Trends in Hearing
2016, Vol. 20: 1–16
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216516654022
tia.sagepub.com
SAGE

Effects of Within-Talker Variability on Speech Intelligibility in Mandarin-Speaking Adult and Pediatric Cochlear Implant Patients

Qiaotong Su[1], John J. Galvin[2], Guoping Zhang[1], Yongxin Li[1], and Qian-Jie Fu[2]

Abstract

Cochlear implant (CI) speech performance is typically evaluated using well-enunciated speech produced at a normal rate by a single talker. CI users often have greater difficulty with variations in speech production encountered in everyday listening. Within a single talker, speaking rate, amplitude, duration, and voice pitch information may be quite variable, depending on the production context. The coarse spectral resolution afforded by the CI limits perception of voice pitch, which is an important cue for speech prosody and for tonal languages such as Mandarin Chinese. In this study, sentence recognition from the Mandarin speech perception database was measured in adult and pediatric Mandarin-speaking CI listeners for a variety of speaking styles: voiced speech produced at slow, normal, and fast speaking rates; whispered speech; voiced emotional speech; and voiced shouted speech. Recognition of Mandarin Hearing in Noise Test sentences was also measured. Results showed that performance was significantly poorer with whispered speech relative to the other speaking styles and that performance was significantly better with slow speech than with fast or emotional speech. Results also showed that adult and pediatric performance was significantly poorer with Mandarin Hearing in Noise Test than with Mandarin speech perception sentences at the normal rate. The results suggest that adult and pediatric Mandarin-speaking CI patients are highly susceptible to whispered speech, due to the lack of lexically important voice pitch cues and perhaps other qualities associated with whispered speech. The results also suggest that test materials may contribute to differences in performance observed between adult and pediatric CI users.

Keywords

speech variations, cochlear implant, speech intelligibility, pediatric, mandarin Chinese

Date received: 1 February 2016; revised: 13 May 2016; accepted: 15 May 2016

Introduction

In the clinic and in research, speech performance is often evaluated using well-enunciated speech produced at a normal rate by a single adult talker. Outside of the clinic, listeners regularly encounter great variability in speech signals, even within a single talker (e.g., different speaking rates, emotional qualities, voice pitch ranges, production levels, etc.). While normal-hearing (NH) listeners are able to largely accommodate within- and across-talker variability, hearing-impaired (HI) listeners have greater difficulty with talker variability (e.g., Kirk, Pisoni, & Miyamoto, 1997; Uchanski, Choi, Braida, Reed, & Durlach, 1996). Because of the limited spectral resolution afforded by their device, cochlear implant (CI) users have even greater difficulty with challenging listening conditions (Shannon, Fu, & Galvin, 2004). In noise,

[1]Department of Otolaryngology, Head and Neck Surgery, Beijing TongRen Hospital, Capital Medical University, Ministry of Education of China, Beijing, People's Republic of China
[2]Department of Head and Neck Surgery, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

Corresponding author:
Qian-Jie Fu, Department of Head and Neck Surgery, David Geffen School of Medicine, UCLA, 2100 West Third Street, Suite 100, Los Angeles, CA 90057, USA.
Email: qfu@mednet.ucla.edu

CI users have greater difficulty with conversational speech (Liu, Del Rio, Bradlow, & Zeng, 2004). CI users also have greater difficulty recognizing vocal emotion (Chatterjee et al., 2015; Luo, Fu, & Galvin, 2007) and discriminating between a question and statement (Peng, Lu, & Chatterjee, 2009) than do NH listeners. Talker variability has been shown to negatively affect CI users' speech understanding (Chang & Fu, 2006; Liu, Galvin, Fu, & Narayanan, 2008). CI users have greater difficulty understanding nonnative talkers than do NH listeners (Ji, Galvin, Chang, Xu, & Fu, 2014). CI users also have greater difficulty with fast speaking rates and whispered speech (Ji, Galvin, Xu, & Fu, 2013; Li et al., 2011). Thus, the highly intelligible materials typically used for clinical evaluation may greatly overestimate CI users' perception of variable speech encountered in everyday life. It is important to understand the effects of speech variations to design signal-processing and rehabilitation techniques to help CI users perform better in the "real world" outside of the clinic.

Eskenazi (1993) defined several sources of speech variation within a single talker. The first type of speech variation is introduced by differences in speaking style. Most speech materials used in clinical or research testing can be classified as "clear" speech, in which the speaking style enhances intelligibility, which in turn may be most appropriate when testing nonnative or HI listeners (Smiljanić & Bradlow, 2008). However, natural productions of conversational speech may be more representative of real-world listening experience. As such, testing with clear speech may underestimate the difficulties in speech understanding some listeners experience outside the clinic or laboratory. Previous studies have shown that clear speech is more intelligible than conversational speech in HI (Picheny, Durlach, & Braida, 1985, 1986, 1989; Uchanski et al., 1996) and CI listeners (Liu et al., 2004).

A second type of speech variation is introduced by differences in speaking rate. In general, NH listeners are able to understand speech regardless of variations in talker, speaking rate, and language context (e.g., Eskenazi, 1993; Sommers, Nygaard, & Pisoni, 1992). Li et al. (2011) investigated the effects of speaking rate on understanding of naturally produced Chinese sentences in Mandarin-speaking adult CI patients. While NH performance with unprocessed speech was largely unchanged across speaking rates, CI performance gradually deteriorated from slow (2.5 words per second, or wps) to fast (5.7 wps) speaking rates. Interestingly, when listening to a four-channel acoustic CI simulation, NH performance also deteriorated as the speaking rate was increased, suggesting that susceptibility to speaking rate may be partially due to the limited spectral resolution afforded by CIs. Ji et al. (2013) measured English sentence recognition in NH and CI listeners with multiple talkers and different speaking rates and with naturally produced and synthetic speech; naturally produced speech was time scaled to achieve the target speaking rates. While there was a significant deficit in performance for NH subjects listening to unprocessed fast-rate (6.6 wps) versus normal-rate speech (3.3 wps), the deficit for fast-rate speech was much greater when NH subjects listened to an eight-channel acoustic CI simulation; the deficit was even greater for real CI subjects.

A third type of speech variation can be defined in terms of voice quality (e.g., breathy, creaky, lax, whispery, tense, etc.). In whispered speech, air is forced through the constricted glottis. While whispered speech does not contain voice frequency information, the noise-like sound is shaped similarly to voiced speech via the larynx and oral cavity, thus preserving much formant information. Whispered speech is typically softer than voiced speech (Traunmüller & Eriksson, 2000) and may differ from voiced speech in other ways such as vowel duration and spectral tilt (e.g., Zhang & Hansen, 2007, 2011). Because of the importance of F0 cues to tonal language perception, the absence of voice pitch information would be expected to negatively impact lexical tone and sentence recognition in tonal language such as Mandarin Chinese. Liang (1963) found that Mandarin-speaking NH listeners could only understand 64.0% of whispered lexical tones due to the missing F0 and harmonic fine structure cues. Li et al. (2011) measured recognition of voiced and whispered Mandarin sentences in adult Chinese CI users. Recognition of whispered speech dropped by nearly 40 percentage points, relative to performance with voice speech. One possible explanation for this large deficit is the missing periodicity cues in whispered speech (Fu & Zeng, 2000; Xu, Tsai, & Pfingst, 2002). The poor performance with whispered speech further highlights the importance of F0 cues to Mandarin tone and sentence recognition by Mandarin-speaking CI users. In nontonal languages such as English, perception of whispered speech is somewhat poorer than for voiced speech (e.g., Freyman, Griffin, & Oxenham, 2012; Ruggles, Freyman, & Oxenham, 2014). However, the deficit with whispered speech is typically less for nontonal languages than for tonal languages.

