Published in final edited form as:
Audiol Neurootol. 2009 ; 14(5): 327–337. doi:10.1159/000212112.

Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners

Shu-Chen Peng[a,b], **Nelson Lu**[a], and **Monita Chatterjee**[b]

[a] US Food and Drug Administration, Rockville

[b] University of Maryland, College Park

Abstract

Cochlear implant (CI) recipients have only limited access to fundamental frequency (F0) information, and thus exhibit deficits in speech intonation recognition. For speech intonation, F0 serves as the primary cue, and other potential acoustic cues (e.g., intensity properties) may also contribute. This study examined the effects of acoustic cues being cooperating or conflicting on speech intonation recognition by adult cochlear implant (CI), and normal-hearing (NH) listeners with full-spectrum and spectrally degraded speech stimuli. Identification of speech intonation that signifies question and statement contrasts was measured in 13 CI recipients and 4 NH listeners, using resynthesized bi-syllabic words, where F0 and intensity properties were systematically manipulated. The stimulus set was comprised of tokens whose acoustic cues, i.e., F0 contour and intensity patterns, were either cooperating or conflicting. Subjects identified if each stimulus is a “statement” or a “question” in a single-interval, two-alternative forced-choice (2AFC) paradigm. Logistic models were fitted to the data, and estimated coefficients were compared under cooperating and conflicting conditions, between the subject groups (CI vs. NH), and under full-spectrum and spectrally degraded conditions for NH listeners. The results indicated that CI listeners’ intonation recognition was enhanced by F0 contour and intensity cues being cooperating, but was adversely affected by these cues being conflicting. On the other hand, with full-spectrum stimuli, NH listeners’ intonation recognition was not affected by cues being cooperating or conflicting. The effects of cues being cooperating or conflicting were comparable between the CI group and NH listeners with spectrally-degraded stimuli. These findings suggest the importance of taking multiple acoustic sources for speech recognition into consideration in aural rehabilitation for CI recipients.

Keywords

cochlear implants; intonation; speech; perception

INTRODUCTION

Cochlear implant (CI) devices provide restored functional hearing to patients with a bilateral severe-profound hearing loss. The reduced spectral resolution via these auditory prosthetic devices, however, makes it challenging to present fundamental frequency (F0, or voice pitch) information to CI listeners [Faulkner et al., 2000; Geurts and Wouters, 2001; Green et al.,

Corresponding Author: Shu-Chen Peng, Ph.D., Division of Ophthalmic and Ear, Nose and Throat Devices, Office of Device Evaluation, Center for Devices and Radiological Health, US Food and Drug Administration (HFZ-460), 9200 Corporate Blvd., Rockville, MD 20850, Phone: (240) 276-4242; Fax: (240) 276-4234, E-mail: shu-chen.peng @fda.hhs.gov.

2002, 2004]. Voice pitch is important for recognition of speech prosodic properties, such as intonation and lexical tones. This limitation has thus resulted in CI listeners' poor recognition of speech intonation [Green et al., 2002, 2004; Peng et al., 2008] and lexical tones [Ciocca et al., 2002; Peng et al., 2004]. Fortunately, even though F0 serves as the primary cue for recognition of speech prosodic contrasts, it is not the only possible cue; additional sources of acoustic cues, such as intensity and duration patterns may also contribute to this recognition [e.g., Lehiste, 1976]. On the other hand, although multiple acoustic cues may contribute, these acoustic dimensions in natural speech may not always be present cooperatively. Listeners generally show augmented identification of speech contrasts when cues are combined cooperatively, but show reduced identification when cues are in conflict or when only a single cue is available [Eilers et al., 1989; Hazan and Rosen, 1991].

The effects of acoustic cues being cooperating or conflicting in acoustic signals may be different for speech recognition, depending upon factors such as the listener's age or linguistic/auditory experience [Eilers et al., 1989; Morrongiello et al., 1984]. Given CI listeners' unique auditory experience with electric hearing, it is plausible that the effects of cues being cooperating or conflicting on their speech recognition would be greater than that observed in normally hearing (NH) listeners with acoustic hearing. It has been demonstrated that via electric hearing, CI listeners have only limited spectral resolution and access to F0 information that is critical for recognition of prosodic contrasts [Faulkner et al., 2000; Geurts and Wouters, 2001; Green et al., 2002, 2004]. However, very little is known about the effects of multiple acoustic cues (e.g., F0 contour and intensity cues) that are cooperating or conflicting on CI recipients' speech intonation recognition.

Maximization of acoustic cues being cooperating and coping of cues being conflicting are important in any aural rehabilitation programs designed for CI listeners. However, the effects of acoustic cues being cooperating or conflicting on listeners' speech perception can not be assessed using any clinically available speech materials. In fact, outcome measures that focus on CI recipients' speech prosodic recognition are rather limited in clinical settings. Recognition of intonation and lexical tones can be measured in CI listeners using recorded, naturally-produced utterances (e.g., words, phrases, or sentences). These naturally-produced utterances, however, almost always contain contextual cues and multiple acoustic sources that collectively contribute to listeners' intonation or lexical tone recognition. As a result, although the results based on these testing materials may inform CI users' real-world functionality, they are limited in providing the information regarding these individuals' utilization of specific acoustic cue (s), as well as cue integration in their perception.

To overcome such limitations, Peng and colleagues developed a set of resynthesized stimuli that were systematically varied in multiple acoustic dimensions (i.e., F0, intensity, and duration patterns). This was performed by following the actual acoustic variations observed in speech intonation production by adult speakers who are native speakers of American English [Peng et al., unpublished data]. These resynthesized stimuli were constructed in a way to permit an in-depth examination of CI listeners' utilization of specific acoustic cues in recognizing speech intonation that signifies question and statement contrasts *(speech intonation* or *intonation*, hereafter). The purpose of this study was to further examine the effects of acoustic cues (F0 contour and intensity patterns) being cooperating or conflicting on adult CI recipients' intonation recognition, as well as on that of NH listeners under full-spectrum and spectrally degraded conditions (i.e., acoustic CI simulations).

Materials and Methods

Subjects

Thirteen adult CI recipients served as subjects. Their mean (±SD) age was 58.92±13.96 years, and their mean (±SD) length of device experience was 6.69±4.61 years. All subjects spoke American English as their first language, and had a minimum of one year of device experience. Two subjects were prelingually deaf (CI-5 and CI-13), and the remaining 11 subjects were postlingually deaf. These subjects used either a Nucleus or a Clarion device. Subjects were mapped with a variety of speech-coding strategies (Table 1). In addition to the CI group, four adult NH listeners in their 20's were also recruited to obtain a reference baseline of adult NH listeners' performance in the same task. All had hearing sensitivity better than 20 dB HL from 250 to 8000 Hz at octave intervals, bilaterally. All subjects gave written informed consent approved by the University of Maryland, College Park Institutional Review Board (ERB) prior to participation.

