
Imitative Production of Rising Speech Intonation in Pediatric Cochlear Implant Recipients

Shu-Chen Peng
J. Bruce Tomblin
Linda J. Spencer
Richard R. Hurtig
University of Iowa,
Iowa City

Purpose: This study investigated the acoustic characteristics of pediatric cochlear implant (CI) recipients' imitative production of rising speech intonation, in relation to the perceptual judgments by listeners with normal hearing (NH).

Method: Recordings of a yes-no interrogative utterance imitated by 24 prelingually deafened children with a CI were extracted from annual evaluation sessions. These utterances were perceptually judged by adult NH listeners in regard with intonation contour type (non-rise, partial-rise, or full-rise) and contour appropriateness (on a 5-point scale). Fundamental frequency, intensity, and duration properties of each utterance were also acoustically analyzed.

Results: Adult NH listeners' judgments of intonation contour type and contour appropriateness for each CI participant's utterances were highly positively correlated. The pediatric CI recipients did not consistently use appropriate intonation contours when imitating a yes-no question. Acoustic properties of speech intonation produced by these individuals were discernible among utterances of different intonation contour types according to NH listeners' perceptual judgments.

Conclusions: These findings delineated the perceptual and acoustic characteristics of speech intonation imitated by prelingually deafened children and young adults with a CI. Future studies should address whether the degraded signals these individuals perceive via a CI contribute to their difficulties with speech intonation production.

KEY WORDS: cochlear implants, speech intonation, speech development, prosody, acoustic analysis

A *cochlear implant (CI)* is an auditory prosthetic device that is surgically implanted in the inner ear and stimulates primary auditory nerve fibers to elicit sound sensation in individuals with a severe-profound sensorineural hearing loss. These devices are fairly successful in facilitating the spoken language development in prelingually deafened children. However, current CI devices are limited in encoding fundamental frequency (F₀), that is, voice pitch information (Faulkner, Rosen, & Smith, 2000; Geurts & Wouters, 2001; Green, Faulkner, & Rosen, 2004). Such voice pitch variation is critical for the recognition of prosodic components of speech that mark linguistic contrasts, such as lexical tones, stress, and speech intonation (Ladd, 1996; Lehiste, 1970, 1976). Because current CI devices provide only restricted access for the recognition of prosodic components of speech that signify linguistic contrasts, these devices can be limited in facilitating the acquisition of the prosodic components, that is, lexical tones and speech intonation in prelingually deafened children who must rely on a CI to develop spoken language.

Prosodic components of speech are referred to as the perceptual and acoustic realizations at the suprasegmental level of speech (Lehiste, 1970). Variation in prosodic components of speech can have various expressive functions in semantic, attitudinal, psychological, and social domains (Crystal, 1979). The most noticeable importance of prosodic aspects of speech is perhaps its linguistic functions, such as lexical tones, contrastive stress, and speech intonation. In tonal languages, variation in lexical tones conveys meanings at the syllable level. For example, in Mandarin Chinese, 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 a nontonal language, such as English, variation in prosodic components of speech may also lead to changes of linguistic meanings at word, phrase, and sentence levels. For example, a word can be contrasted in meaning with different stress patterns on its two syllables (e.g., *sub'ject* vs. *'subject*).

Recognition of speech intonation is associated with the acoustic parameters including F₀, intensity, and duration patterns (Denes, 1959; Denes & Milton-Williams, 1962; Hadding-Koch & Studdert-Kennedy, 1964; Studdert-Kennedy & Hadding, 1973). These acoustic correlates of intonation, when perceived by listeners, can be denoted as voice pitch, loudness, and length, respectively. Fundamental frequency variation is the principal acoustic correlate of the perceived changes in voice pitch. Variation in F₀ contours can lead to changes in prosodic patterns at various levels of linguistic units (e.g., word, phrase, sentence, and discourse). At the sentence level, for example, an interrogative utterance can be distinguished from its declarative form by varying its F₀ patterns. A statement typically has a falling F₀ contour at the terminal position of an utterance, whereas F₀ contour in a yes–no interrogative utterance usually rises at the end (Ladefoged, 2001). Similarly, syllables in finished and unfinished sentences can also have different F₀ variation patterns. The final syllables in unfinished sentences tend to have higher F₀ peaks, smaller F₀ falling slopes, and higher valleys in the fluctuation contour than their finished counterparts (Berkovits, 1984).

Although F₀ contour provides the listener with the most prevailing information for the recognition of speech intonation contrasts that mark utterances to be declarative or interrogative (Cooper & Sorensen, 1981; Ladd, 1996; Lehiste, 1970, 1976), changes in F₀ contours often take place in conjunction with the variation in intensity and duration patterns (Freeman, 1982). Previous studies have also suggested that the acoustic properties of F₀, intensity, and duration patterns can all contribute to the perception of speech intonation contrasts in listeners with normal hearing (NH; Fry, 1955, 1958; Lehiste, 1970, 1976; Lieberman, 1967).

Infants and young children show contrastive use of intonation and other prosodic components of speech in their vocalization or utterances at a very young age (e.g., D'Odorico & Franco, 1991; Furrow, 1984; Galligan, 1987). However, although children with NH develop control over some core features of intonation early in life, mastery of certain prosodic components of speech is associated with an increasing age (e.g., Loeb & Allen, 1993; Snow & Balog, 2002). In other words, development of speech intonation and other prosodic components requires learning or exposure to linguistic inputs. Note that certain prosodic features may require a longer period of time for young language learners to master than others. In particular, rising intonation (as opposed to falling) requires physiological effort and relies upon linguistic experience and learning (Boothroyd, 1982; Lieberman, 1967; Snow, 1998; Vihman, 1996). Examples can be seen in NH children's acquisition of speech intonation (e.g., Snow, 1998) and lexical tones (e.g., Li & Thompson, 1977).

Listeners with NH have full access to temporal envelope, periodicity, and spectral cues from resolved harmonics at low frequencies that are critical for speech intonation recognition (Fu, Zeng, Shannon, & Soli, 1998; Moore, 1997; Rosen, 1989, 1992). On the other hand, because of the limited number of spectral channels, CI recipients are only able to access the weak voice pitch cues and unresolved harmonic structures of speech signals (Ciocca, Francis, Aisha, & Wong, 2002; Faulkner et al., 2000; Geurts & Wouters, 2001; Green, Faulkner, & Rosen, 2002; Green et al., 2004). As a result, these individuals' ability to recognize speech intonation contrasts is likely hindered.

Acquisition of rising intonation relies upon linguistic experience and speech inputs that are not accessible to prelingually deafened children prior to implantation. Because of the device limitation in presenting voice pitch information, acquisition of prosodic aspects of speech is potentially challenging for pediatric CI recipients who speak English (e.g., Green et al., 2004; O'Halpin, 2001). However, there is only limited empirical evidence that supports this assumption. Findings in the literature indicated that prelingually deafened children, with 2 years of CI experience, do not generally show skilled perception and production of intonation and other prosodic components of speech (Osberger, Miyamoto, et al., 1991; Osberger, Robbins, et al., 1991; Tobey et al., 1991; Tobey & Hasenstab, 1991). These earlier studies, however, addressed the performance of pediatric CI recipients during the initial 2 years following implantation. Many of the CI recipients in those studies were mapped with relatively old speech-coding strategies, such as F₀/F₂ or F₀/F₁/F₂ (as opposed to MPEAK or SPEAK; see Method section). The extent to which CI devices with relatively recent speech-coding strategies (MPEAK and SPEAK, as opposed to F₀/F₂ or F₀/F₁/F₂) can facilitate the acquisition of speech

intonation in prelingually deafened children with extended device experience remained unclear.

Unlike the limited number of studies on the acquisition of speech intonation in pediatric CI recipients, the perception and production of lexical tones in prelingually deafened children with CIs who speak a tonal language, such as Mandarin or Cantonese, have been evaluated in several studies (Barry, Blamey, & Martin, 2002; Barry, Blamey, Martin, Lee, et al., 2002; Ciocca et al., 2002; Lee, van Hasselt, Chiu, & Cheung, 2002; Peng, Tomblin, Cheung, Lin, & Wang, 2004; Wei et al., 2000; Xu et al., 2004). Findings of these studies unambiguously suggested that prelingually deafened children with CIs exhibit great deficiencies in perceiving and producing lexical tone contrasts. Moreover, among individual lexical tones, rising tones (e.g., Mandarin Tone 2) have been reported to be more difficult than falling tones (e.g., Mandarin Tone 4) for CI children to produce accurately (Peng, Tomblin, et al., 2004; Xu et al., 2004). These findings are particularly relevant to the present study, as the target utterances to be evaluated involved a rising intonation contour. In summary, the acoustic properties that contribute to the recognition of lexical tone contrasts (at least in Mandarin) comprised F0, intensity, and duration patterns, with F0 as the primary cue (Whalen & Xu, 1992). These acoustic properties are similar to those critical for speech intonation recognition. Hence, it is reasonable to hypothesize that the acquisition of speech intonation can be challenging for prelingually deafened children with a CI.

The purpose of this study was to investigate the extent to which CI devices can facilitate prelingually deafened children's acquisition of rising intonation production. Specifically, the acoustic characteristics of pediatric CI recipients' imitative production of rising intonation were evaluated in relation to the perceptual judgments of adult NH listeners. Using a retrospective, longitudinal examination, utterances produced by a group of 24 pediatric CI recipients were perceptually judged in regard with intonation contour type (non-rise, partial-rise, or full-rise) and contour appropriateness (on a 5-point scale). The F0, intensity, and duration properties of each utterance were also acoustically analyzed. It was anticipated that given the limited voice pitch information provided by CI devices, pediatric CI recipients would not produce rising intonation appropriately. Moreover, it was expected that the acoustic properties pertaining to speech intonation would be associated with the intonation contour appropriateness of pediatric CI recipients' utterances.