The fourth type of speech variation can be situational or emotional. For example, a person may need to shout to be heard or may change their speaking style to convey a target emotion. Shouted and emotional speech conveys both prosodic and linguistic messages. Production, perception, and response to emotional signals are important for social interactions (Mitchell, 2007; Soto & Levenson; 2009, van Rijn et al., 2005) and for language and emotional development (Cooper & Aslin, 1990; Fernald, 1989; Trainor, Austin, & Desjardins, 2000).

Acoustic cues that encode vocal emotion can be categorized into three main types: speech prosody, voice quality, and vowel articulation (Banse & Scherer, 1996; Murray & Arnott, 1993; Yildirim et al., 2004). Luo et al. (2007) found that while spectral envelope, temporal pitch, and overall amplitude cues all contributed to vocal emotion recognition, performance was much poorer for CI (45.6% correct) than for NH listeners (89.8% correct). Similarly, Chatterjee et al. (2015) found that vocal emotion recognition was poorer in pediatric CI users than in their NH peers. While vocal emotion recognition by CI users has received some attention, it is unclear how the variability in F0, speaking rate, and amplitude affects the intelligibility of emotional speech. If it is difficult to identify the targeted emotion in an utterance, the acoustic variations associated with emotional speech may reduce intelligibility. This may be especially true for tonal languages, as F0 patterns within and across syllables may vary substantially for speech produced with different target emotions. For example, the mean F0 can be approximately twice as high for happy than for sad speech (Luo et al., 2007).

Most of the earlier-cited studies have been conducted with postlingually deafened adult CI users listening to English sentences; central representations of speech patterns were likely developed during previous acoustic hearing. Prelingually deafened CI users develop central speech pattern templates only with the impoverished signal provided by their device. When the spectral resolution is limited, as in the CI case, pediatric CI users may require more time to develop robust speech patterns. For NH subjects listening to acoustic CI simulations, significantly poorer performance was observed for 5 to 7 year olds than for 10 to 12 year olds, with no significant difference between the older children and adults (Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000). In China, the large majority of CI recipients have been prelingually deafened children, who develop speech patterns exclusively with electric hearing. Post-lingually deafened adult CI recipients often have an extended or uncertain period of hearing impairment before implantation. Chinese pediatric patients often have a shorter duration of deafness and longer CI experience than do adult patients.

Mandarin Chinese is a tonal language in which covarying F0, amplitude, and duration are lexically meaningful (Liang, 1963; Lin, 1988). In everyday speech, Mandarin-speaking listeners encounter great inter- and intratalker variability, in which important lexical tone cues are embedded within dynamic changes in production, depending on the context. Because F0 information is not well represented, Mandarin-speaking CI users must depend more strongly on covarying amplitude and duration cues to perceive lexical tones. It is unclear how acoustic variation in lexically meaningful F0, amplitude, and duration cues within a talker might affect Mandarin-speaking CI users' speech perception. It is also unclear how differences in duration of deafness, age at testing, age at implantation, and CI experience might affect Mandarin-speaking CI users' understanding of speech produced in different speaking styles. In this study, recognition of Mandarin sentences produced in different speaking styles was measured in adult and pediatric Mandarin-speaking CI users. Given the expected variation in F0, amplitude, and duration cues, it was expected that the different speaking styles would significantly affect performance.

Methods

Participants

A total of 15 postlingually deafened, adult (10 males and 5 females) and 11 pediatric CI patients (7 males and 4 females) participated in this study. All subjects were native speakers of Mandarin Chinese. All were unilateral CI users and none used a hearing aid in conjunction with their CI. Subject demographic information is shown in Table 1. For adult subjects, the mean age at testing was 42.6 years, the mean age at implantation was 40.6 years, the mean duration of deafness was 5.8 years, and the mean CI experience was 2.0 years. For pediatric subjects, the mean age at testing was 9.7 years, the mean age at implantation was 4.5 years, the mean duration of deafness was 3.3 years, and the mean CI experience was 5.2 years. Note that data for slow, normal, fast, and whispered speech for adult CI Subjects A1 to A13 are from a previous study (Li et al., 2011); for these subjects, data for the emotional speech, shouted speech, and speech understanding of Mandarin Hearing in Noise Test (MHINT; Wong, Soli, Liu, Han, & Huang, 2007) sentences were collected for the present study. All subjects were paid for their participation, and all provided informed consent in accordance with the local institutional review board.

Test Materials

Sentence recognition was measured in quiet using Mandarin Speech Perception (MSP) materials (Fu, Zhu, & Wang, 2011). The MSP materials consist of 10 lists, with 10 sentences of easy difficulty within each list. Each sentence includes seven monosyllabic words. Phonetic balancing (across lists) and word familiarity were carefully considered in the development of the MSP test materials. List equivalency (in quiet) was confirmed in NH subjects listening to unprocessed speech or to a four-channel acoustic simulation of CI processing.

All sentences were produced by a single female talker in six different speaking styles, including three different

Table 1. CI Subject Demographics.

Subject	Gender	Age at testing (years)	Age at implantation (years)	Duration of deafness (years)	Device	Etiology
A1	M	22.5	21.3	15.3	AB	Unknown
A2	F	56.2	54.0	Unknown	AB	Unknown
A3	M	33.0	32.5	0.5	Med-EL	Unknown
A4	M	70.8	70.5	25.5	Med-EL	Ototoxicity
A5	F	42.7	39.1	Unknown	Cochlear	Unknown
A6	M	37.4	32.1	Unknown	Cochlear	Unknown
A7	M	38.5	36.6	1.6	AB	Unknown
A8	M	55.4	54.9	0.9	Med-EL	Unknown
A9	F	21.7	18.8	1.8	Med-EL	LVAS
A10	M	51.0	50.4	6.4	Med-EL	Otitis Media
A11	F	31.1	24.1	20.1	Cochlear	Ototoxicity
A12	M	66.6	66.4	1.4	Med-EL	Unknown
A13	M	41.0	40.4	6.4	Med-EL	Noise
A14	F	47.8	46.7	1.7	Cochlear	Unknown
A15	M	22.9	21.5	4.5	Cochlear	Unknown
C1	F	8.1	1.8	0.8	Cochlear	Unknown
C2	M	8.6	5.4	4.1	Cochlear	Unknown
C3	F	6.8	2.9	1.9	Med-EL	Unknown
C4	M	8.8	2.6	1.4	AB	Unknown
C5	F	8.8	6.6	6.6	Cochlear	Unknown
C6	M	12.3	3.5	2.0	Cochlear	LVAS
C7	M	8.6	1.6	0.6	Cochlear	Unknown
C8	M	10.1	1.7	0.7	Cochlear	Unknown
C9	M	16.3	14.6	12.6	Cochlear	Unknown
C10	M	9.7	3.7	2.7	Med-EL	Unknown
C11	F	8.7	4.7	3.2	Med-EL	Unknown

Note. For Subject, A = adult; C = child; M = male, F = female, LVAS = large Vestibular Aqueduct Syndrome; CI = cochlear implant.

speaking rates (slow, normal, and fast), whispered speech, emotional speech (happy), and shouted speech. At the time of recording, the talker had more than 10 years of professional experience as a broadcaster in a radio station. During the recording of sentence materials at different speaking rates, the talker was instructed to produce sentences at a slow, normal, or fast speaking rate. For whispered speech, the talker was instructed to produce speech while whispering at a normal speaking rate. For emotional speech, the talker was instructed to produce speech in a happy and excited manner. For shouted speech, the talker was instructed to produce speech while shouting with a raised voice. Multiple utterances were recorded for each sentence and each style, and the utterance that best represented the targeted speaking style, as determined by the experimenters, was used in the test materials for the present study.