Stimuli

Speech intonation recognition was measured using resynthesized bi-syllabic words, which permitted direct evaluations of listeners' potential use of multiple acoustic cues, as well as their integration of these cues in perception. The ranges of parametric manipulations (Figure 1) were determined on the basis of the observed acoustic characteristics in speech intonation production by multiple adult speakers of American English [Peng et al., unpublished data]. Specifically, a rising F0 contour and a greater intensity ratio between the 2^{nd} and 1^{st} syllables were rather consistently observed in adult speakers' production of questions, whereas a falling F0 contour and a smaller intensity ratio were consistently observed in their production of statements.

Based on these acoustic findings, one bi-syllabic word ("popcorn") was acoustically manipulated in a systematic manner using the Praat software [Boersma and Weenink, 2004]. Bi- syllabic words were adopted, as the 1^{st} syllable could be used as the reference for the parametric variations to the F0 and intensity properties of the 2^{nd} syllable. These stimuli were demonstrated as valid, as the data obtained using these resynthesized stimuli correlate strongly to CI listeners' performance using naturally-uttered sentence materials [Peng et al., unpublished], as well as to CI listeners' psychophysical sensitivity [Chatterjee and Peng, 2008]. The reynthesized tokens were generated by orthogonally varying the two acoustic parameters, i.e., F0 contour and intensity properties, in addition to F0 height and duration properties. Figure 1 illustrates examples of bi-syllabic tokens via parametric manipulations. Of note, F0 contour and intensity properties were the two primary parameters examined in this study, as previous findings indicated that CI listeners who are native speakers of American English exhibit the tendency to use these two parameters more consistently than the other acoustic properties, i.e., duration [Peng et al., unpublished]. Detailed manipulation in each dimension is described as follows:

- **a.** F0 height (Figure 1a): Two flat F0 contours with 120- and 200-Hz initial-F0 heights for the vocalic portions of the two syllables were first generated to simulate male and female speakers.
- **b.** F0 contour (Figure 1b): Each of the F0 height-adjusted tokens was further manipulated to vary the F0 contour, in the form of glides, from the onset to offset of the vocalic portion of each entire bi-syllabic token in nine steps (−1.00, −0.42, −0.19, 0, 0.17, 0.32, 0.58, 1.00, and 1.32 octaves). For the 120-Hz tokens, the target F0s at the offset were 60, 90,105,120,135, 150, 180, 240, and 300 Hz; for the 200-Hz tokens, the target F0s at the offset were 100, 120, 150, 175, 200, 225, 300, 400, and 500 Hz. Note that according to the above-mentioned acoustic findings, while statements are associated

with a falling F0 contour (parameter values < 0), questions are associated with a rising contour (parameter values > 0).

c. Peak intensity ratio (Figure 1c): Each of the F0 height- and variation-adjusted tokens was further varied for the peak intensity ratio of the 2nd syllable. The peak intensity ratio of each token was systematically varied (i.e., −10, −5, 0, 5, and 10 dB), relative to the peak amplitude of the 1st syllable of a reference token (see the panel indicating "0 dB" in Figure 1c). 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 (Figure 1d): Each of the F0 height-, F0 contour-, and intensity-adjusted tokens was further manipulated to vary the duration of the 2nd syllable, relative to a reference token (a neutral token that was naturally produced as part of a question). The duration of the 2nd syllable was shortened or lengthened by a factor of 0.65, 1.00, 1.35, and 1.70 (see the panel indicating "1.00" in Figure 1d). Note that a greater ratio (i.e., longer duration) roughly corresponds to questions, but this parameter is not used by speakers to mark contrasts between statements and questions as consistently as the F0 or intensity parameters.

Reynthesized tokens were generated by varying four acoustic parameters orthogonally, resulting in a total of 360 tokens (1 bi-syllabic word × 2 steps of F0 height × 9 steps of F0 contour × 5 steps of peak intensity ratio × 4 steps of duration ratio). Notably, as above-mentioned, the two major acoustic properties of interest in this study were F0 contour and intensity patterns that were present cooperatively or in conflict. Thus, although in each run of experiment, 360 tokens were played to the listener, the actually number of target tokens that were analyzed for the purpose of this study were comprised of a subset of these tokens.

Specifically, the target stimulus set contained 128 tokens with which F0 contour and intensity cues were either cooperating (N = 80) or conflicting (N = 48). In this study, the cooperating condition referred to the combinations with both F0 contour and intensity cues greater or less than zero, whereas the conflicting condition referred to the combinations with one of these two cues greater than zero, and the other less than zero (Table 2). These definitions were consistent with the observed acoustic characteristics in questions and statements produced by adult speakers of American [Peng et al., unpublished data]. Further, this stimulus set included 40 tokens in one additional condition, where F0 contour was kept neutral (i.e., a flat F0 contour; see F0 contour fixed at "0 octave" in Figure 1b), while the intensity cue was made available (i.e., with varying peak intensity ratios), was also created (Table 2). Results from this condition would presumably confirm the hypothesis that given their limited access to F0 information (particularly, F0 contour), CI listeners' response in intonation recognition would depend primarily on the intensity cue. The resulting target stimulus set, in each experimental run, was comprised of 168 tokens.

Signal processing

All resynthesized stimuli were noise-vocoded to create acoustic CI simulations [Shannon et al., 1995], using software developed by Dr. Qian-Jie Fu (Tigerspeech Technology, House Ear Institute). The speech stimuli were bandpass filtered into 8 and 4 frequency bands (denoted as the 8-channel and 4-channel conditions), as in these conditions NH listeners' speech recognition performance tend to best mimic the range of CI listeners' performance [e.g., Friesen et al., 2001]. The input frequencies ranged from 200 to 7000 Hz; corner frequencies for each analysis band were determined according to Greenwood's frequency-to-place mapping [Greenwood, 1990], as described by Fu and Shannon (1999) assuming a 35 mm long cochlea. The resulting bandwidths of each filter yielded a fixed cochlear distance in mm. Specifically, the corner frequencies were 200, 359, 591, 931, 1426, 2149, 3205, 4748 and 7000 Hz for the

8-channel condition, and were 200, 591, 1426, 3205 and 7000 Hz for the 4-channel condition. These conditions were identical to those described in previous studies [e.g., Fu and Nogaki, 2004]. The temporal envelope from each band was extracted by half-wave rectification and lowpass filtering (cutoff at 400 Hz, −24 dB/octave). The extracted envelope from each band was used to amplitude modulate white noise. For each channel, the modulated noise was spectrally limited by the same bandpass filter used for frequency analysis. The output signals from all channels were summed, and the long-term root mean square (RMS) amplitude was adjusted to match that of the full-spectrum input signals.

