Research Article

Processing of Acoustic Cues in Lexical-Tone Identification by Pediatric Cochlear-Implant Recipients

Shu-Chen Peng,^a Hui-Ping Lu,^b Nelson Lu,^a Yung-Song Lin,^{b,c}
Mickael L. D. Deroche,^d and Monita Chatterjee^e

Purpose: The objective was to investigate acoustic cue processing in lexical-tone recognition by pediatric cochlear-implant (CI) recipients who are native Mandarin speakers.

Method: Lexical-tone recognition was assessed in pediatric CI recipients and listeners with normal hearing (NH) in 2 tasks. In Task 1, participants identified naturally uttered words that were contrastive in lexical tones. For Task 2, a disyllabic word (*yanjing*) was manipulated orthogonally, varying in fundamental-frequency (F0) contours and duration patterns. Participants identified each token with the second syllable *jing* pronounced with Tone 1 (a high level tone) as eyes or with Tone 4 (a high falling tone) as *eyeglasses*.

Results: CI participants' recognition accuracy was significantly lower than NH listeners' in Task 1. In Task 2, CI participants' reliance on F0 contours was significantly less than that of NH listeners; their reliance on duration patterns, however, was significantly higher than that of NH listeners. Both CI and NH listeners' performance in Task 1 was significantly correlated with their reliance on F0 contours in Task 2.

Conclusion: For pediatric CI recipients, lexical-tone recognition using naturally uttered words is primarily related to their reliance on F0 contours, although duration patterns may be used as an additional cue.

Lexical tones convey linguistic meaning in tonal languages such as Mandarin Chinese and Cantonese. In listeners with normal hearing (NH), variation in fundamental frequency (F0) provides the primary source of acoustic cues for lexical-tone recognition, although additional cues, such as intensity or duration characteristics, may provide additional sources of information (e.g., S. Liu & Samuel, 2004; Whalen & Xu, 1992; S. Zhu, Wong, & Chen, 2014). In these languages, the F0 inflections conveying lexical tones are relatively brief, occurring over a few hundred milliseconds within the syllabic duration. In the literature, the average duration of naturally spoken lexical tones in Mandarin is reported to vary widely, from 200–300 ms (e.g., Tao et al., 2015; J. Wang, 1994) to much

longer (e.g., 600–800 ms, reported by S. Wang, Liu, Zhang, et al., 2012). Nonetheless, it is reasonable to assume rapid fluctuations in lexical tones, occurring within the course of single syllables. The actual duration depends on their position in the sentence and the stress pattern of the utterance. The rapid F0 variation occurs within the syllables of the spoken word to encode its meaning, whereas the slower F0 changes, as in English, occur across words or phrases or at the sentence level, and contribute to communicative mood and other prosodic information. Listeners attending to Mandarin Chinese or other tonal languages must simultaneously analyze these inflections at both levels to decode the semantic meaning and the communicative intent or mood of the speaker. This places even greater informational emphasis on F0 processing in tonal languages (Chao, 1968; Howie, 1976; P. Liu & Pell, 2012).

Previous studies have shown that cochlear implants (CIs) do not transmit F0 cues with sufficient spectral resolution, resulting in deficits in speech-intonation recognition as well as melody recognition, voice-emotion recognition, speaker and gender identification, and listening in competing backgrounds (Chatterjee et al., 2015; Chatterjee & Peng, 2008; Cullington & Zeng, 2008; Fu, Chinchilla, & Galvin, 2004; Galvin, Fu, & Nogaki, 2007; Gfeller et al., 2007; Kong, Stickney, & Zeng, 2005; Peng, Lu, & Chatterjee,

^aCenter for Devices and Radiological Health, United States Food and Drug Administration, Silver Spring, MD

^bChi-Mei Medical Center, Tainan, Taiwan

^cTaipei Medical University, Taiwan

^dMcGill University, Montréal, Québec, Canada

^eBoys Town National Research Hospital, Omaha, NE

Correspondence to Monita Chatterjee: monita.chatterjee@boystown.org

Editor: Julie Liss

Associate Editor: Megan McAuliffe

Received February 5, 2016

Revision received July 19, 2016

Accepted October 27, 2016

https://doi.org/10.1044/2016_JSLHR-S-16-0048

Disclosure: The authors have declared that no competing interests existed at the time of publication.

2009; Peng, Tomblin, & Turner, 2008; Qin & Oxenham, 2003). It is well-known that CI listeners also show significant deficits in lexical-tone recognition (e.g., Han et al., 2009; Lee & van Hasselt, 2005; Luo, Fu, Wei, & Cao, 2008; Peng, Tomblin, Cheung, Lin, & Wang, 2004; S. Wang, Liu, Zhang, et al., 2012; Wu & Yang, 2003; Zhou, Huang, Chen, & Xu, 2013; M. Zhu et al., 2011). However, relatively little is known about how these individuals utilize acoustic cues in lexical-tone recognition. Using equal-duration chimeric stimuli composed of temporal fine-structure and temporal envelope cues derived from the different lexical tones, Xu and Pfingst (2003) showed that temporal fine structure dominates lexical-tone recognition in adult NH listeners. Luo and Fu (2004) demonstrated that NH listeners, attending to CI-simulated speech in which the temporal fine structure is replaced by noise, can recognize lexical tones better with a manipulation in which the temporal envelope closely follows the F0 contour. These findings suggest that the temporal envelope might serve as an additional cue for lexical-tone recognition.

A large proportion of CI recipients in China and Taiwan are children with prelingual deafness (i.e., born deaf or becoming deaf before age 5 years). These individuals receive a CI in their early childhood (as young as age 12 months) and develop mastery of a spoken language with a CI device in place. It would be of great interest to discover whether the relatively heavy demands of a tonal language can drive children developing in that environment to be more adept at processing F0 changes at the syllabic level (lexical tones) in addition to the level of phrases and sentences (prosody). In recent work, Deroche, Lu, Limb, Lin, and Chatterjee (2014) tested the hypothesis that linguistic environments might influence children's perceptual F0 resolution. Using broadband harmonic complexes and a psychophysical three-interval forced-choice F0 discrimination task to obtain psychometric functions, they reported no differences in performance measures between children living in the United States or in Taiwan. In both cases, children with CIs showed significantly poorer F0 discrimination than children with NH.