Figure 1 shows the waveforms for an example MSP sentence (你将来要干什么？What do you want to do in the future?) in each of the speaking styles. Stimuli have been normalized according to the long-term root-mean-square (RMS) amplitude for the sentence portion. In this example, durations were similar between normal, whispered, and emotional speech, shortest for fast speech, and relatively long for the slow and shouted speech.

Figure 2 shows spectrograms for the same stimuli shown in Figure 1. For whispered speech, there is no harmonic information, although formant information is coarsely preserved. For emotional speech, upper harmonic information appears to become more diffuse. The shouted speech displays the strongest harmonic content.

Figure 3 shows electrodograms (CI stimulation patterns) for the same stimuli shown in Figures 1 and 2. The electrodograms were created using the default stimulation parameters for Cochlear Corporation's Freedom device: ACE strategy, 900 pulses per second per electrode, default frequency allocation (input frequency

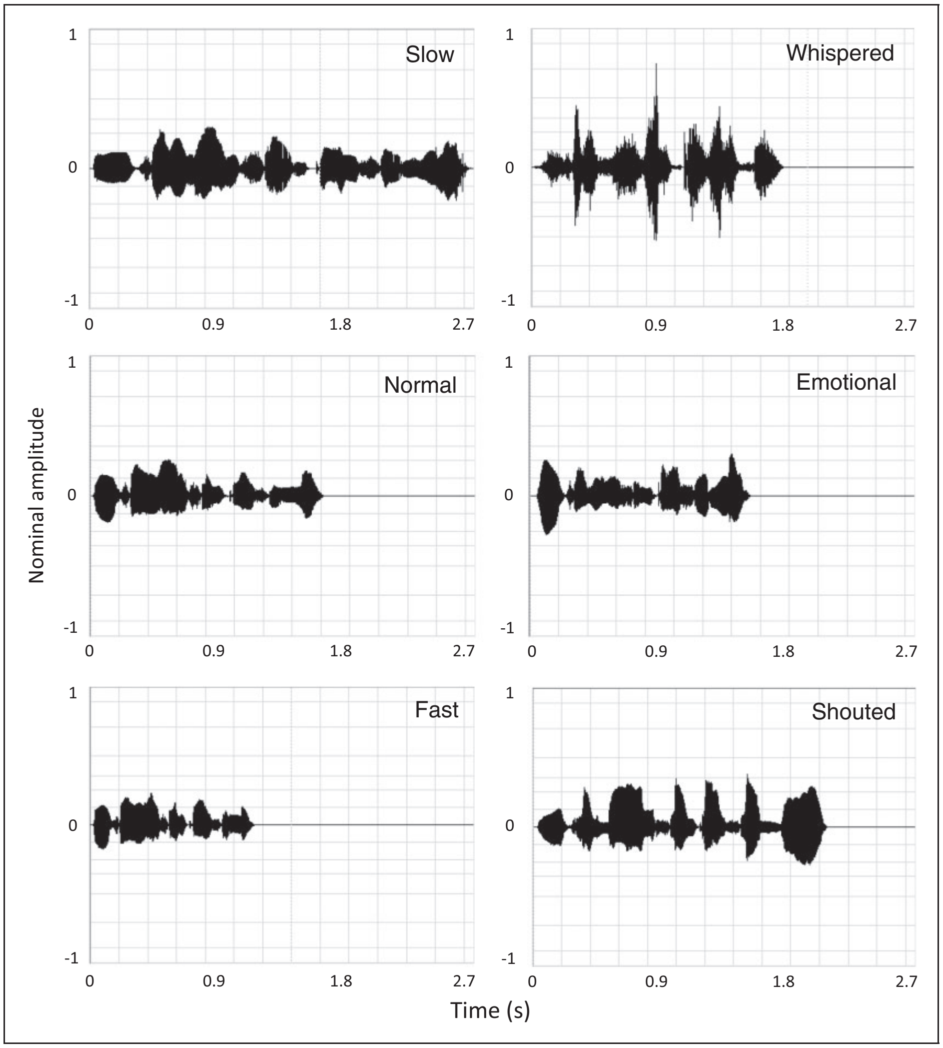

Figure 1. Waveforms for an example MSP sentence (你将来想干什么; English translation: What do you want to do in the future?) produced in six different speaking styles. The x-axis shows time (in seconds) and the y-axis shows nominal amplitude, where 1 and −1 refer to maximum and minimum amplitude, respectively. All stimuli were normalized to have the same long-term RMS amplitude for the sentence portion.

range: 188–7988 Hz), 8 maxima, and so forth. The stimulation patterns are quite similar between the slow, normal, and fast speaking rates, albeit with different sentence durations. For whispered speech, there was more stimulation on basal electrodes due to the lack of low-frequency voice pitch information; the apical stimulation in the formant frequency ranges is also more diffuse, most likely due to the noise-like source of air. The stimulation patterns were quite similar between normal and emotional speech, except for weaker transitions exhibited in the middle electrode region with emotional speech. Shouted speech exhibited much stronger stimulation in the middle and basal electrode regions.

Figure 4 shows the F0 contours for the same stimuli shown in Figures 1 to 3. For illustrative purposes, the duration of each word has been normalized. Note that the F0 transitions between the second and third words and between the sixth and seventh words have been preserved in the analysis. For all but the final word, the F0 contours were quite similar for the slow, normal, and fast speech. For the final word, there appears to be a downward trajectory in F0 for the slow speech that is not observed with the normal and fast speech. The F0 contours are similar for fast and emotional speech until the final word, in which the F0 is elevated and the upward trajectory is much higher for emotional than for fast speech. The F0 contours are greatly elevated for shouted speech, relative to the other speaking styles. For emotional and shouted speech, the range of F0 variation according to tone deviates in inconsistent ways relative to that with slow, normal, and fast speech. The upward shift in overall F0 often resulted (but not always) in less