Method

Participants

Twenty-four prelingually deafened children and young adults, who participated in Peng, Spencer, and

Tomblin's (2004) study, served as participants in the present study. They all received a CI and attended follow-up assessments in the Department of Otolaryngology—Head and Neck Surgery at the University of Iowa Hospital and Clinics. All participants received the Nucleus 22 device (Cochlear, Lane Cove, Australia). Eighteen CI recipients had been mapped with the spectral-peak (SPEAK) speech-coding strategy, and 6 had been mapped with multipeak (MPEAK) strategy as of the 7th year post-implantation. Table 1 provides a summary of each participant's background information. Participant CI-6 received his education in a mainstream, public school setting where only spoken English was used (oral communication; OC). The remaining 23 participants received their education in a mainstream public school setting where both signing exact English and spoken English were used (total communication; TC). Classification of communication methods (OC or TC) was based on parental reports, confirmed by the educational setting of the participant at the 7th year postimplantation. Although the majority of participants in the present study received their instructions in a TC program, all of these individuals had significant exposure to spoken language at school and at home following implantation.

Ten adult native English speakers (9 women and 1 man) between the ages of 22 and 44 years ($M = 28.9$ years) were recruited to perceptually judge the utterances produced by the 24 CI users. None of these listeners reported a history of hearing or speech disorders. Hearing screening prior to perceptual judgments revealed that all listeners' hearing sensitivity was within normal limits (thresholds better than 20 dB HL) at all octave intervals from 250–8000 Hz in both ears. All listeners gave written, informed consent approved by the University of Iowa Institutional Review Board before the task began and were paid for participation.

Preparation of Speech Stimuli

The target utterance, "Are you ready?" was imitatively produced by CI participants following an examiner-modeled utterance via spoken English and signing exact English simultaneously during each of their preimplant and annual follow-up sessions. These utterances were modeled by an experienced female speech-language pathologist (CCC-SPL holder) who evaluated the speech and language development of the present CI participants. The examiner-modeled utterances were always superimposed with a rising intonation contour. The target utterance, "Are you ready?" is one of the 14 short-version sentences of the Short-Long Sentence Test, that is, one subtest of the battery designed to assess pediatric CI recipients' spoken language development. A detailed description of the Short-Long Sentence Test can be found

Table 1. Background information of the 24 CI participants.

ID	Gender	Age at implantation (years)	Speech-coding strategy	Pre-op PTA in better ear (dB HL)	Total utterances available
CI-1	male	2.58	SPEAK	98.3	6
CI-2	male	2.74	SPEAK	112.5†	8
CI-3	male	2.88	SPEAK	90†	8
CI-4	male	3.52	SPEAK	NR	7
CI-5	male	3.57	SPEAK	110	5
CI-6	female	3.74	SPEAK	NR	5
CI-7	male	3.82	SPEAK	110†	10
CI-8	male	3.91	MPEAK	90†	7
CI-9	female	4.24	SPEAK	100†	6
CI-10	female	4.38	SPEAK	NR	10
CI-11	male	4.53	SPEAK	102.5†	9
CI-12	male	4.74	SPEAK	100†	4
CI-13	female	4.84	SPEAK	103.3	9
CI-14	female	5.03	MPEAK	110†	6
CI-15	female	5.16	SPEAK	115†	7
CI-16	female	5.42	MPEAK	113.3	7
CI-17	male	5.55	SPEAK	100†	7
CI-18	male	5.55	MPEAK	95†	7
CI-19	male	5.58	SPEAK	110†	6
CI-20	female	5.75	SPEAK	105†	9
CI-21	female	6.71	MPEAK	92.5†	6
CI-22	female	7.44	MPEAK	100†	7
CI-23	female	9.90	SPEAK	103.3	6
CI-24	male	11.04	SPEAK	97.5†	8

Note. Pre-op PTA = pre-operative pure-tone average; CI = cochlear implant; SPEAK = spectral-peak speech-coding strategy; MPEAK = multipeak speech-coding strategy; NR = no response at audiometer output limits (110 dB HL at 500 Hz, 115 dB HL at 1000 Hz, and 115 dB HL at 2000 Hz).

elsewhere (e.g., Tye-Murray, 1998; Tye-Murray, Spencer, & Woodworth, 1995).

The resulting set comprised a total of 170 utterances produced by the 24 participants. The number of utterances contributed by each participant ranged from 4 to 10 throughout the preimplant and annual follow-up sessions. Table 1 provides a summary of the number of total utterances available from each participant across his or her annual sessions. The exact number of utterances at each test interval is displayed in Figures 1 and 2. Note that because each child contributed one utterance at any test interval, the number of utterances was identical to the number of CI children who had an utterance at each interval.

The audio recordings of these utterances were extracted from videotapes, digitally edited at a sampling rate of 44100 Hz, and stored in a 16-bit format using the sound-analysis software CoolEdit 2000 (Syntrillium Software, Scottsdale, AZ). Each utterance was normalized for long-term rms amplitude to maintain relatively constant sound levels across all utterances. A computer program was developed to present the utterances to each NH listener in random order and to automatically record the

listener's responses. A set of 10 additional utterances (same as the target "Are you ready?") produced by CI children who were not included in this study were adopted as examples. The program and the sound files of all target utterances and examples were loaded onto a laptop (Sony Vaio PCG-R505EL) for perceptual judgments described below.

Perceptual Judgments

The 170 utterances, along with the 14 randomly selected utterances modeled by the examiner were presented to each NH listener binaurally through headphones (Sennheiser HD 25 SP) at a comfortable listening level (approximately 65 dB SPL, A-weighting) in a double-walled IAC sound treated room. Prior to the task, the listener was familiarized with the task using the 10 examples. The listener made the perceptual judgments for each utterance in terms of (a) intonation contour type (i.e., "Which type of intonation contour does the utterance have—non-rise, partial-rise, or full-rise?"), and (b) contour appropriateness (i.e., "How appropriate is the intonation contour of the target utterance?"), judged on a 5-point rating scale ranging from 1 (*completely inappropriate*) to 5 (*absolutely appropriate*).

Figure 1. Distributions of each intonation contour type at each test interval. The total number of utterances at each time interval and the examiner's utterances are shown on the top of each bar. The crosshatched portion refers to the intonation contour type of partial-rise; the solid portion refers to the intonation contour type of full-rise.

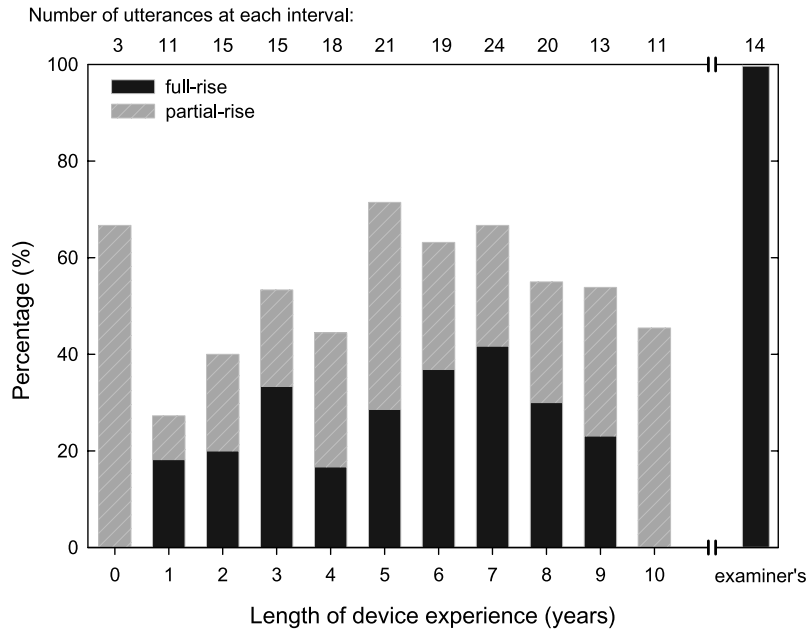
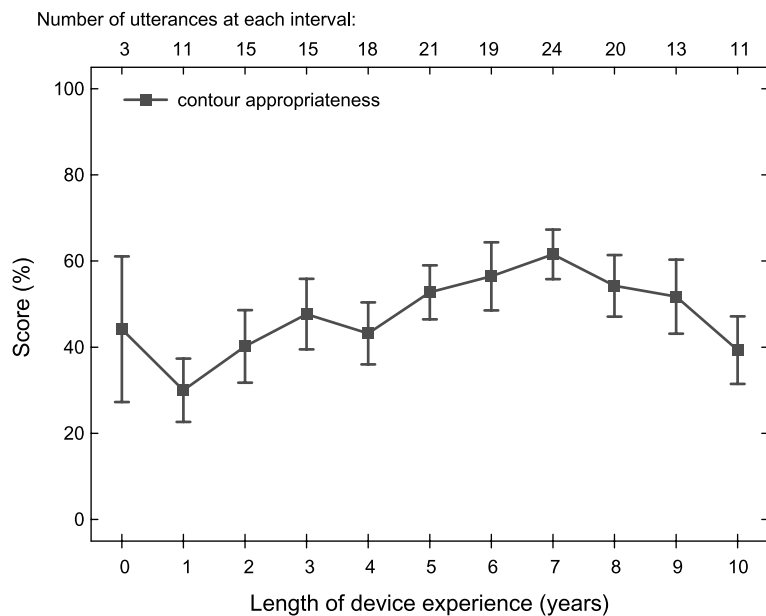


Figure 2. Mean scores of contour appropriateness as a function of number of postimplant years. Error bars display ± 1 SE of the mean score.



