.0069	–1.0834**
CI-7	96.67	–0.8591**	7.1140**	0.0448*	–0.2085
CI-8	65.83	0.5551*	2.4198**	0.0194	0.5151*
CI-9	95.83	3.7448**	4.8787**	0.1713**	–2.4352**
CI-10	93.33	2.1612**	2.6967**	0.2812**	1.2774**
CI-11	58.00	1.4831**	1.1909**	0.2747**	1.2549**
CI-12	85.00	2.7837**	4.3230**	0.0068	–0.7333*
CI-13	75.00	1.9817**	2.2823**	0.1977**	1.8613**
CI group	84.46 +14.68	0.8141**	2.3968**	0.0841**	0.0952
NH-1	100.00	0.5218	5.9949**	0.0191	–0.8503*
NH-2	95.00	1.4088**	5.0620**	0.0095	–0.0733
NH-3	96.67	0.8809**	5.4048**	–0.0065	1.0125*
NH-4	99.17	0.2879	9.2178**	–0.0144	0.6219
NH group	97.71 +2.29	0.7123**	5.2470**	0.0025	0.1294
Acoustic CI simulations (group data only)[c]					
16-channel	91.25 +1.98	NA	NA	NA	NA
8-channel	84.79 +9.75	1.5452**	2.5262**	0.1317**	1.5280**
4-channel	80.63 +10.77	0.6000*	1.9564**	0.1170**	1.4540**
1-channel	69.38 +12.70	NA	NA	NA	NA

Note: CI = cochlear implant; NH = normal hearing; NA = not applicable/not tested.
[a]The equation used to derive the estimated coefficients is as follows: Expected logarithm of odds of question response = Intercept + (A) × F0 height (=1 if 220 Hz; =0 if 120 Hz) + (B) × F0 contour (in octaves) + (C) × peak intensity ratio (in dB) + (D) × log (duration ratio), where (A), (B), (C), and (D) in this equation refer to the estimated coefficient listed in column entitled F0 height, F0 contour, peak intensity ratio, and duration ratio, respectively.
[b]Rather than using duration ratio, which was restricted to positive values, logarithm of duration ratio was adopted since it theoretically took values on all real numbers, and thus, it was more appropriate in a logistic model setting.
[c]Only group data are shown for acoustic CI simulations, as estimated coefficients were quite consistent across NH listeners.
*$p < .05$. **$p < .001$.

full-spectrum stimuli, as well as 8- and 4-channel acoustic CI simulations (conditions that reasonably approximate the range of performance of CI listeners, for example, Friesen et al., 2001). The order of listening conditions (full-spectrum and various spectrally degraded conditions) was randomized across listeners for both sentence and resynthesized stimuli. A small set of stimuli similar to the test stimuli were used to familiarize the listeners before actual data collection began; no feedback was provided during the familiarization session or during actual testing. Practice stimuli were not included in the sets of test stimuli. The test time for each run was approximately 20 min with the sentence stimuli and 45 min with the resynthesized stimuli.

Scoring and Data Analysis

With the sentence stimuli, percentage-correct scores were computed for identification accuracy, under full-spectrum (for CI and NH participants) and under each of the acoustic CI simulation conditions (for NH participants). The performance levels of the CI and NH groups were compared using independent-samples *t* tests, while NH participants' performance in the various listening conditions was analyzed using a one-way repeated measures ANOVA (analysis of variance) incorporating a heterogeneous compound symmetry variance-covariance structure. With the resynthesized stimuli, scores of "one" and "zero" were assigned to "question" and "statement" responses, respectively, and psychometric functions (proportion of "question" responses as a function of the parameter of interest, for instance, F0 change in the contour) derived from the data. Logistic models were fitted to the data for each participant and in each of the various listening conditions. The four acoustic parameters were included as independent variables, and the participant's response as the dependent variable. In these models, the coefficient for each parameter, which corresponds to the steepness of the slope of the psychometric function, was used to approximate the listener's reliance on F0, intensity, and duration cues in question-statement identification. The coefficients for each parameter were compared between participant groups (CI vs. NH full spectrum). Repeated-measures logistic models were fitted to the NH group data to compare each parameter among the various listening conditions (full-spectrum, 8- and 4-channel conditions). The repeated-measures logistic models were fitted using the generalized estimating equation (GEE) method (Liang & Zeger, 1986). The coefficients in all logistic models were estimated using PROC GENMOD of SAS.

Results

Question-Statement Identification With Sentences

The group mean of the overall identification accuracy with sentences was 84.46% (*SD* = 14.68%) for the CI participants; intersubject variability was observed among the identification accuracy of individual participants (range: 58.00%-98.33%). As a group, the CI participants' performance was significantly lower than that of NH participants with full-spectrum stimuli, $t(13.52) = 3.20$, $p = .001$. The group mean of CI participants' accuracy was not significantly different from NH participants' accuracy with acoustic CI simulations (i.e., ranging from 69.38% to 91.25% across 16-, 8-, 4- and 1-channel conditions; all p values > .05). Even under the 1-channel condition, the NH listeners' accuracy was significantly above chance ($p = .0055$ based on a t test). Identification accuracy was dependent on the listening condition (full-spectrum, 16-, 8-, 4- and 1-channel conditions), $F(4, 12) = 6.07$, $p = .007$. As shown in Figure 3, NH participants' question-statement identification performance improved as the number of spectral channels was increased from 1 to 16, and their accuracy with 16 channels was slightly poorer than that with the full-spectrum stimuli. Post hoc pair-wise group mean comparisons (Tukey–Kramer adjustment) revealed a statistically significant difference in the accuracy between 1- and 16-channel conditions ($p = .047$), between 1-channel and full-spectrum conditions ($p = .014$), between 4-channel and full-spectrum conditions ($p = .048$), and between 16-channel and full-spectrum conditions ($p = .011$).

