Production and Perception of Speech Intonation in Pediatric Cochlear Implant Recipients and Individuals with Normal Hearing

Shu-Chen Peng, J. Bruce Tomblin, and Christopher W. Turner

Objectives: **Current cochlear implant (CI) devices are limited in providing voice pitch information that is critical for listeners' recognition of prosodic contrasts of speech (e.g., intonation and lexical tones). As a result, mastery of the production and perception of such speech contrasts can be very challenging for prelingually deafened individuals who received a CI in their childhood (i.e., pediatric CI recipients). The purpose of this study was to investigate (a) pediatric CI recipients' mastery of the production and perception of speech intonation contrasts, in comparison with their age-matched peers with normal hearing (NH), and (b) the relationships between intonation production and perception in CI and NH individuals.**

Design: **Twenty-six pediatric CI recipients aged from 7.44 to 20.74 yrs and 17 age-matched individuals with NH participated. All CI users were prelingually deafened, and each of them received a CI between 1.48 and 6.34 yrs of age. Each participant performed an intonation production task and an intonation perception task. In the production task, 10 questions and 10 statements that were syntactically matched (e.g., "The girl is on the playground." versus "The girl is on the playground?") were elicited from each participant using interactive discourse involving pictures. These utterances were judged by a panel of eight adult listeners with NH in terms of utterance type accuracy (question versus statement) and contour appropriateness (on a five-point scale). In the perception task, each participant identified the speech intonation contrasts of natural utterances in a two-alternative forced-choice task.**

Results: **The results from the production task indicated that CI participants' scores for both utterance type accuracy and contour appropriateness were significantly lower than the scores of NH participants (both $p < 0.001$). The results from the perception task indicated that CI participants' identification accuracy was significantly lower than that of their NH peers (CI, 70.13% versus NH, 97.11%, $p < 0.001$). The Pearson correlation coefficients (r) between CI participants' performance levels in the production and perception tasks were approximately 0.65 ($p = 0.001$).**

Conclusion: **As a group, pediatric CI recipients do not show mastery of speech intonation in their production or perception to the same extent as their NH peers. Pediatric CI recipients' performance levels in the production and perception of speech intonation contrasts are moderately correlated. Intersubject variability exists in pediatric CI recipients' mastery levels in the production and perception of speech intonation contrasts. These findings suggest the importance of addressing both aspects (production and perception) of speech intonation in the aural rehabilitation and speech intervention programs for prelingually deafened children and young adults who use a CI.**

(Ear & Hearing 2008;29;336–351)

INTRODUCTION

Past research has demonstrated that cochlear implants (CIs) are fairly successful in facilitating speech and language development in prelingually deafened children (e.g., Blamey, et al., 2001; Spencer, et al., 1998; Svirsky & Chin, 2000; Svirsky, et al., 2000; Tobey & Hasenstab, 1991; Tye-Murray, et al., 1995). However, current CI devices provide only restricted access for the recognition of pitch-based prosodic components (i.e., fundamental frequency or voice pitch) of speech that signify linguistic contrasts (Faulkner, et al., 2000; Green, et al., 2004; Geurts & Wouters, 2001). Perception of 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). As a result, CI devices are likely to be restricted in facilitating the acquisition of these prosodic properties in prelingually deafened children who must rely on these devices to develop spoken language.

Prosodic components of speech (i.e., suprasegmental properties of speech) can convey several expressive functions in semantic, attitudinal, psychological, and social domains (Crystal, 1979; Lehiste, 1970). Linguistic functions (e.g., lexical tones and speech intonation) are among the most noticeable expressive aspects of prosodic properties of speech. In a tonal language such as Mandarin Chinese, lexical tones are phonemic and can contrast the meanings of syllables or words. For

---

Department of Speech Pathology and Audiology, University of Iowa, Iowa City, IA.

0196/0202/08/2903-0336/0 • Ear & Hearing • Copyright © 2008 by Lippincott Williams & Wilkins • Printed in the U.S.A.

example, when the syllable *ma* is produced with a high-level tone, it refers to mother, but it refers to scold when produced with a high-falling tone. In Mandarin Chinese, fundamental frequency (F0) serves as the major acoustic cue for lexical tone contrasts (Chao, 1968; Howie, 1976; Shih, 1988). However, other acoustic properties (e.g., intensity and duration) may also contribute to the contrasts (Whalen & Xu, 1992).

In a nontonal language such as English, speech prosodic variation can convey linguistic changes in a way similar to that of tonal languages. However, the contrasts may occur at various levels of linguistic units such as words, phrases, or sentences. In spoken English, two sentences with an identical syntactic structure (e.g., "The girl is on the playground?" versus "The girl is on the playground."), but different F0 contours (e.g., rising versus falling) can mark difference in utterance types (question versus statement) (Ladefoged, 2001). Note that although F0 information plays a dominant role in listeners' speech intonation recognition, variation in F0 contours typically takes place in conjunction with variations in intensity and duration patterns (Cooper & Sorensen, 1981; Freeman, 1982; Ladd, 1996; Lehiste, 1970, 1976). Listeners with normal hearing (NH) are able to use F0, intensity, and duration characteristics of utterances to recognize speech intonation contrasts collectively (Fry, 1955, 1958; Lehiste, 1970, 1976; Lieberman, 1967).

Although intensity and duration aspects of speech can be well transmitted to CI listeners via temporal coding of CI devices, voice pitch information is not well presented (Faulkner, et al., 2000; Geurts & Wouters, 2001; Green, et al., 2002, 2004). There are significant limitations in current commercially available CI devices' transmission of voice pitch information (Faulkner, et al., 2000; Geurts & Wouters, 2001). From the temporal perspective, speech-coding strategies such as F0/F2, F0/F1/F2, and spectral peak (SPEAK) permit transmission of only pulse train rates, which do not provide well-defined periodicity information at the frequencies beyond 250 Hz (ASHA Working Group on Cochlear Implant, 2003; Wilson, 2004). Moreover, even with speech-coding strategies that use high stimulation rates, such as advanced combination encoder (ACE), continuous interleaved sampling, or HiResolution, listeners are likely not able to use the temporal envelope cues to recognize voice pitch variations above 300 Hz based on pulse train rate. This is due to the general limitations of the human auditory system in using temporal cues; that is, listeners are highly constrained in their ability to discriminate frequency differences based exclusively on temporal envelope information beyond 300 Hz (Burns & Viemester, 1976; Loizou, 1998; Shannon, 1983). Hence, listeners have only limited access to voice pitch information of certain speakers such as children and female speakers, because the natural voice pitch ranges of these individuals often extend beyond 300 Hz.

From the spectral perspective, several factors may hinder CI listeners' actual benefits from place cues provided by the tonotopic organization (i.e., a "frequency-to-place" mapping) of the basilar membrane of the cochlea. That is, the width of channel allocations, or the number of electrodes designated to deliver spectral information is limited, in particular in the F0 range. For example, with the Nucleus device, only two (of 22) electrodes are normally designated to encode the frequency range relevant to F0. Although spectral information can supplement F0 percepts, listeners are restricted in extracting voice pitch information via a CI because of the device's degraded spectral resolution (Green, et al., 2002, 2004). Moreover, with a small set of available electrodes, many CI listeners show a lack of ability to fully use spectral information, which can be due to poor nerve survival, channel interaction, and warping of the spectral-tonotopic mapping (Friesen, et al., 2001; Wilson, et al., 1988). As a result, voice pitch information relying on place cues is likely not fully resolved (Fu, et al., 1998; Moore, 1997; Rosen, 1989, 1992).

Infants and young children with NH demonstrate the ability to contrast intonation and other prosodic properties of speech in their vocalization or utterances at a very young age (i.e., 1 or 2 yrs; D'Odorico & Franco, 1991; Furrow, 1984; Galligan, 1987). With an increasing age, young children show improvement in their mastery of intonation and other prosodic components of speech (Loeb & Allen, 1993). On the other hand, perception and production of intonation and other prosodic aspects of speech can be challenging to English-speaking children with a CI (O'Halpin, 2001; Green, et al., 2004), due at least partially to the limitations of current CI devices in presenting voice pitch information as well as the general limitations of the human auditory system with electric simulation.

In the literature, only limited empirical evidence has been made available to support this postulate. Previous findings indicated that school-aged children who are prelingually deafened do not generally show mastery in the production or perception of intonation and other prosodic components of speech with 2 yrs of CI experience (Osberger, et al., 1991a,b; Tobey & Hasenstab, 1991; Tobey, et al., 1991). These earlier studies, however, addressed pediatric CI recipients' production or perception performance only during the initial 2 yrs after implantation. In a recent study by Peng et al. (2007), it was reported that pediatric CI users do not consistently produce questions with a proper intonation contour, even with up to 7 yrs of device experience. This study, however, was retrospective in nature and hence was limited in the utterances available; moreover, it evaluated only the production of speech intonation, but not its perception. Nonetheless, many of the CI recipients in the earlier studies were mapped with relatively old speech-coding strategies such as F0/F2 or F0/F1/F2. These older strategies encoded F0 explicitly whereas relatively recent strategies such as SPEAK and ACE implicitly encode F0 only in the amplitude modulations. As such, it is reasonable to anticipate that those older strategies might be better at transmitting F0 information. Nonetheless, with more recent speech-coding strategies, it remains unclear whether or not someone can anticipate better intonation production and perception skills in children with relatively extended CI device experience.