Acoustic Analyses

Measurements of the acoustic correlates of speech intonation were performed using the Praat software program for Windows (Version 4.3; Boersma & Weenink, 2004). The Praat Sound Edit Window provides the visualization of the F0 contour, intensity contour, and duration properties of an utterance. The acoustic correlates of speech intonation—that is, F0, intensity, and duration patterns of each utterance—were examined. The Appendix provides a summary of the F0-related, duration, and intensity parameters examined in this study, as well as a description of the abbreviated forms of each parameter.

Fundamental frequency (F0). The auto-correlation pitch extraction algorithm was adopted to analyze the absolute F0 values (F0_valley_utterance, F0_peak_utterance, F0_onset_final, and F0_offset_final) and average F0 values for different syllables or words (F0_mean_nonfinal1, F0_mean_nonfinal2, F0_mean_final1, and F0_mean_final2). Each utterance was carefully monitored to obtain perceptually valid voice pitch contours. F0-related settings, such as upper and lower limits, might have been adjusted when gross tracking errors (e.g., pitch halving and doubling) occurred. If changed, the settings were recorded.

The F0-related parameters comprised (a) maximal and minimal F0 values, which were recorded at the peak and valley points at the utterance level (F0_peak_utterance and F0_valley_utterance), and (b) onset and offset of the utterance-final word *ready* (F0_onset_final and F0_offset_final). The average F0 values were measured for the vocalic nucleus at each syllable or word position (F0_mean_nonfinal1 and F0_mean_nonfinal2 for the non-utterance-final words *are* and *you*; and F0_mean_final1 and F0_mean_final2 for the two syllables of the utterance-final word *ready*). These F0-related values were calculated, resulting in the set of the following parameters: (a) peak-to-valley F0 difference at the utterance level (F0_peak_utterance – F0_valley_utterance), (b) amount of F0 change from the onset to the offset at utterance-final words (F0_offset_final – F0_onset_final), and (c) rate of F0 change at utterance-final words [(F0_offset_final – F0_onset_final)/DUR_final].

The F0-related values were converted into voice pitch on a logarithmic, semitone scale (measured in cents; 1 semitone = 100 cents) that corresponded more closely to perceived pitch (Burns & Ward, 1982). This conversion also helped account for the substantial intra- and inter-subject F0 variability and permitted comparisons of the quantitative differences in the F0-related values (Allen & Arndorfer, 2000; Burns & Ward, 1982). These voice-pitch-related parameters included the following: (a) voice pitch range for the entire utterance (PITCH_range_utterance), (b) amount of voice pitch change

from the onset to the offset at utterance-final words (Δ PITCH_final), and (c) rate of voice pitch change at utterance-final words (Δ PITCH_rate_final).

Intensity. Peak intensity values were identified at each of the four syllable nuclei along the intensity contour of utterances displayed on the Praat Sound Edit Window. These parameters were referred to as INT_peak_nonfinal1, INT_peak_nonfinal2, INT_peak_final1, and INT_peak_final2. Ratios (in dB) between each of the peak intensity values in each utterance were calculated using INT_peak_nonfinal2 as the reference. These ratios were denoted as follows: INT_ratio_nonfinal1/nonfinal2, INT_ratio_nonfinal2/nonfinal2, INT_ratio_final1/nonfinal2, and INT_ratio_final2/nonfinal2. The second non-utterance-final word (i.e., *you*) was chosen as the reference because this word is a pronoun and a closed-class word. It tends to be unstressed and receives little emphasis in its unmarked form (i.e., when not being contrasted) in English (Quirk, Greenbaum, Leech, & Svartvik, 1985).

Duration. Supplemented by the time waveform and the auditory playback of each utterance, the values for duration patterns were obtained primarily with a wide-band spectrographic display (200 Hz). Cursors were placed at the onset and offset of non-utterance-final words *are you* and the utterance-final word *ready*. The duration of the non-utterance-final words, utterance-final word, and entire utterance was denoted as DUR_nonfinal, DUR_final, and DUR_utterance, respectively. Note that speaking rates varied among the participants and among the utterances produced by the same child. Moreover, a small portion of utterances contained pauses between non-utterance-final and utterance-final words. Thus, the duration patterns examined in this study were primarily the duration ratio of non-utterance-final words to the entire utterance (DUR_ratio_nonfinal/utterance) versus the duration ratio of utterance-final words to the entire utterance (DUR_ratio_final/utterance).

Intra- and Interjudge Reliability

The acoustic findings reported in this study were based on the results derived from the analyses performed by the first author (primary measures). Intra- and interjudge reliability measures of acoustic results were performed with 10% of the utterances ($n = 17$) randomly sampled among all utterances. Note that the reliability measures were limited to only 10% of utterances ($n = 17$) because of the time-consuming nature of the large set of acoustic parameters examined for each utterance. In addition, the means and standard deviations for the acoustic parameters of interest derived from this small sample were comparable with those derived from the primary measures, as can be observed by the minimal differences derived from the inter- and intrajudge

reliability measures (see Table 2). Both the intra- and interjudge reliability measures were performed within 2 months following the completion of the primary measures. For intrajudge reliability measures, comparisons were made between the results of the first author's primary measures and the results derived from the reanalysis of the sampled utterances. The interjudge reliability measures were conducted by having a second trained examiner acoustically analyze the same subset of utterances. This examiner was an undergraduate student (senior standing), was majoring in speech and hearing sciences, and had taken a course in phonetics at the time of performing acoustic measurements. This examiner performed the acoustic analysis of the same subset of utterances independently without access to the participant's background information. These results were then compared with the results derived from the primary measures.

Table 2 provides a summary of the absolute differences between results derived from the first author's primary measures and reanalyses (intrajudge), as well as the absolute differences between the results derived from the first author's primary measures and the second examiner's analyses (interjudge) for the acoustic parameters examined in this study. To illustrate that the values of each acoustic parameter derived from this subset of utterances were comparable with those derived from the primary measures, Pearson correlation coefficients (*r*s)

for each of the intra- and interjudge comparisons are also indicated in Table 2.

Results

This section provides a summary of the findings from both adult listeners' perceptual judgments and acoustic analyses of the utterances produced by the CI individuals. Perceptual judgment results are reported on the basis of (a) the group of adult NH listeners' judgments of intonation contour type and contour appropriateness of utterances from the 24 pediatric CI recipients, and (b) the scores of contour type and contour appropriateness as a function of pediatric CI recipients' length of device experience. The acoustic properties of pediatric CI recipients' utterances are described on the basis of the group of NH listeners' perceptual judgment results. General linear mixed models were extensively adopted to statistically analyze the data in this section, and the analyses were performed using PROC MIXED of SAS.

Perceptual Judgments

First of all, the analyses were conducted to examine whether pediatric CI recipients demonstrated any evidence of superimposing rising intonation contours in their utterances, regardless of these individuals' length of device experience when the utterances were produced.

Table 2. Differences derived from the inter- and intrajudge reliability measures.

Reliability	Intrajudge			Interjudge		
	<i>M</i>	<i>SD</i>	<i>r</i>	<i>M</i>	<i>SD</i>	<i>r</i>
Acoustic parameter						
F0 (Hz)						
FO_valley_utterance	1.01	1.41	.889	1.00	1.41	.890
FO_peak_utterance	1.85	2.61	1.000	1.84	2.61	1.000
FO_onset_final	2.06	2.31	.998	3.78	4.04	.994
FO_offset_final	1.40	1.83	1.000	1.81	2.25	.956
FO_mean_nonfinal1	0.99	0.79	1.000	1.25	1.59	.999
FO_mean_nonfinal2	1.25	1.12	.999	1.16	1.26	.999
FO_mean_final1	2.26	2.05	.998	2.68	2.42	.998
FO_mean_final2	2.03	2.36	.998	2.10	1.72	.999
Duration (ms)						
DUR_nonfinal	16.51	21.88	.990	26.47	25.83	.976
DUR_final	28.09	25.03	.993	31.81	26.20	.991
DUR_utterance	14.39	11.91	1.000	12.36	12.86	.999
Intensity (dB)						
INT_peak_nonfinal1	0.57	0.57	.994	0.46	0.40	.998
INT_peak_nonfinal2	0.24	0.31	.998	0.42	0.69	.991
INT_peak_final1	0.31	0.22	.999	0.28	0.20	.999
INT_peak_final2	0.30	0.29	.999	0.55	0.57	.996

Note. All of the *p* values for the Pearson correlation coefficients (*r* values) are less than .001. F0 = fundamental frequency; DUR = duration; INT = intensity.

To provide an approximation of how adequate the intonation contour of an utterance was likely perceived by average NH listeners, the intonation contour type of each utterance was determined in accordance with the majority of the 10 NH listeners' responses. However, if there was equality in the numbers, then the judgment of full-rise would override that of partial-rise and non-rise, and the judgment of partial-rise would override that of non-rise. By doing so, 45.29% of the utterances ($n = 77$) were classified as non-rise, 28.24% ($n = 48$) as partial-rise, and 26.47% ($n = 45$) as full-rise. Whereas the majority of CI participants demonstrated use of rising intonation contours (partial-rise or full-rise) for at least one of their utterances, 41.67% of these participants (10 of 24) never produced any utterance that was judged to be full-rise.

