Original Article

Music Training Can Improve Music and Speech Perception in Pediatric Mandarin-Speaking Cochlear Implant Users

Xiaoting Cheng[1,2,*], Yangwenyi Liu[1,2,*], Yilai Shu[1,2],
Duo-Duo Tao[3], Bing Wang[1,2], Yasheng Yuan[1,2], John J. Galvin, III[4],
Qian-Jie Fu[5], and Bing Chen[1,2]

Abstract
Due to limited spectral resolution, cochlear implants (CIs) do not convey pitch information very well. Pitch cues are important for perception of music and tonal language; it is possible that music training may improve performance in both listening tasks. In this study, we investigated music training outcomes in terms of perception of music, lexical tones, and sentences in 22 young (4.8 to 9.3 years old), prelingually deaf Mandarin-speaking CI users. Music perception was measured using a melodic contour identification (MCI) task. Speech perception was measured for lexical tones and sentences presented in quiet. Subjects received 8 weeks of MCI training using pitch ranges not used for testing. Music and speech perception were measured at 2, 4, and 8 weeks after training was begun; follow-up measures were made 4 weeks after training was stopped. Mean baseline performance was 33.2%, 76.9%, and 45.8% correct for MCI, lexical tone recognition, and sentence recognition, respectively. After 8 weeks of MCI training, mean performance significantly improved by 22.9, 14.4, and 14.5 percentage points for MCI, lexical tone recognition, and sentence recognition, respectively ($p < .05$ in all cases). Four weeks after training was stopped, there was no significant change in posttraining music and speech performance. The results suggest that music training can significantly improve pediatric Mandarin-speaking CI users' music and speech perception.

Introduction

Cochlear implantation is an effective treatment for severe-to-profound sensorineural hearing loss. While perception of nontonal (e.g., English) and tonal languages (e.g., Mandarin Chinese) is comparable in cochlear implant (CI) users, tonal language perception depends strongly on fundamental frequency (F0) cues, especially for speech segments (Lin, 1988). For the most part, F0 information is weakly represented by CI signal processing and is primarily conveyed in the temporal envelopes used to modulate pulse trains delivered to the relevant intracochlear electrodes. The functional spectral resolution is too limited to provide good pitch perception (Croghan, Duran, & Smith, 2017; Friesen, Shannon, Baskent, & Wang, 2001; Madsen, Whiteford, & Oxenham, 2017; Shannon, Fu, & Galvin, 2004). Indeed, CI users' pitch perception is much poorer than that of normal-hearing (NH) listeners (Brockmeier et al., 2011; Gfeller & Lansing, 1991; Kong, Cruz, Jones, &

Zeng, 2004), which limits CI users' perception of music, prosody, vocal emotion, and tonal language (Brockmeier et al., 2011; Chatterjee & Peng, 2008;

[1]Department of Otology and Skull Base Surgery, Eye and Ear, Nose, and Throat Hospital, Fudan University, Shanghai, China
[2]Key Laboratory of Hearing Medicine, National Health and Family Planning Commission, Shanghai, China
[3]Department of Ear, Nose and Throat, The First Affiliated Hospital of Soochow University, Suzhu, China
[4]House Ear Institute, Los Angeles, CA, USA
[5]Department of Head and Neck Surgery, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

*These authors contributed equally to this work.

Corresponding authors:
Bing Chen and Yilai shu, Department of Otology and Skull Base Surgery, Eye and Ear, Nose, and Throat Hospital, Fudan University, 83 Fenyang Road, Shanghai 200031, China.
Email: bingchen@fudan.edu.cn; yilai_shu@fudan.edu.cn

Gfeller & Lansing, 1991; Fu, Zeng, Shannon, & Soli, 1998; Galvin, Fu, & Nogaki, 2007; Kong et al., 2004; Luo & Fu, 2004; Luo, Fu, & Galvin, 2007). However, music training has been shown to improve CI users' melodic pitch perception (Fu, Galvin, Wang, & Wu, 2015; Galvin et al., 2007; Galvin, Fu, & Shannon, 2009; Galvin, Eskridge, Oba, & Fu, 2012). Such melodic pitch training might also improve tonal language perception, as pitch cues are important for both listening tasks.

Mandarin Chinese includes four distinctive tones: Tone 1 (high-flat), Tone 2 (high-rising), Tone 3 (falling-rising), and Tone 4 (high-falling). While F0 contributes strongly to lexical tone perception (Abramson, 1978), CI users' may make use of other acoustic cues such as duration and amplitude that covary with F0 (Fu, Hsu, & Horng, 2004; Lin, 1988). Luo and Fu (2004) showed that manipulating the amplitude envelope to more closely resemble the F0 contour can improve lexical tone recognition. Exaggerated pitch contours have also been shown to improve CI users' tone recognition (He, Deroche, Doong, Jiradejvong, & Limb, 2016).

In terms of acoustic information, music and speech have much in common, such as pitch (F0), timbre (spectral envelope), and timing (rhythm; Kraus, Skoe, Parbery-Clark, & Ashley, 2009; Patel, 2003; Tzounopoulos & Kraus, 2009). Previous studies with adult Mandarin-speaking CI users have shown significant correlations between melodic pitch and lexical tone perception (Looi, Teo, & Loo, 2015; Wang et al., 2012; Wang, Zhou, & Xu, 2011), suggesting that both listening tasks may share a similar pitch processing mechanism. However, Tao et al. (2015) found no correlation between lexical tone perception and melodic pitch perception in prelingual or postlingual CI users, possibly due to highly variable performance and a wide age range across subjects. If there is a relationship between music and speech perception, music training may benefit Chinese CI users' speech perception.

Previous studies have observed advantages in speech perception for musicians or for people with extensive music training (Kraus et al., 2009; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Patel, 2003; Tzounopoulos & Kraus, 2009). This "musician effect" may show some advantage when listening to spectrotemporally degraded signals (as in CIs), especially for pitch-mediated speech such as vocal emotion (Fuller, Galvin, Maat, Free, & Başkent, 2014). There is evidence that music training may be related to brain plasticity (Hyde et al., 2009; Kraus & Chandrasekaran, 2010). Music training has been shown to generate structural and functional changes in the brain that may benefit development of musical skills as well as general auditory skills (Besson, Schön, Moreno, Santos, & Magne, 2007; Hyde et al., 2009; Kraus & Chandrasekaran, 2010; Musacchia, Sams, Skoe, & Kraus, 2007). Chen et al.

(2010) showed that music training improved pitch perception in prelingual CI children, and that the duration of training was significantly correlated with improvements in pitch perception. Vandali, Sly, Cowan, and van Hoesel (2014) found improved pitch and timbre perception in CI listeners using computer-based music training. Fu et al. (2015) also found improved melodic contour identification (MCI) performance in Mandarin-speaking pediatric CI users after MCI training with different stimuli, similar to outcomes with adult, English-speaking CI users (Galvin et al., 2007, 2009, 2012).