Procedure

All stimuli were presented in soundfield, at approximately 65 dB SPL (A-weighting), via a single loudspeaker (Tannoy Reveal) in a sound-treated booth (IAC). All stimulus tokens in the cooperating, conflicting and neutral conditions were presented to each listener in random order, in two runs. The CI listeners were tested using their clinically assigned speech processors, with clinically recommended volume and sensitivity settings. The NH listeners were tested while listening in the full-spectrum and acoustic CI simulation conditions. A single-interval, 2-alternative forced-choice (2AFC) paradigm was adopted. In this paradigm, a stimulus was presented, and the subject was instructed to choose between two response alternatives, labeled as “statement” and “question” on the computer monitor. The entire stimulus set was tested in one experimental run, and two runs were performed with each CI subject under the full-spectrum listening condition, and with each NH subject under the full-spectrum condition and each of the simulation conditions. The order of listening conditions was randomized among the NH listeners. A Matlab-based user interface was used to control the stimulus presentation, and to record the subject’s responses. Prior to testing, each subject was presented with a small set of stimuli similar to test stimuli for task familiarization. No feedback was provided during testing.

Scoring and data analysis

A score of “one” and “zero” was assigned to “question” and “statement” responses, respectively, in order to derive psychometric functions. The proportion of “question” responses was calculated for each subject, in various listening conditions. Simple logistic regression models were adopted to estimate the regression coefficients for each subject/listening condition (i.e., CI full-spectrum, NH full-spectrum; NH 8-channel; NH 4-channel). In the cooperating and conflicting conditions, two parameters, F0 contour and peak intensity ratio were included as independent variables, while in the neutral condition, peak intensity ratio was included as the independent variable. In all conditions, two additional independent variables, F0 height and the duration properties, were also included as random variables in the models, and the subject’s response was included as the dependent variable. In these models, the estimated coefficient for each parameter was used to approximate the listener’s sensitivity to F0 contour and intensity cues in intonation recognition.

At the group level, data from each single subject were correlated. Thus, logistic models were performed to estimate coefficients of the acoustic parameters, F0 contour and peak intensity ratio, using the generalized estimating equations (GEE) method. The Wald test was used to examine (a) whether the estimated coefficient was different from 0, and (b) whether the estimated coefficient was different across groups. The effect of F0 contour and intensity cues being cooperating or conflicting on listeners’ intonation recognition was defined as the difference in estimated coefficients between the cooperating and conflicting conditions. Statistical analyses were performed using PROC GENMOD of SAS.

RESULTS

As a group, CI subjects' response patterns in intonation recognition were quite different between the cooperating and conflicting conditions (Figure 2a). Specifically, the slope of their psychometric function was steeper in the cooperating condition than that in the conflicting condition. The estimated coefficient for F0 contour was significantly greater for the cooperating condition than that for the conflicting condition ($Z = 4.01$, $p < 0.0001$) (Table 3). That is, CI subjects' certainty of labeling questions and statements in accordance with their target utterance types was considerably higher in the cooperating condition than in the conflicting condition.

The NH group's response patterns were also compared between the cooperating and conflicting conditions (Figure 2b). Unlike the different patterns between these two conditions observed for the CI group, the the slope of the NH group's psychometric function did not appear to differ between the cooperating and conflicting conditions. The estimated coefficient for F0 contour was not found to be significantly different between these two conditions ($Z = 0.23$, $p = 0.8165$). The difference in the effects of acoustic cues being cooperating or conflicting between the CI and NH groups was statistically significant (χ^2 (df = 1) = 9.49, $p = 0.0021$).

Under the 8- and 4-channel conditions, on the other hand, the slope of NH subjects' psychometric function was found to be steeper in the cooperating condition than that in the conflicting condition (Figures 2c and 2d). In both conditions, the estimated coefficient for F0 contour was significantly greater when acoustic cues were cooperating than when these cues were conflicting ($Z = 6.41$ and $Z = 5.24$ for the 8- and 4-channel conditions, respectively; both p-values < 0.0001). That is, when attending to acoustic CI simulations, NH subjects showed significantly higher certainty in labeling questions and statements in accordance with their utterance types in the cooperating condition than in the conflicting condition.

As displayed in Figure 2, the effect of F0 contour and intensity cues being cooperating or conflicting on the group of CI listeners' response patterns in intonation recognition appeared to be fairly similar to that on the group of NH listeners' response patterns in the 8- and 4-channel conditions. The response patterns with cooperating or conflicting cues were compared between the CI group and the NH group with 8- and 4- channel acoustic CI simulations: The differences in the estimated coefficients for F0 contour between the cooperating and conflicting conditions were not found to be statistically significant (all p-values > 0.3298), with the exception that in the conflicting condition, the estimated coefficients for F0 contour between the CI group and the NH 8-channel condition were statistically significantly different ($Z = 2.53$, $p = 0.0113$). Further, the magnitude of this effect (i.e., differences in estimated coefficients) between the two conditions was also not found to be statistically significantly different between the CI group with full-spectrum stimuli and the NH group with spectrally degraded stimuli (CI vs. 8-channel: χ^2 (df = 1) = 2.87, $p = 0.0903$; CI vs. 4-channel: χ^2(df = 1) =1.17, $p = 0.2794$). Together, F0 contour and intensity cues being cooperating or conflicting played a role in the response patterns of intonation recognition by CI subjects with full-spectrum stimuli, as well as by NH subjects under acoustic CI simulations. On the other hand, with full-spectrum stimuli, NH subjects' intonation recognition did not differ when these two cues were cooperating or conflicting.

Data were also obtained from one additional condition, where the intensity cue was systematically manipulated, while F0 contour was kept neutral (i.e., a flat F0 contour). As was mentioned earlier, results from this condition may assist in examining the hypothesis that CI listeners must rely on the intensity cue in intonation recognition, in the absence of the F0 information. Figure 3 illustrates CI and NH subjects' response patterns in intonation recognition as a function of peak intensity ratio in this condition with a flat F0 contour. With

full-spectrum stimuli, the slope of CI subjects' psychometric function was significantly different from zero ($p < 0.0001$), whereas the slope of NH subjects' psychometric function was not found to significantly differ from zero ($p = 0.0600$). This slope was found to be significantly steeper for CI subjects than that for NH subjects, under the full-spectrum listening condition ($p < 0.0001$). Moreover, similar to CI subjects with full-spectrum stimuli, NH subjects' exhibited a positive slope when listening to spectrally degraded stimuli. This slope was found to be significantly different from zero in both the 8- and 4-channel conditions (8-channel: $p < 0.0001$; 8-channel: $p = 0.0013$); both slopes were found to be significantly different from that of NH subjects with full-spectrum stimuli (both p-values < 0.0001). There was no significant difference observed between the 8- and 4-channel conditions ($p = 0.2062$). The difference was also not found to be different between CI subjects in the full-spectrum condition and NH subjects in both spectrally degraded conditions (full-spectrum vs. 8-channel: $p = 0.6487$; full-spectrum vs. 4-channel: $p = 0.8550$).