To evaluate the relationship between NH listeners' judgments of intonation contour type and contour appropriateness, Pearson correlation coefficients (r_s) between the scores of intonation contour type and contour appropriateness were computed for the utterances produced by each CI participant. To do this, a point of "0," "1," and "2" was assigned to the utterance type of non-rise, partial-rise, and full-rise, respectively. Each coefficient was computed on the basis of the 10 adult listeners' average scores of intonation contour type and contour appropriateness for all utterances produced by each participant. The coefficients ranged from .891 to .998 (mean of r values = .959; all p s < .043), suggesting that the NH listeners' judgments of intonation contour type were strongly positively correlated with their judgments of contour appropriateness.

The other set of analyses was performed to determine whether the adequate use of a rising intonation contour by pediatric CI recipients improved with increasing device experience. Figure 1 illustrates the distributions of the CI utterances of which the contour types had a rising component (i.e., partial-rise and full-rise) as a function of length of device experience (in years). The distribution of the contour types of the examiner-modeled utterances ($n = 14$) is also displayed as a reference. All examiner-modeled utterances were classified to be full-rise. However, the distributions of the three contour types for CI utterances at all annual follow-up sessions were very different from examiner-modeled utterances. Even with several years of device experience, there was always a portion of CI participants who did not show signs of using any rising intonation contour. At the 7th year postimplantation, for example, 33.33% of the utterances produced by CI participants were judged to be non-rise.

As illustrated by the filled and crosshatched portions of each bar in Figure 1, although the percentage of partial-rise and full-rise utterances increased until the

7th year postimplantation, it did not increase after that. In fact, this percentage declined following the 7th year postimplantation. Recall that at all annual sessions, there was always a portion of utterances judged by the NH listeners to be non-rise ($M = 46.50\%$, $SD = 12.73\%$). Nevertheless, it is important to note that not all CI participants' utterances were available at every annual session. The fact that not all CI participants contributed to the later observation intervals might be responsible for this declining trend after the 7th year. For example, one might suggest that CI participants with longer device experience were systematically poorer users. This concern was taken into account by performing the statistical analyses as described below.

Figure 2 illustrates the average scores (both in percentage) of contour appropriateness with increasing device experience. The average score of contour appropriateness was the highest at the 7th year postimplantation ($M = 61.56\%$, $SD = 28.18\%$). Note that the scores of intonation contour type and contour appropriateness were highly correlated in all CI participants. Both judgments were seemingly indicative of each CI participant's adequate use of a rising intonation contour with the target utterance. The scores of intonation contour appropriateness, instead of the percentage of intonation contour type, were statistically analyzed as follows.

Data from a subgroup of CI participants ($n = 21$) were used to estimate the improving (or declining) rates for the scores of intonation contour appropriateness using the 7th year postimplantation as the cutoff point. The scores of 3 participants (CI-6, CI-19, and CI-24) were excluded from the analyses because they did not have utterances available after the 7th year, yet a minimum of two data points would be required for estimating improving rates (i.e., slope estimation).

A piecewise regression model, knotted at the 7th year, was fitted to the subgroup of CI participant's scores of intonation contour appropriateness. Two estimated slopes were hence derived for each participant. One estimated the improving rate (percentage points per year) from the preimplant session to the 7th year postimplantation, and the other estimated the improving rate after the 7th year. Signed rank tests were performed to evaluate whether the mean of the improving rate over each time period was different from zero. On average, improvements in the contour appropriateness scores were observed from the preimplant session to the 7th year postimplantation ($M = 5.29\%$ per year, $SD = 6.01\%$ per year; signed rank test statistic = 89.50, $p < .001$). However, a decrease in the average rate of contour appropriateness scores was observed from the 7th to the 10th year. This drop was found to be significantly different from zero ($M = -11.79\%$ per year, $SD = 17.58\%$ per year; signed rank test statistic = -71.50, $p = .009$).

These perceptual judgment results indicated that many utterances imitated by CI participants did not exhibit an appropriate rising intonation contour. These individuals' average scores of intonation contour appropriateness showed steady improvements until the 7th year postimplantation. The average contour appropriateness score reached a plateau at slightly below 62%. Finally, a decline was observed in the average contour appropriateness score following the 7th year postimplantation. Linear regression analyses further revealed that the improving rates from the preimplant session to the 7th year and from the 7th year to the 10th year did not differ as a function of age at implantation (0–7 years, $p = .061$; 7–10 years, $p = .316$).

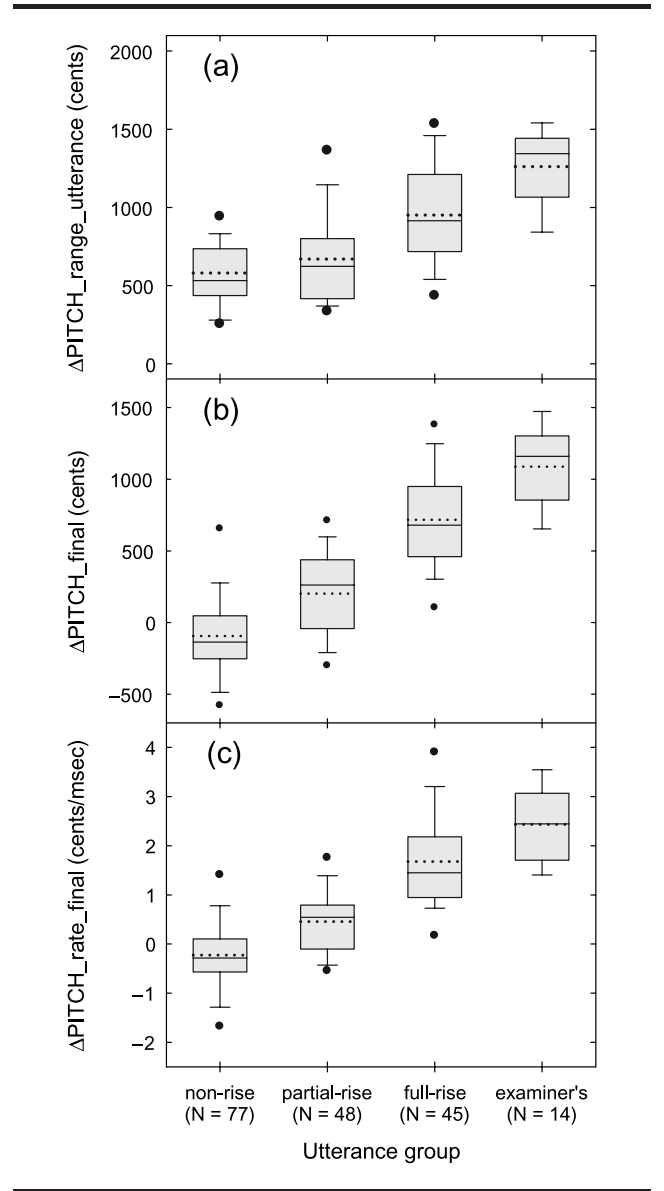
Acoustic Analyses

It was noted in the previous section that 45.29% ($n = 77$), 28.24% ($n = 48$), and 26.47% ($n = 45$) of the present CI participants' utterances were judged to have the intonation contour type of non-rise, partial-rise, and full-rise, respectively. In this section, acoustic properties of utterances were compared in relation with the intonation contour types of utterances judged by adult NH listeners. Additionally, 14 examiner-modeled utterances that were randomly selected from the recordings were also acoustically analyzed to illustrate how utterances could possibly be produced by an adult speaker. The 170 utterances of the three intonation contour types and the examiner-modeled utterances were assigned to four utterance groups—non-rise, partial-rise, full-rise, and examiner's. The acoustic properties of F0, intensity, and duration patterns of these utterances are described below.

Fundamental frequency (F0). The F0-related analyses were first performed to examine peak-to-valley voice pitch range at the utterance level (PITCH_range_utterance). Figure 3a shows the distributions of PITCH_range_utterance for the four utterance groups. The mean voice pitch range was the greatest for the examiner's utterances, followed by that of the full-rise, partial-rise, and non-rise groups. A one-way analysis of variance (ANOVA)¹ revealed that not all of the means among utterance groups were equal, $F(3, 180) = 32.26, p < .001$. Post hoc pairwise comparisons of group means (Tukey–Kramer adjustment) indicated that all pairs of the means were significantly different (all $ps < .001$), except for the difference between the means of the non-rise and partial-rise groups ($p = .339$).

¹A mixed model in which group (i.e., four utterance groups) was treated as fixed effect and subject (i.e., CI participant and examiner) was treated as random effect was first fitted to data to test for the equivalence of the group means. If the subject effect was nonexistent, it was dropped from the model, and a one-way ANOVA model was used. This principle was applicable to the group mean comparisons for all other acoustic parameters.

Figure 3. F0-related parameters versus utterance groups. The x-axis displays the four utterance groups (non-rise, partial-rise, full-rise, and examiner's); the number of utterances in each utterance group is indicated in the parentheses. (a) The voice pitch range at the utterance level (PITCH_range_utterance). (b) The amount of voice pitch change at the utterance-final word (Δ PITCH_final). (c) The rate of voice pitch change at the utterance-final word (Δ PITCH_rate_final). 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 first and third quartiles; the end of the whiskers are located at ± 1.25 SD away from the mean; and the filled circles represent the 5th and 95th percentiles bounds, if they are outside of the end of whisker.



Note that differences in PITCH_range_utterance among utterance groups did not indicate whether the voice pitch variation involved direction change (e.g., rising vs. falling). In the subsequent analyses of the F0-related parameters, the directions of voice pitch change

at utterance-final words were considered. The differentials of the mean voice pitch between each of the four positions, using F0_mean_final1 as the reference in each utterance, were first calculated. The F0 values at these positions were then converted into its voice pitch equivalence (measured in cents). Figure 4 displays the voice pitch contours for the four utterance groups. Each contour was plotted by connecting the means of voice pitch differentials of the same group. Noticeably, the magnitude of contour variation was the most evident at the position of utterance-final words.