The above studies show some relationship between music and speech, especially where pitch cues are important (e.g., vocal emotion, lexical tones, etc.). Musical training can improve music perception in NH and CI listeners. However, it is not clear whether music training can also improve CI users' speech perception. Besides the importance of pitch cues for lexical tone perception, pitch cues are also important for language development (e.g., infant-directed speech; Trainor & Desjardins, 2002); as such, music training may be especially beneficial for pediatric Mandarin-speaking CI users. In this study, the effect of music training on music and speech perception was studied in young Mandarin-speaking CI users. Given the importance of pitch cues to both listening tasks, we hypothesized that the MCI training would improve both music and speech perception.

Materials and Methods

Ethics Statement

The study and its consent procedure were approved by the local ethics committee (Ethics Committee of Eye and Ear, Nose, and Throat Hospital of Fudan University, approval number: KY2012-009) and written informed consent was obtained from children's parents before participation.

Participants

Sixteen (5 females and 11 males) Mandarin-speaking CI children were recruited from the Shanghai Eye and Ear, Nose, and Throat Hospital, Fudan University, China. The inclusion criteria were that all pediatric participants be prelingually deaf and diagnosed with severe-to-profound sensorineural hearing loss at ≤1 year old. The exclusion criteria were previous formal music training experience as well as any cognitive, visual, or intelligence disorders. The mean age at testing was 6.3 years (range: 4.5 to 9.3 years), the mean age at implantation was 3.4 years (range: 1.7 to 6.1 years), and the mean CI experience was 2.8 years (range: 0.8 to 6.0 years). Twelve

Table 1. CI Subject Demographic Information.

Subject	Gender	Age at testing (years)	Age at CI (years)	CI exp. (years)	CI device	CI strategy
T1	M	5.8	2.8	3.0	Cochlear N-24	ACE
T2	F	6.8	6.0	0.8	Cochlear N-24	ACE
T3	M	5.8	3.3	2.5	Cochlear N-24	ACE
T4	M	9.3	6.1	3.2	Cochlear N-24	ACE
T5	M	6.4	1.6	4.8	Cochlear N-24	ACE
T6	M	8.8	2.8	6.0	AB HiRes 90K	F120
T7	F	7.8	5.8	2.0	Cochlear N-24	ACE
T8	F	6.3	2.3	4.0	MED-EL Pulsar	FSP
T9	M	4.5	1.7	2.8	MED-EL Pulsar	FSP
T10	M	5.5	3.0	2.5	Cochlear N-24	ACE
T11	M	5.6	3.1	2.5	AB HiRes 90K	F120
T12	M	5.5	3.5	2.0	Cochlear N-24	ACE
T13	F	5.5	2.5	3.0	MED-EL Pulsar	FSP
T14	M	5.5	3.8	1.7	Cochlear N-24	ACE
T15	M	5.3	2.8	2.5	Cochlear N-24	ACE
T16	F	5.6	3.6	2.0	MED-EL Pulsar	FSP
AVE		6.3	3.4	2.8		
(SE)		(0.3)	(0.4)	(0.3)		

Note. CI exp. = amount of CI experience; N-24 = Nucleus 24; AB = Advanced Bionics; ACE = Advanced combination encoder; F120 = Fidelity 120; FSP = Fine-structure processing.

subjects used hearing aids for nearly 6 months prior to cochlear implantation. After implantation, none of the subjects used a hearing aid on the contralateral ear. The mean unaided pure-tone average threshold across 500, 1000, and 2000 Hz was 105 dB HL for the nonimplanted ear. Relevant demographic information is shown in Table 1. Twenty-two (11 females and 11 males) Mandarin-speaking NH children were also tested to ensure that the listening tasks were age appropriate. The inclusion criterion was normal hearing, and exclusion criteria included previous formal music training as well as any cognitive, visual, and intelligence disorders. The mean age at testing for NH subjects was 6.2 years (range: 4.5 to 9.3 years).

All subjects were trained and tested at the Xinsheng rehabilitation center in Jiangsu province, China. The Xinsheng center offers special education services for hearing impaired people, including pediatric CI users. However, the center offers little direct auditory habilitation for CI users, and speech perception is largely learned during regular class offerings (e.g., mathematics, Chinese language, etc.). During training and testing, all stimuli were presented in sound field at 65 dBA via a single loudspeaker positioned 1 m away from the subject, who directly faced the speaker. For the closed-set MCI and tone recognition tasks, children were instructed how to use the computer interface; for the open-set Mandarin

Speech Perception (MSP) sentence recognition, an experimenter scored subject responses. Parents or supervisors accompanied their children for all test and training sessions.

MCI Stimuli and Testing

MCI stimuli and testing were similar to that in Galvin et al. (2007, 2009). Stimuli consisted of nine melodic contours (rising, rising-flat, rising-falling, flat-rising, flat, flat-falling, falling-rising, falling-flat, or falling), composed of five notes of equal duration (250 ms, with 50 ms of silence between each note). The lowest note in any contour was A4 (440 Hz). The spacing between successive notes in each contour was varied to be 1, 2, 3, or 5 semitones. As such, a rising contour might range from A4 (440 Hz) to D4 (587 Hz) with 1-semitone spacing, or from A4 (440 Hz) to F6 (1397 Hz) with 5-semitone spacing. The instrument used for the contour was a piano sample, as in Galvin, Fu, and Oba (2008). Thus, the stimulus set consisted of 36 stimuli (9 melodic contours × 4 semitone spacing), and all 36 stimuli were presented during each test run.

MCI testing was performed using a closed-set, nine-alternative forced-choice (9AFC) procedure. Prior to formal testing, a practice session was conducted to familiarize subjects with the stimuli, task, and procedures. During testing, a contour would be randomly selected from the stimulus set and presented to the subject, who responded by clicking on one of the response boxes shown on the computer screen, labeled with a picture of the contour as well as Chinese text describing the pitch direction. Subjects were allowed to repeat the stimulus up to three times; no trial-by-trial feedback was provided. A minimum of two test runs was conducted for each subject, and performance was averaged across test runs.

Lexical Tone Stimuli and Testing

Lexical tone stimuli consisted of four tonal patterns spoken by two males and two females, taken from the Standard Chinese Database (Wang, 1993). The four tonal patterns included Tone 1 (high-level), Tone 2 (high-rising), Tone 3 (falling-rising), and Tone 4 (high-falling) produced for four monosyllables (/ba/, /bo/, /bu/, /bi/). Thus, the stimulus set consisted of 64 stimuli (4 tone × 4 monosyllables × 4 talkers), and all 64 stimuli were presented during each test run. Tone contour features such as F0, duration, and amplitude were preserved. The mean duration was 273 ms for Tone 1, 340 ms for Tone 2, 410 ms for Tone 3, and 213 ms for Tone 4. Table 2 shows the change in F0 (in semitones) for the four lexical tones produced in the four vowel contexts by the four talkers.

Table 2. Change in F0 (in Semitones) Over Vowels for Lexical Tones Produced by Two Female (F1, F2) and Two Male (M1, M2) Talkers.