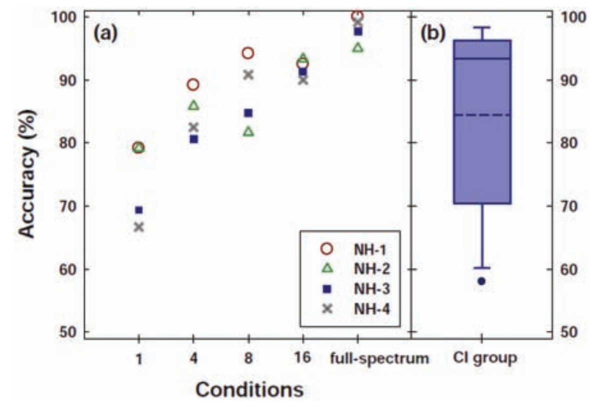

Figure 3. Question-statement identification accuracy with naturally uttered stimuli, by NH and CI listeners. Panels (a) and (b) display the data for NH and CI listeners, respectively. In Panel (a), the individual data of four NH listeners are shown with different symbols. The x-axis for Panel (a) displays the spectrally degraded (1-, 4-, 8-, and 16-channel) and the full-spectrum conditions. For Panel (b), the mean and median are displayed by the dotted and solid lines across each box, respectively. For both panels, the y-axis displays the overall identification accuracy. The upper and lower bounds of each box represent the quartiles, the whisker away from the box bounds showed the ±1.25 *SD* of the mean, and the filled circles represent the 5th and 95th percentiles bounds, if they are outside of the end of whisker.

Question-Statement Identification With Resynthesized Tokens

As with the sentence stimuli, question-statement identification was measured using resynthesized stimuli in the same CI and NH participants. Figure 4(a) displays the mean proportion of "question" responses, as a function of the acoustic parameter—F0 height (120-Hz and 200-Hz initial F0 heights)—for the CI group (full spectrum) and the NH group (full-spectrum, 8- and 4-channel acoustic CI simulations). The overall proportion of "question" responses was consistently higher with the 200-Hz initial F0 stimuli than that with the 120-Hz initial F0 stimuli, for both CI and NH groups with the full-spectrum stimuli, as well as under the 8- and 4-channel acoustic simulations. The estimated coefficients are summarized in Table 2, for the CI and NH groups as well as for each individual CI participant. Among the 13 CI participants, 9 participants' proportion of "question" responses was significantly different with the 120- and 200-Hz initial F0 stimuli (all p values < .005). Of these participants, eight participants' proportion of "question" responses was significantly higher with the 200-Hz initial F0 stimuli than that with the 120-Hz initial F0 stimuli, and one participant (CI-7) showed the opposite trend in performance.

The mean proportion of "question" responses is displayed in Figure 4(b), as a function of the acoustic parameter—F0 contour (from –1.00 to 1.32 octaves)—for both CI and NH groups with the full-spectrum stimuli as well as under the 8- and 4-channel acoustic simulations. Data are collapsed across the other acoustic dimensions. The overall proportion of "question" responses increased with increments along the psychometric function of F0 contour, for both CI and NH groups with the full-spectrum stimuli as well as under the 8- and 4-channel conditions for the NH group. The results plotted in Panel (b) are very similar to those in Figure 5b in Chatterjee and Peng, but note that the data were obtained