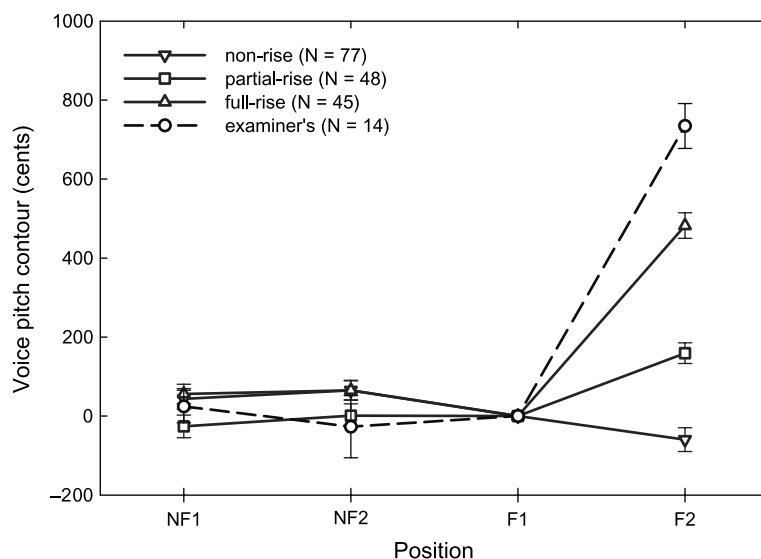
The magnitude and direction of voice pitch variation at utterance-final words were further compared among the utterance groups with respect to the amount of voice pitch change from the onset of the utterance-final word to the offset (Δ PITCH_final). Figure 3b displays the distributions of Δ PITCH_final for the four utterance groups. The mean Δ PITCH_final was the greatest for the examiner's utterances, followed by the means for utterances of the full-rise, partial-rise, and non-rise groups. Results of a one-way ANOVA indicated that not all of the means of Δ PITCH_final were equal among the four utterance groups, $F(3, 180) = 88.45, p < .001$. Post hoc pairwise comparisons of group means (Tukey-Kramer adjustment) indicated that all pairs of the group means were significantly different ($p = .002$ for full-rise & examiner's; all other $ps < .001$). Note that Δ PITCH_final was positive for the majority of the full-rise and examiner's utterances, but it appeared to be less than zero for

some of the partial-rise utterances. In addition, whereas the mean Δ PITCH_final of the non-rise utterances was less than zero, Δ PITCH_final was greater than zero for a proportion of these non-rise utterances.

The last set of acoustic analyses regarding F0-related parameters examined the rate of voice pitch change at utterance-final words (Δ PITCH_rate_final). Figure 3c illustrates the distributions of Δ PITCH_rate_final for the four utterance groups. In general, a positive number for Δ PITCH_rate_final was consistent with a rising voice pitch contour, and a negative number or zero was consistent with a falling or flat voice pitch contour. The mean Δ PITCH_rate_final for the four utterance groups ranged from -0.23 to 2.43 cents/ms, and was the greatest for the examiner's utterances, followed by that for utterances of the full-rise, partial-rise, and non-rise groups. A one-way ANOVA indicated that not all of the means of Δ PITCH_rate_final were equal among the four utterance groups, $F(3, 180) = 70.41, p < .001$. Post hoc pairwise comparisons of group means (Tukey-Kramer adjustment) revealed a significant difference between all pairs of utterance group means ($p = .021$ for full-rise & examiner's; all other $ps < .001$).

In summary, the above-mentioned acoustic analyses revealed systematic patterns for F0-related parameters among the intonation contour types (i.e., non-rise, partial-rise, and full-rise), as were judged by NH listeners. The mean peak-to-valley voice pitch range of non-rise utterances was more reduced than that of partial- and full-rise

Figure 4. Voice pitch contour versus utterance groups. The x-axis displays the four positions (NF1 for nonfinal1, NF2 for nonfinal2, F1 for final1, and F2 for final2) in the target utterance. The y-axis shows the differential of mean voice pitch at each position from final1. Each line represents one utterance group; the number of utterances of each group is indicated in the parentheses (see legend). Error bars display ± 1 SE of the mean score.



utterances. Similarly, the amount of the voice pitch change from the onset to the offset at utterance-final words of non-rise utterances was more reduced when compared with that of partial-rise and full-rise utterances. Similar results were obtained with the rate of voice pitch change at utterance-final words. That is, the average rate was more reduced for utterances of the non-rise group than the rates for utterances of the remaining utterance groups (partial-rise, full-rise, and examiner's). Moreover, many of non-rise utterances exhibited a negative rate of voice pitch change (66.23%). On the other hand, there were fewer utterances of which the rate of voice pitch change was less than zero for partial-rise utterances (33.33%). None of the full-rise and examiner-modeled utterances showed a rate that was less than zero. Altogether, the patterns of voice pitch variation of utterances were systematically associated with the intonation contour type and the appropriateness of intonation contour judged by the adult NH listeners.

Intensity. The intensity properties were examined with respect to the peak intensity values for the vocalic nuclei at four positions (nonfinal1, nonfinal2, final1, and final2) of each utterance. The peak intensity values at these positions were first normalized with reference to the peak intensity at the non-final word *you* (nonfinal2). The normalized peak intensity values, that is, differentials of the peak intensity at each position from INT_peak_nonfinal2, were denoted as INT_ratio_nonfinal1/nonfinal2, INT_ratio_nonfinal2/nonfinal2, INT_ratio_final1/nonfinal2, and INT_ratio_final2/nonfinal2. Panels a–d of Figure 5 illustrate the distributions of the normalized peak intensity at each position for the four utterance groups. The number in each panel displays the total number of utterances for each utterance group; one utterance in the partial-rise group was excluded from this set of analysis because only one syllable (as opposed to two) was identifiable at the utterance-final position of this utterance.

A one-way ANOVA was performed, respectively, to compare the group means of each of the normalized peak intensity values (i.e., INT_ratio_nonfinal1/nonfinal2, INT_ratio_final1/nonfinal2, and INT_ratio_final2/nonfinal2). The differences in group means were not found to be statistically significant for INT_ratio_nonfinal1/nonfinal2, $F(3, 180) = 1.96, p = .122$. However, the group means were found to be significantly different for INT_ratio_final1/nonfinal2, $F(3, 180) = 4.82, p = .003$, and for INT_ratio_final2/nonfinal2, $F(3, 180) = 13.97, p < .001$.

The group means of INT_ratio_final1/nonfinal2 ranged from 1.33 dB (non-rise, $SD = 4.78$ dB) to 5.76 dB (examiner's, $SD = 4.92$ dB) for the four utterance groups. The group means of INT_ratio_final2/nonfinal2 ranged from -3.42 dB (non-rise, $SD = 4.92$ dB) to 1.98 dB (examiner's, $SD = 4.74$ dB) for the four utterance groups. Post hoc pairwise comparisons of group means (Tukey–Kramer

adjustment) indicated that for INT_ratio_final1/nonfinal2, the means between the non-rise and examiner's groups were significantly different ($p = .003$); for INT_ratio_final2/nonfinal2, the mean of the non-rise group was significantly lower than each of the other three utterance groups (all $ps < .005$). In summary, the normalized peak intensity values of each of the two syllables at utterance-final words (final1 and final2) were significantly different between the non-rise group and at least one of the other utterance groups (partial-rise, full-rise, and examiner's). These results demonstrate that relative peak intensity patterns are associated with the intonation contour types (i.e., non-rise, partial-rise, or full-rise) judged by adult NH listeners.

Duration. The duration ratio of the non-utterance-final word to the entire utterance (DUR_ratio_nonfinal/utterance) and the duration ratio of the utterance-final word to the entire utterance (DUR_ratio_final/utterance) were derived and compared among the four utterance groups. Figure 6 illustrates these two types of ratios (nonfinal vs. final) for the utterance groups of non-rise, partial-rise, full-rise, and examiner's. Results of paired t tests indicated a significantly greater mean ratio of DUR_ratio_final/utterance than that of DUR_ratio_nonfinal/utterance for the examiner's group, $t(13) = 2.53, p = .025$; for the full-rise group, $t(44) = 3.78, p < .001$; and for the partial-rise group, $t(47) = 2.88, p = .006$. The difference in the mean ratios, however, was not found to be statistically significant for the non-rise group, $t(76) = 0.91, p = .365$.

The duration ratio difference between the nonfinal and final portions of utterances was further evaluated. The standard deviation of the duration ratio difference for the non-rise, partial-rise, full-rise, and examiner's group was .1593, .1495, .1243, and .0897, respectively. A likelihood ratio test, which assessed the equivalence of the standard deviations among the groups, revealed that not all of the standard deviations were equal, $\chi^2(3) = 8.90, p = .031$. F tests that were performed to test for the equivalence of the standard deviation between each pair among the groups indicated that the distribution of the examiner's utterances spread out significantly less than the distributions of the non-rise and partial-rise utterances, $F(76, 13) = 3.16, p = .012$, for non-rise versus examiner's; $F(47, 13) = 2.78, p = .024$, for partial-rise versus examiner's. Figure 7 displays the distributions of the duration ratio difference between the nonfinal and final portions of utterances for the non-rise, partial-rise, full-rise, and the examiner's groups.