To develop mastery in a tonal language, CI recipients need to learn to accommodate to the deficit in F0 representation in their auditory system so that they can distinguish and identify lexical tones in everyday speech. This may be partially achieved by redundancy in speech, which allows many aspects of speech information to be successfully reconstructed by CI listeners from the spectro-temporally degraded signal. In previous work, we have reported that in a question-statement identification task, English-speaking CI users may partially compensate for the deficit in voice-pitch information by relying instead on intensity or duration changes in the signal that covary with the F0 cue (Peng et al., 2009; Peng, Chatterjee, & Lu, 2012). Recent studies on phonetic discrimination also show evidence for such altered weighting under various listening and viewing conditions (Winn, Chatterjee, & Idsardi, 2012, 2013). These studies have demonstrated cue-weighting changes for specific phonemic contrasts in speech. Previous studies on phoneme recognition

and speaker-gender recognition, however, indicate that NH and CI listeners both rely more strongly on temporal envelope cues when spectral cues are degraded (Fu et al., 2004; Xu, Thompson, & Pfingst, 2005). As a group, these studies suggest that adult listeners shift their listening strategies under altered acoustic conditions without training. Pediatric CI recipients with prelingual deafness have less experience with sound perception in general than adult CI recipients. It remains unclear whether these individuals would show similar changes in cue weighting as adult CI listeners, and if so, then to what extent their length of CI experience would play a role in their cue-weighting patterns for prosodic aspects of speech recognition, such as lexical tones.

The purpose of the present study was to examine pediatric CI recipients' ability to identify lexical-tone contrasts, in relation to their reliance on the acoustic cues that may contribute to this identification. Twenty-eight CI recipients with prelingual deafness participated, ranging in age from 6 to 21 years. They all received a CI before age 5 years. In addition, 11 individuals with NH were recruited, ranging in age from 4 to 15 years. Each participant performed two lexical-tone identification tasks, one to determine accuracy of lexical-tone identification and the other to determine reliance on acoustic variations that may contribute to lexical-tone recognition.

The first task involved lexical-tone recognition using a corpus of naturally uttered words. In Mandarin, lexical tones 1 through 4 are distinguished primarily by variations in F0 height and F0 contour. To be specific, Tones 1 and 4 are, acoustically speaking, relatively simple: Tone 1 is represented by a relatively high, flat F0 contour, whereas Tone 4 is represented by a sharply falling F0 contour and is markedly shorter in duration. In contrast, Tones 2 and 3 are more complex. Tone 2 is represented by a slowly rising F0 contour, whereas Tone 3 (as spoken in Taiwan) contains a slowly falling dipping F0 contour, typically with a small rise toward the end of a syllable. Furthermore, F0 contours for Tones 2 and 3 tend to start at a somewhat low F0 value relative to Tones 1 and 4 (e.g., Fon, Chiang, & Cheung, 2004). Previous studies have reported that Tones 1 and 4 are relatively easy for CI users to identify and Tones 2 and 3 are most difficult to distinguish (e.g., Chen, Wong, Chen, & Xi, 2014; Tao et al., 2015; W. Wang, Zhou, & Xu, 2011).

The covarying duration cues for these tones, although they do not vary consistently in naturally uttered speech, have been shown to contribute to lexical-tone recognition (e.g., Lin & Repp, 1989). In previous studies of tone recognition, Xu and colleagues have deliberately equalized the durations of the stimuli to ensure that listeners did not attend to duration as a confounding cue (e.g., Xu & Pfingst, 2003; Xu, Tsai, & Pfingst, 2002). Across conditions of vocoding intended to simulate hearing with CIs, Xu et al. (2002) reported significant reductions in lexical-tone recognition when they equalized tone durations (mean reduction across lexical tones was 15.1 percentage points, with the greatest reductions observed in the recognition of Tones 3 and 4,

which are respectively the longest and shortest tones in Mandarin speech). This underscores the contribution of the duration cue to lexical-tone recognition.

In the present study, however, one of the specific objectives was to examine the potential role of duration as an additional acoustic cue. Although syllabic and vocalic durations are variable and difficult to track in natural lexical-tone recognition, we believe their contribution may be more observable under relatively controlled conditions in which F0 and duration are explicitly and systematically altered. In the second task, therefore, we examined pediatric CI recipients' relative reliance on the F0 variation and duration cues in a Tone 1 versus Tone 4 identification task using carefully controlled, resynthesized stimuli. Given that F0 variation is the primary cue and is readily accessible by NH listeners, we hypothesized that NH listeners would show strong reliance on the F0 variation and less reliance on the duration cue. Because the CI allows severely limited access to F0 information but retains duration cues, we hypothesized that CI listeners would demonstrate less reliance on F0 variation and greater reliance on the duration cue. We further hypothesized that age at implantation and length of device experience would positively correlate with CI listeners' overall performance in lexical-tone recognition and with their reliance on the duration cue.

The study was also designed to examine relationships between results obtained by the same listeners in the two tasks. We hypothesized, for instance, that CI participants' reliance on the F0 contour cue, as revealed in Task 2, would be predictive of their overall lexical-tone recognition performance in Task 1, on the basis that lexical-tone recognition performance depends heavily on the quality of F0 information. In addition, we hypothesized that CI listeners' reliance on the duration cue in Task 2 might predict their performance in Task 1—but perhaps differentially for Tone 4 (the shortest tone) versus Tones 1, 2, and 3, consistent with findings from Peng et al. (2004). Lexical-tone recognition using a nonprimary cue (i.e., duration patterns) might come at a cost, given that it does not appear useful or consistent between certain lexical-tone pairs, such as Tones 2 and 3. On the other hand, a greater reliance on duration would presumably lead to better perceptual separation between Tone 4 and the remaining lexical tones.

Chatterjee and Peng (2008) showed a significant relation between English-speaking adult CI listeners' psycho-physical sensitivity to an F0-related temporal envelope cue and their reliance on the F0 contour in speech-intonation recognition. Examining the same relation here with Mandarin-speaking pediatric CI recipients was made possible because 24 of the 27 CI listeners in the present study had also participated in a previously published study measuring F0 discrimination in harmonic complexes (Deroche et al., 2014). In that study, listeners' sensitivity to F0 differences in broadband harmonic complexes was quantified by deriving thresholds and slopes of psychometric functions. Reference F0 was either 100 or 200 Hz. Most of the participants with CIs were able to complete the task with the 100-Hz F0 reference but found the task more difficult when a 200-Hz

F0 reference was used. As the data for the present study were obtained at the same time as the data for that study, we additionally tested the hypothesis that F0 discrimination measured in the previous study is related to performance in the lexical-tone recognition tasks of the present study.