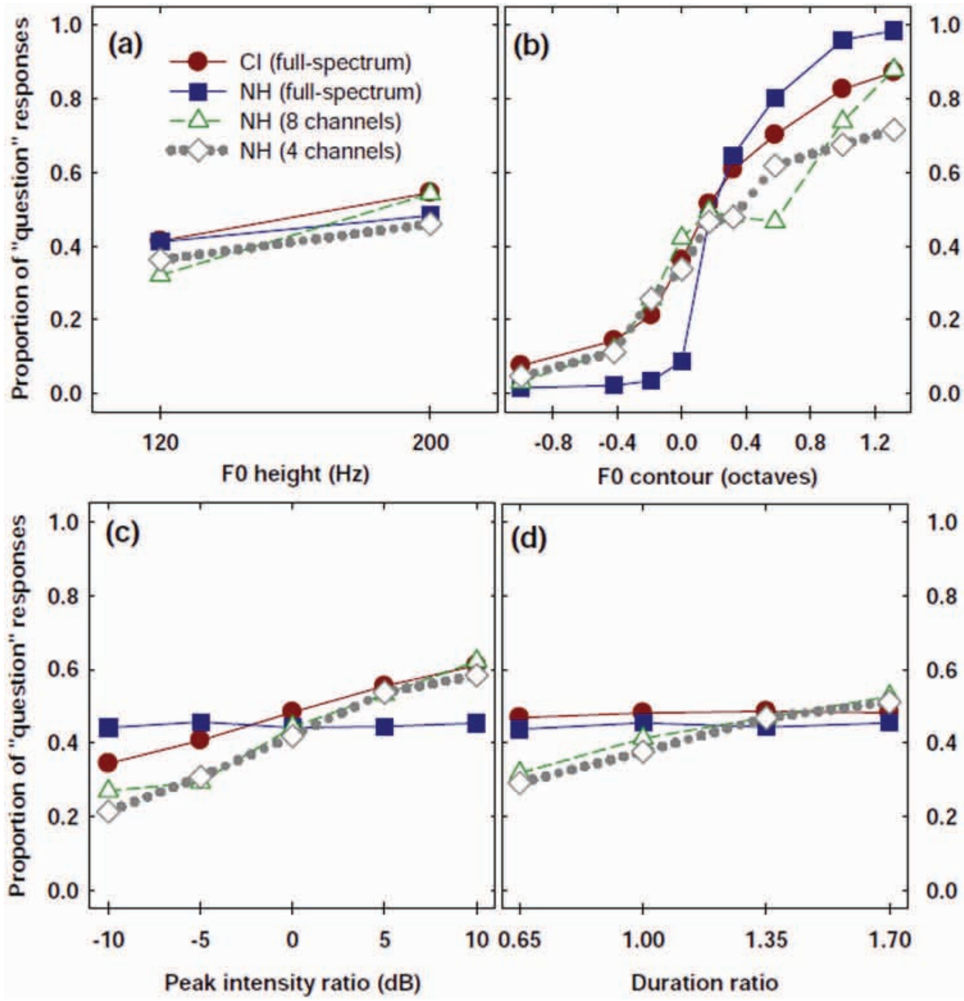

Figure 4. Group mean proportions of question judgments as a function of the manipulation in each acoustic dimension. Panels (a) through (d) display the data for the acoustic dimensions, that is, F0 height, F0 contour, peak intensity ratio, and duration ratio, respectively. The x-axis in each panel indicates the steps of the manipulation specific to each acoustic parameter; the y-axis indicates the proportions of question judgments. The data for two groups (CI vs. NH) as well as for different listening conditions (full-spectrum, 8-, and 4-channel conditions) are shown with different symbols; data points at different steps are linked with a solid line (for the full-spectrum condition) or a dashed line (for the 8- and 4-channel conditions).

from different numbers of CI participants (N = 13 in this study; N = 10 in Chatterjee & Peng, 2008).

As shown in Table 2, the estimated coefficients were all significantly greater than zero (all p values < .005). The mean estimated coefficient for F0 contour was significantly smaller for the CI group than that for the NH group (p < .005). Within the NH group, this estimated coefficient was also significantly smaller for the 8- and 4-channel conditions than that for the full-spectrum condition (both p values < .005). There was no significant difference between this estimated coefficient for the CI group and that for the NH group in the 8- and 4-channel conditions (both p values > .05). As revealed from the estimated coefficients, NH participants were the most sensitive to F0 contour in question-statement identification when listening to the full-spectrum stimuli, and this sensitivity became weaker with the spectrally degraded stimuli.

Furthermore, the CI participants' F0 contour sensitivity was significantly lower with the 200-Hz initial F0 stimuli than that with the 120-Hz initial F0 stimuli (p < .005). This trend was also observed in NH listeners with the 4-channel spectrally degraded stimuli (p < .005). However, when attending to the full-spectrum and 8-channel stimuli, NH listeners' F0 contour sensitivity was not found to be statistically different between the 120- and 200-Hz initial F0 stimuli (full-spectrum: p = .152; 8-channel: p = .688).

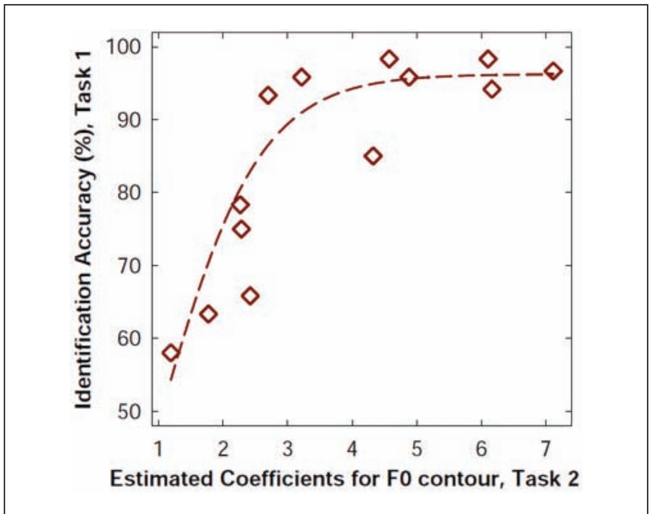

Figure 5. Relationship between CI listeners' question-statement identification accuracy using naturally uttered stimuli and the estimated coefficient for F0 contour. The x-axis displays the estimated coefficient, as revealed from Task 2, while the y-axis displays the overall identification accuracy, as revealed from Task 1. The dotted line represents the regression line fitted to all data points ($r^2 = .8472$, $p < .001$).

Figure 4(c) displays the mean proportion of "question" responses, as a function of the acoustic parameter—peak intensity ratio (from –10 to 10 dB)—for both CI and NH groups with the full-spectrum stimuli as well as under the 8- and 4-channel acoustic simulations. Data are collapsed across the other acoustic dimensions. For the CI group and for the NH group with the 8- and 4-channel stimuli, the overall proportion of "question" responses became higher, when peak intensity ratio was increased from –10 to 10 dB. However, for the NH group with the full-spectrum stimuli, the overall proportion of "question" responses did not differ as a function of peak intensity ratio. As shown in Table 2, among the 13 CI participants, 10 had estimated coefficients for peak intensity ratio that were significantly different from zero ($p < .05$ for CI-1 and CI-7; all other p values $< .005$); these individuals' estimated coefficients were all positive in value, except for one (CI-3). Under the 8- and 4-channel conditions, all NH participants' estimated coefficients were significantly different from zero (all p values $< .005$) and were positive in value. However, NH participants' estimated coefficients were not significantly different from zero (all p values $> .05$) with the full-spectrum stimuli.