Talker	Vowel	Change in F0 (semitones)			
		Tone 1	Tone 2	Tone 3	Tone 4
F1	/ba/	1.0	7.2	12.3	10.1
	/bi/	0.9	7.3	6.1	14.5
	/bo/	2.0	7.0	19.9	16.9
	/bu/	0.6	7.1	9.7	14.9
F2	/ba/	2.5	8.3	6.6	25.9
	/bi/	2.7	8.2	10.7	20.7
	/bo/	2.5	8.9	8.2	24.3
	/bu/	2.2	9.1	11.3	26.0
M1	/ba/	1.3	6.6	8.1	15.4
	/bi/	0.8	8.9	8.3	19.4
	/bo/	1.4	8.5	8.3	17.5
	/bu/	2.1	10.1	8.4	18.7
M2	/ba/	2.0	8.2	7.6	16.6
	/bi/	2.8	9.0	7.5	18.0
	/bo/	3.7	8.0	10.4	18.7
	/bu/	1.8	8.3	9.3	17.1
AVE		1.9	8.2	9.5	18.4
(SE)		(0.2)	(0.2)	(0.8)	(1.1)

Tone recognition was measured in quiet as in Tao et al. (2015), using a four-alternative forced-choice procedure. Stimuli were presented in sound field at 65 dBA. During testing, a stimulus would be randomly selected from the stimulus set and presented to the subject, who responded by clicking on one of the four response boxes shown on the computer screen, labeled in Chinese as Tone 1—flat, Tone 2—rising, Tone 3—falling-rising, and Tone 4—falling. Subjects were allowed to repeat the stimulus up to three times; no trial-by-trial feedback was provided. A minimum of two test runs was conducted for each subject, and performance was averaged across test runs.

Sentence Stimuli and Testing

Sentence recognition in quiet was measured using sentences from the MSP test (Fu, Zhu, & Wang, 2011; Zhu, Wang, & Fu, 2012). Lists of 20 sentences developed for testing CI listeners were used (Li, Wang, Su, Galvin, & Fu, 2016; Su, Galvin, Zhang, Li, & Fu, 2016). The MSP sentences are relatively easy and can be used to test pediatric CI users.

Sentence recognition was measured using an open-set paradigm. During testing, a sentence was randomly selected from the list and presented to the subject, who repeated as many words as possible. The experimenter

scored the correctly identified words, and then a new sentence was presented. One MSP list was presented for each test session.

MCI Training

After baseline measures were completed, subjects began MCI training at the Xinsheng rehabilitation center in Jiangsu, using customized training software loaded onto the rehabilitation center computers. All subjects were required to train for 5 days per week, for 8 weeks. Subjects completed three to six sessions each training day; the average time for each training session was 15 min (range = 12 to 18 min).

MCI training was similar to that in previous studies (Fu et al., 2015; Galvin et al., 2007, 2012). The lowest note of the training stimuli was randomly varied from trial to trial to be any note between A3 (220 Hz) and A5 (880 Hz) except A4 (440 Hz), thus avoiding direct training of the pitch range used for testing. Each training run contained 25 stimuli. Contours were presented with either 3 to 4 or 5 to 6 semitone spacing (i.e., 3 semitones between successive notes, 4 semitones between successive notes, etc.). These two spacing conditions were randomly assigned across training runs. During training, a contour was randomly selected from the stimulus set and the contour would not be presented again in the remaining training procedure. Subjects responded by clicking on one of the nine response boxes shown onscreen. If the subject responded correctly, a new contour would be presented. If not, audio and visual feedback was provided allowing subjects to compare their response to the correct response, after which a new contour was presented. MCI, tone, and sentence recognition were remeasured after 2, 4, and 8 weeks of training. Four weeks after training was stopped, MCI, tone, and sentence recognition were remeasured to observe whether any training benefits were retained.

Results

Figure 1 shows boxplots of MCI scores for each semitone spacing condition, as a function of test week; Test Week 0 = baseline, Test Weeks 2, 4, and 8 = posttraining, and Test Week 12 = follow-up measures 4 weeks after training was stopped. A two-way repeated measures analysis of variance (RM ANOVA) with semitone spacing (1, 2, 3, 5) and test week (0, 2, 4, 8, 12) as factors was performed on the data shown in Figure 1. Results showed a significant effect for test week, $F(4, 168) = 17.5$; $p < .001$, but not for semitone spacing, $F(3, 168) = 2.4$; $p = .081$; there was no significant interaction, $F(12, 168) = 0.3$; $p = .986$. Post hoc Bonferroni pairwise comparisons showed that performance was significantly better at Weeks 4, 8, and 12 than at Week 0 ($p < .05$ in all
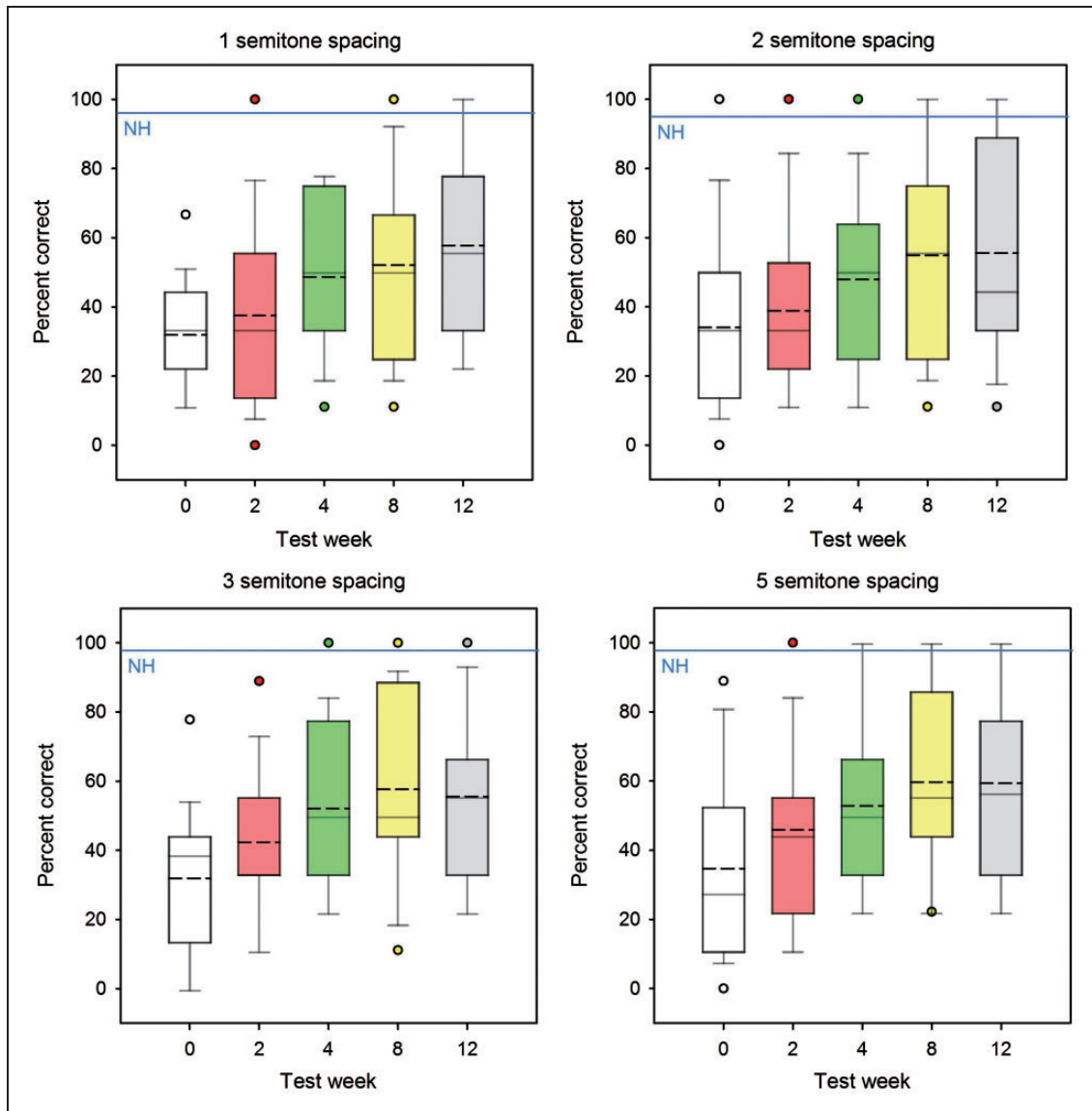
Figure 1. Boxplots of CI users' MCI scores for each semitone spacing as a function of test week; Week 0 = baseline, Weeks 2, 4, and 8 = posttraining, and Week 12 = follow-up performance 4 weeks after training was stopped. The solid horizontal line shows mean performance for a group of NH peers. The boxes show the 25th to 75th percentiles, the error bars show the 5th and 95th percentiles, the circles show outliers, the solid horizontal lines show median performance, and the dashed horizontal lines shows mean performance.