Inter-subject variability was examined for the individual CI and NH subjects. Although there were only four subjects in the NH group, these listeners' response patterns were rather consistent, with both full-spectrum and spectrally degraded stimuli. On the other hand, substantial inter-subject variability was observed among the 13 CI recipients (Figure 4). Several CI listeners' response patterns in intonation recognition did not appear to differ between the cooperating and conflicting conditions, whereas the other CI listeners' response patterns appeared to be different between the two conditions. The effect of F0 contour and intensity cues being cooperating or conflicting (i.e., differences in estimated coefficients between the cooperating and conflicting conditions) was derived for each CI subject (Table 3). This effect was found to be statistically significant (i.e., different from zero) for eight subjects ($p < 0.0001$ for CI-2, CI-4, CI-5, CI-9, CI-10, CI-11 and CI-13; $p = 0.0262$ for CI-7).

The estimated coefficients for F0 contour in both cooperating and conflicting conditions, and the effects of cues being cooperating or conflicting were examined in relation to individual CI listeners' identification accuracy measured with naturally-produced sentences (see Table 3, last column for detailed scores for CI subjects). These listeners' identification accuracy with naturally-produced stimuli was found to be significantly correlated with the estimated coefficient for F0 contour in the conflicting condition ($r = 0.686$; $p = 0.010$), but not with that in the cooperating condition ($r = 0.528$; $p = 0.064$). Further, the correlation between CI listeners' identification accuracy with naturally-produced stimuli and the effect of cues being cooperating or conflicting was also not found to be statistically significant ($r = 0.210$; $p = 0.491$).

DISCUSSION

Although multiple acoustic cues may contribute to the perceptual identity of a speech contrast in natural speech, these cues are not always available to listeners cooperatively. Listeners' integration of acoustic cues has been shown to depend on the degree to which acoustic cues cooperate or conflict [Eilers et al., 1989; Kazan and Rosen, 1991]. This study examined the effects of the two acoustic cues, F0 contour and intensity cues being cooperating or conflicting on speech intonation recognition via electric hearing (as with CI listeners and NH listeners with CI acoustic simulations) and acoustic hearing (as with NH listeners under the full-spectrum condition). The effect of acoustic cues (F0 contour and intensity patterns) being cooperating or conflicting was found to be significant on speech intonation recognition by CI listeners with full-spectrum stimuli, as well as by NH listeners under spectrally degraded conditions (acoustic CI simulations). On the other hand, whether or not cues being cooperating or conflicting was not found to have any effect on intonation recognition by NH listeners with full-spectrum stimuli. These findings suggest that different auditory experience (i.e., electric hearing via a CI or acoustic CI simulations vs. acoustic hearing by NH listeners) contributes

to the extent to which listeners' intonation recognition would be affected by acoustic cues being cooperating or conflicting. This is consistent with previous findings on speech perception in relation to acoustic cues being cooperating vs. conflicting cues [e.g., Eilers et al., 1989]. These findings collectively suggest that auditory experience (or the type of auditory inputs) plays an important role in speech perception, and this study provides evidence that such effects are extended to recognition of prosodic properties speech, particularly speech intonation that signifies question and statement contrasts.

Only four listeners in the NH group participated in this study. These four NH listeners, however, exhibited fairly uniform response patterns, with both full-spectrum and spectrally degraded stimuli. Specifically, these NH listeners' intonation recognition was adversely affected under spectral degradation (as seen in the 8- and 4-channel conditions) when the F0 contour and intensity cues were conflicting, but was not affected at all when they attended to full-spectrum stimuli. These results are consistent with previous findings regarding the adverse effect of cues being conflicting on speech perception [Eilers et al., 1989; Kazan and Rosen, 1991], and further suggest that the effect of cues being cooperating or conflicting may be relatively evident when in challenging conditions, particularly under spectral degradation.

Recognition of prosodic properties of speech, such as lexical tones and intonation has been reported to be one challenging aspect in speech perception for CI listeners. This can be attributed to the limited spectral resolution critical for F0 information via the speech coding of contemporary CI processors [Faulkner et al., 2000; Geurts and Wouters, 2001; Green et al., 2002, 2004]. The present results indicated that under spectral degradation (as with CI listeners and NH listeners under acoustic CI simulations), speech intonation recognition is adversely affected when F0 contour and intensity cues are present in conflict than cooperatively. In naturally-produced utterances, it is not uncommon that multiple acoustic dimensions become available to listeners in conflict [Eilers et al., 1989; Kazan and Rosen, 1991]. These results suggest that conflicting cues present in speech may possibly make it extremely challenging for CI listeners' recognition of intonation that signifies question and statement contrasts.

In the conflicting condition, the intensity patterns are deviated from the primary cue, F0 contour that listeners rely upon in intonation recognition. Thus, in this condition, any listener who continues utilizing F0 contour in intonation recognition would suggest that the listener relies primarily upon the F0 information, discarding any intensity-related information. The present results indicated an estimated coefficient that was significantly different from zero for CI listeners as a group with full-spectrum stimuli, as well as for NH listeners as a group under the 8-and 4-channel conditions. Nonetheless, the estimated coefficients for CI listeners with full-spectrum stimuli, and NH listeners under spectral degradation were found to be significantly lower than the coefficients for the NH listeners with full-spectrum stimuli. These results suggest that F0 contour continues serving as the primary cue when it conflicts with the intensity pattern, but in this conflicting condition, the extent of listeners' reliance upon F0 contour in speech intonation recognition is relatively limited.

Fundamental frequency servers as the primary cue in intonation recognition, even in the conflicting condition. What happens when the F0 information becomes obscured? The present results indicated that with a flat F0 contour, utilization of the intensity patterns in intonation recognition was observed in both CI listeners with full-spectrum stimuli and NH listeners with spectrally degraded stimuli, but not in NH listeners with full-spectrum stimuli. Together, these results are consistent with the previous findings that indicated the fact that F0 serves as the primary cue in speech intonation recognition, and intensity may also contribute to this recognition [e.g., Lehiste, 1976]. The present findings further suggest that with full-spectrum stimuli, whereas CI listeners would rely upon the intensity cue in intonation recognition, NH listeners would not use this cue in intonation recognition, even when the F0 contour information

is obscured. Further, variations in intensity patterns are conveyed primarily via temporal envelope fluctuations [Rosen, 1992]. The fact that CI listeners utilize not only F0 contour but also intensity cues to recognize speech intonation is consistent with the notion that CI listeners are generally capable of using the temporal envelope information in speech signals [Green et al., 2002, 2004]. Also, interestingly, listeners' potential use of both F0 contour and intensity in speech intonation recognition is similar to the previous findings on Mandarin lexical tone recognition; that is, F0 contour serves as the primary cue, while intensity properties may also contribute [Whalen and Xu, 1992].