The mean estimated coefficient for peak intensity ratio was significantly greater for the CI group than that for the NH group under the full-spectrum condition ($p < .005$). Within the NH group, this estimated coefficient was also significantly greater for the 8- and 4-channel conditions than that for the full-spectrum condition (both p values $< .005$). There was no significant difference between this estimated coefficient for the CI group and that for the NH group in the 8- and 4-channel conditions (both p values $> .05$). These results indicated that CI participants were sensitive to changes in peak intensity ratio in question-statement identification; these individuals utilized this peak intensity ratio as a cue in a rather consistent fashion (i.e., estimated coefficients mostly positive, with only one exception). Furthermore, NH participants did not appear to utilize peak intensity ratio as a cue in identification when listening to the full-spectrum stimuli, but they did when listening to the spectrally degraded stimuli. Of note, even with the most extreme intensity ratios (+10 dB and –10 dB), the proportion of utterances listeners judged as questions (at +10 dB) or statements (at –10 dB) were rather limited.

Figure 4(d) displays the mean proportion of "question" responses as a function of duration ratio (0.35, 0.70, 1.00, 1.35), for both CI and NH groups. For both CI and NH groups with the full-spectrum stimuli, the overall proportion of "question" responses did not differ, when duration ratio was increased from 0.35 to 1.35. However, for the NH group with the 8- and 4-channel stimuli, the overall proportion of "question" responses became higher as a function of duration ratio. As shown in Table 2, the estimated coefficients appeared to be relatively unpredictable for CI participants. That is, although several CI participants (10 out of 13) had estimated coefficients for duration ratio that were significantly different from zero ($p < .05$ for CI-8 and CI-12, all other p values $< .005$), the estimated coefficients were either positive ($n = 6$) or negative ($n = 4$) in value.

The mean estimated coefficient for duration ratio was not significantly different between the CI and NH groups with the full-spectrum stimuli ($p = .93$). For the two spectrally degraded conditions, the differences between the estimated coefficient for the CI group and that for the NH group were statistically significant (both p values $< .005$). These results indicated that when listening to the full-spectrum stimuli, neither group used duration ratio as a cue in question-statement identification. However, NH participants showed consistent use of duration cue in question-statement identification when listening to the spectrally degraded stimuli.

Relationships Between the CI participants' Performance With Sentences and Resynthesized Stimuli

The correlations between CI participants' identification accuracy with naturally produced stimuli) and the estimated coefficients derived from the logistic models for each of the acoustic parameters, F0 height, F0 contour, peak intensity ratio, and duration ratio using the resynthesized stimuli were evaluated. The correlations between CI participants' question-statement identification accuracy with the sentence stimuli and their sensitivity to the acoustic parameters, F0 height, peak intensity ratio, and duration ratio with the resynthesized stimuli were not found to be statistically significant (all p values > .05). However, there was a moderately strong, positive correlation ($r = .891$, $p < .005$) between CI participants' performance with the sentence stimuli and their sensitivity to F0 contour increments. Note that these correlations may be affected by ceiling effects observed in their performance with the sentence stimuli. Figure 5 illustrates the CI listeners' identification accuracy with sentences as a function of the estimated coefficient for F0 contour with the resynthesized stimuli.

Discussion

The acoustic analysis of the naturally produced sentences indicates that the F0 contour is the most reliable cue for intonation, followed by the intensity cue. The duration cue appears to be the least reliable, being used most variably by individual speakers in the sample. Consistent with the acoustic features of the question-statement contrast, the perceptual results suggest that, under ideal listening conditions, the intensity and duration cues are ignored by listeners. However, under conditions of spectral degradation (or listening with CIs), listeners change their listening strategies and rely more heavily on the secondary and tertiary cues.

The overall intonation identification accuracy was evaluated using naturally produced sentences. Seven of the 13 adult CI users achieved high performance levels (i.e., 90% and above). Nonetheless, as a group, CI listeners' overall performance was significantly poorer than that of NH listeners. The results with spectrally degraded stimuli indicated that NH listeners' question-statement identification based on prosodic information was significantly poorer in the noise-band-vocoded conditions, when compared to that in the full-spectrum conditions. These results were consistent with our own previous findings (Chatterjee & Peng, 2008) and those of others regarding the effects of spectral resolution on speech perception, specifically recognition of phonemes (e.g., Dorman, Loizou, & Rainey, 1997; Fu, Shannon, & Wang, 1998; Fu, Zeng et al., 1998; Munson & Nelson, 2005; Shannon et al., 1995), voice gender identification (Fu et al., 2004, 2005), prosodic cue processing (Green et al., 2002, 2004), and lexical tone recognition (e.g., Xu et al., 2002).

Contrary to previous findings on speech intonation and lexical tone recognition by pediatric CI recipients (e.g., Ciocca et al., 2004; Peng et al., 2004, 2008), the present findings suggest that question-statement identification based on prosodic information is not prohibitively challenging to about half of adult CI users. Note that previous studies in this area generally involved pediatric CI users who were prelingually deaf, while the majority of the CI participants (11 out of 13) in the present study were postlingually deaf. It is possible that question-statement identification based on prosodic information is not as challenging for these postlingually deaf adult CI users as with the prelingually deaf individuals who received a CI in their childhood (Peng et al., 2008). These results might underscore the importance of linguistic experience (prior to deafness) for question-statement identification based on prosodic information.