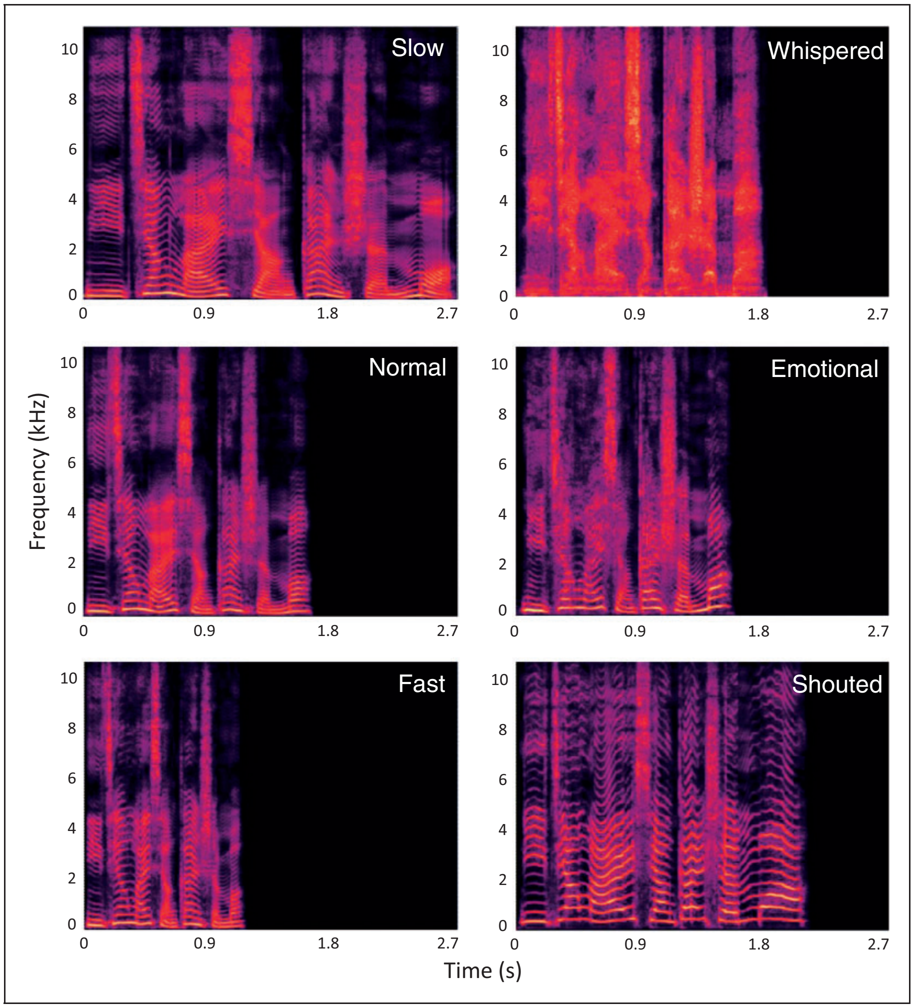

Figure 2. Spectrograms for the same stimuli shown in Figure 1. The x-axis shows time (in seconds) and the y-axis shows frequency in kHz.

F0 variation for tones. Thus, lexical tone information may be distorted with emotional and shouted speech.

Figure 5 shows boxplots for duration and F0 estimated for each sentence, for each of the speaking styles. Duration gradually reduced as the speaking rate increased; F0 also slightly increased with speaking rate. The durations for normal, whispered, emotional, and shouted speech were comparable; F0 was much higher for emotional and shouted speech, compared with slow, normal, or fast speech. There also appears to be greater variation in duration and F0 for emotional and shouted speech. Sentence recognition was also measured with the MHINT materials. Table 2 shows mean sentence duration, mean F0, and mean wps for all experimental test materials. Note that the MSP materials used a single female talker, while the MHINT materials used a single male talker.

Test Procedures

All CI subjects were tested with their clinical processors and settings; these were not changed during the course of testing. For all CI subjects, stimuli were presented in the sound field at 65 dBA via a single loudspeaker; subjects were seated directly facing the loudspeaker at a 1 -m distance. During testing, a sentence list was randomly selected, and sentences were randomly selected from within the list (without replacement) and presented to the subject, who repeated each sentence as accurately as possible. The experimenter calculated the percent of words correctly identified in sentences. All words in the MSP materials were scored, resulting in a total of 70 keywords for each list. For pediatric CI subjects, a parent or guardian was present during testing to reduce subject anxiety. For each subject, the total amount of time for testing was 40 to 60 min. All test materials

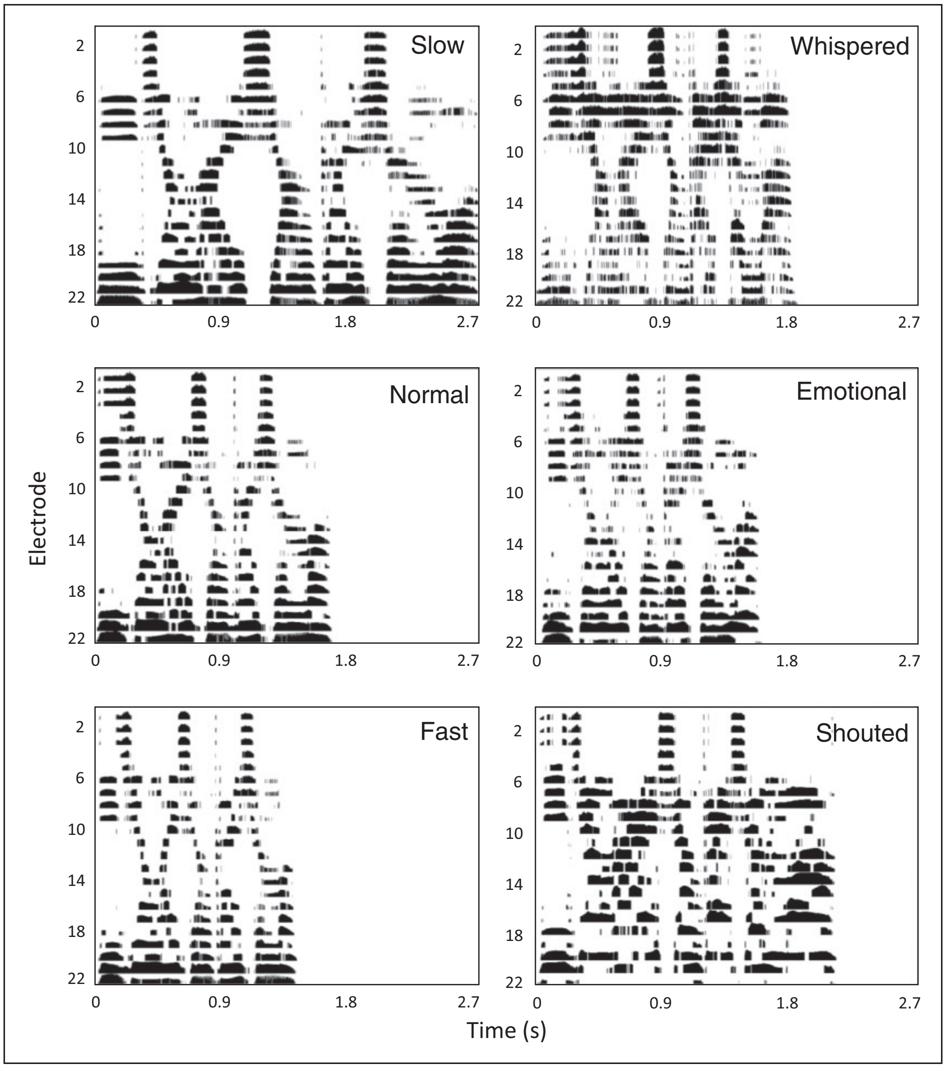

Figure 3. Electrodograms for the same stimuli shown in Figures 1 and 2. The x-axis shows time (in seconds) and the y-axis shows the electrode number from most apical (22) to most basal (1).

were presented once (without repeat), and no feedback was provided regarding the correctness of the response. Testing with the different speaking styles was blocked (i.e., performance was measured separately for each speaking style). The test order for the different speaking styles was randomized within and across subjects. Subjects were allowed to take breaks any time they felt fatigued.

Results

For both subject groups, mean performance was best for the slow speech and gradually worsened as the speaking rate was increased. Performance sharply declined for whispered speech for both subject groups, and performance for emotional and shouted speech was similar to that for the normal speech. For adult subjects, mean performance was 87.9%, 81.0%, 70.4%, 40.7%, 71.9%, and 76.7% correct for slow, normal, fast, whispered, emotional, and shouted speech, respectively. For pediatric subjects, mean performance was 85.4%, 84.7%, 74.8%, 53.7%, 81.5%, and 80.3% correct for slow, normal, fast, whispered, emotional, and shouted speech, respectively.