As a group, CI listeners' intonation recognition was affected by F0 contour and intensity cues being cooperating or conflicting. However, inter-subject variability was observed in the effects of cues being cooperating or conflicting on CI listeners' intonation recognition. Among 13 CI listeners, the effect was found to be significant for eight individuals (CI-2, CI-4, CI-5, CI-7, CI-9, CI-10, CI-11, and CI-13). Notably, the effect being weak indicates that the listener use F0 contour as the primary cue, whereas such effect being strong indicates that the listener also use secondary cues such as intensity patterns.

In this study, the effect of cues being cooperating or conflicting, as well as each CI listener's reliance on F0 contour in both cooperating and conflicting conditions was also examined in relation to each listener's accuracy of intonation recognition with naturally-produced stimuli. A significant, positive correlation was observed between CI listeners' identification accuracy with naturally-produced stimuli and the estimated coefficient for F0 contour in the conflicting condition. This correlation was, on the other hand, was not found to be significant in the cooperating condition. Together, these results suggest that CI listeners' ability to utilize F0 contour in intonation recognition, particularly in the absence of other secondary cues such as intensity patterns (as in the conflicting condition), may be the most critical predictor for their intonation recognition performance with natural utterances. Nonetheless, in this condition, high levels of performance in speech recognition with naturally-produced stimuli were observed in a small proportion of CI listeners with a relatively small coefficient for F0 contour (e.g., CI-10). Together, the results in the conflicting condition can reasonably explain the general performance levels of CI listeners' intonation recognition with naturally-produced utterances.

Not all CI listeners' accuracy in intonation recognition with naturally-produced utterances can be well predicted by the extent of reliance on F0 contour in the conflicting condition. Evidently, intonation recognition with naturally-produced stimuli may depend on other factors, such as the amount of each cue and whether or not they are cooperating. These results in fact highlight the importance of assessing how individual CI listeners would utilize multiple sources of information in speech intonation recognition. In other words, these findings suggest that assessment of CI listeners' overall performance levels in intonation recognition using natural utterances would not permit any understanding of these listeners' processing of the degraded signals via a CI device in speech intonation recognition. It is important to recognize that CI listeners may achieve similar levels of performance, but the way these individuals achieve a certain performance level may be very different. This is consistent with the previous studies that suggested the importance of examining individual difference in listeners' integration of acoustic cues in speech perception [Hazan and Rosen, 1991].

In summary, intonation recognition with naturally-produced utterances can be achieved by CI listeners' combined utilization of F0 contour and intensity cues. Whether or not these cues being cooperating or conflicting becomes particularly important for CI listeners as these listeners are not capable of fully utilizing F0 contour as a cue in this recognition. The present findings have important implications for developing aural rehabilitation programs for CI recipients. Specifically, it is critical to take multiple acoustic sources for speech recognition into consideration when a (re)habilitation plan is tailored for these individuals. For example,

the clinician may work with CI listeners and their conversation partners, developing strategies to take advantage of cooperating cues, as well as to reduce any adverse effects when cues are present in conflict. Future studies should address how acoustic cues being cooperating or conflicting may affect CI listeners' recognition of phonemes and other speech prosodic contrasts (e.g., lexical tones), in quiet and in challenging listening conditions (e.g., in competing noise).

Acknowledgments

We thank all cochlear implant users and listeners with normal hearing for their participation. We also thank Dr. Qian-Jie Fu for sharing software, Shubhasish B. Kundu for additional software support, and Kelly Hoffard for her assistance with data collection. This research was funded by NIH/NIDCD grant no. R01DC04786 (PI: Dr. Monita Chatterjee).

References

Boersma P, Praat Weenink D. 2004(Computer Software, Version 4.3)

Chatterjee M, Peng S. Processing fundamental frequency contrasts with cochlear implants: psychophysics and speech intonation. Hear Res 2008;235:145–156.

Ciocca V, Francis AL, Aisha R, Wong L. The perception of Cantonese lexical tones by early-deafened cochlear implantees. J Acoust Soc Am 2002;111:2250–2256. [PubMed: 12051445]

Eilers R, Oller DK, Urbano R, Moroff D. Conflicting and cooperating cues: Perception of cues to final consonant voicing by infants and adults. J Speech Lang Hear Res 1989;32:307–316.

Faulkner A, Rosen S, Smith C. Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants. J Acoust Soc Am 2000;108:1877–1887. [PubMed: 11051514]

Fu QJ, Nogaki G. Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. J Assoc Res Otolaryngol 2004;6:19–27. [PubMed: 15735937]

Fu QJ, Shannon RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. J Acoust Soc Am 1999;105:1889–1990. [PubMed: 10089611]

Friesen LM, Shannon RV, Baskent D, Wang XS. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. J Acoust Soc Am 2001;110:1150–1163. [PubMed: 11519582]

Geurts L, Wouters J. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am 2001;109:713–726. [PubMed: 11248975]

Green T, Faulkner A, Rosen S. Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. J Acoust Soc Am 2002;112:2155–2164. [PubMed: 12430827]

Green T, Faulkner A, Rosen S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. J Acoust Soc Am 2004;116:2298–2310. [PubMed: 15532661]

Greenwood DD. A cochlear frequency-position function for several species – 29 years later. J Acoust Soc Am 1990;87:2592–2605. [PubMed: 2373794]

Hazan V, Rosen S. Individual variability in the perception of cues to place contrasts in initial stops. Percept Psychophys 1991;49:187–200. [PubMed: 2017355]

Lehiste, I. Suprasegmental features of speech. In: Lass, NJ., editor. Contemporary Issues in Experimental Phonetics. Academic Press; 1976. p. 225-239.

Morrongiello BA, Robson RC, Best CT, Clifton RK. Trading relations in the perception of speech by 5-year-old children. J Exp Child Psychol 1984;37:231–250. [PubMed: 6726113]

Peng S, Tomblin JB, Cheung H, Lin YS, Wang LS. Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. Ear Hear 2004;25:251–264. [PubMed: 15179116]

Peng S, Tomblin JB, Turner CW. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. Ear Hear 2008;29:336–351. [PubMed: 18344873]

Rosen S. Temporal information in speech: Acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 1992;336:367–373. [PubMed: 1354376]

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303–304. [PubMed: 7569981]

Whalen DH, Xu Y. Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica 1992;49:25–47. [PubMed: 1603839]