The present results indicated that the acoustic parameter, F0 height, was utilized by CI listeners as a perceptual cue in question-statement identification. Furthermore, this cue was also used by NH listeners with the full-spectrum stimuli, as well as with the spectrally degraded stimuli. As only two initial F0 heights were used in the present study, the tendency of listeners to use one over the other may be viewed as a "bias" rather than the greater weighting of an acoustic cue. However, there is no clear evidence in favor of using initial F0 height as a bias. In some languages (e.g., Cantonese), initial F0 height is an important acoustic cue for lexical tone recognition: The same may be true in American English. Until further evidence becomes available, we interpret and analyze the results assuming that initial F0 height serves as an acoustic cue to the listener.

Using the same set of resynthesized stimuli and most of the same participants ($N = 10$), Chatterjee and Peng (2008) reported that adult CI listeners' F0 contour sensitivity in a question-statement identification task (identical to Task 2 in the present study) was significantly lower with the 200-Hz initial F0 stimuli than with the 120-Hz initial F0 stimuli. They attributed the CI listeners' poorer performance to the declining usefulness of the temporal pitch cue at high F0s, which was also demonstrated by Green and colleagues (Green et al., 2002, 2004). Consistent with findings of Chatterjee and Peng, the present data also suggest that CI listeners' higher F0 contour sensitivity with the lower 120-Hz initial F0 stimuli was consistent with the NH listeners' pattern of results with the 4-channel stimuli but not with the 8-channel stimuli. It is plausible that the usefulness of temporal pitch cues is inversely related to the extent of spectral degradation. This finding was consistent with the findings of Xu et al. (2002), where trade-off was found between the temporal and spectral cues for lexical tone recognition under acoustic CI simulations. It is also important to note that temporal envelope periodicity cues would be more available in the 4-channel case, with more harmonics of voice speech falling into the broader filters.

Among the four acoustic parameters examined in this study, F0 contour (the dominant acoustic cue signaling the contrast) was utilized consistently by all CI and NH listeners in question-statement identification. That is, the overall proportion of "question" responses became higher along the increment in F0 contour between –1.00 and 1.32 octaves, for both participant groups with the full-spectrum stimuli and under the spectrally degraded conditions for NH listeners. Nonetheless, as revealed from the estimated coefficients in the logistic models, CI listeners' sensitivity to F0 contour was significantly lower than that of NH listeners with the full-spectrum stimuli but was comparable to that of NH listeners with spectrally degraded stimuli. Similarly, for NH listeners, the sensitivity to F0 contour was also lower for the 8- and 4-channel stimuli than for the full-spectrum stimuli.

These findings are consistent with the weak representation of voice pitch via CIs.

Note, however, that considerable intersubject variability was observed among CI listeners in their sensitivity to the F0 contour. As shown in Table 2, the group of CI listeners' estimated coefficients for F0 contour ranged from 1.19 to 7.11. Although these estimated coefficients were all significantly greater than zero, some CI listeners appeared to be more sensitive to the increments in F0 contour than others. We further examined the relationships between the estimated coefficients for each of the acoustic dimensions (F0 height, F0 contour, peak intensity ratio, and duration ratio), on the basis of individual CI listeners' data obtained using the resynthesized stimuli and their performance with naturally produced sentences. The results indicated that CI listeners' identification accuracy with sentences significantly correlated with their F0 contour sensitivity but not with their sensitivity to changes in other acoustic dimensions. These results are again consistent with the notion that the F0 contour provides the primary cue for identifying questions and statements matched for their syntactic structures, whereas the other cues might be available to listeners with relatively limited utility (Cooper & Sorensen, 1981; Ladd, 1996; Lehiste, 1970, 1976). These findings also suggest that the use of highly controlled (i.e., resynthesized) stimuli is informative about listeners' real-world performance (Figure 5). The fact that the CI listeners' performance with the sentence stimuli did not correlate with their use of duration or intensity cues with the resynthesized stimuli is likely because these cues were not available as strongly or reliably in Task 1 as they were in Task 2. The results obtained with Task 2 clearly indicate that when these alternative cues to F0 are made available, listeners may utilize them under conditions of spectral degradation. Furthermore, these results imply that CI speech-processing strategies may benefit from enhancing these cues for CI listeners. The results presented here suggest that some CI listeners are able to use intensity or duration cues to their advantage but others are not. Two possible factors may underlie this observation: (a) Those CI listeners who use alternative cues are also more sensitive to those cues, or (b) all CI listeners are sufficiently sensitive to these alternative cues, but not all of them can adjust their listening strategy to attend to them. In both cases, targeted training may benefit the listener. Adult CI patients today are provided with limited aural rehabilitation/auditory training. Furthermore, even when patients do have access to training, such training rarely emphasizes listening strategies relating to acoustic cue integration. Thus, improved training in these areas may benefit CI patients in speech intonation recognition as well as in speech perception in general.