Taken together, unlike those in the other utterance groups, the non-rise utterances did not show a greater duration ratio of the utterance-final word to the entire utterance than the ratio of the non-utterance-final word to the entire utterance. Moreover, the data points of the

Figure 5. Distributions of the normalized peak intensity at each position, using the peak intensity at nonfinal2 as the reference. The x-axis displays the four positions (NF1 for nonfinal1, NF2 for nonfinal2, F1 for final1, and F2 for final 2). The y-axis displays the peak intensity ratio (INT_ratio) measured in dB. Each panel displays the distributions for each of the following utterance groups: non-rise, partial-rise, full-rise, and examiner's. The number of utterances in each group is indicated in the parentheses. Error bars display ± 1 SE of the mean score.

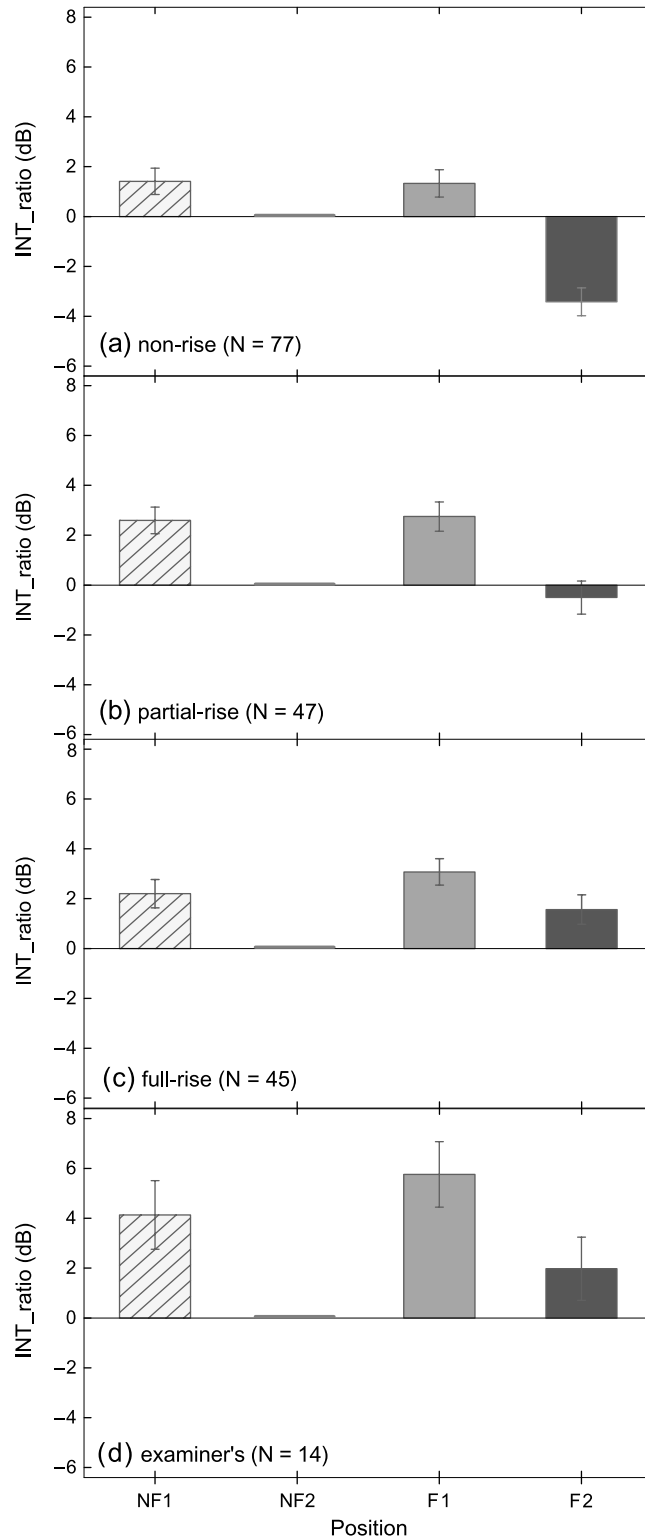


Figure 6. Means of the duration (DUR) ratio for the nonfinal versus final portions of utterances. The x-axis displays the four utterance groups (non-rise, partial-rise, full-rise, and examiner's). The y-axis displays the duration ratio. The crosshatched portion denotes DUR_ratio_nonfinal/utterance; the solid portion denotes DUR_ratio_final/utterance. Error bars display ± 1 SE of the mean score.

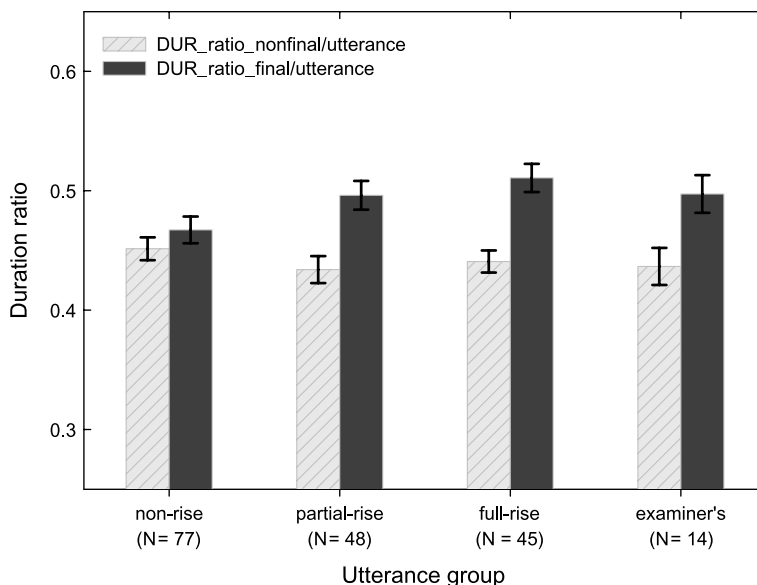
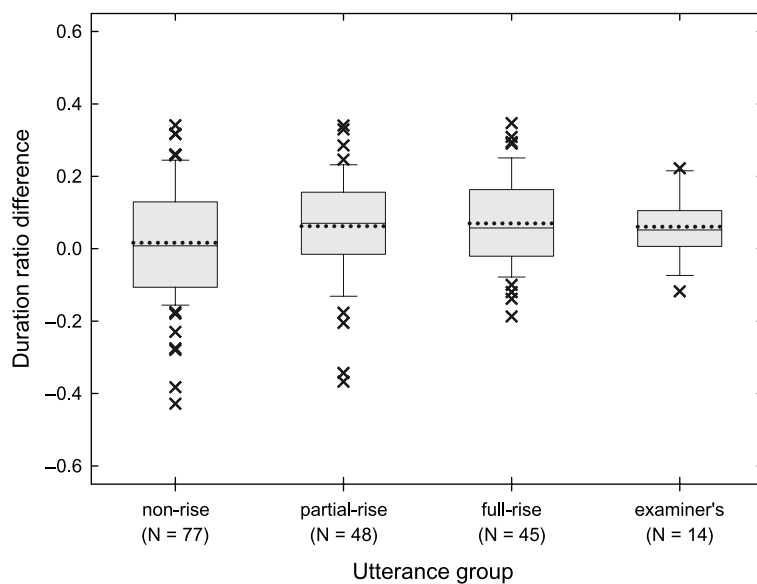


Figure 7. Duration ratio difference between the nonfinal and final portions of utterances. The x-axis displays the four utterance groups (non-rise, partial-rise, full-rise, and examiner's). The y-axis displays the duration ratio difference. The mean and median are displayed by the dotted and solid lines across each box, respectively. The upper and lower bounds of each box represent the first and third quartiles; both ends of the whiskers are located at ± 1.25 SD away from the mean. The x symbols display the data points outside the ends of whiskers.



utterances of the non-rise and partial-rise groups tend to spread out more than the examiner-modeled utterances. These results indicate that duration patterns are different among utterances of some intonation contour types.

Discussion

The present study depicted pediatric CI recipients' performance in producing rising speech intonation, using both perceptual judgments and acoustic analyses. The perceptual judgment results indicate that the pediatric CI recipients in this study did not consistently use appropriate intonation contours when imitating a yes–no question. The acoustic results provided in-depth descriptions of how acoustic properties pertaining to speech intonation, that is, F0, intensity, and duration properties were utilized by pediatric CI recipients in their production. With this combined approach, the study provided objective quantifications of how each of the acoustic correlates pertaining to speech intonation was associated with pediatric CI recipients' utterances judged by adult NH listeners to be different in intonation contour types (non-rise, partial-rise, and full-rise).

Perceptual Judgments

The perceptual judgment results provided limited evidence suggesting the consistent use of rising intonation in a yes–no interrogative utterance, at least for the pediatric CI recipients in this study. At all test intervals following implantation, there was a portion of utterances that were judged to be non-rise. Even though these CI users showed improvement in their use of rising intonation from the preimplant session to the 7th year postimplantation, no further improvement was observed in the adequate use of a rising intonation contour in these individuals following 7 years of device experience.

Noticeably, the longitudinal data displayed in Figure 2 indicate a drop in the contour appropriateness scores from the preimplant to 1-year postimplantation. This drop was likely rooted from data variability: Data points were available from only 3 of the participants at the preimplant interval. The contour appropriateness scores were low for 2 of the participants at this interval (22.5% and 32.5%), but the score was relatively high for 1 participant (77.5%). In fact, the participant with the score of 77.5% also showed high contour appropriateness scores at later follow-up intervals (all scores > 82.5%) but not the other 2 participants (all scores < 77.5%).