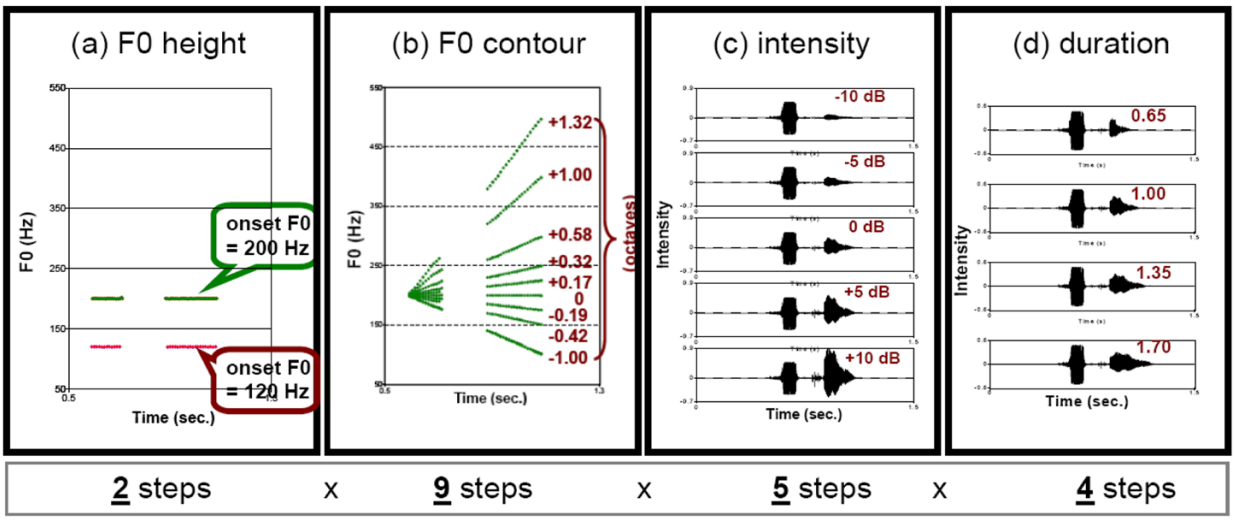

Figure 1.
Illustrations of acoustic dimensions that were systematically manipulated – Panels (a) through (d) display the steps specific to each of the acoustic dimensions, including F0 height, F0 contour, peak intensity ratio, and duration ratio. The number of steps for each dimension is indicated below each panel.

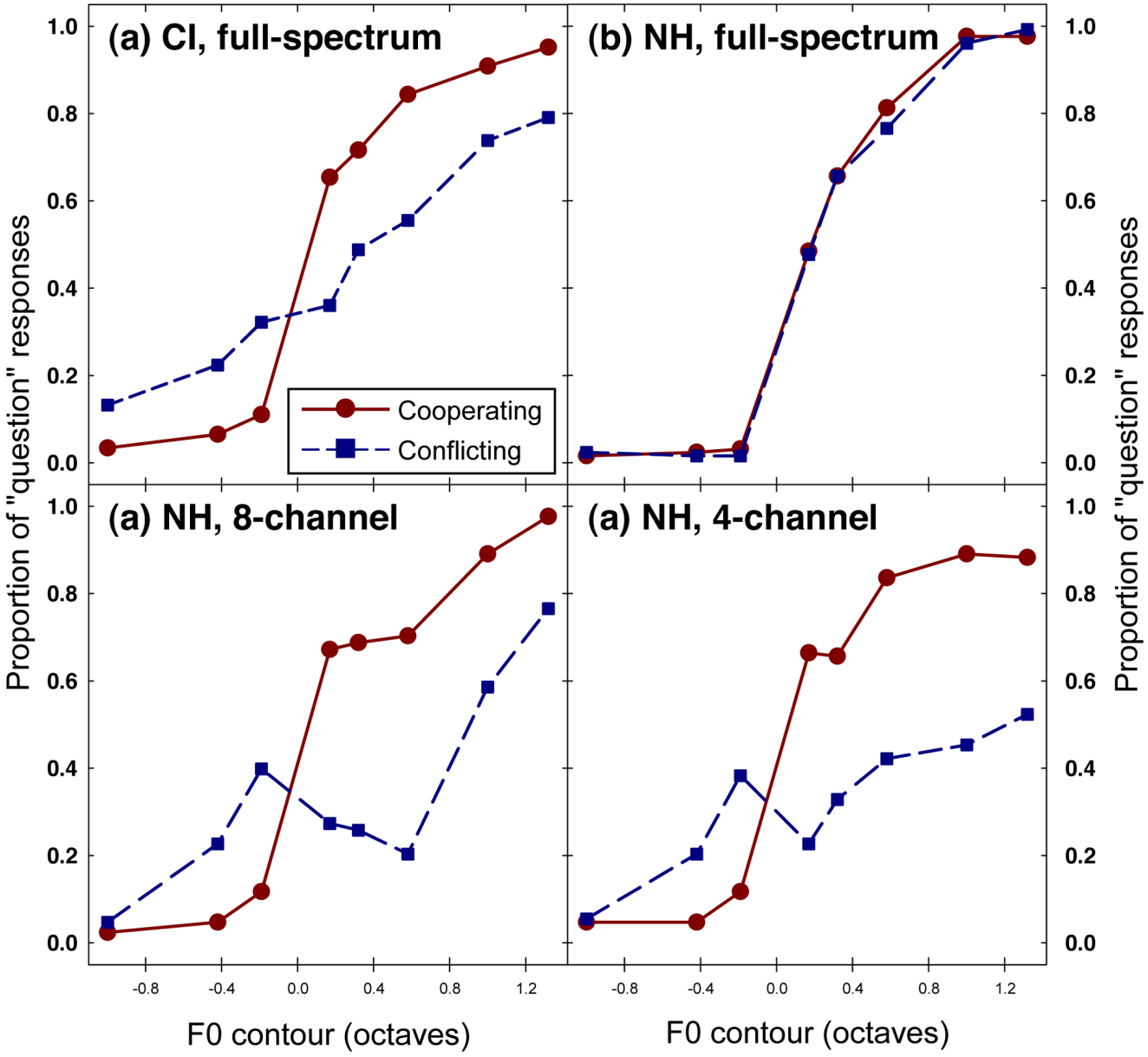

Figure 2.
Group mean proportions of question judgments as a function of the F0 increments in the cooperating vs. conflicting conditions. Panels (a) through (d) display the data for the CI group, the NH group under the full-spectrum condition, the NH group under the 8-channel condition, and the NH group under the 4-channel condition, respectively. The x-axis in each panel indicates the steps of F0 contour increments; the y-axis indicates the proportions of question judgments. Data points in the cooperating and conflicting conditions are linked with solid and dashed lines, respectively.