The F0 contour of speech is critical for the identification of questions and statements matched for their syntactic structures, and CI listeners' question-statement identification accuracy can be moderately predicted by their sensitivity to this acoustic parameter. Chatterjee and Peng (2008) demonstrated that there was a strong correlation between CI listeners' modulation frequency sensitivity, as measured using direct electric stimulation (i.e., bypassing their own speech processor) and their F0 contour sensitivity, as measured using the identical set of resynthesized stimuli in Task 2 of the present study. In fact, among the 13 CI participants in this study, 8 who were users of a Nucleus 22, 24, or Freedom device also served as the participants in Chatterjee and Peng. It is possible that the psychophysical capability to process subtle differences in temporal envelope cues underlies CI listeners' F0 contour sensitivity. It is to be noted that some of the CI listeners in the present study used the SPEAK processing strategy, which transmits information at about 250 pulses/sec on average, per channel. This low carrier rate precludes the proper transmission of temporal envelope information above 125 Hz. It is interesting, therefore, to observe that some of these listeners (CI-9, CI-10, and CI-11) were still able to use F0 cues in the present study. Chatterjee and Peng commented on this, speculating that these listeners may have learned to attend to subsampled temporal cues in the signal, or (perhaps less likely) subtle F0-based spectral differences that might arise from the filtering of the harmonic structure. What cues these listeners were actually using to perform these tasks is as yet unclear.

Consistent with the notion that voice pitch is only weakly presented via CI devices, on average CI listeners' F0 contour sensitivity was reduced relative to that of NH listeners. It was anticipated that CI listeners would be more sensitive to increments in peak intensity and duration ratios, as these are transmitted by the device relatively well. The results indicated that CI listeners' overall proportion of "question" responses became higher, as peak intensity ratio was increased from –10 to 10 dB. This trend was consistent with the results of acoustic analyses of naturally uttered sentences, where questions were observed to have a positive peak intensity ratio (Figure 1). This response pattern was also observed in NH listeners with spectrally degraded stimuli, but not with full-spectrum stimuli, indicating a shift in listening strategy with spectral degradation.

Analyses of results obtained with variations in the duration ratio showed that, as a group, CI listeners' overall proportion of "question" responses did not differ when duration ratio was increased from 0.35 to 1.35. This might be related to the fact that among CI listeners who used this parameter as a cue in question-statement identification (n = 10), the proportion of "question" responses became higher with a greater duration ratio in some CI listeners (n = 6), whereas the proportion became lower with a smaller duration ratio in the others (n = 4). In other words, this acoustic parameter appeared to be utilized in a variable way. Among NH listeners with spectrally degraded stimuli, duration ratio was used as a cue in question-statement identification in a more consistent manner (i.e., the overall proportion of "question" responses became higher when duration ratio was increased from 0.35 to 1.35).

Different response patterns were observed in CI listeners' question-statement identification with changes in intensity and duration patterns. Specifically, compared to duration ratio, peak intensity ratio was used by CI listeners in question-statement identification more consistently. That is, the estimated coefficients for peak intensity ratio were mostly positive in value when they were significantly different from zero, whereas the coefficients for duration ratio were either positive or negative in value (Table 2). It is interesting that this difference between CI listeners' reliance on the intensity and on duration cues was found to be consistent with the observed acoustic properties in adult speakers' production (Figure 1). That is, among F0 contour, intensity, and duration characteristics, the duration pattern tends to be used by speakers to mark question versus statement contrasts in the most variable way. The results from NH listeners also suggest that F0 contour serves as the dominant cue in question-statement identification, but these individuals may utilize cues in other dimensions (i.e., intensity and duration patterns) in those listening conditions in which only limited spectral resolution is available. These results also suggest that although multiple acoustic cues exist, NH listeners may not necessarily utilize (or need) some of these cues under ideal listening conditions. Note that this does not conflict with what we know about how acoustic cues may contribute to NH listeners' identification of questions and statements matched for their syntactic structures. Rather, it provides further evidence that these listeners' patterns of acoustic cue integration in question-statement identification may vary, depending on the listening conditions (e.g., full-spectrum or spectrally degraded).

All four NH participants in the present study were in their 20s, but the majority of the present CI participants were relatively older in age. To address this potential confounding factor, we performed the two tasks with identical sets of full-spectrum and spectrally degraded stimuli with two additional NH listeners, aged 60 and 72, respectively. These two listeners' performance levels in question-statement identification decreased with spectrally degraded stimuli, when compared to their performance levels with full-spectrum stimuli. Their performance patterns in perceptual cue weighting were not observed to be remarkably different from those observed in the four NH listeners in their 20s.

Previous studies that compared the speech perception performance between older adult patients with a CI (e.g., 65 years of age and older) and younger patients (e.g., below 60 years of age) revealed no significant difference in the performance between the two participant groups, when the speech stimuli were presented auditorily (e.g., Haensel et al., 2005; Hay-McCutcheon, Pisoni, & Kirk, 2005). However, Blamey et al. (1996) found that age at implantation did contribute to the postimplantation outcomes in adult CI recipients. In addition, recent studies have shown significant effects of age on hearing adults' ability to process spectrally degraded speech (e.g., Schvartz, Chatterjee, & Gordon-Salant, 2008; Sheldon, Pichora-Fuller, & Schneider, 2008; Souza & Boike, 2006). Recent work by Schvartz and colleagues (Schvartz, 2010; Schvartz & Chatterjee, 2012) suggests that older NH listeners have more difficulty with processing temporal pitch cues than younger NH listeners. It is possible that effects of age may extend to listeners' performance in question-statement identification using prosodic information. This possibility was explored by examining the data obtained with the CI listeners for a relationship between chronological age and (a) the performance levels in Task 1 and (b) the estimated coefficient for each acoustic parameter in Task 2. None of the Pearson correlation coefficients was observed to be statistically significant (r values ranged from .005 to .332, all p values > .05). These results suggest a lack of relationship between the normal aging process and question-statement identification based on prosodic information. Nonetheless, further research with a study design focusing on the effects of aging on prosodic perception will be required before final conclusions are reached.