Method

Participants

Two groups of participants were recruited for this study. The CI group comprised 28 CI recipients, ranging in age from 6.6 to 21.4 years ($M = 12.4$ years). All CI participants had prelingual deafness (i.e., onset of deafness at age 5 years or younger). They received an implant at a mean age of 2.9 years, ranging from 1.6 to 4.8 years. The mean length of their device experience was 9.5 years, ranging from 2.1 to 15.4 years. Table 1 displays CI participants' background information. Of the 28 CI participants, one (CM36) completed both Tasks 1 and 2. His data from Task 1, however, were corrupted and thus could not be included in analysis.

In addition, Tasks 1 and 2 were administered to 11 participants with NH, ranging in age from 4.6 to 15.6 years ($M = 10.3$ years). The NH participants were recruited so that their range of chronological range would match CI participants' range of "hearing age" ($M = 9.5$ years), to account for the fact that the CI participants in this study did not receive an implant until age 1.6 to 4.8 years ($M = 2.9$ years). The hearing sensitivity of all NH listeners was within normal limits (thresholds better than 20 dB HL) at all octave intervals from 250 to 8000 Hz, bilaterally. All CI and NH participants provided written informed consent, which was reviewed and approved by the institutional review board at the Chi Mei Medical Center.

Stimuli

Test stimuli consisted of the naturally produced stimuli (Task 1) and resynthesized stimuli (Task 2). The natural stimuli were recorded using 10 monosyllabic words; each was naturally produced with four different lexical tones (Tones 1 through 4) by three male and three female talkers (total number of words = 240). These 10 monosyllables were the following: *ba*, *ma*, *da*, *tao*, *ke*, *hui*, *qi*, *xie*, *zhu*, and *shu*. All recorded words, normalized for their long-term root-mean-square, were identified by each participant in Task 1. Time waveforms and spectrograms (with the F0 contours overlaid) of recordings of the word *ba* uttered in the four tones by one male and one female talker are shown in Figure 1. The stimuli used in Task 2 comprised 96 tokens, which were generated from a single naturally produced, disyllabic stimulus *yanjing* via resynthesis using the Praat software (Boersma & Weenink, 2004), which uses the Pitch Synchronous Overlap Add algorithm to resynthesize stimuli. In the resynthesis process, the disyllabic stimulus *yanjing* was kept fixed for its first syllable *yan* as the first part of Tone 3, and the second syllable *jing* was varied in F0 and duration properties orthogonally to vary between Tone 1 (which

Table 1. Background information of participants with cochlear implants.

Participant	Gender	Age at testing (years)	Device experience (years)	Age at implantation (years)	Device model	Speech-coding strategy
CG01	F	12.0	8.7	3.2	Nucleus 24CS	ACE
CM01	F	18.5	16.2	2.3	Nucleus 22M	SPEAK
CG02	F	6.4	2.1	4.3	Nucleus 24 Freedom	ACE
CG03	F	10.9	9.3	1.6	Nucleus 24CS	ACE
CG04	F	10.0	6.2	3.8	Nucleus 24CS	ACE
CG05	F	11.7	8.3	3.4	Nucleus 24CS	ACE
CG06	F	12.7	10.1	2.5	Nucleus 24M	ACE
CG07	F	12.1	8.9	3.2	Nucleus 24CS	ACE
CG08	F	7.1	4.6	2.5	Nucleus 24 Freedom	ACE
CM08	M	12.9	9.9	3.0	Nucleus 24M	ACE
CG09	M	11.5	9.9	1.6	Nucleus 24M	ACE
CM09	F	21.3	16.0	5.3	Nucleus 22M	SPEAK
CG10	F	13.2	10.8	2.4	Nucleus 24M	ACE
CG11	M	9.8	7.5	2.3	Nucleus 24CS	ACE
CG12	F	11.0	8.4	2.6	Nucleus 24CS	ACE
CM14	M	19.6	14.8	4.8	Nucleus 22M	ACE
CM26	M	15.7	13.9	1.9	Nucleus 24M	ACE
CM27	M	17.5	14.7	2.8	Medel Pulsar (reimplanted)	FSP
CM28	F	11.7	10.8	0.9	Nucleus 24M	ACE
CM29	M	21.4	17.0	4.4	Nucleus 22M	SPEAK
CM30	F	16.2	13.1	3.1	M-C40+	CIS
CM31	F	11.8	9.9	1.9	Nucleus 24M	ACE
CM32	M	11.5	9.4	2.1	M-C40+	CIS
TZ33	F	9.0	6.7	2.3	Hi Rates 90K	HiRes-P/Fidelity 120
TZ34	F	7.5	4.2	3.3	M-Sonata	FSP
TZ35	M	6.7	2.9	3.8	M-Sonata	FSP
CM36	M	19.7	15.1	4.6	M-C40+	CIS
CG37	M	6.6	4.3	2.3	Nucleus 24 Freedom	ACE

represents the lexical meaning *eyes*) and Tone 4 (which represents the lexical meaning *eyeglasses*). These two lexical tones were chosen primarily because they do not involve complex contour changes as with Tones 2 and 3; further, the vocalic portion of Tone 4 is typically shorter in duration relative to those of the other three lexical tones. Their intrinsic properties make it possible to permit manipulations in both F0 and duration properties for the purpose of this study. The process resulted in two, eight, and six steps of variations in the acoustic dimensions—that is, F0 height, F0 contour, and duration ratio, respectively. The entire stimulus set consisted of 96 tokens (1 disyllabic stimulus $\times$ 2 F0 heights $\times$ 8 F0 variations $\times$ 6 duration ratios). Figure 2 illustrates the systematic manipulation in each acoustic dimension. Note that the manipulation of F0 cues occurs at the level of the acoustic signal, and does not explicitly alter the nature of F0 information available to listeners (e.g., temporal envelope periodicity or harmonic structure) in any way. Thus, NH and CI participants listened to the resynthesized stimuli using the same cues that are typically available to them.