Toreduce ceiling and floor performance effects, all scores were transformed into rationalized arcsine units (RAU) (Studebaker, 1985). Figure 6 shows boxplots of RAU scores for the different speaking styles for adult (left panel) and pediatric CI subjects (right panel). A split-plot repeated measures analysis of variance (RM ANOVA), with speaking style (slow, normal, fast, whispered, emotional, and shouted) as the within-subject factor and subject group (adults, children) as the between-subject factor was performed on the RAU

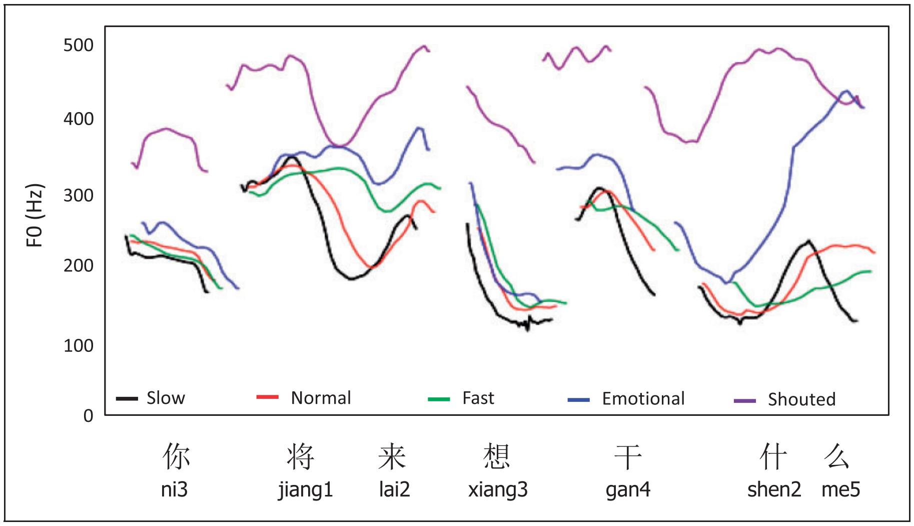

Figure 4. F0 contours extracted for the stimuli shown in Figures 1 to 3, except for whispered speech. For illustration purposes, the contours were normalized in terms of duration. The Chinese characters and tones for each monosyllable are shown below the x-axis. The y-axis shows frequency in Hz.

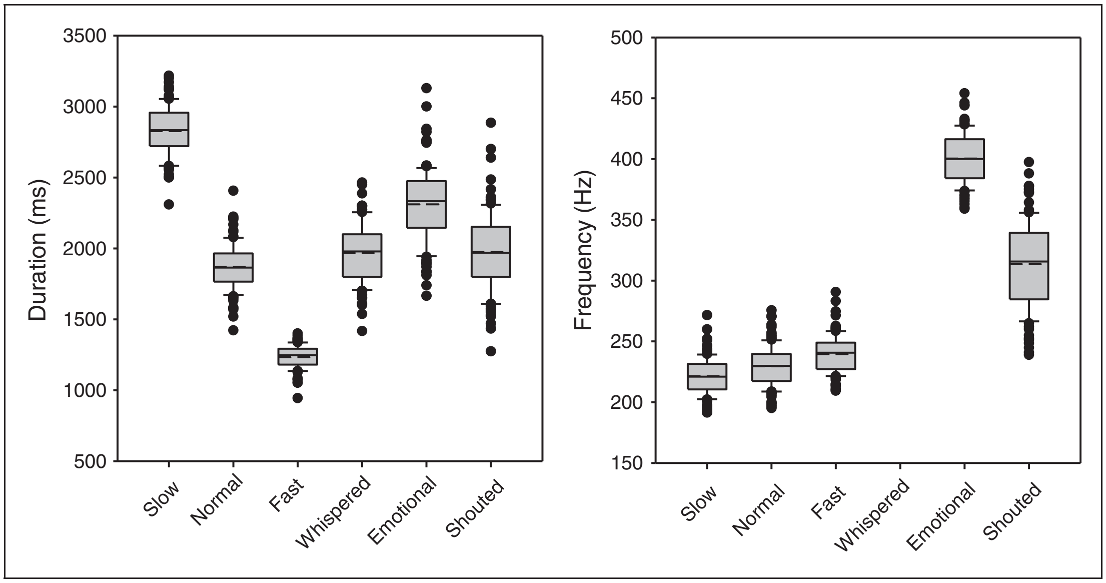

Figure 5. Boxplots of duration (left panel) and F0 (right panel) for all sentences in each of the speaking styles. The boxes show the 25th and 75th percentiles, the solid line shows median performance, the dashed line shows mean performance, the error bars show the 10th and 90th percentiles, and the circles show outliers.

scores. Results showed a significant effect for speaking style, $F(5,120) = 67.7$, $p < .001$; there was a significant interaction, $F(1,24) = 2.4$, $p = .041$. There was no significant difference between subject groups, $F(1,24) = 0.3$, $p = .860$. Because there was a significant interaction in the split-plot ANOVA, separate one-way RM ANOVAs were performed on the RAU scores for adults and children, with speaking style as the factor. For adults, results showed a significant effect of speaking style, $F(5,70) = 42.3$, $p < .001$. Post hoc Bonferroni pairwise comparisons showed that performance with whispered speech was significantly poorer than that with slow, normal, fast, emotional, or shouted speech ($p < .05$ in all cases). Performance was significantly better with slow than with fast, emotional, or shouted speech ($p < .05$ in all cases). For children, results showed a significant effect of speaking style, $F(5,50) = 28.1$, $p < .001$. Post hoc Bonferroni pairwise comparisons showed that performance with whispered speech was significantly poorer than that with slow, normal, fast, emotional, or shouted speech ($p < .05$ in all cases) and significantly poorer with fast than with slow or normal speech ($p < .05$ in both cases).

Figure 7 shows boxplots of the difference in RAU scores between different speaking styles for adult (left panel) and pediatric CI subjects (right panel). While the mean difference in RAU scores was similar between adults and children, there was much greater variability in RAU scores between slow and fast speech for adults (range = −2.7 to 63.6) than for children (range = −6.5 to 30.0). The greatest overall variability in RAU difference scores was observed between normal and whispered speech (adult range = 18.0–98.9; pediatric range = 11.2–60.2). The variability in RAU difference scores was more comparable for emotional (adult range = −25.1 to 36.0; pediatric range = −13.2 to 25.1) and shouted speech (adult range = −25.1 to 31.7; pediatric range = −23.7 to 33.5).

For adults, mean performance was 81.0% and 69.2% correct with the MSP sentences at the normal rate and the MHINT sentences, respectively. For children, mean performance was 84.7% and 50.0% correct with the MSP sentences at the normal rate and MHINT sentences, respectively. As aforementioned, all scores were converted to RAU units (Studebaker, 1985). Figure 8 shows boxplots of RAU scores with the MSP (normal rate) and MHINT sentences for adult (left panel) and pediatric CI subjects (right panel). A split-plot ANOVA was performed on the RAU scores shown in Figure 8, with test material (MSP, MHINT) as the within-subject factor and subject group as the between-subject factor. Note that data for pediatric subjects C1 and C3 were excluded because they were unable to complete the MHINT testing due to time constraints. Results showed a significant effect for test materials, $F(1,22) = 72.7$, $p < .001$; there was a significant interaction, $F(1,22) = 14.9$, $p = .001$. There was no significant difference between subject groups, $F(1,22) = 0.8$, $p = .382$. Because there was a significant interaction, separate one-way RM ANOVAs were performed on the adult and pediatric data shown in Figure 8, with test material as the factor. Results showed that performance

Table 2. Mean Sentence Duration, Mean F0, and Mean Words Per Second for the Experimental Speech Materials.

Test material	Sentence duration (ms)	F0 (Hz)	Words per second
MSP slow	2,826	221	2.48
MSP normal	1,870	230	3.74
MSP fast	1,234	240	5.67
MSP whispered	1,968	N/A	3.56
MSP emotional	1,974	314	3.55
MSP shouted	2,310	400	3.03
MHINT	2,242	131	4.46

Note. MHINT = Mandarin Hearing in Noise Test; MSP = Mandarin Speech Perception. Note that the MSP materials contained seven words per sentence while the MHINT material contained 10 words per sentence.