As mentioned earlier, F0 serves as the primary acoustic cue for both speech intonation and lexical tones. Although the research studies addressing pediatric CI recipients' production and perception of speech intonation are considerably limited, production and perception of lexical tones in pediatric CI users who are native speakers of Mandarin Chinese or Cantonese have been evaluated in several studies (e.g., Barry, et al., 2002; Ciocca, et al., 2002; Lee, et al., 2002; Wei, et al., 2000). These studies consistently suggested that prelingually deafened children with a CI show difficulty in perceiving lexical tone contrasts. Peng et al. (2004) further examined lexical tone production in Mandarin-speaking pediatric CI users. The authors indicated that with 1.5 to 6.5 yrs of device experience, the majority of prelingually deaf children with a CI do not master Mandarin tone production. However, the authors reported that several children with a CI achieve high levels of performance in production in addition to perception, and those who exhibit exceptional performance in tone production also tend to perform well in tone identification (but not vice versa).

In studies of normal spoken language development, there is a weak link between the investigations of perception and production (Vihman, 1996). Investigations of perception alone do not explicitly provide integrated information about children's production ability. Additionally, production studies have not routinely concerned with the children's ability to perceive speech contrasts. The mechanisms (i.e., fundamental processes) involved in speech perception and speech production may be different. For example, the phonological systems for the perception and production of speech contrasts are relatively independent at early developmental stages (Ferguson, 1978). However, speech perception and production in infants and young children may share similar mechanisms, and both are associated with biological predisposition and linguistic experience (Vihman, 1996).

The relationships between speech perception and production have been appraised in different models, and the basic belief is "perception precedes production" (Edwards, 1974). However, this principle may sometimes fail to account for speech development in young children. For example, poor performance in production may occur despite the child's mastery of perception (Velleman, 1988). As a result, when children fail to produce certain speech sounds or sound sequences successfully, it is unclear to which extent the difficulty originates from perception, production, or both. In the present study, we were interested in identifying the sources of difficulty with the production of speech intonation contrasts experienced by individuals who received a CI during their childhood. A combined examination of the production and perception skills of these individuals may provide noteworthy information. Hence, the purposes of this present study were to (a) evaluate pediatric CI recipients' mastery of the production and perception of speech intonation contrasts, in comparison with their age-matched peers with NH, and (b) determine the relationships between the production and perception skills in both CI and NH individuals.

MATERIALS AND METHODS

Participants

Three groups of individuals were recruited as participants, designated as CI group, NH group, and adult listeners. All participants were native speakers of English. Individuals in the CI and NH groups participated in the production and perception tasks. The third group of individuals (i.e., adult listeners) served as judges to evaluate the CI and NH participants' speech intonation production. Below is a description of these individuals.

The CI group comprised 26 prelingually deafened individuals, ranging from 7.44 to 20.74 yrs of age (mean, 13.87 yrs). They received a CI between 1.48 and 6.34 yrs of age, and had used a CI between 5.32 and 16.83 yrs at test time. All CI participants were users of Nucleus 22 or 24 devices (Cochlear Americas, Denver, CO). Fifteen CI users had been mapped with the SPEAK speech-coding strategy and 11 had been mapped with the ACE strategy. Nineteen participants received their education in a mainstream, public school setting where both signing exact English and spoken English were used [total communication (TC)], and the other seven received education in a mainstream, public school setting where only spoken English was used [oral communication (OC)]. Classification of communication methods (OC or

TABLE 1. Background information of all CI participants

ID	Gender	Etiology of deafness	Device type	Speech-coding strategy	Age at testing (yrs)	Device use (yrs)	Age at implantation (yrs)	Preop PTA in better ear (dB HL)	Comm. mode
CI-1	Male	Meningitis	N 22	SPEAK	16.52	13.94	2.58	98.3	TC
CI-2	Male	Unknown	N 22	SPEAK	15.53	9.98	5.55	100 ↑	TC
CI-3	Male	Unknown	N 22	SPEAK	20.74	16.83	3.91	90 ↑	OC
CI-4	Female	Unknown	N 22	SPEAK	19.53	14.69	4.84	106.67	TC
CI-5	Male	Genetic	N 22	SPEAK	15.39	9.04	6.34	100	OC
CI-6	Female	Meningitis	N 22	SPEAK	17.61	12.45	5.16	115 ↑	OC
CI-7	Male	Meningitis	N 22	SPEAK	15.34	11.52	3.82	110 ↑	TC
CI-8	Female	Unknown	N 24	ACE	7.44	5.85	1.59	107.5 ↑	TC
CI-9	Male	Unknown	N 24	ACE	10.68	5.39	5.29	108.3	TC
CI-10	Female	Usher's	N 24	ACE	8.61	6.98	1.63	NR	TC
CI-11	Male	Unknown	N 22	SPEAK	13.42	9.18	4.24	100 ↑	TC
CI-12	Female	Unknown	N 24	ACE	12.25	6.01	6.24	90	TC
CI-13	Male	Unknown	N 24	ACE	8.69	5.32	3.37	113.3	TC
CI-14	Female	Unknown	N 22	SPEAK	20.02	14.28	5.75	105 ↑	TC
CI-15	Male	Genetic	N 22	SPEAK	15.75	12.36	3.39	105 ↑	TC
CI-16	Male	Genetic	N 22	SPEAK	13.59	10.85	2.74	105 ↑	TC
CI-17	Male	Unknown	N 24	ACE	9.44	6.08	3.36	96.7	TC
CI-18	Female	Unknown	N 22	SPEAK	18.84	14.46	4.38	110 ↑	OC
CI-19	Male	Meningitis	N 24	ACE	8.46	6.98	1.48	NR	TC
CI-20	Female	Unknown	N 24	ACE	11.88	6.20	5.68	100	TC
CI-21	Female	Unknown	N 22	SPEAK	15.12	11.59	3.53	115 ↑	TC
CI-22	Female	Unknown	N 24	ACE	8.25	5.98	2.27	96.7	TC
CI-23	Male	Meningitis	N 22	SPEAK	18.76	15.24	3.52	NR	OC
CI-24	Female	Unknown	N 24	ACE	10.40	8.06	2.34	NR	TC
CI-25	Male	Genetic	N 24	ACE	9.63	7.04	2.59	95	OC
CI-26	Female	Unknown	N 22	SPEAK	18.65	14.76	3.89	NR	OC
Mean					13.87	10.04	3.83		
SD					4.22	3.71	1.45		

PTA, pure-tone average thresholds (at 500, 1000, and 2000 Hz); 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); N22, Nucleus 22; N24, Nucleus 24; Comm. mode, Communication mode.

TC) was based on parental reports, confirmed by the participant's educational setting at test time. Note that all CI participants in a TC setting had significant exposure to spoken language at school and at home after implantation.

The NH group comprised 17 children and teenagers, ranging from 6.33 to 19.98 yrs of age (mean, 11.52 yrs). The age range of NH participants was approximately matched to that of CI participants. No statistically significant difference was observed in the chronological age between the CI and NH groups [$t(40) = 1.671$, $p = 0.098$]. None of the NH participants presented a clinical history of speech-language impairments. The hearing sensitivity of all NH participants was screened in both ears. All these participants' hearing sensitivity was better than 20 dB HL at octave intervals from 250 to 8000 Hz, bilaterally. All participants gave written informed consent (or assent, if younger than 10 yrs of age) approved by the University of Iowa Institutional Review Board before the task. One parent (or guardian) also gave parental informed consent if the participant was younger than 18 yrs of age at test time. All participants were paid. Tables 1 and 2 provide a summary of the background information of the CI and NH participants.*

Finally, a panel of eight adult listeners with NH (six females and two males) were recruited to perceptually judge the set of utterances elicited from the CI and NH individuals in the production task (see next section). These adult listeners ranged from 22.03 to 38.22 yrs of age (mean, 25.72 yrs). Before judgments, the hearing sensitivity of the listener was screened in both ears. The hearing sensitivity of all listeners was within normal limits (thresholds better than 20 dB HL) at all octave intervals from 250 to 8000 Hz, bilaterally. All adult listeners gave written informed consent that was reviewed and approved by the Institutional Review Board at the University of Maryland – College Park and were paid for participation.

---

*In the production task, recordings of utterances from two participants were not obtained because of the equipment problems (CI-23) or time constraints (NH-13). One participant (CI-18) did not perform the perception task because of limited appointment time. The task descriptions are detailed in the remainder of this section.