cases), and significantly better at Weeks 8 and 12 than at Week 2 ($p < .05$ in both cases); there were no significant differences among the remaining test weeks ($p > .05$ in all cases).

Figure 2 shows boxplots of recognition scores for each lexical tone as a function of test week. A two-way RM ANOVA, with tone (1, 2, 3, 4) and test week as factors, showed significant effects for tone, $F(3, 168) = 4.4$; $p = .009$, and test week, $F(4, 168) = 13.5$; $p < .001$; there was no significant interaction, $F(12, 168) = 0.7$; $p = .735$. Post hoc Bonferroni pairwise comparisons showed that recognition was significantly better with Tones 1 and 4 than with Tone 2 ($p < .05$ in both cases); there were no

significant differences among the remaining tones. Post hoc Bonferroni pairwise comparisons also showed that performance was significantly better at Weeks 2, 4, 8, and 12 than at Week 0 ($p < .05$ in all cases); there were no significant differences among the remaining test weeks ($p > .05$ in all cases).

Figure 3 shows boxplots of MCI (averaged across all semitone spacing conditions), tone recognition (averaged across all tones), and sentence recognition scores, as a function of test week. One-way RM ANOVAs were performed on the MCI, tone, and sentence recognition data shown in Figure 3, with test week as the factor. There was a significant effect of test week for MCI, $F(4, 59) = 16.1$;
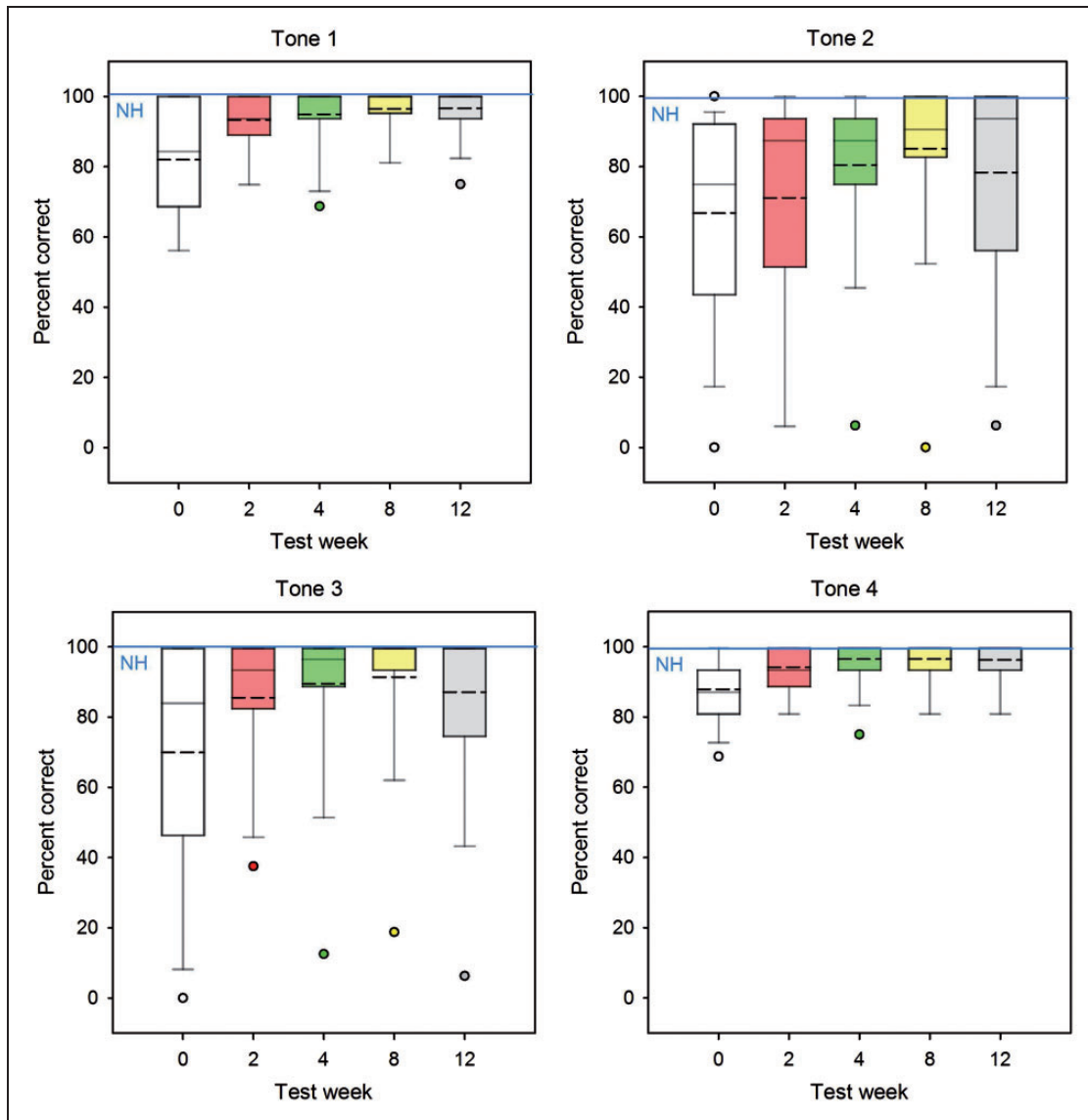
Figure 2. Boxplots of CI users' recognition scores for individual lexical tones, as a function of test week. The solid horizontal line shows mean performance for a group of NH peers. The boxes show the 25th to 75th percentiles, the error bars show the 5th and 95th percentiles, the circles show outliers, the solid horizontal lines show median performance, and the dashed horizontal lines shows mean performance.