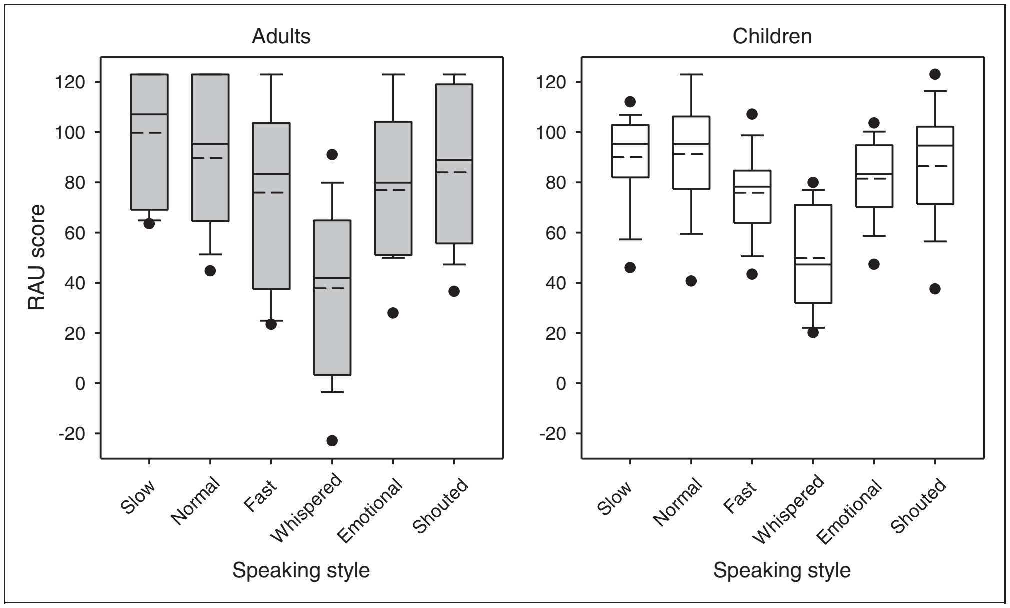

Figure 6. Boxplots of RAU scores for MSP sentences with the different speaking styles, for adult (left panel) and pediatric CI subjects (right panel). The boxes show the 25th and 75th percentiles, the solid line shows median performance, the dashed line shows mean performance, the error bars show the 10th and 90th percentiles, and the circles show outliers.

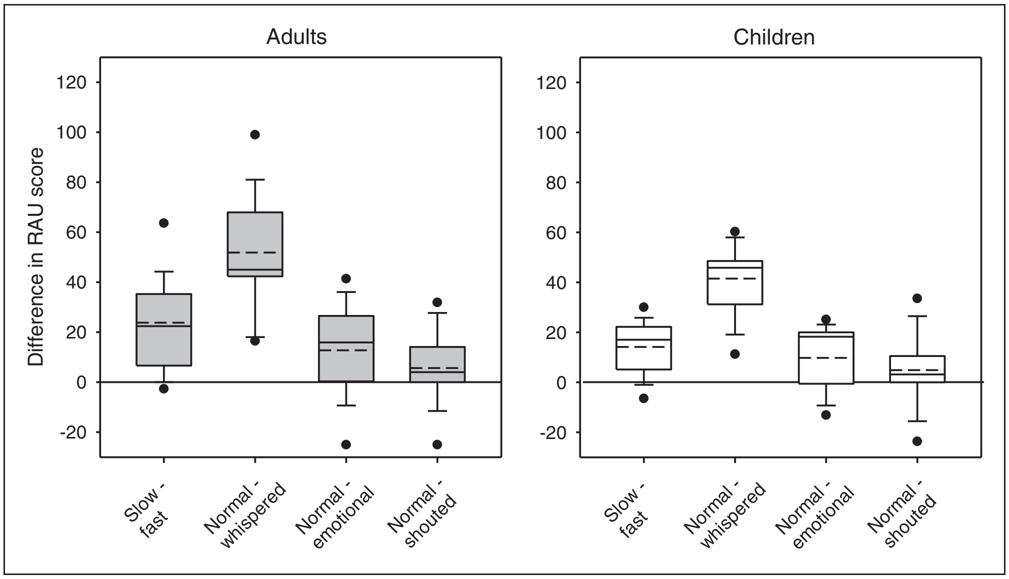

Figure 7. Boxplots of difference in RAU scores between various speaking styles, for adult (left panel) and pediatric CI subjects (right panel). The boxes show the 25th and 75th percentiles, the solid line shows median performance, the dashed line shows mean performance, the error bars show the 10th and 90th percentiles, and the circles show outliers.

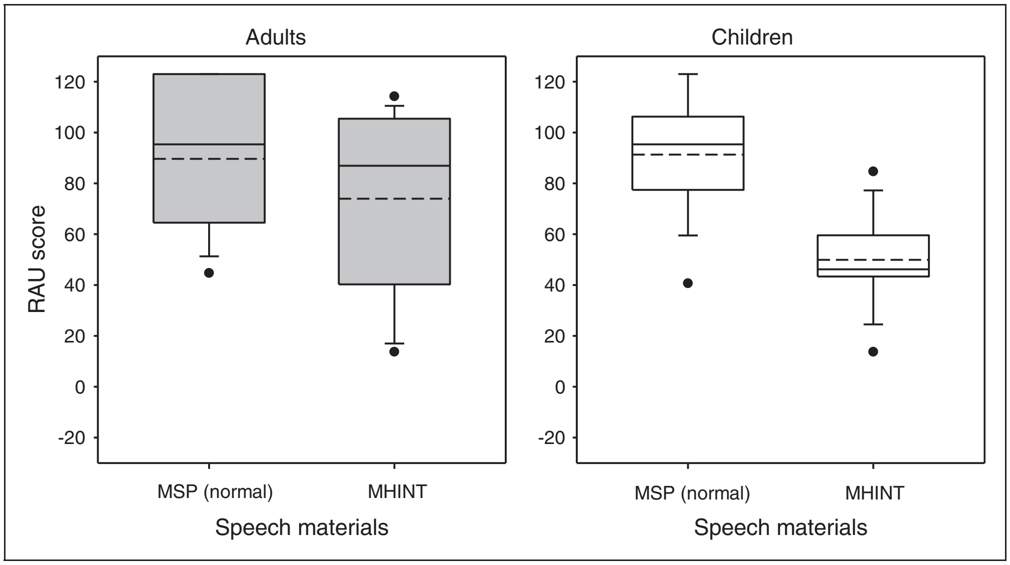

Figure 8. Boxplots of RAU scores for MSP (normal rate) and MHINT sentences, for adult (left panel) and pediatric CI subjects (right panel). The boxes show the 25th and 75th percentiles, the solid line shows median performance, the dashed line shows mean performance, the error bars show the 10th and 90th percentiles, and the circles show outliers.

was significantly better with the MSP than with the MHINT for adults, $F(1,14) = 10.9$, $p = .005$, and children, $F(1,8) = 146.4$, $p < .001$.

Table 3 shows r^2 values for correlations among the various speech materials; percent correct data were used for the correlations. In all cases, correlations were significant ($p < .05$). For both adult and pediatric subjects, r^2 values were generally small between whispered speech and the slow, normal, and fast speaking rates. Interestingly, r^2 values were relatively high between whispered speech and emotional or shouted speech for adult subjects but relatively low for pediatric subjects. Between the MSP and MHINT materials, r^2 values were lowest for the slow speech, with relatively high r^2 values for the rest of the MSP speaking styles.

Table 3. Correlations Among Performance With the Experimental Sentence Materials (r^2).