TABLE 2. Background information of all NH participants

ID	Gender	Age (yrs)	PTA in left ear	PTA in right ear
NH-1	Male	12.83	8.33	8.33
NH-2	Female	17.51	3.33	5.00
NH-3	Female	17.36	1.67	1.67
NH-4	Male	9.84	8.33	3.33
NH-5	Male	11.68	8.33	0.00
NH-6	Male	9.03	5.00	1.67
NH-7	Female	12.04	11.67	5.00
NH-8	Female	12.96	6.67	5.00
NH-9	Male	7.00	5.00	1.67
NH-10	Female	6.33	11.67	10.00
NH-11	Male	9.24	10.00	11.67
NH-12	Male	7.32	11.67	10.00
NH-13	Male	19.98	6.67	6.67
NH-14	Male	10.10	3.33	6.67
NH-15	Female	10.10	5.00	1.67
NH-16	Male	8.30	8.33	6.67
NH-17	Female	14.29	10.00	10.00
Mean		11.52	7.35	5.59
SD		3.93	3.12	3.58

PTA, pure-tone average thresholds at 500, 1000, and 2000 Hz.

Production Task

Speech materials • A set of 20 pictures was used to elicit 10 questions and 10 statements from each CI or NH participant. The vocabularies were familiar to even the youngest CI participants. The target names (locations or objects) were illustrated on the pictures. The length of each target utterance ranged from 5 to 7 words; all utterances were syntactically simple (Appendix A). The 20 pictures were presented to the participant in pairs (e.g., pictures of a fish tank and a jar). For each pair of pictures, one served as the target and the other served as the competing item that was intended to make the interactive discourse pragmatically appropriate. With each pair, four utterances were expected to be produced by the examiner and the participant in a turn-taking task. The participant was first instructed that s/he was going to play a role-playing game with the examiner. The participant was informed that in the game, the examiner would take turns with him/her asking and answering questions.

Procedure • Each participant performed the production task in a quiet testing room (ambient noise level = 40 dB SPL; long-term averaged level, A-weighting). The production task in the role-playing format always involved the same interactive discourse between the participant and the examiner. An example of this exchange with one set of utterances using the pictures of a fish tank and a jar would be (i) Where is the fish? (ii) *The fish is in the fish tank.* (iii) Really? *The fish is in the fish tank?* (iv) Yes, the fish is in the fish tank. The italicized utterances in (ii) (a statement) and (iii) (a question) were the target utterances for the elicitation purpose. The role of the participant in this game was either the inquisitor who started with the questions (i) and (iii) or the respondent who provided answers (ii) and (iv). During the game these roles were altered between the examiner and the participant. The elicitation began after a practice session (for familiarization), where two sets of utterances with identical discourse exchange (i.e., following (i)–(iv)) and with similar picture materials were used.

The utterances elicited from each participant were recorded onto a digital audio tape (DAT) recorder (Sony TCD-D100) through a Telex WT-700 microphone that was attached to the participant's clothing about 4.5 in below his/her mouth. These recordings were then extracted from the DATs, digitally sampled at a rate of 44,100 Hz, and stored as .wav format files in a 16-bit format using the Adobe Audition software program (version 1.5; Adobe Systems Inc., San Jose, CA). The long-term RMS amplitude of each utterance was normalized to maintain relatively constant sound levels across utterances. The resulting utterances were stored onto the hard drive of a laptop (Sony Vaio PCG-R505EL).

Utterance judgments • A total of 820 utterances were elicited from CI and NH participants (N = 500 from the 25 CI participants; N = 320 from the 16 NH participants). Each of these utterances was perceptually judged by a panel of eight adult listeners in terms of (a) utterance type, that is, if the utterance is a question or a statement, and (b) intonation contour appropriateness, that is, how appropriate the intonation contour of the utterance is based on the response in (a). The listener judged the utterance type in a two-alternative forced-choice format, and the contour appropriateness on a five-point rating scale, where "1 " was denoted to be completely inappropriate and "5 " to be absolutely appropriate. The listener was instructed to indicate, by clicking on a check box on the screen if s/he found an utterance being so noisy that it might affect the judgments s/he made. The utterances judged by the listener were excluded from analyses whenever this box was checked. This occurred in less than 0.88% of all utterances judged by the panel of listeners.

The utterances were presented to each adult listener via a loudspeaker (Tanny Reveal) at a comfortable listening level (60–70 dB SPL; long-term averaged level, A-weighting) in a double-walled sound-treated room. The presentation was divided into five blocks, each lasting approximately 45 to 60 min. All listeners completed the five blocks on 2 or 3 days within 2 wk. The utterances produced by CI versus NH participants were evenly assigned to the five blocks. The presentation order of blocks was randomized across all listeners, and was also randomized among the utterances within each block. A practice session where a small set of utterances similar to the target utterances of the same CI and NH participants were used to familiarize each of the eight adult listeners with the task format. Each listener made judgments for each utterance in the two aspects described above (i.e., utterance type and intonation contour appropriateness), using a custom software program. This program was developed using Microsoft Visual C^{++}, and permitted automatic playback of all utterances in random order and automatic recording of the listener's responses.

Perception Task

Speech materials • The speech stimuli for the perception task comprised 120 natural utterances, that is, 10 statements and 10 questions produced by each of six adult speakers (between 22 and 47 yrs of age; three per gender). These utterances were paired so that each pair had both interrogative and declarative forms that were syntactically matched (Appendix B). To increase the some variety in the syntactic structures of target sentences, we used sentences elicited using both scripts and pictorial materials. That is, five pairs of the target utterances (1s–5s and 1q–5q) were recorded from each adult speaker using the scripts, which involved the conversation between two children about a topic (e.g., the circus), developed by Allen and Arndorfer (2000). The other five pairs of utterances (6s–10s and 6q–10q) were elicited from each speaker using pictorial materials, using the same procedure in the intonation production task.

Before the recording, the speaker practiced using provided scripts and pictures as much as s/he desired. The speaker was instructed to read out the scripts loud and to produce the target utterances following the pictures as naturally as possible. If any utterance was produced differently from that on the script, s/he was instructed to repeat until the utterance was produced consistently with the targeted form (indicated in Appendix B).

The utterances from each speaker were recorded using a DAT recorder (Sony TCD-D100) through a stereo condenser microphone (Aiwa CM-TS22) attached to the speaker's clothing about 4.5 in below his/her mouth. The sessions were recorded in a double-walled sound-treated room. The recordings were then extracted from the tapes and digitally sampled at a rate of 44,100 Hz. The target utterances were stored as .wav format files in a 16-bit format using the Adobe Audition software program. The long-term RMS amplitude of each utterance was equalized to maintain relatively constant sound levels across utterances. All target utterances were then stored onto the hard drive of a laptop (Sony Vaio PCG-R505EL) for later computerized presentation.

Procedure • This task was performed in a quiet testing room (same as that for the production task). The order of the production and perception tasks was randomized among the participants in the CI and NH groups. Each participant practiced before the perception task, by listening to 12 questions and statements that were similar to the target utterances. The participant judged whether each utterance was a question ("someone asking") or a statement ("someone telling") in a two-alternative forced-choice task. Target utterances were presented in random order to each participant via a loudspeaker (Altec ACS41) at approximately 65 dB SPL (long-term averaged level; A-weighting, measured using a sound level meter, RadioShack) using computerized presentation. No feedback was provided during the task. Twenty CI participants listened to all 120 target utterances. The other five participants (CI-9, CI-10, CI-13, CI-14, and CI-17) attended to a half list of utterances ($N = 60$), which comprised only the even numbered utterances (i.e., 2s, 2q, 4s, 4q, 6s, 6q, 8s, 8q, 10s, and 10q in Appendix B). A half list was used because of either (i) limited appointment time (CI-14), or (ii) an original concern that the task with the use of a full list of utterances would tax the relatively limited attention span of children at a young age (CI-9, CI-10, CI-13, and CI-17).[†] The total amount of time for each participant to complete the task between 30 and 50 min, depending on whether a full list (N = 120) or a half list (N = 60) was used as well as the participant's response rate. Breaks were generously provided whenever needed. The responses were automatically recorded as a "question" or "statement" by the custom software program (developed using Microsoft Visual C^{++}). Identification accuracy was computed as percent correct.

RESULTS

Production Task

The utterance type accuracy (in % correct) was derived by averaging the judgments of each utterance by the panel of eight listeners. Note that among the panel of adult listeners, the utterance type judgment of the set of utterances from one listener was always significantly correlated with the judg-

---

[†] The use of a half list was later abandoned because an increasing number of stimuli would permit higher confidence levels for the participant's performance. The concern regarding limited attention span was taken care of by providing extended breaks to the participant whenever needed.

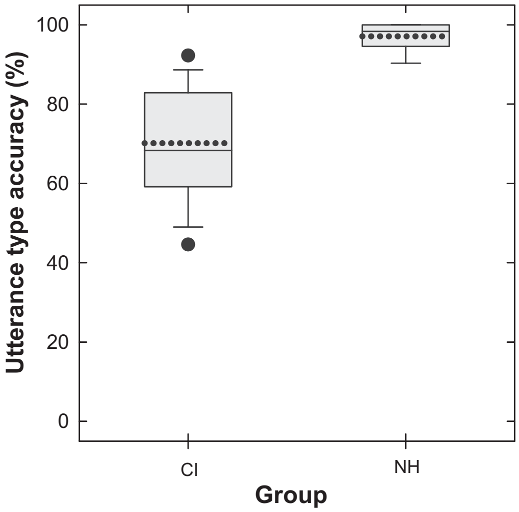

Fig. 1. Distributions of the overall utterance type accuracy in the production task for CI vs. NH groups. The x axis displays the CI vs. NH groups; the y axis displays the overall utterance type accuracy. The mean and median are displayed by the dotted and solid lines across each box, respectively. The upper and lower bounds of each box represent the quartiles, the whisker away from the box bounds showed the ±1.25 SD of the mean, and the filled circles represent the 5th and 95th percentiles bounds, if they are outside the end of whisker.