$p < .001$, tone recognition, $F(4, 59) = 12.3$; $p < .001$, and sentence recognition, $F(4, 59) = 13.2$; $p < .001$. Post hoc Bonferroni pairwise comparisons showed that MCI was significantly better at Weeks 4, 8, and 12 than at Week 0 ($p < .05$ in all cases), and significantly better at Weeks 8 and 12 than at Week 2 ($p < .05$ in both cases); there were no significant differences among the remaining test weeks ($p > .05$ in all cases). Tone recognition was significantly better at Weeks 2, 4, 8, and 12 than at Week 0 ($p < .05$ in all cases); there were no significant differences among the remaining test weeks ($p > .05$ in all cases). Performance was significantly better at Week 8 relative to Week 0 for Tones 1, 2, and 3 ($p < .05$ in all cases). Sentence

recognition was significantly better at Weeks 4, 8, and 12 than at Week 0 ($p < .05$ in all cases), and significantly better at Weeks 8 and 12 than at Week 2 ($p < .05$ in both cases); there were no significant differences among the remaining test weeks ($p > .05$ in all cases).

Demographic factors such as age at testing, age at implantation, and CI experience were compared with MCI, tone recognition, and sentence recognition performance at the different test intervals; results of Pearson correlations are shown in Table 3. After Bonferroni adjustment for multiple comparisons, significant correlations were observed between CI experience and MCI at Week 0 ($p = .012$), between age at implantation and tone
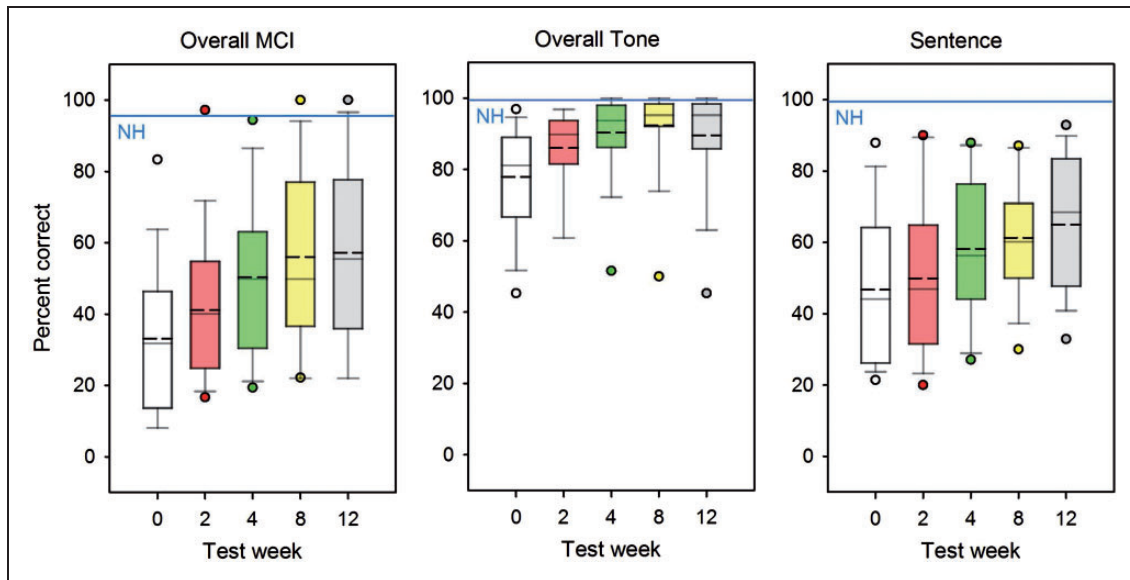
Figure 3. Boxplots of CI users' MCI (across all semitone spacings), tone recognition (across all lexical tones), and sentence recognition scores as a function of test week. The solid horizontal line shows mean performance for a group of NH peers. The boxes show the 25th to 75th percentiles, the error bars show the 5th and 95th percentiles, the circles show outliers, the solid horizontal lines show median performance, and the dashed horizontal lines shows mean performance.

Table 3. Results of Pearson Correlations Between Demographic Factors and Music and Speech Perception for CI Users at Each Test Week, and for NH Listeners.

	Week		MCI		Tone		Sentence	
			r	p	r	p	r	p
CI	0	Age at test	0.26	0.335	−0.34	0.203	−0.03	0.919
		Age at CI	−0.30	0.259	−0.36	0.165	−0.30	0.264
		CI exp.	0.61	0.012*	0.05	0.839	0.30	0.352
	2	Age at test	0.30	0.263	−0.31	0.240	0.10	0.721
		Age at CI	−0.14	0.599	−0.31	0.249	−0.23	0.389
		CI exp.	0.47	0.063	0.02	0.956	0.36	0.169
	4	Age at test	0.04	0.869	−0.56	0.023	−0.01	0.992
		Age at CI	−0.35	0.179	−0.75	<0.001*	−0.26	0.330
		CI exp.	0.44	0.084	0.24	0.362	0.30	0.277
	8	Age at test	0.10	0.709	−0.70	0.006*	0.01	0.984
		Age at CI	−0.20	0.457	−0.57	0.022	−0.25	0.349
		CI exp.	0.33	0.209	−0.06	0.837	0.29	0.288
	12	Age at test	0.07	0.797	−0.44	0.105	0.01	0.964
		Age at CI	−0.36	0.163	−0.54	0.037	−0.17	0.547
		CI exp.	0.48	0.069	0.15	0.599	0.20	0.470
NH		Age at test	0.41	0.057	0.05	0.857	0.46	0.030

Note. The asterisks indicate significant correlations after Bonferroni correction for multiple comparisons. Age at CI = age at cochlear implantation; CI exp. = amount of CI experience; MCI = melodic contour identification.

recognition at Week 4 ($p < .001$), and between age at testing and tone recognition at Week 8 ($p = .006$).

The mean amount of time spent training over the 8-week period was 16.3 hr (range: 12.7 to 20.0). After

8 weeks of training, the mean improvement was 22.9 (range: −5.7 to 47.2), 14.5 (range: 4.7 to 32.8), and 14.5 (−1.5 to 34.3) percentage points for MCI, tone recognition, and sentence recognition, respectively. Pearson

correlations showed no significant relationships between the time spent training and training benefit for MCI, tone recognition, or sentence recognition ($p > .05$ in all cases). There were no significant correlations observed in terms of training benefits (Week 8—Week 0) among MCI, tone, and sentence recognition ($p > .005$ in all cases). After Bonferroni correction for multiple comparisons, Pearson correlation showed a significant relationship between MCI and tone recognition only at Week 12 ($r = .65$, $p = .009$). There were no other significant correlations among MCI, tone, and sentence recognition at the other test intervals ($p > .05$ in all cases).