Procedure

Two tasks, using the naturally uttered stimuli (Task 1) and resynthesized stimuli (Task 2) described previously, were administered to each CI and NH participant using Superlab 4.0 for experimental control. In Task 1, each stimulus was presented once to the participant using a single-interval

four-alternative forced-choice paradigm. The participant was instructed to identify which of four words was just perceived, with an identical syllabic structure (i.e., same consonants and vowels) but different lexical tones, by selecting with the computer mouse one of the four response choices with corresponding Chinese characters and phonetic notation (*zhuyin fuhao*, the phonetic notation system used in Taiwan) displayed on the screen. Children in kindergarten are taught to recognize all phonetic notations in *zhuyin fuhao* as part of preparation for starting elementary school (typically before age 6 years). *Zhuyin fuhao* is formally taught during the first 3 months of first-grade education in Taiwan, and children are expected to fully master the system. In Task 2, each stimulus was presented three or four times to the participant using a single-interval two-alternative forced-choice paradigm. The participant was instructed to indicate whether the stimulus referred to eyes (*jing* pronounced with Tone 1, a high level tone) or eyeglasses (*jing* pronounced with Tone 4, a high falling tone) by selecting with the computer mouse one of the two response choices displayed on the screen. In both tasks, stimuli were delivered from a laptop computer with an external sound card (Soundmax Integrated HD Audio) at an approximate level of 65 dB SPL at the listener's ear (the listener was seated approximately one meter from the Audio Pro loudspeakers). In both tasks, the participant was familiarized with the task paradigms, using a small set of words (for Task 1) and resynthesized tokens (for Task 2) that were similar to those adopted in this study. They were randomly presented to the participant prior to each task.

Figure 1. Examples of amplitude waveforms, spectrograms, and pitch contours (yellow lines overlaid on the spectrograms) of the word *ba* spoken by a female talker (upper panels) and a male talker (lower panels) using the four lexical tones (left to right panels). F0 = fundamental frequency.

No feedback in terms of response accuracy was provided to the participant during task familiarization. Both tasks were conducted at three sites: Chi Mei Medical Center in Tainan, Taiwan; Chang Kung Memorial Hospital in Taoyuan, Taiwan; and Dalin Tzu Chi Hospital in Jiayi, Taiwan. At all study sites, both tasks were performed using the same equipment in a sound-treated booth. Note that in both tasks, the CI participants listened through their implant (monaurally), whereas the NH participants performed the tasks via binaural presentation. This was to reflect CI and NH participants' real-world listening experience, and is not expected to influence the results. The total amount of time for each participant to complete both tasks was

approximately 60 min, not including breaks. Breaks were generously provided as needed.

Scoring and Data Analysis

For Task 1, the responses were recorded as Tone 1, Tone 2, Tone 3, or Tone 4. One point was given when the response matched the target lexical tone. Identification accuracy for the 240 items was computed as percent correct. For Task 2, in order to derive psychometric functions, a score of 1 or 0 was assigned to responses of Tone 1 and Tone 4, respectively. The proportion of Tone 1 responses was calculated for each participant, and this proportion

Figure 2. The systematic manipulation in each acoustic dimension, using Praat Version 4.3 (Boersma & Weenink, 2004). One disyllabic word, *yanjing*—in which *yan* was fixed with Tone 3 and *jing* was varied in fundamental frequency (F0) and duration properties to represent lexical tone contrasts for Tone 1 (meaning eyes) and Tone 4 (meaning eyeglasses)—was resynthesized to generate two, eight, and six steps of variations, respectively, in (a) F0 height, (b) F0 contour, and (c) duration ratio. The entire stimulus set comprised 96 tokens (1 disyllabic stimuli $\times$ 2 F0 heights $\times$ 8 F0 variations $\times$ 6 duration ratios). As a reference, the naturally uttered tokens are most closely represented with the F0 variation values of 0 and -1.0 octave (i.e., corresponding with Tone 1 and Tone 4, respectively), and with the duration ratio values of 0.8 – 1.2 and 0.6 – 0.8 (i.e., corresponding with Tone 1 and Tone 4, respectively). For detailed acoustic depictions of lexical tones in Mandarin Chinese, please refer to Fon et al. (2004), Fon and Chiang (1999), and Tao et al. (2015).

was included as the dependent variable in simple logistic-regression models adopted to estimate the regression coefficients for each participant. Two variables, F0 contour and log duration ratio, were included as the independent variables. One additional independent variable—F0 height—was included as a random variable in the models. In these models, the estimated coefficient for each parameter was used to approximate the participant's reliance on changes in F0 contour and duration ratio in lexical-tone identification. Of note, the interaction terms of the main effects were investigated, and the analysis results indicated that the estimated coefficients would not be altered much by incorporating these interaction terms in the model. Therefore, only the results from the model with the main effect are reported in this study. Correlation analyses were conducted on the data from individual participants in the CI and NH groups, respectively. Logistic models were thus used to estimate coefficients of the acoustic parameters F0 contour and log duration ratio using the method of generalized estimating equations.

The Wald test was used to examine if the estimated coefficient was different from zero. Statistical analyses were performed using Proc Genmod in SAS. Hereafter, we refer to the parameter of log duration ratio as the *duration ratio* or *duration pattern*.

Methods Used by Deroche et al. (2014)

The reader is referred to Deroche et al. (2014) for details of the experimental methods. In brief, F0 discrimination

was measured in a three-alternative forced-choice procedure, using the method of constant stimuli. Stimuli were broadband harmonic complexes with reference F0s of 100 or 200 Hz. Psychometric functions were analyzed using a maximum-likelihood procedure (Wichmann & Hill, 2001). The researchers found that results obtained for the CI listeners using the 100- and 200-Hz reference stimuli were significantly correlated with each other. The majority of the participants with CIs were able to complete the task with the 100-Hz F0 reference; the task with the 200-Hz reference was more difficult, and fewer participants with CIs were able to complete it. To minimize the number of comparisons and maximize the number of available samples, slopes and thresholds of the psychophysical functions obtained with only the 100-Hz F0 reference were used in the present analyses. The data for the Deroche et al. (2014) study and the present study were obtained in the same period of time.