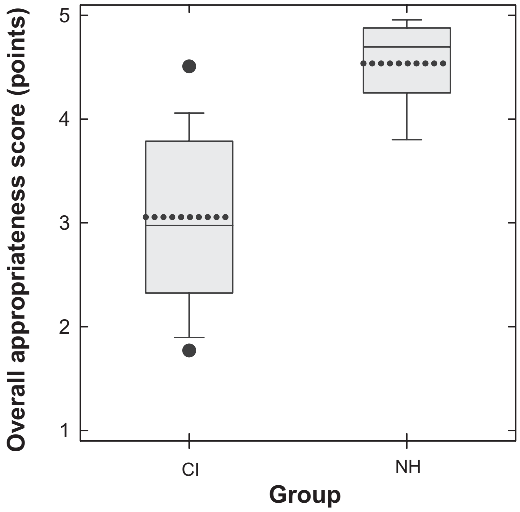

Fig. 2. Distributions of the overall contour appropriateness scores in the production task for CI vs. NH groups. The x axis displays the CI vs. NH groups; the y axis displays the overall contour appropriateness scores. The mean and median are displayed by the dotted and solid lines across each box, respectively. The upper and lower bounds of each box represent the quartiles, the whisker away from the box bounds showed the ±1.25 SD of the mean, and the filled circles represent the 5th and 95th percentiles bounds, if they are outside the end of whisker.

ment from another [Pearson correlation coefficients (r) ranging from 0.675 to 0.906; all $p < 0.001$]. Figure 1 illustrates the distributions of the overall utterance type accuracy for CI versus NH groups. The chance level for this task was 50%. The overall mean accuracy was 73.38% (SD = 19.60%) for the CI group and 97.31% (SD = 4.57%) for the NH group. The mean accuracy of the CI group was significantly lower than that of the NH group (Wilcoxon two-sample test statistic = 352.50, $p < 0.001$).

The five-point rating value was assigned for each utterance by each adult listener, and the contour appropriateness score for each participant was calculated by taking the average of the panel of eight listeners' judgments. A score of 1 point (i.e., lowest along the five-point scale) was assigned if the utterance type of any utterance was misjudged by the listener (i.e., incorrect utterance type). This is because our goal was to evaluate the participant's ability to appropriately produce a sentence in accordance with its utterance type (i.e., question or statement). Figure 2 displays the distributions of the overall contour appropriateness scores for CI versus NH groups. The overall mean score was 3.06 points (SD = 0.82 points) for the CI group and 4.52 points (SD = 0.45 points) for the NH group. The average score of the CI group was significantly lower than that of the NH group (Wilcoxon two-sample test statistic = 353.50, $p < 0.001$).

Among CI participants, some were fitted with the SPEAK speech-coding strategy (N = 14) and the others with the ACE strategy (N = 11). Similarly, some of these individuals were in a TC setting (N = 19) whereas the others were in an OC setting (N = 6). The utterance type accuracy and contour appropriateness scores were compared between the individuals in the SPEAK versus ACE subgroups, as well as between those in the TC versus OC subgroups. The utterance type accuracy was 73.53% (SD = 19.63%) for the SPEAK subgroup and 73.20% (SD = 20.51%) for the ACE subgroup. The contour appropriateness score was 3.15 points (SD = 0.83 points) for the SPEAK subgroup and 2.93 points (SD = 0.84 points) for the ACE subgroup. No significant difference was found in either the utterance type accuracy or the contour appropriateness score between the CI participants who were mapped with the SPEAK versus ACE speech-coding strategies [$t(23) = 0.040, p = 0.968$ for utterance type accuracy; $t(23) = 0.657, p = 0.518$ for contour appropriateness score].

The mean utterance type accuracy was 71.41% (SD = 19.94%) for the TC subgroup and 79.64% (SD = 18.70%) for the OC subgroup. The contour appropriateness score was 2.99 points (SD = 0.85 points) for the TC subgroup and 3.27 points (SD = 0.77 points) for the OC subgroup. No significant difference was observed in these two production scores of the participants in the TC versus OC settings [$t(23) = 0.894$, $p = 0.381$ for utterance type score; $t(23) = 0.732$, $p = 0.472$ for contour appropriateness score].

Taking all CI participants as a group, neither the utterance type accuracy nor the contour appropriateness score was found to be significantly correlated with the variables of age at implantation and length of device experience (r ranging from -0.275 to 0.189; all $p \geq 0.183$). However, both scores were negatively correlated with the variables of chronological age and age at implantation among those who were fitted with the ACE strategy (r ranging from -0.674 to -0.775; all $p \leq 0.023$). Note that the chronological age of individuals in the ACE subgroup was highly correlated with their age at implantation ($r = 0.879$, $p < 0.001$). Hence, the correlations between these individuals' production scores and age at implantation were reexamined, controlling for chronological age. The correlations were not found to be significant when controlling for age (utterance type: $r = 0.080$, $p = 0.825$; contour appropriateness: $r = 0.339$, $p = 0.338$). No significant correlations were found for those who were fitted with the SPEAK strategy, or for the TC or OC subgroups.

Perception Task

Data from the five CI participants who identified the half list of 120 utterances were first compared with the data of the remaining 20 CI participants who identified the full list of 120 utterances. The two data sets were combined for the subsequent analyses given that (i) the average identification accuracy of the two CI subgroups was not found to be statistically significant, and (ii) the responses of the odd and even numbered utterances were highly correlated in the 20 participants who judged the full set of 120 utterances (for details see Appendix C).

The identification accuracy (in % correct) of CI participants was first assessed to evaluate if these individuals' responses were biased to "question" or "statement" using the measurement B''_D, which estimated the participants' response bias (Donaldson, 1992). In this measure, B''_D is bounded by ± 1, and it equals zero if no bias response exists. To evaluate whether or not the CI participants had a dominant direction of bias response, a one-sample t test was performed to examine if the mean B''_D was significantly different from zero. The results indicated that as a group, the CI participants[‡] did not exhibit an overall dominant bias toward a certain response [mean difference = 0.04; $t(24) = 0.60$, $p = 0.556$]. Hence, original % scores, rather than adjusted scores of any sort were adopted for the comparison purpose.

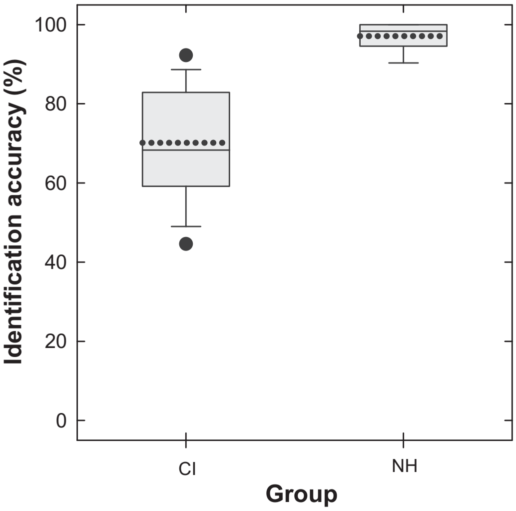

Fig. 3. Distributions of the overall identification accuracy in the perception task for CI vs. NH groups. The x axis displays the CI vs. NH groups; the y axis displays the overall accuracy. The mean and median are displayed by the dotted and solid lines across each box, respectively. The upper and lower bounds of each box represent the quartiles, the whisker away from the box bounds showed the ±1.25 SD of the mean, and the filled circles represent the 5th and 95th percentiles bounds, if they are outside the end of whisker.

Figure 3 displays the distributions of overall identification accuracy for CI versus NH groups. Average accuracy was 70.13% (SD = 14.46%) for the CI group and 97.11% (SD = 3.73%) for the NH group. The chance level for this task was 50%. Although CI participants' overall accuracy was above chance, the average accuracy of the CI group was significantly lower than that of the NH group (Wilcoxon two-sample test statistic = 572, $p < 0.001$).

As a group, the identification accuracy of CI participants was not found to be significantly correlated with the variable of age at implantation ($r = 0.176$, $p = 0.400$). However, the accuracy was positively correlated with chronological age and length of device experience. The overall identification accuracy was plotted for each CI participant as a function of chronological age in Figure 4a. The identification accuracy was also plotted for each NH participant to provide a reference. The partial Pearson correlation coefficient (r) was 0.505 between the overall accuracy of CI participants and their chronological age ($p = 0.012$; controlled for age at implantation), and was 0.656 between the overall accuracy of NH participants and their chronological age ($p = 0.004$). Figure 4b illustrates the overall identification of CI participants against their length

---

[‡] The bias response is evidently not a concern for the NH participants; these participants received high levels of identification accuracy (all were 88.33% and above).