Discussion

Consistent with our hypothesis, 8 weeks of music training significantly improved music and speech perception in Mandarin-speaking pediatric CI users. Training benefits were largely retained 4 weeks after training was stopped. However, posttraining MCI, tone, and sentence recognition remained much poorer for CI users than for NH peers. We next discuss the results in greater detail.

Baseline Music and Speech Performance

Baseline MCI performance for CI subjects was generally poor (mean = 33.2% correct) and highly variable (range: 8.3 to 83.3% correct), similar to previous reports with Mandarin-speaking CI users (Fu et al., 2015), but poorer than observed in previous studies with adult postlingual English-speaking CI users (Galvin et al., 2007, 2009). Mean MCI recognition in this study was better than the 18.5% correct observed in Tao et al. (2015). In that study, the mean age at testing for prelingual subjects was 10.8 years (range: 6 to 16 years), compared with 6.1 years (range: 5 to 9 years) in this study. The mean age at implantation for prelingual subjects was 4.3 years (range: 2 to 12 years) in Tao et al. (2015), compared with 3.5 years (range: 1.2 to 6.5 years) in this study. Earlier implantation and shorter duration of deafness may have contributed to differences in MCI performance and the relationship between speech and music perception observed between Tao et al. (2015) and the present study.

There was no significant effect of semitone spacing on baseline performance, different from previous studies with adult CI users that showed better MCI performance with increased semitone spacing (Galvin et al., 2007, 2009). However, the present results are consistent with Tao et al. (2015), who showed no significant differences among semitone spacing, except between 1 and 6 semitones. In this study, some melodic contours (rising-flat, flat-rising, flat, flat-falling, and falling-flat) were more easily identified than others (rising, rising-falling, falling-rising, and falling), which might have offset advantages associated with larger spacings. Differences in subject age, duration of deafness, and previous acoustic hearing experience may have also contributed to differences in MCI performance observed between this and previous studies.

Mean baseline tone recognition in this study (77.9% correct) was similar to that of prelingual CI subjects in Tao et al. (2015; 80.9% correct). Tone 2 recognition was significantly poorer than Tone 1 or Tone 4 recognition, consistent with previous studies (Su et al., 2016). In this study, all covarying cues (F0, amplitude, and duration) were preserved. Given the weak coding of F0 cues in CI signal processing, differences in duration may have contributed to the present pattern of results, with Tone 2 having a relatively long duration and Tone 4 having a relatively short duration (Wei, Cao, & Zeng, 2004).

MCI performance was significantly correlated with tone recognition at Week 12 (follow-up measure), but not at baseline. The correlation at Week 12 is consistent with previous CI studies involving tonal languages (Looi et al., 2015; Wang S et al., 2012; Wang W et al., 2011), suggesting that music and lexical tone perception may share perceptual mechanisms most likely related to pitch perception. However, the present data at Week 0 (baseline) were consistent with Tao et al. (2015), who found no correlation between MCI performance and lexical tone perception in pre- or postlingual Mandarin-speaking CI users. This suggests that music training may be necessary to observe relationships between speech and music perception.

Mean baseline MSP sentence recognition was only 45.8% correct, much poorer than that of NH subjects (97.5% correct). The present performance was also much poorer than that observed for MSP sentences with pediatric CI users (84.7% correct in Su et al., 2016), observed for MSP sentences with NH adult subjects listening to four-channel acoustic CI simulations (90.1% correct in Fu et al., 2011), and poorer than observed with the similar MEST sentences for adult CI users (82% correct in Li et al., 2016). It is possible than differences in age at testing may have contributed to the discrepancies in sentence recognition across studies. In Fu et al. (2011) and Li et al. (2016), adult subjects were tested. In Su et al. (2016), the mean age at testing for pediatric CI subjects was 9.7 years, compared with 6.1 years in this study. Eisenberg, Shannon, Martinez, Wygonski, and Boothroyd (2000) found significantly better performance for older (10 to 12 years) than for younger NH children (5 to 7 years) listening to acoustic CI simulations, suggesting some developmental contribution to perception of spectrally degraded speech. The lower sentence recognition performance, compared with tone recognition, may have also been due to different testing paradigms (closed-set for tones and open-set for sentences).

After Bonferroni correction for multiple comparisons, the amount of CI experience was significantly correlated with baseline MCI performance ($p < .05$), but not with baseline tone or sentence recognition ($p > .05$ in both cases). There were no significant correlations between age at testing and MCI, tone, or sentence recognition ($p > .05$ in all cases). Thus, CI experience appears to have contributed more strongly to baseline music perception than to speech perception. Perception of melodic contours may have depended more strongly on adaptation to the spectrotemporal resolution of the CI, rather than development of specific patterns, as would be required for speech perception. Also, prelingual pediatric CI users must develop speech patterns with the limited spectrotemporal resolution available with their device. Perception of melodic contours requires attention to pitch cues, while lexical tone perception depends on F0, duration, and amplitude cues, which may interact with development (age at testing and age at implantation). Note that CI experience was not significantly correlated with MCI performance (or with tone or sentence recognition) at any of the other test intervals ($p > .05$ in all cases), suggesting that the MCI training may have been more of a factor than CI experience for MCI performance.

Benefits of Music Training

Eight weeks of MCI training significantly improved MCI performance ($p < .05$). The range of posttraining gains for MCI was quite large (−5.7 to 47.2 percentage points). Some subjects with relatively low baseline scores (e.g., T8, T15) received relatively small training benefits (<12 percentage points), while others (e.g., T14, T16) received large benefits (>38 percentage points). Posttraining gains were retained 4 weeks after training was stopped and did not appear to be due to selective improvements for different semitone spacings, as there was no significant difference among the different spacings after 8 weeks of MCI training. The present subjects trained only with the 3 to 4 and 5 to 6 semitone spacings, which seemed to improve performance for all semitone spacing conditions.

Although baseline MCI performance was poorer in this study than in previous studies with adult, postlingual English-speaking CI users (Galvin et al., 2007, 2009, 2012), the mean training benefit was comparable across studies (approximately 25 percentage points). The present training benefit was much less than reported by Fu et al. (2015) for six pediatric Mandarin-speaking CI users. Note that subjects in Fu et al. (2015) trained with a relatively simple harmonic complex, compared with the more spectrotemporally complex piano sound used in this study. The mean age of this study was also lower than that of previous studies, which may have influenced the training benefit.