Results

Accuracy and Confusion Matrices for Lexical-Tone Recognition in Task 1

Table 2 displays the identification accuracy of lexical tones by CI participants ($N = 27$). The identification accuracy by NH participants ($N = 11$) is also listed for reference. Among CI participants, the mean identification score was 57.5% ($SD = 13.3$). The mean scores for lexical tones

Table 2. Identification accuracy (%) of lexical tones by participants with cochlear implants (CI) and normal hearing (NH).

Tone	CI (<i>N</i> = 27)		TH (<i>N</i> = 11)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	69.3	25.2	97.0	2.7
2	44.5	14.9	92.0	5.3
3	38.2	14.9	87.7	2.9
4	78.0	18.6	98.2	1.7
Overall	57.5	13.3	93.7	1.7

1, 2, 3, and 4 were, respectively, 69.3% (*SD* = 25.2), 44.5% (*SD* = 14.9), 38.2% (*SD* = 14.9), and 78.0% (*SD* = 18.6). A repeated-measures analysis of variance, calculated by means of SAS Proc Mixed, confirmed a significant effect of lexical tone, $F(3, 27) = 16.43$, $p < .001$. Identification accuracy was significantly higher for Tones 1 and 4 than for Tones 2 and 3, as revealed by post hoc Tukey adjusted *t* tests ($ps < .001$ for Tone 1 vs. Tone 2, Tone 1 vs. Tone 3, Tone 4 vs. Tone 2, and Tone 4 vs. Tone 3). All mean scores were significantly higher than the 25% chance level (all $ps < .001$).

Figure 3a illustrates the distribution of individual CI participants' identification accuracy for each lexical tone, with individual NH participants' performance plotted in Figure 4 for reference. The identification accuracy of all individual tones varied strongly among the CI participants. In contrast, the NH participants' identification accuracy was more homogeneous across all four tones. In addition, CI participants achieved lower scores with Tones 2 and 3 than with Tones 1 and 4. Only 70.4% (*N* = 19) and 40.7% (*N* = 11) of the 27 participants' accuracy in Tone 2 and Tone 3 identification, respectively, were significantly above chance (25%). On the other hand, 85.2% (*N* = 23) and 100% (*N* = 27) of the 27 CI participants' Tone 1 and Tone 4 identification accuracy, respectively, were significantly above chance. A response is determined as significantly

above chance when a subject correctly identified at least 22 out of 60 tokens (36.67%), on the basis of the binomial exact test at one-sided $\alpha = .025$.

Table 3 displays a confusion matrix constructed from results obtained for CI participants (*N* = 27). Among CI participants, the error patterns were as follows: Tone 1 was mainly misidentified as Tone 2 (55.8% of the errors), Tone 2 was mainly misidentified as Tone 3 (60.9% of the errors), and Tone 3 was mainly misidentified as Tone 2 (38.2% of the errors) and Tone 4 (49.0% of the errors). Last, Tone 4, which was identified with the fewest errors, was misidentified as Tones 1, 2, and 3 with similar probabilities (all less than 10%). In contrast—and consistent with expectations—the error rate for NH participants was fairly low for individual lexical-tone identification (all less than 8%), as shown in Table 4. Among the low error rates for NH listeners, the error rate was somewhat higher for the identification of Tones 2 and 3 (about 7% on average).

Contribution of Individual Tones to Overall Accuracy in Lexical-Tone Recognition in Task 1

It is of interest to consider the potential contribution of individual lexical tones to overall lexical-tone recognition for the two groups of listeners. Table 5 displays the Pearson correlation coefficients (*r*s) between the identification accuracy for the overall score and individual lexical tones for the CI group. For CI participants, the overall accuracy of lexical-tone identification was significantly associated with the accuracy of Tone 1 ($r = .90$, $p < .0001$), Tone 2 ($r = .69$, $p < .0001$), and Tone 4 ($r = .80$, $p < .0001$), but not with that of Tone 3 ($r = .35$, $p = .069$). The Pearson correlation coefficients are also shown for the NH group (*N* = 11) in Table 6 for reference. For NH participants, the overall accuracy of lexical-tone identification was only significantly associated with the accuracy of Tone 2 ($r = .80$, $p = .003$).

Figure 3. Identification accuracy of individual participants with (a) cochlear implants and (b) normal hearing for individual lexical tones.

Figure 4. The proportion of *yanjing* stimuli identified as Tone 1, plotted as a function of the changes in acoustic properties. (a) Variation in fundamental frequency; (b) duration ratio. CI = cochlear implant; NH = normal hearing.

Performance in Task 2: Use of Acoustic Cues

Figure 4 displays the proportion of *yanjing* stimuli identified as Tone 1, plotted as a function of the step increase in F0 variation (see Figure 4a) and the duration ratio (see Figure 4b). A steeper slope of these functions results in a larger best-fitting coefficient in the repeated-measures logistic-regression analysis and indicates greater reliance on the variation in a specific acoustic property of interest. The estimated coefficient for F0 variation was 2.29 for CI participants and 16.22 for NH participants. The coefficient for F0 variation of CI participants was statistically significantly smaller than that of NH participants ($p < .0001$). On the other hand, CI participants' estimated coefficient for the duration parameter was 2.59, whereas NH participants' was 1.31. The coefficient for duration of CI participants was significantly larger than that of NH participants ($p = .002$). Of the 27 CI participants, 20 (74.1%) used F0 variation as a cue in identifying Tone 1 versus Tone 4 contrast to some extent ($p < .025$) for the coefficients for F0 variation—that is, slopes significantly larger than zero on the basis of the logistic-regression analysis. All 11 NH participants (100%) used F0 variation as the dominant cue. As for duration ratio, 26 CI participants (96.3%)

versus eight NH participants (72.7%) used this cue in identifying the Tone 1 versus Tone 4 contrast. Nonetheless, the reliance on the duration parameter was more variable among CI participants as compared with NH participants.