	Normal	Fast	Whispered	Emotional	Shouted	MHINT
Adult						
Slow	.82	.67	.42	.71	.80	.53
Normal		.82	.59	.69	.85	.75
Fast			.61	.81	.74	.87
Whispered				.73	.70	.75
Emotional					.80	.78
Shouted						.73
Pediatric						
Slow	.78	.77	.37	.73	.73	.68
Normal		.56	.48	.74	.70	.74
Fast			.43	.79	.79	.72
Whispered				.58	.42	.60
Emotional					.70	.72
Shouted						.70

Note. MHINT = Mandarin Hearing in Noise Test. In all cases, $p < .05$.

Performance (in percent correct) with the different MSP speaking styles and the MHINT sentences was compared with CI subject demographic factors age at testing, age at implantation, and duration of deafness; correlation analyses were performed separately for adults and children. Note that adult Subjects A2, A5, and A6 were excluded from the correlation analyses with duration of deafness because of uncertainty regarding the onset of severe-to-profound deafness. Note also that pediatric Subjects C1 and C3 were excluded from the correlations with the MHINT sentences as they were unable to complete this task due to time constraints. Results showed no significant correlations between any demographic variables and any of the speech tests in adults or in children ($p > .05$ in all cases).

Discussion

The present results show that both adult and pediatric Mandarin-speaking Chinese CI users' speech recognition can be affected by different speaking styles. For both subject groups, performance was poorest with whispered speech. While there was no significant difference in performance with the MSP materials between adult and pediatric CI subjects, performance was significantly better for adult CI subjects with the MHINT materials. Later, we discuss the results in greater detail.

Effects of Different Speaking Styles on Sentence Recognition

The present data were collected in CI subjects only. As such, there are no NH data for these particular stimuli. However, a related study showed that Mandarin-speaking NH listeners scored 100% correct with the same slow, normal, fast, and whispered MSP sentences (Li et al., 2011). The generally good performance with emotional and shouted speech (mean > 75% correct in both cases, across all subjects) for the present CI subjects suggests that NH performance would have been very good, if not perfect. Thus, we would expect a deficit in CI performance for all speech materials relative to NH listeners.

Speaking Rate

For the relatively easy slow-rate MSP sentences, recognition scores were quite variable for both adult and pediatric CI subjects. For the adult CI subjects, excellent performance (>90% correct) was observed in 9 of 15 subjects, with scores for the remaining 6 subjects ranging from 64% to 86% correct. For the pediatric CI subjects, excellent performance was observed in 8 of 11 subjects, with scores for the remaining 3 subjects ranging from 46% to 79% correct.

The mean performance deficit (17.5 percentage points) with fast Mandarin speech was less than reported by Ji et al. (2013) for fast English speech (40.0 percentage points). The mean rate for fast speech in the present study was 5.7 wps, considerably slower than the 6.6 wps for fast speech in Ji et al. (2013). It is possible that a further deficit might have been observed in this study with faster speaking rates. It is also possible that tonal language may be less susceptible to speaking rate differences due to covarying F0, amplitude, and duration cues. As shown in Figures 1 to 4, the amplitude cues, harmonic patterns, stimulation patterns, and F0 contours differed little among slow, normal, and fast speech. Even though CI users may not have been able to access F0 and harmonic information, covarying amplitude and relative duration cues may have been readily available. The present pediatric CI subjects seem to have made better use of these cues, possibly because of longer experience with the CI. Pre-lingually deafened pediatric CI users may also have learned to weight these cues more strongly during their development with electric hearing. Postlingually deafened adult CI users may have weighted F0 cues more strongly during acoustic hearing experience. After cochlear implantation, postlingually deafened adult CI users can no longer depend on F0 cues, as they are not well represented by the CI; as such these listeners must depend on duration and amplitude cues that may have been less weighted during acoustic hearing. It is possible that with longer CI experience, postlingually deafened adult CI users may use these cues as effectively as the present prelingually deafened pediatric CI users.

It is unclear why performance with fast speech was more variable with adults than with children.

As shown in Figure 7, the variability in the deficit with fast speech relative to slow speech was much greater in adult subjects, with RAU scores dropping by as much as 63.6 points (Subject A11); the maximum deficit in children was only 30 points (C1). The standard deviation across RAU scores for fast speech was more than twice as large for adults (35.9) than for children (17.8). For slow speech, the standard deviation in RAU scores was more comparable between adults (26.0) and children (19.5). Pediatric CI subjects had more than twice as much experience with their device (mean = 5.2 years) than did adult CI subjects (mean = 2.0 years), which may have contributed to the greater variability in performance with fast speech. It is possible that with greater CI experience, adult CI users may better accommodate fast speech.

Whispered Speech

Across all subjects, mean recognition of whispered speech dropped by 37.6 percentage points, relative to normal speech. As discussed in Li et al. (2011), the large deficit with whispered speech may have been due to the unavailability of periodicity cues. When only amplitude envelope cues (2–50 Hz) are available, lexical tone recognition has been shown to be approximately 60% correct for Mandarin-speaking NH subjects listening to whispered (Liang, 1963) or vocoded speech (Fu, Zeng, Shannon, & Soli, 1998). When temporal periodicity cues are available (50–500 Hz), lexical tone recognition greatly improves, which in turn significantly improves Chinese sentence recognition (Fu et al., 1998). The poor performance with whispered speech in the present study further highlights the importance of F0 cues to Chinese tone and sentence recognition by Mandarin-speaking CI users, even if these cues are poorly received.

The mean performance deficit with whispered speech (37.8 percentage points across all subjects) was more than twice as large as that reported by Fu et al. (1998) for NH subjects listening to four-channel vocoded Mandarin Chinese when periodicity cues were removed (15.4 percentage points) by reducing the envelope cutoff frequency 500 Hz to 50 Hz. While no periodicity cues were available with the 50 Hz envelope filter in Fu et al. (1998), amplitude contour cues were still available, which may explain some of the advantage over whispered speech in this study since amplitude envelope contour may have been flattened in the whispered speech, reducing an important cue for lexical tones when F0 information is unavailable (Fu & Zeng, 2000). The present deficit is also nearly twice as large as reported for NH musicians and nonmusicians listening to voiced versus whispered English speech by Ruggles et al. (2014), suggesting that the present subjects may have relied more strongly on the availability of F0 cues in normal speech (even if poorly represented). However, the unavailability of F0 information may not fully explain the present pattern of results. As shown in Figure 3, the formant information and transitions are somewhat preserved with whispered speech, but the spectral tilt is flatter (i.e., less stimulation in the apical region and more stimulation in the basal region than with normal speech). Stimulation on Electrodes 6 and 7 is nearly continuous throughout the sentence for whispered speech, as opposed to better defined stimulation patterns with normal speech. Ito, Takeda, and Itakura (2005) found an upward shift in vowel formant frequencies for whispered speech, compared with normal speech. Voiced consonants in whispered speech have lower energy below 1.5 kHz, with greater spectral flatness compared with normal speech. The authors also found that training with whispered speech significantly improved recognition of whispered speech recognition for the trained talker, suggesting that training may be one approach toward improving perception of voiceless speech.