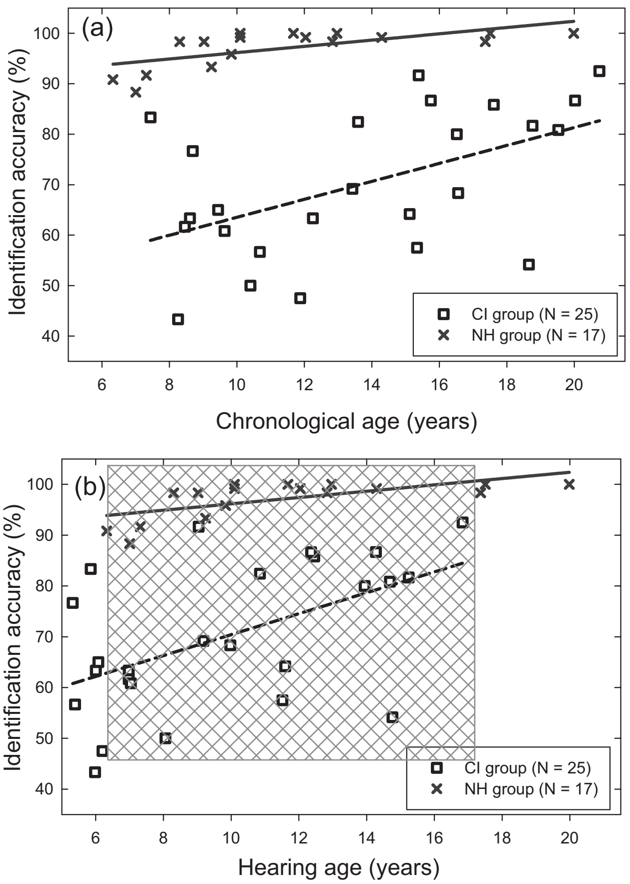

Fig. 4. Distributions of CI and NH participants' overall identification accuracy in the perception task, as a function of chronological age (Panel a), and hearing age (Panel b). On Panel a, the x axis displays the chronological age for all participants; on Panel b, the x axis displays the "hearing age" for participants (i.e., chronological age for NH participants; length of device experience for CI participants). On both panels, the y axis displays the overall identification accuracy. The scores of CI and NH participants are marked with open squares and x's, respectively. The dashed and solid lines display the best-fitted linear regression lines for CI and NH groups, respectively. In Panel b, data points within the coarsed region were further compared (see text for details).

of device experience ("hearing age"). The overall accuracy of NH participants was also plotted against their chronological age to provide a reference. The Pearson correlation coefficient (r) was 0.509 between the overall accuracy of CI participants and their length of device experience ($p = 0.011$; controlled for age at implantation).

These results indicated that pediatric CI recipients' identification accuracy of speech intonation contrasts was significantly lower than that of their NH peers. Both CI and NH participants demonstrated higher identification accuracy with an increasing chronological age. The identification accuracy of CI participants was positively associated with extended device experience. The overall accuracy of CI participants, as a group, remained much lower than that of their NH peers even when their length of device experience was approximately matched to NH participants' chronological age[§] (for CI group, mean = 73.19%, SD = 13.61%; for NH group, mean = 96.67%, SD = 3.98%; Wilcoxon two-sample test statistic = 351.5, $p < 0.001$).

The partial Pearson correlation coefficient between CI participants' identification accuracy and their age at implantation was not found to be statistically significant ($r = -0.114$, $p = 0.594$; controlled for chronological age). That is, the identification accuracy of CI participants was positively correlated with increasing device experience, but was not correlated with age at implantation.

The overall identification accuracy was 67.72% (SD = 13.38%) for the TC subgroup and 77.78% (SD = 16.34%) for the OC subgroup. The difference in the group mean accuracy was not found to be statistically significant between the CI users in TC versus OC settings [$t(23) = 1.527$, $p = 0.140$]. However, the identification accuracy was 77.26% (SD = 12.39%) for the SPEAK subgroup and 61.06% (SD = 11.86%) for the ACE subgroup. This difference in the group mean accuracy was found to be statistically significant [$t(23) = 3.305$, $p = 0.003$]. Note that the individuals in the SPEAK group, on average, had a significantly longer period of device experience than those in the ACE group [the SPEAK group: mean = 12.74 yrs, SD = 2.36 yrs; the ACE group: mean = 6.35 yrs, SD = 0.82 yrs; $t(24) = 8.56$, $p < 0.001$]. In addition, there was a statistically positive correlation between CI participants' overall identification accuracy and length of device experience.

Relationships Between Production and Perception

Twenty-four CI participants and 16 NH participants completed both production and perception tasks. The relationships between these individuals' performance in the production and perception of speech intonation contrasts were evaluated. Figure 5 illustrates the overall utterance type accuracy as a function of the overall identification accuracy for CI and NH participants. Similarly, Figure 6 illustrates

---

[§] The identification accuracy was compared between a subgroup of 18 CI participants and a subgroup of 14 NH participants. The data of CI participants whose length of device was less than 6.33 yrs (the youngest chronological age of NH participants) and the data of NH participants whose chronological age was more than 16.83 yrs (the longest device experience of CI participants) were excluded from this set of analysis. The matched ages of the CI and NH subgroups were not observed to be statistically different [CI group, mean =11.43, SD = 3.12; NH group, mean = 10.08, SD = 2.42; $t(30) = 1.341$, $p = 0.190$].

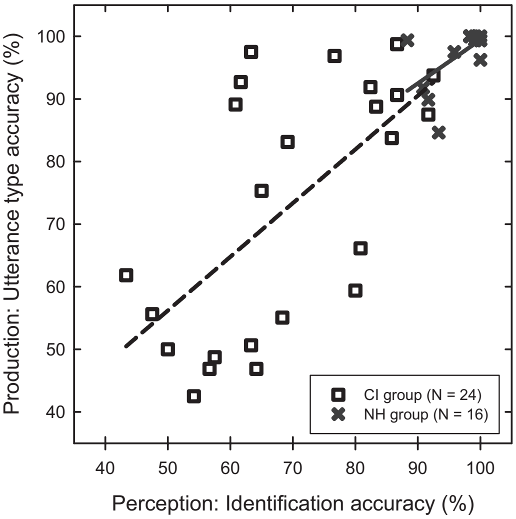

Fig. 5. Distributions of CI and NH participants' overall utterance type accuracy in the production task and overall identification accuracy in the perception task. The x and y axes display the overall identification accuracy and overall utterance type accuracy, respectively. The scores of CI and NH participants are marked with open squares and x's, respectively. The dashed and solid lines display the best-fitted linear regression lines for CI and NH groups, respectively.

the same CI and NH individuals' overall contour appropriateness scores as a function of the overall identification accuracy.

Pearson correlation coefficients (r) were computed between the production and perception scores for each group. The Pearson correlation coefficient (r) between the utterance type accuracy and identification accuracy was 0.627 ($p = 0.001$) and 0.575 ($p = 0.020$) for CI and NH groups, respectively. The Pearson correlation coefficient (r) between the contour appropriateness score and identification accuracy was 0.696 ($p < 0.001$) and 0.426 ($p = 0.100$) for the CI and NH groups, respectively.[¶] Taken together, CI participants' production scores (for both utterance type and contour appropriateness) were observed to be positively correlated with their overall identification accuracy. This correlation was statistically significant. There was a moderately positive correlation between the overall utterance type accuracy and overall identification accuracy in NH individuals. On the other hand, no statistically significant correlation was found between NH individuals' overall contour appropriateness scores and identification accuracy.

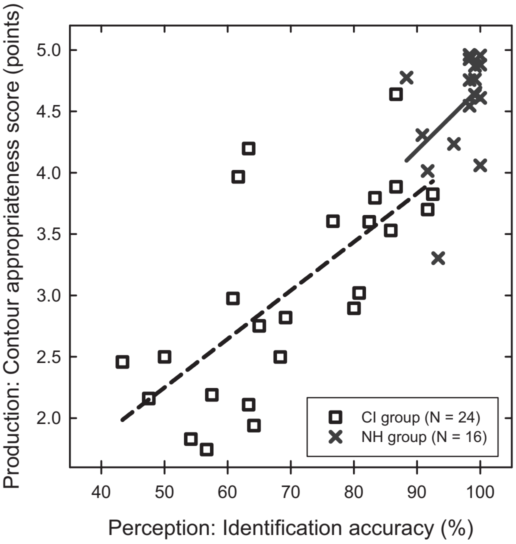

Fig. 6. Distributions of CI and NH participants' overall intonation contour appropriateness scores in the production task and overall identification accuracy in the perception task. The x and y axes display the identification accuracy and overall contour appropriateness scores, respectively. The scores of CI and NH participants are marked with open squares and x's, respectively. The dashed and solid lines display the best-fitted linear regression lines for CI and NH groups, respectively.

DISCUSSION

Production Task

According to the panel of adult listeners' judgments, the utterances of pediatric CI recipients were not only lower in utterance type accuracy, but also less appropriate in speech intonation contours than the utterances of their NH peers. Among CI individuals, no significant difference was found for the utterance type accuracy or contour appropriateness score between the individuals who were fitted with the SPEAK and ACE speech-coding strategies, or between those received their education in the TC and OC settings.