Data of a subgroup of 21 CI participants were used to estimate the improving (or declining) rates for the scores of intonation contour appropriateness. This was because

not every CI participant's utterances were available at all annual sessions. Moreover, 3 participants did not have any utterances available after 7 years of CI experience. The analysis revealed that the proportions of pediatric CI recipients' utterances involving a rising intonation contour showed a decline beyond the 7th year postimplantation. Many of the utterances produced by these individuals were not judged as having a rising component (partial-rise or full-rise) in the intonation contours of their utterances.

Previous studies in the literature indicated that pediatric CI users did not develop mastery of producing contrasts in intonation and other prosodic components of speech during the initial 2 years postimplantation (Tobey et al., 1991; Tobey & Hasenstab, 1991). Moreover, some of pediatric CI recipients demonstrated improved production of contrasts in the prosodic components of speech during the 1st year following implantation, but this improvement did not continue to grow after 1 year of device use experience. Note that the pediatric CI users in the earlier studies (e.g., Tobey et al., 1991; Tobey & Hasenstab, 1991) had less than 2 years of device experience and were mapped with relatively older speech-coding strategies (F0/F2, F0/F1/F2, and MPEAK). Consistent with the findings of the studies of Tobey and colleagues, the perceptual results of this study indicated that pediatric CI recipients did not show mastery of speech intonation production. Although pediatric CI recipients showed improvement in their appropriate use of a rising intonation contour with the use of an implant, a plateau was observed in the CI participants' performance at around the 7th year postimplantation. In other words, with increasing device experience, the appropriate use of rising intonation in the pediatric CI recipients in this study showed steady improvements only during the initial several years postimplantation (see Figure 2).

Acoustic Analyses

The acoustic results indicated that each of the F0, intensity, and duration properties was distinctive in the utterances of different intonation contour types (i.e., non-rise, partial-rise, and full-rise) judged by adult NH listeners. First, the voice pitch properties were consistently associated with the intonation contour types. The voice pitch range at the utterance level (PITCH_range_utterance), and the direction and rate of the voice pitch change at utterance-final words (Δ PITCH_final and Δ PITCH_rate_final) were found to be significantly different among the utterances of different intonation contour types. Similarly, the intensity properties were discernible among the utterances of different types. Specifically, the pattern of peak intensity distribution was found to be dissimilar between the non-rise utterances and the partial-rise and full-rise utterances. With regard to

the duration properties, the non-rise utterances did not exhibit greater duration ratio for the utterance-final than for the non-utterance-final components, whereas the partial-rise and full-rise utterances did. In addition, the data points of the duration ratio difference for the non-rise utterances spread out more than the data points of the examiner-modeled utterances.

In the present study, the adult listeners made perceptual judgments based on a totality of acoustic properties. Although this study was not intended to evaluate the relative importance of each of various acoustic dimensions to the perceptual judgments of NH listeners, the consistent acoustic distinctions in the F0, intensity, and duration patterns observed among various contour types can provide clinical implications. For example, in the aural rehabilitation programs designed for pediatric CI recipients and other hearing-impaired individuals, multiple acoustic dimensions (as opposed to any single dimension) should be addressed because all dimensions may contribute to the intonation contours perceived by NH listeners.

The patterns of acoustic properties among the utterances of different intonation types (i.e., non-rise, partial-rise, and full-rise) are consistent with our understanding of speech intonation in terms of how its acoustic correlates correspond to its perception in NH listeners. That is, F0-related properties—such as F0 contour shape, rate of F0 change, endpoint F0, and direction and amount of F0 change at the terminal position of utterances—contribute substantially to NH listeners' perception of speech intonation contrasts (Hadding-Koch & Studdert-Kennedy, 1964; Lehiste, 1970, 1976; Studdert-Kennedy & Hadding, 1973). Similarly, other acoustic dimensions—that is, intensity and duration properties—may also contribute to NH listeners' perception of speech intonation contrasts (Fry, 1955, 1958; Lehiste, 1970, 1976; Lieberman, 1967).

Taken together, the present acoustic findings indicated that the F0, intensity, and duration properties of the pediatric CI recipients' utterances are different among utterances of different intonation contour types judged by adult NH listeners. Moreover, the perceptual judgments of intonation contour type and contour appropriateness were positively correlated. That is, full-rise utterances were judged to be more appropriate than non-rise utterances. In contrast to those of partial-rise and full-rise utterances, the amount and rate of voice pitch change at utterance-final words of non-rise utterances were significantly reduced in their magnitude. Similarly, the intensity and duration properties exhibited different patterns between non-rise utterances and utterances of partial-rise and full-rise types. Note that the adult listeners' perceptual judgments in this study were made on the basis of a totality of acoustic properties. The relative

importance to intonation contour appropriateness of each acoustic dimension is to be determined in future research.

Nonetheless, in the present study, many of the target utterances produced by the group of prelingually deafened children with a CI were superimposed with an intonation contour of non-rise. These findings are consistent with the acoustic findings in Mandarin-speaking CI children, who tend to neutralize tone pattern (i.e., flat contours) in their production (Xu et al., 2004). Recall rising speech intonation is physiologically effortful and requires linguistic experience (Boothroyd, 1982; Lieberman, 1967; Snow, 1998; Vihman, 1996). Rising contours of both speech intonation and lexical tones tend to be more difficult than their falling counterparts for children with NH to acquire (intonation, Snow, 1998; lexical tones, Li & Thompson, 1977). The CI children's lack of mastery of the production of rising intonation contours, even with extended (7–10 years) device experience also suggests that such intonation contours may be particularly demanding for prelingually deafened children to master.

Because of the retrospective nature of this study, some limitations were unavoidable. For instance, only one utterance was available from each participant at each annual session. Moreover, not all utterances from the group of pediatric CI recipients were available from every single session. Another limitation arose from the target utterance's syntactic structure. A rising intonation contour of yes–no questions can sometimes be optional in oral speech communication (Levis, 1999). For that reason, there was no guarantee that the nonuse of a rising intonation contour in this target yes–no question indeed reflected the CI users' breakdown of producing rising intonation. On the other hand, it is also important to note that on the basis of the panel of adult NH listeners' perceptual judgments, yes–no questions without a rising intonation contour tended to receive lower scores of contour appropriateness than those with a rising intonation contour. The superimposition of rising intonation contours, hence, is at least important for NH listeners' appropriateness judgments of yes–no questions.

Note that the present study was not intended to be conclusive but rather served as a gateway that allows researchers and clinicians to identify and verify the potential limitations of CI devices in facilitating the acquisition of prosodic components of speech in prelingually deafened children. Future studies should address (a) whether the degraded signals these individuals perceive via a CI contribute to their difficulties with speech intonation production, and (b) whether more recent CI technology leads to improved utilization of acoustic properties in pediatric CI recipients' perception and production of speech intonation. Findings of these studies may facilitate a better understanding of the acquisition of

intonation and other prosodic properties of speech in this population as well as these individuals' potential sources of difficulties with contrastive use of the acoustic correlates of speech intonation in their perception and production.

Conclusion

Pediatric CI recipients' ability to appropriately use rising intonation in their imitative speech production was perceptually and acoustically evaluated in this study. The perceptual results indicate that the present pediatric CI recipients did not show mastery of using rising intonation in their imitative speech production, although these CI users exhibited some progress in their production of appropriate rising intonation contours with increasing device experience. However, this improvement was limited and did not continue to consistently increase after 7 years of device experience. Instead, signs of decline with prolonged device experience (i.e., beyond 7 years) were observed. The acoustic findings delineated the systematic patterns of acoustic properties for speech intonation in accordance with adult NH listeners' perceptual judgments. The acoustic properties of F0, intensity, and duration were all found to be distinguishable among the utterances that were judged to be non-rise, partial-rise, and full-rise. Note that adult NH listeners' judgments of utterances types were highly positively correlated with their judgments of contour appropriateness. Unlike the full-rise utterances, the non-rise utterances were less likely to be judged as appropriate. Hence, even though the acoustic findings have been described with respect to the intonation contour type of utterances, systematic variation of each parameter among utterances of non-rise, partial-rise, and full-rise contour types reflect the adequacy of the suprasegmental components of target utterances perceived by adult NH listeners.

Acknowledgments

Portions of this article were presented at the 8th International Cochlear Implant Conference in Indianapolis, Indiana (2004) and at the American Speech-Language-Hearing Association Convention in Philadelphia, Pennsylvania (2004). Funding of this project was provided by National Institute on Deafness and Other Communication Disorders Grant P50 DC00242. This article was primarily based on one study of the first author's doctoral dissertation submitted to the University of Iowa. We would like to express our gratitude to all cochlear implant recipients, their parents, and adult normal-hearing listeners who participated in the present study. We would also like to thank Chris Turner, Arthur Boothroyd, Kay Gfeller, Sandie Bass-Ringdahl, and Monita Chatterjee for their helpful feedback and comments on

earlier versions of this article. We appreciate Arik Wald for providing programming support and Nelson Lu for his assistance in statistical analysis.