Relation Between Performance in the Two Tasks

In preliminary graphing, CI and NH listeners' accuracy (Task 1) and their coefficient of F0 variation (Task 2) were found to be related (see Figure 5a), but the data were also fitted after a log transform of the absolute value of the coefficient of F0 variation, which helped homogenize the variance across the CI subjects (see Figure 5b). All values of the coefficient for F0 variation were negative with the exception of one CI subject (CM29), for whom the value was slightly positive (0.18): This subject's data were excluded for the analysis in Figure 5b. The correlations were significant for the CI group for both the raw data (see Figure 5a; $r = .66$, $p < .001$) and log-transformed data (see Figure 5b; $r = .75$, $p < .0001$). For the NH group, only the correlation obtained with the log-transformed data was significant ($r = .64$, $p = .033$). To be more specific, among CI participants the correlations were significant between the coefficient

Table 3. Confusion matrix for participants with cochlear implants ($N = 27$).

Stimulus	Response (%)			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	69.3	17.2	9.3	4.3
Tone 2	18.0	44.5	33.8	3.8
Tone 3	7.8	23.6	38.2	30.3
Tone 4	6.9	6.9	8.2	78.0
Overall accuracy	57.5			

Note. Boldface indicates correct (accurate) responses.

Table 4. Confusion matrix for participants with normal hearing ($N = 11$).

Stimulus	Response (%)			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	97.0	2.4	0.5	0.2
Tone 2	1.8	92.0	6.2	0.0
Tone 3	0.3	7.4	87.7	4.5
Tone 4	0.3	0.3	1.2	98.2
Overall accuracy	93.7			

Note. Boldface indicates correct (accurate) responses.

Table 5. Pearson correlation coefficients for the group with cochlear implants ($N = 27$).

	Overall	Tone 1	Tone 2	Tone 3	Tone 4
Overall	—				
Tone 1	.90**	—			
Tone 2	.69**	.54*	—		
Tone 3	.35	.08	.10	—	
Tone 4	.80**	.72**	.36	.03	—

Note. Probability $> |r|$ under H_0 : $p = 0$.

* $p < .05$. ** $p < .0001$.

for F0 variation and their identification of Tones 1 ($r = .78$, $p < .001$), 2 ($r = .54$, $p = .016$), and 4 ($r = .58$, $p = .008$), but not Tone 3 ($r = .15$, $p = .46$). Note that the Bonferroni correction was applied to the p values in all cases. The correlations between the identification accuracy and the other acoustic parameter (duration ratio) were also examined. None of the correlations was found to be statistically significant, for either the CI or NH groups.

Relations Between Age-Related Factors and F0 Sensitivity Measures and Outcomes in Tasks 1 and 2

No significant correlations were observed between the chronological age of CI participants or duration of experience with CIs and any of the perceptual measures (average lexical-tone recognition accuracy or cue-weighting patterns) among CI participants. In contrast, age of implantation was significantly correlated with the overall accuracy in Task 1 ($r^2 = .28$, $p = .005$; see Figure 6a). In Task 2, age of implantation was significantly correlated with the coefficient for log duration ratio ($r^2 = .33$, $p = .002$; see Figure 6b), but not with the coefficients for F0 variation or F0 height (Bonferroni correction applied, $\alpha = .017$). Note that the correlation with the coefficient for log duration ratio remained significant even after eliminating one extreme data point ($r^2 = .24$, $p = .010$; see Figure 6b, brown squares and solid line). Among NH participants, chronological age was not significantly correlated with overall tone recognition in Task 1. In Task 2, chronological age was strongly correlated with the coefficient of F0 variation ($r^2 = .72$, $p < .001$; see Figure 6c) but not with the coefficients

Table 6. Pearson correlation coefficients for the group with normal hearing ($N = 11$).

	Overall	Tone 1	Tone 2	Tone 3	Tone 4
Overall	—				
Tone 1	.25	—			
Tone 2	.80*	-.07	—		
Tone 3	.52	-.26	.19	—	
Tone 4	.28	.07	-.12	.20	—

Note. Probability $> |r|$ under H_0 : $p = 0$.

* $p < .05$.

obtained with the remaining acoustic cues. Last, no significant correlations were found between lexical-tone recognition performance in Task 1, or participants' reliance on F0 variation or F0 height in Task 2, and the same participants' F0 discrimination thresholds or slopes obtained in the Deroche et al. (2014) study.

Discussion

The objectives of the present study were to investigate acoustic-cue processing in lexical-tone recognition by pediatric CI recipients who speak Mandarin Chinese. In Task 1, lexical-tone recognition was assessed using a set of naturally uttered words. In Task 2, we determined the participants' acoustic cue-weighting patterns in a Tone 1 vs. Tone 4 identification task, in which F0 and duration patterns of a resynthesized word (*yanjing*) were systematically manipulated. As expected, the present CI listeners showed a significant deficit in lexical-tone recognition in Task 1, with large individual variation in both overall performance and error patterns. Their poorer performance was similar to that reported in previous studies with CI patients (Han et al., 2009; Luo et al., 2008; Luo, Fu, Wu, & Hsu, 2009; S. Wang, Liu, Dong, et al., 2012; W. Wang et al., 2011; Zhou et al., 2013). However, these individuals' performance levels in the present study were poorer than those reported by Tao et al. (2015) and Chen et al. (2014). Because the different studies used different stimuli and had different subject characteristics, and some of them controlled for duration whereas others did not, it is not easy to compare performance directly. Regardless, all studies have consistently shown significant deficits among pediatric CI recipients relative to NH listeners.

The results from Task 1 showed that Tones 2 and 3 were most difficult for both NH and CI listeners to identify, although NH participants made relatively few errors overall and achieved near-ceiling performance with all four lexical tones. For CI participants, the frequency of errors was high: Tone 2 was most frequently confused with Tones 3 and 1, and Tone 3 was most frequently confused with Tones 2 and 4. Recall that Tones 2 and 3 both tend to have a long dipping F0 contour, making them more easily confused. Similar confusions were reported in native Mandarin speakers from China in a younger cohort of CI children by Chen et al. (2014) and in adult CI users tested by W. Wang et al. (2011). Also, as reported by Whalen and Xu (1992), Mandarin-speaking listeners with NH may benefit from amplitude cues in lexical-tone recognition. Results from those experiments revealed that NH listeners are able to differentiate Tones 2, 3, and 4 on the basis of amplitude cue alone (i.e., with both F0 and duration cues removed). Together, these findings are consistent with the notion that amplitude cues intrinsic to these lexical tones might be utilized by Mandarin-speaking listeners (with NH or through a CI) in lexical-tone recognition, especially when other cues are obscured or unavailable to them.