Under normal listening conditions, the secondary and tertiary cues of interest may not always be changing in the same direction as the primary cue. How might listeners' judgments in the question-statement task be influenced by conflicts between the different cues? Peng et al. (2009) examined a subset of the data presented here for listeners' response patterns when the F0 and intensity cues were specifically cooperating or conflicting with each other. The results demonstrated that when the full complement of spectrotemporal F0 cues were available to the listener, responses were independent of conflicts with the intensity cue. When the F0 cue was degraded, however, listeners placed greater weight on the intensity cue, to the detriment of performance when the intensity cue conflicted with the primary (F0) cue.

In summary, the use of resynthesized stimuli in the present study has allowed a deeper evaluation of CI and NH listeners' integration of multiple acoustic cues in question-statement identification than would otherwise be possible. These findings have broadened our understanding of the perceptual basis for question-statement identification based on prosodic information, via both electric and acoustic hearing. The findings reported here have clinical implications, as they confirm the importance of enhancing CI users' acoustic cue integration in aural rehabilitation/auditory training programs. Future studies should address issues regarding how listening in challenging conditions (e.g., in competing noise) may reshape listeners' acoustic cue integration.

Conclusions

1. Acoustic analyses indicated that the F0 contour was the most reliable cue signaling the question-statement contrast, followed by intensity change. Duration change was the least reliable cue. The bisyllabic word in the sentence-final position

appeared to carry the dominant cues (F0, duration, and intensity patterns) for the contrast.

2. With naturally produced sentences, CI listeners' overall accuracy in question-statement identification was significantly poorer than that of NH listeners with the full-spectrum stimuli. Furthermore, reduced spectral resolution (as with acoustic CI simulations) resulted in poorer question-statement identification accuracy in NH listeners' performance in the task. These results are entirely consistent with those reported by Chatterjee and Peng (2008).
3. With resynthesized stimuli, NH listeners utilized the F0 information, particularly F0 contour, as the primary cue in question-statement identification when listening to the full-spectrum stimuli. They did not utilize peak intensity ratio or duration as cues in recognition with the full-spectrum stimuli. This indicates that secondary cues are ignored when the dominant/primary cue is strongly represented.
4. When listening to the resynthesized stimuli that were spectrally degraded, NH listeners exhibited reduced F0 contour sensitivity. This is consistent with results reported by Chatterjee and Peng (2008). The novel finding in the present study is that of an increased reliance on intensity and duration cues under conditions of spectral degradation. That is, trade-off relationships were observed between NH listeners' utilization of F0 cues and intensity/duration cues under full-spectrum versus spectrally degraded conditions.
5. Unlike NH listeners with the resynthesized stimuli in the full-spectrum condition, CI listeners were sensitive to changes in peak intensity ratio in question-statement identification; they utilized this peak intensity ratio as a cue in a fairly consistent fashion. Furthermore, CI listeners' sensitivity to this cue was comparable to that of NH participants with spectrally degraded stimuli. This finding is consistent with the finding in Peng et al. (2009), in which it was shown that CI listeners used the F0 contour and intensity cues in a combined fashion in intonation recogntion. This is also consistent with the acoustic analyses showing that, after F0, the intensity cue is the second most reliable cue signaling the question-statement contrast.
6. With the resynthesized stimuli in the full-spectrum condition, both NH and CI listeners were variable in their utilization of duration ratio as a cue in question-statement identification. This is consistent with the acoustic analyses showing that the duration cue is the least reliable cue signaling the question-statement contrast. With spectrally degraded stimuli, however, NH listeners used this acoustic parameter as a cue rather consistently. This suggests the important role for training in CI patients' rehabilitation.
7. CI listeners' F0 contour sensitivity derived from the resynthesized stimuli reasonably predicts their question-statement identification using the naturally produced sentences. While using naturally produced stimuli permitted evaluations of listeners' actual capability in identifying questions and statements matched for their syntactic structures, using resynthesized stimuli shed further light on listeners' integration of specific acoustic cues in such tasks.

Acknowledgments

We are grateful to all participants in the present study. We thank Kelly Hoffard and Kara Schvartz for their assistance in some of the data collection, Shubhasish B. Kundu for developing custom software programs for this study, and Dr. Qian-Jie Fu for allowing us to use his software for generating acoustic simulations as well as for controlling the consonant and vowel recognition tasks. We thank the reviewers for their constructive comments on an earlier version of the article.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the NIH/NIDCD (R01DC004786, PI: M. Chatterjee and P30DC004664, PI: R.Dooling). The views presented in this article do not necessarily reflect those of the Food and Drug Administration.

References

Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G.,... Whitford, L. (1996). Factors affecting auditory performance of post-linguistically deaf adults using cochlear implants. *Audiology and Neurotology, 1*, 293-306.

Boersma, P., & Weenink, D. (2004). Praat [Computer Software], Version 4.3. Amsterdam, Netherlands: Institute of Phonetic Sciences, University of Amsterdam.

Chao, Y.-R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chatterjee, M., & Peng, S. (2008). Processing fundamental frequency contrasts with cochlear implants: Psychophysics and speech intonation. *Hearing Research, 235*, 145-156.

Ciocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *Journal of the Acoustical Society of America, 111*, 2250-2256.

Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production*. New York, NY/Berlin, Germany: Springer-Verlag.

Cruttenden, D. (1986). *Intonation.* Cambridge, UK: Cambridge University Press.