References

- Allen, G. D., & Arndorfer, P. M. (2000). Production of sentence-final intonation contours by hearing-impaired children. *Journal of Speech, Language, and Hearing Research, 43*, 441–455.
- Barry, J. G., Blamey, P. J., & Martin, L. F. (2002). A multidimensional scaling analysis of tone discrimination ability in Cantonese-speaking children using a cochlear implant. *Clinical Linguistics and Phonetics, 16*, 101–113.
- Barry, J. G., Blamey, P. J., Martin, L. F. A., Lee, K. Y. S., Tang, T., Ming, Y. Y., & van Hasselt, C. A. (2002). Tone discrimination in Cantonese-speaking children using a cochlear implant. *Clinical Linguistics and Phonetics, 16*, 79–99.
- Berkovits, R. (1984). Duration and fundamental frequency in sentence-final intonation. *Journal of Phonetics, 12*, 255–265.
- Boersma, P., & Weenink, D. (2004). Praat (Version 4.3) [Computer software]. Amsterdam: Institute of Phonetic Sciences, University of Amsterdam.
- Boothroyd, A. (1982). *Hearing impairments in young children*. Englewood Cliffs, NJ: Prentice-Hall.
- Burns, E. M., & Ward, W. D. (1982). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (pp. 241–269). New York: Cambridge University Press.
- Ciocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *The Journal of the Acoustical Society of America, 111*, 2250–2256.
- Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.
- Crystal, D. (1979). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 33–48). Cambridge, England: Cambridge University Press.
- Denes, P. (1959). A preliminary investigation of certain aspects of intonation. *Language and Speech, 2*, 106–122.
- Denes, P., & Milton-Williams, J. (1962). Further studies in intonation. *Language and Speech, 5*, 1–14.
- D'Odorico, L., & Franco, F. (1991). Selective production of vocalization types in different communication contexts. *Journal of Child Language, 18*, 475–499.
- Faulkner, A., Rosen, S., & Smith, C. (2000). Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants. *The Journal of Acoustic Society of America, 108*, 1877–1887.
- Freeman, F. J. (1982). Prosody in perception, production, and pathologies. In D. E. Yoder (Ed.), *Speech, language, and hearing: Pathologies of speech and language* (Vol. 2, pp. 652–672). Philadelphia: W. B. Saunders.
- Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America, 27*, 765–768.

- Fry, D. B.** (1958). Experiments in the perception of stress. *Language and Speech, 1*, 126–152.
- Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D.** (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America, 104*, 505–510.
- Furrow, D.** (1984). Young children's use of prosody. *Journal of Child Language, 11*, 203–213.
- Galligan, R.** (1987). Intonation with single words: Purposive and grammatical use. *Journal of Child Language, 14*, 1–21.
- Geurts, L., & Wouters, J.** (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *The Journal of the Acoustical Society of America, 109*, 713–726.
- Green, T., Faulkner, A., & Rosen, S.** (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *The 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. *The Journal of the Acoustical Society of America, 116*, 2298–2310.
- Hadding-Koch, K., & Studdert-Kennedy, M.** (1964). An experimental study of some intonation contours. *Phonetica, 11*, 175–185.
- Ladd, D. R.** (1996). *Intonational phonology*. Cambridge, England: Cambridge University Press.
- Ladefoged, P.** (2001). *A course in phonetics* (4th ed.). Orlando, FL: Harcourt Brace.
- Lee, K. Y. S., van Hasselt, C. A., Chiu, S. N., & Cheung, D. M. C.** (2002). Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *International Journal of Pediatric Otorhinolaryngology, 63*, 137–147.
- 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: Academic Press.
- Levis, J. M.** (1999). The intonation and meaning of normal yes/no questions. *World Englishes, 18*, 373–380.
- Li, C. N., & Thompson, S. A.** (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*, 185–199.
- Lieberman, P.** (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.
- Loeb, D. F., & Allen, G. D.** (1993). Preschoolers' imitation of intonation contours. *Journal of Speech and Hearing Research, 36*, 4–13.
- Moore, B. C. J.** (1997). Aspects of auditory processing related to speech perception. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic science* (pp. 539–565). Cambridge, MA: Blackwell.
- O'Halpin, R.** (2001). Intonation issues in the speech of hearing impaired children: Analysis, transcription, and remediation. *Clinical Linguistics and Phonetics, 15*, 529–550.
- Osberger, M. J., Miyamoto, R. T., Zimmerman-Phillips, S., Kemink, J. L., Stroer, B. S., Firszt, J. B., & Novak, M. A.** (1991). Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear and Hearing, 12*, 66S–80S.
- Osberger, M. J., Robbins, A. M., Miyamoto, R. T., Berry, S. W., Myres, W. A., Kessler, K. S., & Pope, M. L.** (1991). Speech perception abilities of children with cochlear implants, tactile aids, or hearing aids. *The American Journal of Otolaryngology, 12*, 105S–115S.
- Peng, S., Spencer, L. J., & Tomblin, J. B.** (2004). Speech intelligibility of pediatric cochlear implant recipients with seven years of device experience. *Journal of Speech, Language, and Hearing Research, 47*, 1227–1236.
- Peng, S., 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.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J.** (1985). *A comprehensive grammar of the English language*. New York: Longman.
- Rosen, S.** (1989, June). *Temporal information in speech and its relevance for cochlear implants*. Paper presented at the Cochlear Implant: Acquisitions and Controversies Meeting, Toulouse, France.
- Rosen, S.** (1992). Temporal information in speech: Acoustic, auditory, and linguistic aspects. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 336*, 367–373.
- Snow, D.** (1998). Children's imitations of intonation contours: Are rising tones more difficult than falling tones? *Journal of Speech, Language, and Hearing Research, 41*, 576–587.
- Snow, D., & Balog, H. L.** (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua, 112*, 1025–1058.
- Studdert-Kennedy, M., & Hadding, K.** (1973). Auditory and linguistic processes in the perception of intonation contours. *Language and Speech, 16*, 293–313.
- Tobey, E. A., Angelette, S., Murchison, C., Nicosia, J., Sprague, S., Staller, S., et al.** (1991). Speech production performance in children with multichannel cochlear implants. *The American Journal of Otolaryngology, 12*, 165S–173S.
- Tobey, E. A., & Hasenstab, M. S.** (1991). Effects of a Nucleus multichannel cochlear implant upon speech production in children. *Ear and Hearing, 12*, 48S–54S.
- Tye-Murray, N.** (1998). Speech, language, and literacy development. In N. Tye-Murray & W. Clark (Eds.), *Foundations of aural rehabilitation: Children, adults, and their family members* (pp. 415–446). San Diego, CA: Singular Publishing Group.
- Tye-Murray, N., Spencer, L., & Woodworth, G. G.** (1995). Acquisition of speech by children who have prolonged cochlear implant experience. *Journal of Speech and Hearing Research, 38*, 327–337.
- Vihman, M. M.** (1996). *Phonological development: The origins of language in the child*. Cambridge, MA: Blackwell.
- Wei, W. I., Wong, R., Hui, Y., Au, D. K., Wong, B. Y., Ho, W. K., et al.** (2000). Chinese tonal language rehabilitation following cochlear implantation in children. *Acta Otolaryngologica, 120*, 218–221.

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

Xu, L., Li, Y., Hao, J., Chen, X., Xue, S. A., & Han, D. (2004). Tone production in Mandarin-speaking children with cochlear implants: A preliminary study. *Acta Oto-laryngologica*, 124, 363–367.

Received September 8, 2006

Accepted February 26, 2007

DOI: 10.1044/1092-4388(2007)085

Contact author: Shu-Chen Peng, Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742. E-mail: speng@hesp.umd.edu.

Appendix. Acoustic parameters examined in this study.

Parameter	Descriptions
F0/voice pitch (Hz/cents)	
F0_valley_utterance	Minimal F0 at utterance level
F0_peak_utterance	Maximal F0 at utterance level
F0_onset_final	F0 at onset of utterance-final word
F0_offset_final	F0 at offset of utterance-final word
F0_mean_nonfinal1	Mean F0 of first non-utterance-final word <i>are</i>
F0_mean_nonfinal2	Mean F0 of second non-utterance-final word <i>you</i>
F0_mean_final1	Mean F0 of first syllable of utterance-final word <i>rea-</i>
F0_mean_final2	Mean F0 of second syllable of utterance-final word <i>dy</i>
PITCH_range_utterance	Peak-to-valley voice pitch range at utterance level
ΔPITCH_final	Amount of voice pitch change from onset to offset of utterance-final word
ΔPITCH_rate_final	Rate (slope) of voice pitch change at utterance-final word
Duration (ms)	
DUR_utterance	Duration of entire utterance <i>are you ready</i>
DUR_final	Duration of utterance-final word <i>ready</i>
DUR_nonfinal	Duration of non-utterance-final words <i>are you</i>
DUR_ratio_nonfinal/utterance	Duration ratio of non-utterance-final word to entire utterance
DUR_ratio_final/utterance	Duration ratio of utterance-final word to entire utterance
Intensity (dB)	
INT_peak_nonfinal1	Peak intensity at first non-utterance-final word <i>are</i>
INT_peak_nonfinal2	Peak intensity at second non-utterance-final word <i>you</i>
INT_peak_final1	Peak intensity at first syllable of utterance-final word <i>rea-</i>
INT_peak_final2	Peak intensity at second syllable of utterance-final word <i>dy</i>
INT_ratio_nonfinal1/nonfinal2	Peak intensity ratio of first non-utterance-final word to second non-utterance-final word
INT_ratio_nonfinal2/nonfinal2	Peak intensity ratio of second non-utterance-final word to second non-utterance-final word
INT_ratio_final1/nonfinal2	Peak intensity ratio of first utterance-final word to second non-utterance-final word
INT_ratio_final2/nonfinal2	Peak intensity ratio of second utterance-final word to second non-utterance-final word

Note. The capital letters—F0, (Δ)PITCH, DUR, and INT—denote the acoustic parameters, that is, F0, voice pitch (change), duration, and intensity. The component that follows each refers to the specific measurement made, for example, at a single point (valley, peak, onset, and offset) or by taking the mean, range, or ratio. The last part of the abbreviated form refers to the location where the measurement is made (utterance-final word, non-utterance-final words, and utterance).