Correlation analyses in our study showed that for NH listeners, performance with Tone 2 was significantly predictive of overall performance. In contrast, for CI listeners

Figure 5. (a) Overall Task 1 identification accuracy of individual participants with cochlear implants (CI) and normal hearing (NH) for lexical tones as a function of their coefficients for variation in fundamental frequency (F0) in Task 2. The data points are shown as squares and circles for individual CI and NH participants, respectively. The solid and dotted lines represent the regression line best fitted to the CI and NH groups, respectively. (b) The same data (excluding one data point; see text), with the abscissa showing log-transformed absolute values of the coefficients of F0 variation in Task 2.

overall performance was significantly correlated with recognition of Tones 1 and 4, although a weaker, yet statistically significant, relationship was observed with their recognition of Tone 2. In Task 2, a set of highly controlled stimuli was adopted to examine cue weighting by CI and NH listeners in lexical-tone recognition. As hypothesized, CI listeners showed stronger reliance on the duration pattern and weaker reliance on the F0 variation cue than NH listeners. These results underscore differences in the success of listening strategies used by the two populations. Thus, auditory training of CI participants in lexical-tone recognition would need to take such findings into consideration. For example, rehabilitation could emphasize certain lexical-tone contrasts (e.g., Tone 2 vs. Tone 3).

The results obtained in Task 2 also suggest that children with CIs could exhibit better lexical-tone recognition in the presence of salient duration patterns. This suggests that targeted training—for example, with a focus on clear speech with salient cues secondary to F0 variation—might benefit this population in lexical-tone recognition.

Among CI participants, the coefficient associated with the duration pattern in Task 2 was not significantly correlated with the overall accuracy in lexical-tone recognition in Task 1. However, the coefficient of F0 variation was significantly correlated with overall accuracy in lexical-tone recognition as well as with individual recognition of Tones 1, 2, and 4 (the three tones that also correlated most strongly with overall tone recognition). Taken together, these results demonstrate the importance of the F0 contour in tone recognition—even though the cue is greatly degraded in CIs, the listeners' reliance on the F0 contour remains a dominant predictor of performance with natural speech. In Task 2, with highly controlled stimuli, pediatric CI recipients clearly relied upon the additional duration parameter in identifying contrasts between Tones 1 and 4.

Age at implantation was a significant predictor of performance in Task 1. In Task 2, CI listeners' weighting of the duration pattern was also significantly correlated with age of implantation. For NH listeners, the weighting of their dominant cue (F0 variation) was significantly correlated with age.

Neither chronological age nor length of device experience was predictive of performance by CI listeners in this study. This suggests that, contrary to our hypothesis, experience with their device in their developmental years postimplantation does not play as important a role in shaping pediatric CI recipients' ability to detect lexical-tone contrasts. On the other hand, earlier age of implantation is presumably associated with a more plastic auditory system in the initial stages of adaptation to the device, and results in better tone-recognition performance (Task 1) and greater reliance on the duration pattern (Task 2). The coefficient associated with the duration pattern, however, was not significantly correlated with the CI group's performance in Task 1. Thus, although CI recipients are able to use duration cues in the Tone 1 versus Tone 4 identification task when these cues are made explicitly available, they may not be helpful in natural speech recognition in the absence of targeted rehabilitation.

No significant correlations were observed between the present results and F0 discrimination measured by Deroche et al. (2014) in the previous study with the same listeners. The F0 discrimination thresholds were obtained with static-F0 stimuli, whereas tone recognition requires the listener to attend to dynamic changes in F0. It is possible that listeners' sensitivity to dynamic F0 cues might be more predictive of performance in tone-recognition tasks. In a recent study, however, Tao et al. (2015) found no significant relation between CI listeners' performance in a musical contour-identification (MCI) task and lexical-tone recognition. The

Figure 6. (a) Average tone recognition (% correct score) in Task 1, plotted against age at implantation for individual participants with cochlear implants (CI). The solid line shows the regression. (b) Coefficient of log duration ratio in Task 2, plotted against CI participants' age at implantation. The brown squares and solid line represent the correlation obtained without including the single outlier. (c) Coefficient of variation in fundamental frequency (F0) in Task 2, plotted against chronological age of individual participants with normal hearing.

F0 changes in the MCI task used by those authors occur across, rather than within, successive notes in the musical contours. It is unknown to what extent CI listeners' sensitivity to more rapid changes in F0 might relate to their ability to recognize lexical-tone contrasts. Further, performance on the MCI task by the CI listeners in that study was very poor, whereas performance in the lexical-tone recognition task by the same listeners was very good, suggesting that very different cues were used in the two tasks. On the other hand,

W. Wang et al. (2011) and S. Wang, Liu, Dong, et al. (2012) found significant correlations between adult CI users' performance in discrimination tasks with pitch, musical interval, and melody and their lexical-tone recognition. It is unclear whether the differences in these findings have to do with different methodologies or task difficulty or with the ages and audiological histories (children vs. adults; prelingual vs. postlingual) of the listener groups. It is possible that musical-interval discrimination requires higher-level processing of F0 information than either the simple F0 discrimination task used here or the MCI task of Tao et al. (2015), and that such higher-level F0 processing is similarly involved in lexical-tone perception.

Conclusion

Consistent with previous findings, the results of Task 1 in the present study showed that pediatric CI recipients with prelingual deafness who speak Mandarin Chinese have specific difficulties in word recognition on the basis of contrasts in lexical tones. The more controlled stimuli in Task 2 were used to investigate the possible use of covarying F0 and duration cues. The results showed that pediatric CI recipients placed significantly greater weight on duration patterns than their age-matched NH peers, whereas the latter group placed significantly greater weight on F0 variation cues. Performance in natural word recognition on the basis of lexical-tone contrasts among CI listeners was significantly correlated with their reliance on the F0 contour but not their reliance on the duration pattern in Task 2. Thus, the CI listeners' reliance on the duration cue may have been only a form of perceptual compensation within the constraints of the experimental paradigm of Task 2, and, as the results showed, is unrelated to their performance in tone recognition in natural settings. However, the fact that they were able to use the cue suggests that they have some implicit knowledge of its role as a secondary cue. Perhaps their experience with the degraded input of the CI serves to provide some form of natural training. It is possible that in the future some form of explicit training that capitalizes on the CI users' ability to use the secondary cue will be beneficial.