Significant improvements were observed for Tones 1, 2, and 3 after 8 weeks of MCI training, relative to baseline ($p < .05$); improvement in Tone 4 recognition may have been limited by ceiling performance effects. The greatest posttraining improvement was observed for Tone 3 (rising-falling), which has greater changes in F0 than in the other tones. The improvement in MSP sentence recognition may have been due to improved tone recognition, as lexical tone recognition has been shown to significantly contribute to sentence recognition (Chen, Wong, Chen, & Xi, 2014; Fu et al., 1998). The music training benefit for tone recognition is consistent with previous pediatric CI studies that showed a benefit for music training in terms of speech prosody perception, for which pitch cues are important (Good et al., 2017; Torppa et al., 2014). While many studies have shown a musician advantage for speech perception (Besson et al., 2007; Fuller et al., 2014; Musacchia et al., 2007; Parbery-Clark et al., 2009), others have not (Deroche, Limb, Chatterjee, & Gracco, 2017; Madsen, Whiteford, & Oxenham, 2017; Ruggles, Freyman, & Oxenham, 2014). The variability in musician effects for speech recognition may be due to the importance of pitch cues to the listening task. In this study, the improvements in speech performance were for a speech task where pitch cues are lexically meaningful (as opposed to segregation of competing talkers or perception of voiced vs. whispered speech).

While the improved melodic pitch perception via MCI training appeared to generalize to improved speech perception where voice pitch cues are critical, it is worth noting that the time scale of pitch changes in the MCI task (1,500 ms) was much greater than for lexical tones (309 ms, averaged across all tones). Interestingly, the range of changes in F0 was comparable between the MCI and lexical tone stimuli (Table 2). There seemed to be a global benefit for the MCI training, as recognition significantly improved for all semitone spacings in the MCI task and for three of the four tones in the tone recognition task. It could be that subjects' functional spectrotemporal resolution was improved as a result of the melodic pitch training. Alternatively, subjects' attention, memory, and cognitive processing were improved merely by participating in the training, in which case, the music training did not necessarily improve specific aspects of auditory perception. However, Oba, Galvin, and Fu (2013) found significant improvements in speech understanding in noise perception with auditory training, and not with a similar visual training task, suggesting that auditory training is needed to improve auditory perception.

While there was no proper control group for this study, MCI, tone, and sentence recognition were each repeatedly measured in a group of six Mandarin-

speaking pediatric CI users at the same test intervals as for the above CI users who received MCI training. One-way RM ANOVAs showed no significant effect of test interval on MCI, tone, or sentence recognition performance ($p > .05$), suggesting that the benefit of observed in subjects who received MCI training may not have been due to procedural learning. Given the benefit of MCI training observed in this study, further study with a rigorously matched control group and experimental blinding seems warranted.

Finally, note that the present MCI training is not at all similar to the instrument training experienced by musicians in studies that show musician advantages in speech performance and auditory perception (Besson et al., 2007; Fuller et al., 2014; Kraus and Chandrasekaran, 2010; Kraus et al., 2009; Musacchia et al., 2007; Parbery-Clark et al., 2009). Active participation in music and learning to play an instrument may further improve the benefits of music training for CI users. Patel (2014) presented preliminary data collected with Galvin and coworkers showing improved MCI training benefits when subjects were instructed to play the contours on a musical keyboard, rather than just listen to the contours as in previous MCI studies (Fu et al., 2015; Galvin et al., 2007, 2009, 2012). Previous studies have shown that MCI training generalized to improved familiar melody recognition, and that MCI training with one instrument or pitch range can generalize to better MCI performance with other instruments or pitch ranges (Galvin et al., 2007, 2009, 2012). While these generalized performance gains are encouraging, it is unclear whether MCI training might generalize to other music perception tasks involving more complex listening (e.g., polyphonic music perception). Still, the present data suggest that music training can significantly benefit Mandarin-speaking pediatric CI users' melodic pitch and speech perception.

Conclusions

In this study, the benefits of music training for speech and music perception were studied in young, pediatric, Mandarin-speaking CI users. The results suggest that melodic pitch training can improve melodic pitch perception as well as lexical tone and sentence recognition. Major findings include:

1. Baseline MCI performance was poor, while tone and sentence recognition were moderately good.
2. After 8 weeks of MCI training, significant improvements were observed for MCI, tone, and sentence recognition. However, posttraining performance for CI users remained much poorer than that of NH peers. Posttraining gains were largely retained 4 weeks after training was stopped.

3. Significant correlations were observed between baseline MCI performance and CI experience, between tone recognition and age at implantation after 4 weeks of training, and between age at testing and tone recognition after 8 weeks of training. Training benefits for MCI, tone, and sentence recognition were not significantly correlated with age at testing, age at implantation, or CI experience.

Author Contributions

X. C., Y. S., Q. F., and B. C. conceived and designed the experiments; X. C., Y. S., D. T., Y. L., B. W., and Y. Y. performed the experiments; X. C., Y. L., Q. F., and J. G. analyzed the data; Q. F. and J. G. contributed reagents or materials or analysis tools; X. C., J. G., and Y. S. wrote the paper. X. C. and Y. L. contributed equally to this work.

ORCID iD

Bing Chen http://orcid.org/0000-0002-0988-9534.

References

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *21*, 319–325. doi:10.1177/002383097802100406.

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurological Neuroscience*, *25*(3–4), 399–410. doi:10.1162/jocn.2010.21585.

Brockmeier, S. J., Fitzgerald, D., Searle, O., Fitzgerald, H., Grasmeder, M., Hilbig, S.,...Arnold, W. (2011). The MuSIC perception test: A novel battery for testing music perception of cochlear implant users. *Cochlear Implants International*, *12*(1), 10–20. doi:10.1179/146701010X126778 99497236

Chatterjee, M., & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and

speech intonation recognition. *Hearing Research*, *235*(1–2), 143–156. doi:10.1016/j.heares.2007.11.004.

Chen, J. K., Chuang, A. Y., McMahon, C., Hsieh, J. C., Tung, T. H., & Li, L. P. (2010). Music training improves pitch perception in prelingually deafened children with cochlear implants. *Pediatrics*, *125*(4), e793–e800. doi:10.1542/peds.2008-3620.

Chen, Y., Wong, L. L., Chen, F., & Xi, X. (2014). Tone and sentence perception in young Mandarin-speaking children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, *78*(11), 1923–1930. doi:10.1016/j.ijporl.2014.08.025

Croghan, N. B. H., Duran, S. I., & Smith, Z. M. (2017). Re-examining the relationship between number of cochlear implant channels and maximal speech intelligibility. *The Journal of the Acoustical Society of America*, *142*, EL537–EL543. doi:10.1121/1.5016044

Deroche, M. L. D., Limb, C. J., Chatterjee, M., & Gracco, V. L. (2017). Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. *The Journal of the Acoustical Society of America*, *142*(4), 1739–1755. doi:10.1121/1.5005496.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America*, *107*(5 Pt 1), 2704–2710. doi:10.1121/1.428656.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163. doi:10.1121/1.1381538.