Children and young adults in the NH group achieved high levels of utterance type accuracy. This is not surprising, as infants and young children are capable of using intonation and other prosodic features of speech in their production for communicative or pragmatic purposes or in a grammatical manner (D'Odorico & Franco, 1991; Furrow, 1984; Galligan, 1987). For example, 4 to 8 mo old infants are capable of manipulating melodic patterns in their vocalization. They tend to produce relatively

---

[¶] These correlation coefficients and statistics reported here, as well as the data shown in Figure 6 were based on the correlation between (i) the identification accuracy and (ii) contour appropriateness scores that were adjusted for erroneous utterance type judgments (for details, please refer to the results reported for the production task). That is, when a score of one was not assigned, the Pearson correlation coefficient (r) between (i) and (ii) was 0.651 ($p = 0.001$) for the CI group (as opposed to 0.696, $p < 0.001$ when a score of one was assigned), and was 0.321 ($p = 0.226$) for the NH group (as opposed to 0.426, $p = 0.100$ when a score of one was assigned). The conclusion was consistent with either way of analyses.

high voice pitch when the vocalization demands some involvement by the caregiver (D'Odorico & Franco, 1991). Similarly, children demonstrate extensive grammatical use of intonation during the second year of life, and are able to vary the intonation of words to mark contrasts in meaning in accordance with the grammar of their ambient language (Galligan, 1987).

On the other hand, when the ranges of NH individuals' performance between utterance type accuracy (Fig. 1) and intonation contour appropriateness scores (Fig. 2) were compared, it was evident that the distribution of NH individuals' contour appropriateness scores tended to spread out more than that of utterance type accuracy. These results suggest that even though NH children and young adults may be able to produce accurate utterance types consistently, their speech intonation contours may not always be judged as highly appropriate. Because mastery of intonation and other prosodic components of speech is age-dependent (Loeb & Allen, 1993), the contour appropriateness scores of the NH participants were evaluated in relation to their chronological age. The results indicated that these individuals' contour appropriateness scores were not significantly correlated with their chronological age ($r = 0.409$, $p = 0.116$). The NH participants' contour appropriateness scores tended to be relatively narrow in range, which might contribute to the lack of a significant correlation between their scores and chronological age. Moreover, it is important to note that the youngest age of the participants in the NH group was 6.33 yrs. Because speech intonation tends to be acquired during the initial few years after birth, the lack of significant correlation might be associated with the fact that by the age of around six, the majority of NH children are able to produce relatively appropriate speech intonation contours. Verification of these speculations, however, requires additional evidence (e.g., obtaining longitudinal data from NH children at preschool ages), and is beyond the scope of the present study.

In the present study, many prelingually deafened children and young adults with a CI exhibited reduced utterance type accuracy and intonation contour appropriateness compared with their NH peers. Inadequate contours of speech intonation may coexist with reduced speech intelligibility (i.e., the extent to which a speaker can be understood), or affect how natural the speaker may sound (McGarr & Osberger, 1978; Parkhurst & Levitt, 1978). In hearing-impaired speakers, prosodic errors are highly correlated with low speech intelligibility (Hudgins & Numbers, 1942). Similarly, the production accuracy of speech prosodic components can be indicative of the speech intelligibility in children with hearing impairments (McGarr & Osberger, 1978). Based on these previous findings in the literature, reduced accuracy or appropriateness of pediatric CI recipients' speech intonation production may adversely affect their speech intelligibility. Future studies should address the effects of reduced utterance type accuracy and intonation contour appropriateness on pediatric CI recipient's speech intelligibility.

Findings in previous studies have suggested that pediatric CI users' speech production skills are related to their communication mode (TC versus OC). For example, Osberger et al. (1994) compared speech intelligibility in CI recipients in the OC versus TC settings based on adult listeners' perceptual judgments (Osberger, et al., 1994). In that study, both age at implantation and length of device experience were matched between the individuals in the OC and TC settings (all implanted by age five, with at least 2 yrs of CI experience). The authors indicated that with 3.5 yrs of device experience, the average intelligibility score for individuals in the OC setting was 48%, ranging from 14% to 93%, and for those in the TC setting was 21%, ranging from 4% to 59%.

There are several additional variables that may contribute to the substantial intersubject variability in prelingually deafened children's postimplant speech production performance, for example, age at implantation, length of device experience, and device-related variables such as advancements in speech-coding strategy (Fryauf-Bertschy, et al., 1997; Nikolopoulos, et al., 1999; Osberger & Fisher, 2000; Osberger, et al., 1994). Most authors agree that postimplant speech advancements are positively associated with a younger age at implantation, extended device experience, and reliance on oral communication.

On the other hand, findings in some other studies drew different conclusions regarding the effects of these variables on pediatric CI recipients' speech production performance. For example, Connor et al. (2000), reported that the speech (consonant) production skills did not differ between those in the TC and OC settings, as long as the implant users received a CI at no later than 5 yrs of age. Consistent with the findings of Connor et al., none of the variables in the present study were found to be associated with the production performance of prelingually deafened individuals who received a CI at an average age of 3.83 yrs (SD = 1.45). Note that in the present study, individuals who were fitted with the ACE speech-coding strategy were on average younger in age than those with the SPEAK strategy. Moreover, among those in the ACE subgroup, individuals who were younger in age at test time tended to receive a CI at a younger age. These confounding factors made it impossible to draw conclusions regarding the actual effects of these variables on pediatric CI recipients' performance in the production of speech intonation contrasts.

Perception Task

The overall identification accuracy of speech intonation contrasts was significantly above the chance level (50%) in approximately 80% of pediatric CI recipients. However, the overall mean accuracy of CI individuals was significantly poorer than that of their NH peers (70.13% for CI group; 76.00% for the CI participants whose performance levels were above the chance level, as opposed to 97.11% for NH group). That is, whereas NH children as young as 6 yrs of age are able to consistently identify speech intonation contrasts, prelingually deafened individuals who received a CI in their childhood are less capable of accurately identifying such contrasts.

The findings regarding NH individuals' high performance levels in the identification of speech intonation contrasts were consistent with the findings in the literature. The ability to perceive contrasts in prosodic components of speech emerges at a very young age in infants and children with NH (Hsu, et al., 2000; Jusczyk, et al., 1993; Marcos, 1987; Morgan & Saffran, 1995; Tonkova-Yampol'skaya, 1973). Unlike those with NH, children with severe–profound hearing impairments who did not receive a CI show significant difficulty with speech intonation perception (e.g., Most & Frank, 1994). The present findings indicated that with a CI, many prelingually deafened children and young adults may still exhibit difficulty with accurate identification of speech intonation contrasts.

The poor performance of the present CI participants in the perception task might be associated with multiple variables such as age at implantation, length of device experience, communication mode, and device-related variables. The findings of this investigation revealed no significant difference in the identification accuracy between pediatric CI recipients in the TC versus OC settings. However, the mean identification accuracy of the CI participants in the SPEAK subgroup was significantly higher than that of those in the ACE subgroup. Note that the individuals in the SPEAK group had significantly longer period of device experience than those in the ACE group. Because of this confounding factor, it remains unclear if the difference in overall identification accuracy in those mapped with different speech-coding strategies (SPEAK versus ACE) originates from the device-related factors, or length of device experience.

Age at implantation is considered to be an important variable in the postimplant development of many aspects of spoken language in prelingually deafened children (e.g., Miyamoto, et al., 1999; Osberger, et al., 2002). However, in this study, no direct relationships were found between pediatric CI recipients' identification accuracy and these individuals' age at implantation. In this study, all CI participants received an implant between 1.48 and 6.34 yrs of age. The present results suggest that accurate identification of speech intonation contrasts using suprasegmental information of speech can be challenging for prelingually deafened children including those who received a CI at as young as 1.5 yrs of age. Alternatively, children who received a CI at a relatively young age might ultimately develop mastery of speech intonation perception, but it would require greater amounts of device experience than they already had at test time. Given its cross-sectional nature, this study is constrained in permitting a full evaluation of this speculation; further investigations are required before conclusions can be reached.

On the other hand, the present findings indicated that the identification accuracy of both CI and NH individuals was positively correlated with their chronological age. That is, the ability to perceive speech intonation contrasts is age-dependent, and likely associated with the linguistic inputs these individuals exposed as their chronological age increases. As can be seen from the data in Figure 4b, when the "hearing age" (i.e., length of device experience) of CI participants was approximately matched to that (i.e., chronological age) of NH participants, the trend remained similar to that shown in Figure 4a. These results indicated that higher identification accuracy of pediatric CI recipients was associated with extended device experience. Noticeably, the overall identification accuracy of pediatric CI recipients remained much lower than that of NH individuals when CI participants' "hearing age" was matched to that of their NH peers.

Relationships Between Production and Perception

In studies of normal spoken language development, there is a weak link between the speech production and speech perception (Vihman, 1996). There are limitations when production and perception are examined in separate studies. For example, it is unclear if the limitation in the perception of certain speech contrasts leads to a failure for the individual to produce such contrasts, and vice versa. Similarly, the ability to perceive certain speech contrasts does not assure the individual's mastery of producing such contrasts. In this study, we assessed both production and perception of speech intonation contrasts in the same groups of CI and NH participants, and evaluated the relationships between these individuals' production and perception performance. As can be seen from the data in Figures 5 and 6, there was a moderately strong correlation between the overall utterance type accuracy and overall identification accuracy for both CI and NH participants, and between the overall contour appropriateness score and overall identification accuracy for CI participants.