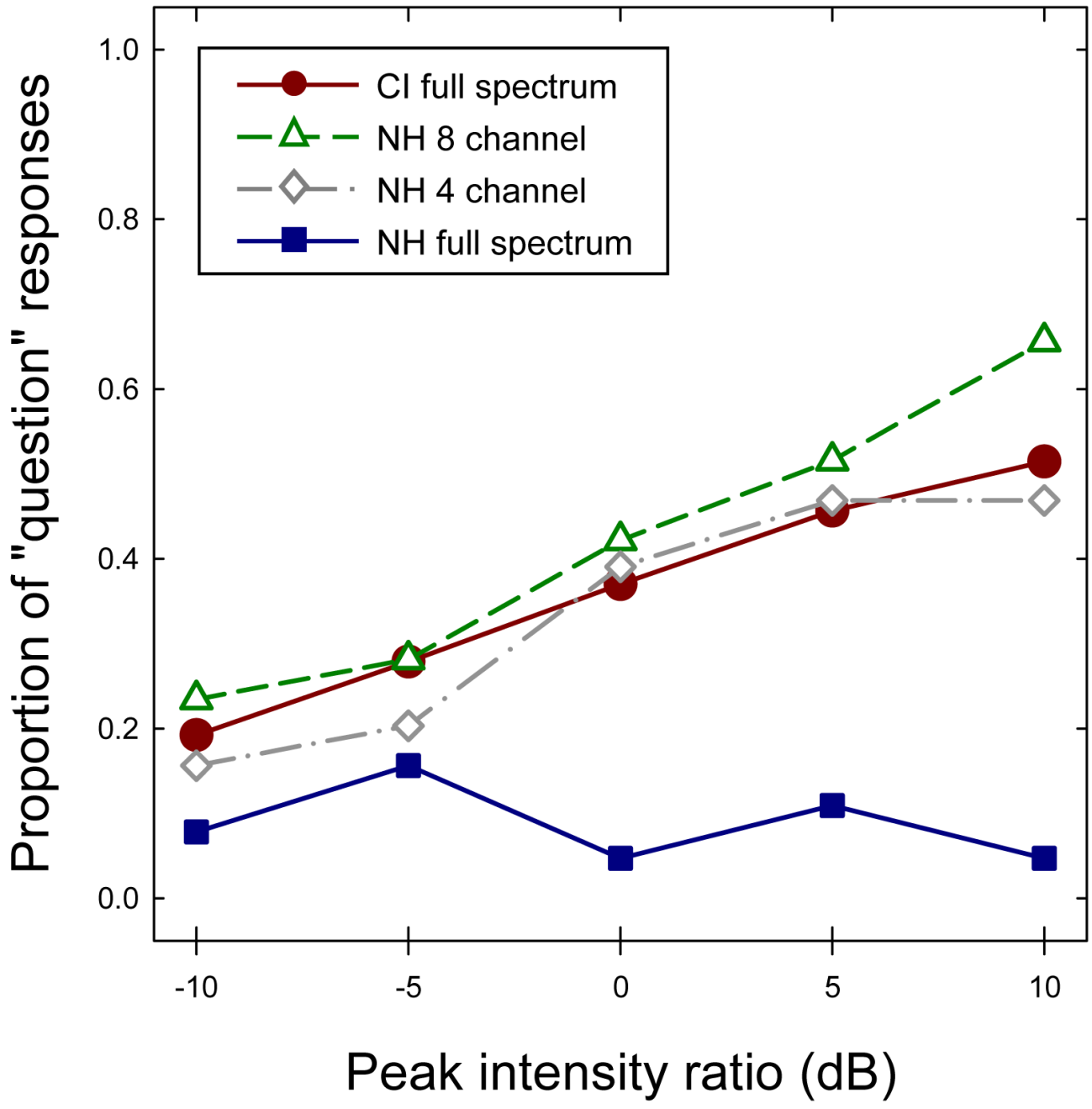

Figure 3.
Group mean proportions of question judgments as a function of the increment in peak intensity ratio, in the neutral condition (i.e., with a flat F0 contour). The x-axis indicates the steps of peak intensity ratio increments; the y-axis indicates the proportions of question judgments. Data points for the CI and NH groups, under the full-spectrum condition are displayed with circles and squares, respectively, linked with a solid line. Data points for the NH group under the 8- and 4-channel conditions are displayed with triangles and diamonds, respectively, linked with a dashed line.

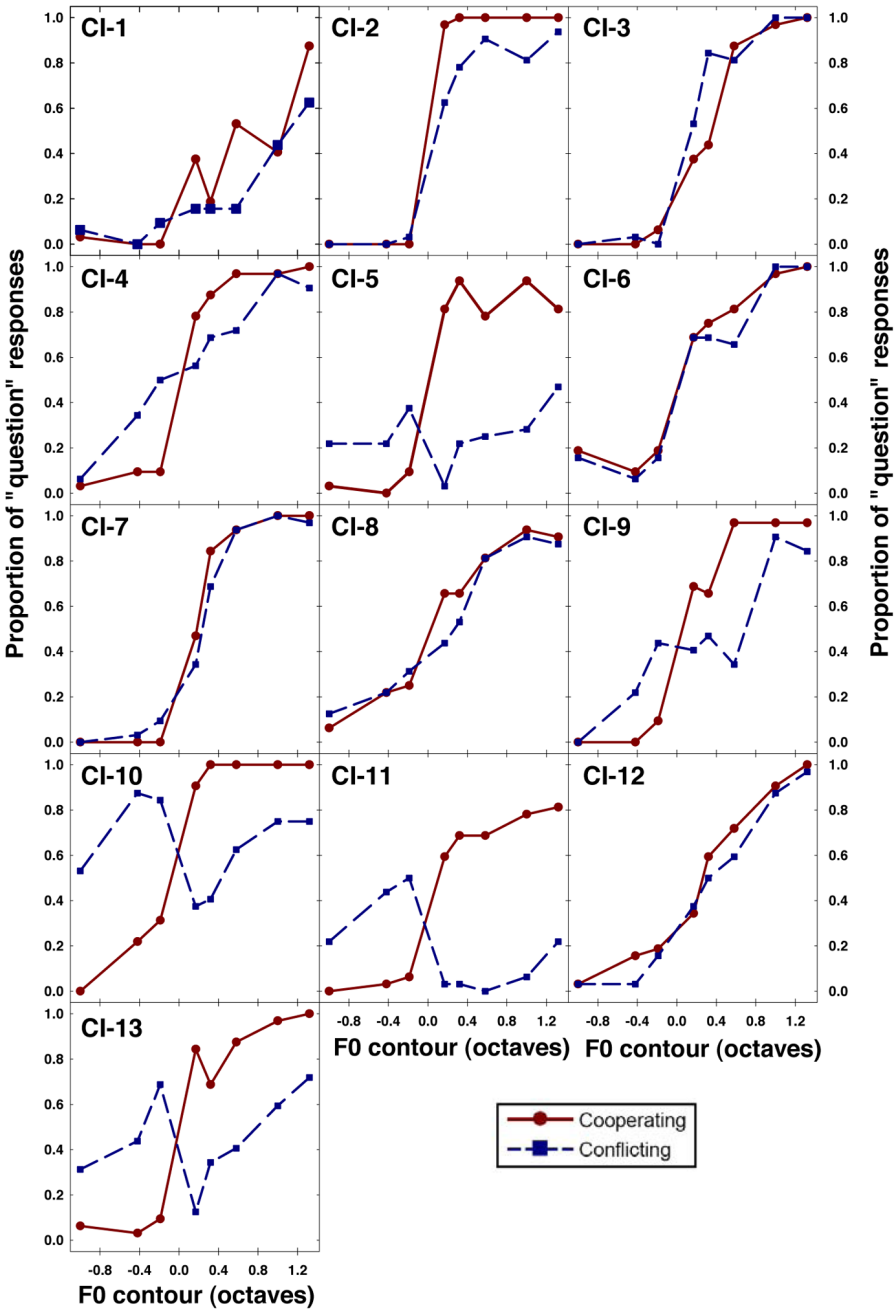

Figure 4.