Fu, Q. J., Galvin, J. J., 3rd, Wang, X., & Wu, J. L. (2015). Benefits of music training in Mandarin-speaking pediatric cochlear implant users. *Journal of Speech, Language, and Hearing Research*, *58*(1), 163–169. doi:10.1044/2014_JSLHR-H-14-0127

Fu, Q. J., Hsu, C. J., & Horng, M. J. (2004). Effects of speech processing strategy on Chinese tone recognition by Nucleus-24 cochlear implant users. *Ear and Hearing*, *25*(5), 501–508. doi:10.1097/01.aud.0000145125.50433.19.

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America*, *104*(1), 505–510. doi:10.1121/1.413004.

Fu, Q. J., Zhu, M., & Wang, X. (2011). Development and validation of the Mandarin speech perception test. *The Journal of the Acoustical Society of America*, *129*(6), EL267–E–273. doi:10.1121/1.3590739

Fuller, C. D., Galvin, J. J., 3rd, Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, *8*, 179. doi:10.3389/fnins.2014.00179

Galvin, J. J., Eskridge, E., Oba, S., & Fu, Q. J. (2012). Melodic contour identification training in cochlear implant users with and without a competing instrument. *Seminars in Hearing*, *33*(4), 399–409. doi:10.1055/s-0032-1329227

Galvin, J. J., 3rd, Fu, Q. J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear and Hearing*, *28*(3), 302–319. doi:10.1097/01.aud.0000261689.35445.20.

Galvin, J. J., 3rd, Fu, Q. J., & Oba, S. (2008). Effect of instrument timbre on melodic contour identification by cochlear implant users. *The Journal of the Acoustical Society of America*, *124*(4), EL189–EL195. doi:10.1121/1.2961171

Galvin, J. J., 3rd, Fu, Q. J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, *1169*, 518–533. doi:10.1111/j.1749-6632.2009.04551.x

Gfeller, K., & Lansing, C. R. (1991). Melodic, rhythmic, and timbral perception of adult cochlear implant users. *Journal of Speech and Hearing Research*, *34*(4), 916–920. doi:10.1044/jshr.3404.916.

Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., . . . Russo, F. A. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing*, *38*(4), 455–464. doi:10.1097/AUD.0000000000000402

He, A., Deroche, M. L., Doong, J., Jiradejvong, P., & Limb, C. J. (2016). Mandarin tone identification in cochlear implant users using exaggerated pitch contours. *Otology and Neurotology*, *37*(4), 324–331. doi:10.1097/MAO.0000000000000980

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, *29*(10), 3019–3025. doi:10.1523/JNEUROSCI.5118-08.2009

Kong, Y. Y., Cruz, R., Jones, J. A., & Zeng, F. G. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing*, *25*(2), 173–185. doi:10.1097/01.AUD.0000120365.97792.2F.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. doi:10.1038/nrn2882

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing: Implications for language and music. *Annals of the New York Academy of Sciences*, *1169*, 543–557. doi:10.1111/j.1749-6632.2009.04549.x

Li, Y., Wang, S., Su, Q., Galvin, J. J., & Fu, Q. J. (2016). Validation of list equivalency for Mandarin speech materials to use with cochlear implant listeners. *International Journal of Audiology*, *56*(Suppl. 2): S31–S40. doi:10.1080/14992027.2016.1204564

Lin, M. C. (1988). The acoustic characteristics and perceptual cues of tones in standard Chinese. *Chinese Yuwen*, *205*, 182–193.

Looi, V., Teo, E. R., & Loo, J. (2015). Pitch and lexical tone perception of bilingual English-Mandarin-speaking cochlear implant recipients, hearing aid users, and normally hearing listeners. *Cochlear Implants International*, *16*(Suppl 3): S91–S104. doi:10.1179/1467010015Z.000000000263

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*(6), 3659–3667. doi:10.1121/1.1783352.

Luo, X., Fu, Q. J., & Galvin, J. J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant

users. *Trends in Amplification*, *11*, 301–315. doi:10.1073/pnas.0701498104.

Madsen, S. M. K., Whiteford, K. L., & Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports*, *7*(1), 12624. doi:10.1038/s41598-017-12937-9

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the USA, 104*(40), 15894–15898. doi:10.1073/pnas.0701498104.

Oba, S. I., Galvin, J. J. III, & Fu, Q. J. (2013). Minimal effects of visual memory training on the auditory performance of adult cochlear implant users. *Journal of Rehabilitation Research and Development*, *50*(1), 99–110. doi:10.1682/JRRD.2011.12.0229.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661. doi:10.1097/AUD.0b013e3181b412e9

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. doi:10.1038/nn1082.

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108. doi:10.1016/j.heares.2013.08.011

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PloS One*, *9*(1), e86980. doi:10.1371/journal.pone.0086980

Shannon, R. V., Fu, Q. J., & Galvin, J. J. 3rd. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngologica Supplement*, *522*, 50–54. doi:10.1080/03655230410017562.

Su, Q., Galvin, J. J., Zhang, G., Li, Y., & Fu, Q. J. (2016). Effects of within-talker variability on speech intelligibility in mandarin-speaking adult and pediatric cochlear implant patients. *Trends in Hearing*, 20. pii:2331216516654022. doi:10.1177/2331216516654022

Tao, D., Deng, R., Jiang, Y., Galvin, J. J., 3rd, Fu, Q. J., & Chen, B. (2015). Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users. *Ear and Hearing*, *36*(1), 102–110. doi:10.1097/AUD.0000000000000086

Torppa, R., Faulkner, A., Huotilainen, M., Järvikivi, J., Lipsanen, J., Laasonen, M., & Vainio, M. (2014). The perception of prosody and associated auditory cues in early-implanted children: The role of auditory working memory and musical activities. *International Journal of Audiology*, *53*(3), 182–191. doi:10.3109/14992027.2013.872302

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340. doi:10.3758/BF03196290.

Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. *Neuron*, *62*(4), 463–469. doi:10.1016/j.neuron.2009.05.002

Vandali, A., Sly, D., Cowan, R., & van Hoesel, R. (2014). Training of cochlear implant users to improve pitch perception in the presence of competing place cues. *Ear and Hearing*, *36*(2), e1–e13. doi:10.1097/AUD.0000000000000109

Wang, R. H. (1993). *The standard Chinese database*. Anhui, China: University of Science of Technology of China, Internal Materials.

Wang, S., Liu, B., Dong, R., Zhou, Y., Li, J., Qi, B., Chen, X., Han, D., & Zhang, L. (2012). Music and lexical tone perception in Chinese adult cochlear implant users. *Laryngoscope*, *122*(6), 1353–1360. doi:10.1002/lary.23271

Wang, W., Zhou, N., & Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. *International Journal of Audiology*, *50*(4), 270–278. doi:10.3109/14992027.2010.542490

Wei, C. G., Cao, K., & Zeng, F. G. (2004). Mandarin tone recognition in cochlear-implant subjects. *Hearing Research*, *197*(1–2), 87–95. doi:10.1016/j.heares.2004.06.002\.

Zhu, M., Wang, X., & Fu, Q. J. (2012). Development and validation of the Mandarin disyllable recognition test. *Acta Otolaryngologica*, *132*(8), 855–861. doi:10.3109/00016489.2011.653668