The lack of a significant correlation between the overall utterance type accuracy (in the production task) and identification accuracy (in the perception task) in NH individuals was possibly related to the narrow range of data distribution of these individuals' overall identification accuracy. Although NH individuals demonstrated high overall identification accuracy (all above 88%; mean, 97.11%), their utterance type accuracy and contour appropriateness scores were quite different in distributions (Figs. 5 and 6). As a group, NH participants' utterance type accuracy was relatively high (all above 84%; mean, 97.31%). On the other hand, these individuals' contour appropriateness scores tended to spread out more in the data range (3.30–4.96; mean, 4.54).

Findings regarding the relationships between the production and perception performance in the NH participants suggest that although both utterance type accuracy and intonation contour appropriateness are indicative of a speakers' mastery of speech intonation production, they are somewhat different in nature. When utterance type accuracy is considered, perception and production of speech intonation develop in a parallel fashion. With regard to intonation contour appropriateness, perception may precede production: mastery of accurate identification of speech intonation contrasts is a required condition for the mastery of the production of appropriate intonation contours. That is, relatively speaking, the identification accuracy (in the perception task) for all NH children was quite high (~90–100%), yet their contour appropriateness scores (in the production task) tended to spread out more (see Fig. 6). This was not true when their utterance type scores were considered (see Fig. 5).

Among children with hearing impairments, those who are able to better perceive contrasts in speech intonation also tend to better able to produce intonation contrasts in their utterances (Most & Frank, 1994). Most and Frank investigated the perception and imitative production of intonation in children with severe–profound hearing impairments. A moderate, but significant correlation was found between these individuals' performance levels in the two tasks ($r = 0.58, p < 0.01$). The present findings suggest that the relationship between the perception and production of speech intonation in pediatric CI recipients is similar to that in hearing-impaired children who did not receive a CI (Most & Frank, 1994).

There are some limitations in the present investigation. For example, when a participant failed to perceive or produce speech intonation contrast, it is possible that this failure was not because of the limitation to the speech signal through a CI, but because of their limited linguistic knowledge regarding using speech intonation to contrast utterance types (question versus statement). Future studies should address the relationship between speech intonation (in terms of both production and perception) and linguistic knowledge (and/or language development) in pediatric CI recipients. Moreover, it is important to note that although F0 serves as an important source of acoustic information for speech intonation recognition, other acoustic dimensions such as intensity and duration cues can also contribute (Fry, 1955, 1958; Lehiste, 1970, 1976; Lieberman, 1967). The present study examined the production and perception of speech intonation of CI and NH individuals based on a totality of acoustic properties that are available in natural utterances. Future studies should address pediatric CI recipients' usage of various acoustic dimensions associated with speech intonation (i.e., F0, intensity, and duration patterns) in both production and perception.

CONCLUSIONS

The present findings indicated that, as a group, pediatric CI recipients demonstrate poorer scores in both production and perception of speech intonation contrasts than their NH peers. Moreover, the performance levels in the perception and production tasks are moderately interrelated in both CI and NH individuals. These results, along with the similar correlations found in hearing-impaired children who did not receive a CI (Most & Frank, 1994), collectively suggest that mastery of the perception of speech intonation contrasts is related to mastery of the production of such contrasts. Note that intersubject variability exists in pediatric CI recipients' production and perception performance. However, variables such as age at implantation, length of device experience, device-related factors, and communication mode can not fully account for the observed intersubject variability. Some of these CI users are able to achieve high performance levels in perception, but not in production (and vice versa). Hence, it is important to address both aspects, i.e., production and perception of speech intonation in prelingually deafened children and young adults who received a CI in their childhood. These findings have implications when pediatric CI recipients' acquisition of intonation and other prosodic aspects of speech are considered in the speech intervention and the aural (re)habilitation programs.

ACKNOWLEDGMENTS

The authors are grateful to all participants in this study, and Arthur Boothroyd, Kay Gfeller, and Sandie Bass-Ringdahl for their feedback on this study. Thanks are also due to Linda Spencer for assisting in data collection; Arik Wald for developing custom software programs; and Nelson Lu for providing advice in statistical analysis. They also thank all reviewers and editors for their constructive comments on the previous versions of this manuscript.

This work was funded by the NIH/NIDCD (P50 Grant DC00242) and the Department of Speech Pathology and Audiology at the University of Iowa.

Portions of this paper were presented at the 2005 American Auditory Society (AAS) Annual Meeting in Scottsdale, AZ and the 2005 Conference on Implantable Auditory Prostheses (CIAP) in Asilomar, CA.

This paper was based on the first author's Ph.D. dissertation submitted to the University of Iowa, Iowa City, IA.

Address for correspondence: Shu-Chen Peng, Ph.D., Department of Hearing and Speech Sciences, University of Maryland – College Park, 0100 LeFrak Hall, College Park, MD 20742. E-mail: speng@hesp.umd.edu.

Received October 27, 2006; accepted October 15, 2007.

Appendix I. A summary of target utterances produced by each participant

No.	Statement	No.	Question
1s	The girl is on the school bus.	1q	The girl is on the school bus?
2s	The girl is in the bedroom.	2q	The girl is in the bedroom?
3s	The dog is in the doghouse.	3q	The dog is in the doghouse?
4s	The mouse is in the trash can.	4q	The mouse is in the trash can?
5s	The fish is in the fish tank.	5q	The fish is in the fish tank?
6s	The dog likes the popcorn.	6q	The dog likes the popcorn?
7s	The sheep likes the mushrooms.	7q	The sheep likes the mushrooms?
8s	The boy likes the ice cream.	8q	The boy likes the ice cream?
9s	The dad likes the hot dog.	9q	The dad likes the hot dog?
10s	The boy likes the sandwich.	10q	The boy likes the sandwich?

Appendix II. Target statements and questions recorded from each of six adult speakers in the perception task

No.	Statement	No.	Question
1s	They all went to the circus.	1q	They all went to the circus?
2s	He didn't take their tickets.	2q	He didn't take their tickets?
3s	He rode a bike in circles.	3q	He rode a bike in circles?
4s	He took the ball from the tigers.	4q	He took the ball from the tigers?
5s	He gave the ball to the monkeys.	5q	He gave the ball to the monkeys?
6s	The boy likes the sandwich.	6q	The boy likes the sandwich?
7s	The cat is in the kitchen.	7q	The cat is in the kitchen?
8s	The girl is on the playground.	8q	The girl is on the playground?
9s	The mom likes the popcorn.	9q	The mom likes the popcorn?
10s	The mouse likes the pizza.	10q	The mouse likes the pizza?

Appendix III. A summary of each CI participant's overall identification accuracy of the even and odd numbered target utterances

Items ID	Even (N = 60)	Odd (N = 60)	Difference* (odd-even)	All (N = 120)
CI-1	78.33	81.67	3.33	80.00
CI-2	65.00	71.67	6.67	68.33
CI-3	90.00	95.00	5.00	92.50
CI-4	76.67	85.00	8.33	80.83
CI-5	88.33	95.00	6.67	91.67
CI-6	86.67	85.00	−1.67	85.83
CI-7	60.00	55.00	−5.00	57.50
CI-8	85.00	81.67	−3.33	83.33
CI-9	56.67	NA	NA	56.67
CI-10	63.33	NA	NA	63.33
CI-11	68.33	70.00	1.67	69.17
CI-12	63.33	63.33	0.00	63.33
CI-13	76.67	NA	NA	76.67

(*Continued*)

Appendix III. Continued

Items ID	Even (N = 60)	Odd (N = 60)	Difference* (odd-even)	All (N = 120)
CI-14	86.67	NA	NA	86.67
CI-15	83.33	90.00	6.67	86.67
CI-16	78.33	86.67	8.33	82.50
CI-17	65.00	NA	NA	65.00
CI-19	61.67	61.67	0.00	61.67
CI-20	51.67	43.33	-8.33	47.50
CI-21	70.00	58.33	-11.67	64.17
CI-22	46.67	40.00	-6.67	43.33
CI-23	78.33	85.00	6.67	81.67
CI-24	46.67	53.33	6.67	50.00
CI-25	51.67	70.00	18.33	60.83
CI-26	58.33	50.00	-8.33	54.17
Mean	69.47	71.08	1.67	70.13
SD	13.50	17.20	7.39	14.46

** The difference refers to that between the accuracy for the even and odd numbered utterances identified by those who performed the perception task with a full list (N = 20). The group difference was not found to be statistically significant (t(23) = 0.080, p = .938). The Pearson correlation coefficient (r) between these 20 CI participants' accuracy with the even and odd numbered utterances was considerably high (r = .907, p < .001). "NA" = Not available.*

REFERENCES

Allen, G. D., & Arndorfer, P. M. (2000). Production of sentence-final intonation contours by hearing-impaired children. *J Speech Lang Hear Res*, *43*, 441–455.