Dorman, M. F., Hannley, M., McCandless, G., & Smith, L. M. (1988). Acoustic/phonetic categorization with the Symbion multichannel cochlear implant. *Journal of the Acoustical Society of America, 84*, 501-510.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America, 102*, 2403-2411.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112*, 259-271.

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

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*, 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Langauge and Speech, 1*, 126-152.

Fu, Q. J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice-gender recognition by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology, 5*, 253-260.

Fu, Q. J., Chinchilla, S., Nogaki, G., & Galvin, J. J. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *Journal of the Acoustical Society of America, 118*, 1711-1718.

Fu, Q. J., & Nogaki, G. (2004). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology, 6*, 19-27.

Fu, Q. J., & Shannon, R. V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *Journal of the Acoustical Society of America, 105*, 1889-1990.

Fu, Q. J., Shannon, R. V., & Wang, X. S. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *Journal of the Acoustical Society of America, 104*, 3586-3596.

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America, 104*, 505-510.

Gfeller, K., Olszewski, C., Rychener, M., Sena, K., Knutson, J.F., Witt, S., & Macpherson, B. (2005). Recognition of "real-world" musical excerpts by cochlear implant recipients and normal-hearing adults. *Ear and Hearing, 26*, 237-250.

Gfeller, K., Turner, C., Woodworth, G., et al. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International, 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. *Journal of the Acoustical Society of America, 112*, 2155-2164.

Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *Journal of the Acoustical Society of America, 116*, 2298-2310.

Green, T., Faulkner, A., Rosen, S., & Macherey, O. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *Journal of the Acoustical Society of America, 118*, 375-385.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *Journal of the Acoustical Society of America, 87*, 2592-2605.

Haensel, J., Ilgner, J., Chen, Y. S., Thuermer, C., & Westhofen, M. (2005). Speech perception in elderly patients following cochlear implantation. *Acta Oto-Laryngologica, 125*, 1272-1276.

Hay-McCutcheon, M. J., Pisoni, D. B., & Kirk, K. I. (2005). Audio-visual speech perception in elderly cochlear implant recipients. *Laryngoscope, 115*, 1887-94.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, UK: Cambridge University Press.

Iverson, P., Smith, C. A., & Evans, B. G. (2006). Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration. *Journal of the Acoustical Society of America, 120*, 3998-4006.

Jusczyk, P. W., Hirsch-Pasek, K., Kemler Nelson, D. G., Kennedy, L.J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*, 252-293.

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

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, UK: Cambridge University Press.

Laneau, J., Wouters, J., & Moonen, M. (2006). Improved music perception with explicit pitch coding in cochlear implants. *Audiology and Neurotology, 11*, 38-52.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1976). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225-239). New York, NY: Academic Press.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 12-22.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech, 29*, 3-11.

Luo, X., Fu, Q. J., Wei, C. G., & Cao, K. L. (2008). Speech recognition and temporal amplitude modulation processing by

Mandarin-speaking cochlear implant users. *Ear and Hearing, 29*, 957-970.

Munson, B., & Nelson, P. B. (2005). Phonetic identification in quiet and in noise by listeners with cochlear implants. *Journal of the Acoustical Society of America, 118*, 2607-2617.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America, 112*, 711-719.

Nittrouer, S. (2005). Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *Journal of the Acoustical Society of America, 118*, 1072-1088.

Peng, S. C. (2005). *Perception and production of speech intonation in pediatric cochlear implant recipients and children with normal hearing*. PhD dissertation, University of Iowa, Iowa City.

Peng, S. C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech recognition by cochlear implant users and normal hearing listeners. *Audiology and Neurotology, 14*, 327-337.

Peng, S. C., Tomblin, J. B., Cheung, H., Lin, Y. S., & Wang, L. S. (2004). Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear and Hearing, 25*, 251-264.

Peng, S. C., Tomblin, J. B., & Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear and Hearing, 29*, 336-351.

Qin, M., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and Hearing, 26*, 451-460.

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

Schvartz, K. C. (2010). *Effects of aging on voice-pitch processing: The role of spectral and temporal cues*. PhD dissertation, University of Maryland, College Park.

Schvartz, K. C., & Chatterjee, M. (2012). Gender identification in younger and older adults: Use of spectral and temporal cues in noise-vocoded speech. *Ear and Hearing* (published online ahead of print January 2012). doi:10.1097/AUD.0b013e31823d78dc).

Schvartz, K. C., Chatterjee, M., & Gordon-Salant, S. (2008). Recognition of spectrally degraded phonemes by younger, middle-aged and older normal-hearing listeners. *Journal of the Acoustical Society of America, 124*, 3972-3988.

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

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Effects of age, presentation method, and learning on identification of noise-vocoded words. *Journal of the Acoustical Society of America, 123*, 476-488.

Soderstrom, M. Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*, 249-267.

Souza, P. E., & Boike, K. T. (2006). Combining temporal-envelope cues across channels: Effects of age and hearing loss. *Journal of Speech, Language, and Hearing Research, 49*, 138-149.

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

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*, 53-71.

Wei, C. G., Cao, K. L., & Zeng, F. G. (2004). Mandarin tone recognition in cochlear-implant subjects. *Hearing Research, 197*, 87-95.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*, 25-47.

Winn, M. B. (2011). *The use of acoustic cues in phonetic perception: Effects of spectral degradation, limited bandwidth and background noise*. PhD dissertation, University of Maryland, College Park.

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *Journal of the Acoustical Society of America, 131*, 1465-1479.

Xu, L., Thompson, C. S., & Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *Journal of the Acoustical Society of America, 117*, 3255-3267.

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