Acknowledgments

This work was supported by National Institutes of Health Grants R01 DC004786-08S1, R21 DC011905, and R01 DC014233 (awarded to Monita Chatterjee) and Chi-Mei Research Foundation and Taiwan National Science Council Grant 100-2314-B-038-037 (awarded to Yung-Song Lin). We are grateful to the research participants for their support of this work.

References

- Boersma, P., & Weenink, D. (2004). Praat (Version 4.3) [Computer software]. Amsterdam, the Netherlands: Institute of Phonetic Sciences, University of Amsterdam.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

- Chatterjee, M., & Peng, S.-C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, 235, 143–156.
- Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., . . . Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research*, 322, 151–162.
- Chen, Y., Wong, L. L. N., Chen, F., & Xi, X. (2014). Tone and sentence perception in young Mandarin-speaking children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 78, 1923–1930.
- Cullington, H. E., & Zeng, F.-G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *The Journal of the Acoustical Society of America*, 123, 450–461.
- Deroche, M. L. D., Lu, H.-P., Limb, C. J., Lin, Y.-S., & Chatterjee, M. (2014). Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. *Frontiers in Neuroscience*, 8, 282. <https://doi.org/10.3389/fnins.2014.00282>
- Fon, J., & Chiang, W.-Y. (1999). What does Chao have to say about tones?—A case study of Taiwan Mandarin. *Journal of Chinese Linguistics*, 27(1), 13–37.
- Fon, J., Chiang, W.-Y., & Cheung, H. (2004). Production and perception of the two dipping tones (Tone 2 and Tone 3) in Taiwan Mandarin. *Journal of Chinese Linguistics*, 32, 249–281.
- Fu, Q.-J., Chinchilla, S., & Galvin, J. J., III. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 5, 253–260.
- Galvin, J. J., III, Fu, Q.-J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear and Hearing*, 28, 302–319.
- Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., & Olszewski, C. (2007). Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing*, 28, 412–423.
- Han, D., Liu, B., Zhou, N., Chen, X., Kong, Y., Liu, H., . . . Xu, L. (2009). Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users. *Ear and Hearing*, 30, 169–177.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, United Kingdom: Cambridge University Press.
- Kong, Y.-Y., Stickney, G. S., & Zeng, F.-G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *The Journal of the Acoustical Society of America*, 117, 1351–1361.
- Lee, K. Y. S., & van Hasselt, C. A. (2005). Spoken word recognition in children with cochlear implants: A five-year study on speakers of a tonal language. *Ear and Hearing*, 26(Suppl. 4), 30S–37S.
- Lin, H.-B., & Repp, B. H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, 32, 25–44.
- Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavioral Research Methods*, 44, 1042–1051.
- Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, 47, 109–138.
- Luo, X., & Fu, Q.-J. (2004). Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, 116, 3659–3667.
- Luo, X., Fu, Q.-J., Wei, C.-G., & Cao, K.-L. (2008). Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. *Ear and Hearing*, 29, 957–970.
- Luo, X., Fu, Q.-J., Wu, H.-P., & Hsu, C.-J. (2009). Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hearing Research*, 256, 75–84.
- Peng, S.-C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends in Hearing*, 16, 67–82.
- Peng, S.-C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology and Neurotology*, 14, 327–337.
- Peng, S.-C., Tomblin, J. B., Cheung, H., Lin, Y.-S., & Wang, L.-S. (2004). Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear and Hearing*, 25, 251–264.
- Peng, S.-C., Tomblin, J. B., & Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear and Hearing*, 29, 336–351.
- Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, 114, 446–454.
- Tao, D., Deng, R., Jiang, Y., Galvin, J. J., III, Fu, Q.-J., & Chen, B. (2015). Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users. *Ear and Hearing*, 36, 102–110.
- Wang, J. (1994). Syllable duration in Mandarin. *Proceedings of the Fifth Australian International Conference on Speech Science and Technology*, 1, 322–327.
- Wang, S., Liu, B., Dong, R., Zhou, Y., Li, J., Qi, B., . . . Zhang, L. (2012). Music and lexical tone perception in Chinese adult cochlear implant users. *The Laryngoscope*, 122, 1353–1360.
- Wang, S., Liu, B., Zhang, H., Dong, R., Mannell, R., Newall, P., . . . Han, D. (2012). Mandarin lexical tone recognition in sensorineural hearing-impaired listeners and cochlear implant users. *Acta Oto-Laryngologica*, 133, 47–54.
- Wang, W., Zhou, N., & Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. *International Journal of Audiology*, 50, 270–278.
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25–47.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness of fit. *Perception & Psychophysics*, 63, 1293–1313.
- Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *The Journal of the Acoustical Society of America*, 131, 1465–1479.
- Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research*, 56, 1097–1107.
- Wu, J. L., & Yang, H. M. (2003). Speech perception of Mandarin Chinese speaking young children after cochlear implant use: Effect of age at implantation. *International Journal of Pediatric Otorhinolaryngology*, 67, 247–253.

-
- Xu, L., & Pfingst, B. E.** (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception. *The Journal of the Acoustical Society of America*, 114, 3024–3027.
- Xu, L., Thompson, C. S., & Pfingst, B. E.** (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *The Journal of the Acoustical Society of America*, 117, 3255–3267.
- Xu, L., Tsai, Y., & Pfingst, B. E.** (2002). Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *The Journal of the Acoustical Society of America*, 112, 247–258.
- Zhou, N., Huang, J., Chen, X., & Xu, L.** (2013). Relationship between tone perception and production in prelingually deafened children with cochlear implants. *Otology & Neurotology*, 34, 499–506.
- Zhu, M., Fu, Q.-J., Galvin, J. J., III, Jiang, Y., Xu, J., Xu, C., . . . Chen, B.** (2011). Mandarin Chinese speech recognition by pediatric cochlear implant users. *International Journal of Pediatric Otorhinolaryngology*, 75, 793–800.
- Zhu, S., Wong, L. L. N., & Chen, F.** (2014). Development and validation of a new Mandarin tone identification test. *International Journal of Pediatric Otorhinolaryngology*, 78, 2174–2182.