Mean proportions of question judgments as a function of the F0 increments in the cooperating vs. conflicting conditions for individual CI listeners. Each panel displays the data obtained from each CI listener. In each panel, the x-axis indicates the steps of F0 contour increments; the y-axis indicates the proportions of question judgments. Data points in the cooperating and conflicting conditions are linked with solid and dashed lines, respectively.

Table 1

Background information of 13 CI subjects.

Subject	Etiology	Age at testing	Device experience (years)	Gender	Device	Processing strategy
CI-1	Unknown	60	3	Male	Nucleus 24	ACE[1]
CI-2	Unknown	52	7	Female	Clarion S	MPS[2]
CI-3	Genetic	63	3	Female	Nucleus 24	ACE
CI-4	Possibly genetic	56	9	Female	Nucleus 24	ACE
CI-5	Unknown	52	4	Female	Clarion CII	HiRes[3]
CI-6	Unknown	65	6	Male	Clarion S	MPS
CI-7	Possibly genetic	64	1	Female	Nucleus Freedom	ACE
CI-8	Unknown	83	5	Male	Nucleus 24	ACE
CI-9	Trauma	49	14	Male	Nucleus 22	SPEAK[4]
CI-10	Genetic	70	13	Female	Nucleus 22	SPEAK
CI-11	Unknown	65	14	Male	Nucleus 22	SPEAK
CI-12	German measles/Ototoxicity	64	1	Female	Nucleus Freedom	ACE
CI-13	Unknown	23	7	Male	Nucleus 24	ACE

[1] Advanced Combination Encoders,

[2] Monopolar Pulsatile Stimulation,

[3] HiResolution,

[4] Spectral Peak

Table 2

F0 contour and intensity characteristics in the cooperating, conflicting, and neutral conditions.

Parameter Condition	F0 contour (octaves)	Peak intensity ratio (dB)	Number of items in one run[1]
Cooperating (A1)	> 0 (5 steps)	> 0 (2 steps)	80
Cooperating (A2)	< 0 (3 steps)	< 0 (2 steps)	48
Conflicting (B1)	> 0 (5 steps)	< 0 (2 steps)	80
Conflicting (B2)	< 0 (3 steps)	> 0 (2 steps)	48
Neutral (N)	= 0 (1 step)	−10 to 10 (5 steps)	40

[1] Number of items was derived by multiplying the step numbers for F0 contour and for peak intensity ratio, by the fixed step numbers for the other two acoustic parameters that were roved (i.e., 2 steps for F0 height & 4 steps for duration ratio) [A1 & B1: 5 * 2 * 2 * 4 = 80; A2 & B2: 3 * 2 * 2 * 4 = 48; N: 1 * 5 * 2 * 4 = 40]. The stimulus presentations were performed in two runs, in each listening condition (CI: full-spectrum; NH: full-spectrum, 8- and 4-spectral channels). Thus, under each listening condition, the total number of tokens played to each listener was 160 [i.e., 2 * 80], 96 [i.e., 2 * 48)], and 80 [i.e., 2 * 40] in the cooperating, conflicting, and neutral condition, respectively. In all conditions, the two acoustic parameters, F0 height and duration ratio, were controlled in the logistic models that examine the estimated coefficient for F0 contour or peak intensity ratio in each of the cooperating, conflicting, and neutral conditions.

Table 3

Estimated coefficients for F0 contour in the logistic models fitted to individual data (for CI subjects) and group data (for both CI and NH subjects).

Group, listening condition (Subject ID for individual data)	(A) Cooperating (A1 & A2)	(B) Conflicting (B1 & B2)	(C) Difference [(A)−(B)]	Test statistic[1]	p-value for (C)	Intonation recognition accuracy[2]
			Group data			
CI, Full-spectrum	3.77	1.63	2.15	4.01	<.0001	87.01±14.52
NH, Full-spectrum	4.97	4.87	0.10	0.23	0.8165	97.70±2.29
NH, 8-channel	3.83	1.47	2.36	6.41	<0.0001	84.79±9.75
NH, 4-channel	3.13	0.96	2.16	5.24	<0.0001	80.63±10.77
			Individual data for CI subjects[3]			
CI-1	2.44	1.89	0.55	1.51	0.2197	78.33
CI-2	19.14	4.91	14.23	21.27	<0.0001	94.17
CI-3	5.88	6.52	−0.63	0.26	0.6115	98.33
CI-4	6.63	2.56	4.07	22.40	<0.0001	98.33
CI-5	3.16	0.38	2.78	40.04	<0.0001	63.33
CI-6	3.18	3.21	−0.03	0	0.9584	95.83
CI-7	10.45	5.96	4.49	4.94	0.0262	96.67
CI-8	2.58	2.09	0.49	1.37	0.2417	65.83
CI-9	7.52	2.70	4.82	27.00	<0.0001	95.83
CI-10	8.12	0.12	8.00	48.62	<0.0001	93.33
CI-11	2.98	−0.88	3.87	75.07	<0.0001	58.00
CI-12	4.06	4.09	−0.03	0	0.9584	85.00
CI-13	4.96	0.57	4.39	47.48	<0.0001	75.00

[1] Z statistics for group data; χ^2 for individual data

[2] Intonation accuracy was measured using 120 naturally-produced utterances, i.e., 10 statements and 10 questions recorded from multiple adult speakers who are native speakers of American English. These utterances were paired so that each pair contained syntactically matched questions and statements (e.g., *"The girl is on the playground."* vs. *"The girl is on the playground?"*). The task was performed in a single-interval, 2-alternative forced-choice paradigm. Utterances were randomly selected by a custom computer program and the presentation level was fixed at 65 dB SPL (A- weighting). The program recorded the subject's responses and the percent-correct score was derived for each subject.

[3] Only group data are shown for acoustic CI simulations, as estimated coefficients were quite consistent across NH listeners.