American Speech-Language-Hearing Association (ASHA). (2003). Technical report: cochlear implants. *ASHA Suppl 24.*

Barry, J. G., Blamey, P. J., Martin, L. F. A., et al. (2002). Tone discrimination in Cantonese-speaking children using a cochlear implant. *Clin Linguist Phon*, *16*, 79–99.

Blamey, P. J., Barry, J. G., & Jacq, P. (2001). Phonetic inventory development in young cochlear implant users 6 years postoperation. *J Speech Lang Hear Res*, *44*, 73–79.

Burns, E. M., & Viemester, N. F. (1976). Nonspectral pitch. *J Acoust Soc Am*, *60*, 863–869.

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

Ciocca, V., Francis, A. L., Aisha, R., et al. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *J Acoust Soc Am*, *111*, 2250–2256.

Connor, C. M., Hieber, S., Arts, H. A., et al. (2000). Speech, vocabulary, and the education of children using cochlear implants: oral or total communication? *J Speech Lang Hear Res*, *43*, 1185–1204.

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: Cambridge University Press.

D'Odorico, L., & Franco, F. (1991). Selective production of vocalization types in different communication contexts. *J Child Lang*, *18*, 475–499.

Donaldson, W. (1992). Measuring recognition memory. *J Exp Psychol Gen*, *121*, 275–277.

Edwards, M. L. (1974). Perception and production in child phonology: the testing of four hypotheses. *J Child Lang*, *1*, 205–219.

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. *J Acoust Soc Am*, *108*, 1877–1887.

Ferguson, C. A. (1978). Learning to pronounce: the earliest stages of phonological development in the child. In F. D. Minifie, & L. L. Lloyd (Eds.), *Communicative and cognitive abilities—early behavioral assessment* (pp. 273–297). Baltimore: University Park Press.

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, PA: W. B. Saunders.

Friesen, L. M., Shannon, R. V., Baskent, D., et al. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*, *110*, 1150–1163.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *J Acoust Soc Am*, *27*, 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Lang Speech*, *1*, 126–152.

Fryauf-Bertschy, H., Tyler, R. S., Kelsay, D. M. R., et al. (1997). Cochlear implant use by prelingually deafened children: the influences of age at implant and length of device use. *J Speech Lang Hear Res*, *40*, 183–199.

Fu, Q. J., Zeng, F. G., Shannon, R. V., et al. (1998). Importance of tonal envelope cues in Chinese speech recognition. *J Acoust Soc Am*, *104*, 505–510.

Furrow, D. (1984). Young children's use of prosody. *J Child Lang*, *11*, 203–213.

Galligan, R. (1987). Intonation with single words: purposive and grammatical use. *J Child Lang*, *14*, 1–21.

Geurts, L., & Wouters, J. (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am*, *109*, 713–726.

Green, T., Faulkner, A., & Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *J Acoust Soc Am*, *112*, 2155–2164.

Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am*, *116*, 2298–2310.

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

Hsu, H. C., Fogel, A., & Cooper, R. B. (2000). Infant vocal development during the first 6 months: speech quality and melodic complexity. *Infant Child Dev*, *9*, 1–16.

Hudgins, C. V., & Numbers, F. C. (1942). An investigation of the intelligibility of the speech of the deaf. *Genet Psychol Monogr*, *25*, 289–392.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Dev*, *64*, 675–687.

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

Ladefoged, P. (2001). *A course in phonetics* (4th ed.), Orlando: Harcourt Brace.

Lee, K. Y. S., van Hasselt, C. A., Chiu, S. N., et al. (2002). Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *Int J Pediatr Otorhinolaryngol*, *63*, 137–147.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge: 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.

Lieberman, P. (1967). *Intonation, perception, and language.* Cambridge: The MIT Press.

Loeb, D. F., & Allen, G. D. (1993). Preschoolers' imitation of intonation contours. *J Speech Hear Res*, *36*, 4–13.

Loizou, P. C. (1998). Introduction to cochlear implants. *IEEE Eng Med Biol Mag*, *18*, 32–42.

Marcos, H. (1987). Communicative functions of pitch range and pitch direction in infants. *J Child Lang*, *14*, 255–268.

McGarr, N. S., & Osberger, M. J. (1978). Pitch deviancy and intelligibility of deaf speech. *J Comm Disord*, *11*, 237–247.

Miyamoto, R. T., Kirk, K. I., Svirsky, M. A., et al. (1999). Communication skills in pediatric cochlear implant recipients. *Acta Otolaryngol*, *119*, 219–224.

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: Blackwell Publishers.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in speech segmentation. *Lang Cogn Process*, *11*, 69–106.

Most, T., & Frank, Y. (1994). The effects of age and hearing loss on tasks of perception and production of intonation. *Volta Rev*, *96*, 137–149.

Nikolopoulos, T. P., O'Donoghue, G. M., & Archbold, S. (1999). Age at implantation: its importance in pediatric cochlear implantation. *Laryngoscope*, *109*, 595–599.

O'Halpin, R. (2001). Intonation issues in the speech of hearing impaired children: analysis, transcription and remediation. *Clin Linguist Phon*, *15*, 529–550.

Osberger, M. J., & Fisher, L. (2000). Preoperative predictors of postoperative implant performance in children. *Ann Otol Rhinol Laryngol*, *185,* 44S–46S.

Osberger, M. J., Miyamoto, R. T., Zimmerman-Phillips, S., et al. (1991a). Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear Hear*, *12*, 66S–80S.

Osberger, M. J., Robbins, A. M., Miyamoto, R. T., et al. (1991b). Speech perception abilities of children with cochlear implants, tactile aids, or hearing aids. *Am J Otol*, *12,* 105S–115S.

Osberger, M. J., Robbins, A. M., Todd, S. L., et al. (1994). Speech-intelligibility of children with cochlear implants. *Volta Rev*, *96*, 169–180.

Osberger, M. J., Zimmerman-Phillips, S., & Koch, D. B. (2002). Cochlear implant candidacy and performance trends in children. *Ann Otol Rhinol Laryngol*, *189,* 62S–65S.

Parkhurst, B. G., & Levitt, H. (1978). The effect of selected prosodic errors on the intelligibility of deaf speech. *J Commun Disord*, *11*, 249–256.

Peng, S., Tomblin, J. B., Cheung, H., et al. (2004). Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear Hear*, *25*, 251–264.

Peng, S., Tomblin, J. B., Spencer, L. J., et al. (2007). Imitative production of rising speech intonation in pediatric cochlear implant recipients. *J Speech Lang Hear Res*, *50*, 1210–1227.

Rosen, S. (1989). *Temporal information in speech and its relevance for cochlear implants.* Paper presented at the cochlear implant: acquisitions and controversies, Toulouse, France. June 9–10, 1989.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci*, *336*, 367–373.

Shannon, R. V. (1983). Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hear Res*, *11*, 157–189.

Shih, C. -L. (1988). Tone and intonation in Mandarin. *Working papers of the Cornell Phonetics Laboratory,* 3, 83–109.

Spencer, L. J., Tye-Murray, N., & Tomblin, J. B. (1998). The production of English inflectional morphology, speech production and listening performance in children with cochlear implants. *Ear Hear*, *19*, 310–318.

Svirsky, M. A., & Chin, S. B. (2000). Speech production. In S. B. Waltzman, & N. L. Cohen (Eds.), *Cochlear implants* (pp. 293–309). New York: Thieme Medical Publishers.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., et al. (2000). Language development in profoundly deaf children with cochlear implants. *Psychol Sci*, *11*, 153–158.

Tobey, E. A., Angelette, S., Murchison, C., et al. (1991). Speech production performance in children with multichannel cochlear implants. *Am J Otol*, 12:165S–173S.

Tobey, E. A., & Hasenstab, M. S. (1991). Effects of a Nucleus multichannel cochlear implant upon speech production in children. *Ear Hear*, 12:48S–54S.

Tonkova-Yampol'skaya, R. V. (1973). Development of speech intonation in infants during the first two years of life. In D. I. Slobin (Ed.), *Studies of child language development* (pp. 128–138). New York: Holt Rinehart and Winston.

Tye-Murray, N., Spencer, L., & Woodworth, G. G. (1995). Acquisition of speech by children who have prolonged cochlear implant experience. *J Speech Hear Res*, *38*, 327–337.

Velleman, S. L. (1988). The role of linguistic perception in later phonological development. *Appl Psycholinguist*, *9*, 221–236.

Vihman, M. M. (1996). Phonological development: the origins of language in the child. Cambridge: Blackwell Publishers.

Wei, W. I., Wong, R., Hui, Y., et al. (2000). Chinese tonal language rehabilitation following cochlear implantation in children. *Acta Otolaryngol*, *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.

Wilson, B. S. (2004). Engineering design of cochlear implants. In F.- G. Zeng, A. N. Popper, & R. R. Fay (Eds.), *Cochlear implants: auditory prostheses and electric hearing* (pp. 14–52). New York: Springer-Verlag.

Wilson, B. S., Finley, C. C., Lawson, D. T., et al. (1988). Speech processors for cochlear prostheses. *IEEE Eng Med Biol Mag*, *76*, 1143–1154.