Although there was no significant difference between subject groups, mean performance with whispered speech was better in children than in adults, as was performance for all speaking styles except for slow speech, reflecting perhaps a slight overall advantage for the present pediatric CI subjects. The correlations among the different speaking styles in Table 3 revealed consistently lower r^2 values between whispered speech and the other speaking styles for children than for adults. This pattern of results may indicate that children were more dependent on voice pitch cues, or that adults made better use of envelope cues. In some ways, it is difficult to explain such performance differences in adult and pediatric CI users. Some adult CI subjects were implanted after long periods of hearing impairment or auditory deprivation. Unfortunately, information regarding initial diagnosis, extent, and severity of hearing loss was not consistently available for the adult subjects, due to poor clinical documentation. In contrast, pediatric CI users (the large majority of CI recipients in China) have developed with electric hearing only. It is unclear whether they have developed better use of pitch or envelope cues (or both). Early implanted Chinese pediatric CI users are continually exposed to the “pitchiness” of “motherese,” as well as to the lexical tones in Mandarin, which may have resulted in different central patterns for speech that may have been somewhat more susceptible to the absence of voice cues with whispered speech. Recent work by Volkova, Trehub, Schellenberg, Papsin, and Gordon (2014) showed that pediatric CI users’ melody recognition with pitch cues alone remained near chance level, despite years of development with the CI alone. This suggests that pediatric CI users may not develop better pitch representations with electric hearing and are similarly limited by the poor spectral resolution of the CI device as are adult CI users.

Emotional and Shouted Speech

As shown in Table 2, the mean speaking rates were comparable between normal (3.77 wps) and emotional speech (3.55 wps). However, mean F0 differed greatly between normal (230 Hz) and emotional speech (314 Hz). As shown in Figure 3, the stimulation patterns for normal and emotional speech are quite similar, although the shifted F0 and a weaker stimulation of the middle electrodes can be observed for emotional speech. For adults, mean performance was poorer with emotional (71.9% correct) than with normal speech (81.0% correct). However, the deficit with emotional speech was quite variable, ranging from−10.0 to 37.1 percentage points. For children, the difference in mean performance was smaller between emotional (82.0% correct) and normal speech (84.7% correct), as was the variability in difference scores (range = −7.1 to 17.1 percentage points). Mean pediatric performance was 10.1 percentage points better than that of adults. Pediatric subjects appeared to be more robust to the shifted F0 information and other distortions to the stimulation pattern (relative to normal speech) than did adult subjects. It could be that longer experience with the CI (5.2 years for pediatric patients versus 2.0 years for adult patients) may have helped in accommodating emotional speech patterns.

The mean speaking rate for shouted speech (3.03 wps) was slightly slower than for normal speech (3.77 wps) but faster than for the slow speech (2.48 wps). Similar to emotional speech, the mean F0 for normal speech (230 Hz) was lower than for shouted speech (400 Hz); the mean F0 was 86 Hz higher for shouted than for emotional speech. As shown in Figure 3, the stimulation pattern for shouted speech contained less apical stimulation and greater mid-basal stimulation than observed for normal speech. For adults, mean performance was poorer with shouted (76.7% correct) than with normal speech (81.0% correct); similarly, mean performance for children was poorer with shouted (80.3% correct) than with normal speech (84.7% correct). The deficit with shouted speech was similar between subject groups, ranging from −.6 to 27.1 percentage points for adults and from −15.7 to 22.1 percentage points for children. Given that the stimulation patterns were more distorted for shouted than for emotional speech, relative to normal speech, it is unclear why performance was similar between shouted and emotional speech or why performance seemed more variable (especially for adults) with emotional speech. In general, the present CI subjects seemed able to accommodate the emotional and shouted speaking styles.

Effects of Test Materials

The MSP sentence materials used in the present study have been used with adult and pediatric NH and CI listeners in many other research studies (e.g., Chen, Wong, & Wong, 2013; Gao et al., 2016; Li et al., 2011; Meng, Zheng, & Li, 2016; Tao et al., 2014; Zhang, Xie, Li, Chatterjee, & Ding, 2014; Zhu et al., 2011). The MHINT sentence materials used in this study have also been used in previous studies but primarily with adult listeners (e.g., Chen et al., 2013; Song, Li, & Wang, 2011; Stuart, Zhang, & Swink, 2010; Zhang et al., 2010). Because the materials were not developed with children in mind, the MHINT sentences cannot be not be used to evaluate pediatric CI patients because the sentences are too difficult for children to understand or too long for children to remember. Also, the MHINT materials were not explicitly phonetically balanced across lists but rather balanced in terms of intelligibility in noise. Although the MHINT materials have been used in many CI studies, the lists were never validated in CI users or with a CI simulation. In contrast, the MSP materials were phonetically balanced across lists, and lists were balanced in intelligibility with both unprocessed speech and vocoded speech.

For the adult CI subjects in the present study, mean recognition of MHINT sentences (69.2% correct) was poorer than that of MSP sentences at the normal rate (81.0% correct) but comparable to that of MSP sentences with the fast rate (70.4% correct). For pediatric CI subjects, mean recognition of MHINT sentences (50.0% correct) was poorer than that of MSP sentences at the normal (84.7% correct) or fast rate (74.8% correct). The speaking rate of MHINT sentences (4.5 wps) was between the normal-rate (3.7 wps) and fast-rate of MSP sentences (5.7 wps). While speaking rate may explain differences in performance for adult subjects between the test materials, it does not explain the much poorer performance with the MHINT materials for the pediatric patients. One major difference between the MHINT and MSP sentences is the number of words in each sentence (10 words for MHINT vs. 7 words for MSP). Tao et al. (2014) found that auditory working memory may contribute to pediatric CI users' difficulties in speech understanding. In that study, the mean score for forward digit span was 6.1 numbers for pediatric CI users, suggesting that pediatric CI users can store approximately 6 to 7 keywords in the short-term memory. This corresponds to the number of words in the MSP sentences but is much less than the 10 words in the MHINT sentences used in this study. Thus, auditory working memory may have played a role in the deficit observed with the MHINT for pediatric CI subjects in this study. Another difference between the MSP and MHINT materials was the talker, with the MSP using a female talker (mean F0 = 230 Hz) and the MHINT using a male talker (mean F0 = 131 Hz). It is possible that the female talker better aligned with the pediatric subjects' speech pattern templates developed while wearing the CI. As there was no significant difference between subject groups, the MSP materials appear to be more appropriate for evaluating adult and pediatric Mandarin-speaking Chinese CI users, at least in quiet.

Effects of Demographic Factors

There were no significant correlations among demographic factors age at testing, age at implantation, or duration deafness, and any of the speech materials, in adults or in children. In analyzing demographic factors, Mandarin Chinese CI users present a different set of circumstances than for Western (United States, Europe) CI users. Historically, pediatric patients were the first to receive CIs in large numbers in China, whereas adults (typically postlingually deafened) were the first CI recipients in Western countries. Recently, more adults have been implanted in China but often with a longer duration of deafness than typical for postlingually deafened Western adult CI recipients. In this study, the mean duration of deafness was 5.8 years, and ranged from 0.1 to 25.5 years. In general, inconsistent and nonstandard hearing screening has been a long-standing issue for Chinese audiometry (Ma, McPherson, & Ma, 2013) and may have contributed to potential over- or under-estimations of duration of deafness in this study. The pediatric CI users in this study had more than twice as much as experience with their device (mean = 5.2 years) compared with the adults CI users (mean = 2.0 years). It is unclear whether pediatric subjects' longer CI experience offset the lack of acoustic hearing experience. It is possible that pediatric CI subjects may have shown an advantage in some measures due to greater experience with different speaking styles via electric hearing.

Clinical Implications

In clinical practice, speech perception is typically evaluated using clear speech materials. Relatively few CI studies have examined the effects of speaking style on speech understanding, such as clear versus conversational (Liu et al., 2004), slow versus fast speaking rates (Ji et al., 2013), and whispered versus voiced speech (Li et al., 2011). For voiced speech, the present data suggest that speaking rate may affect CI performance and should be carefully considered when standardizing tests for clinical evaluation. Similarly, the present data suggest that auditory working memory should be considered when standardizing clinical speech materials, especially if pediatric patients are to be tested and performance compared with adult patients.

In this study, performance with the different speaking styles was measured in quiet. In noise, performance may deteriorate rapidly for variable speech. As noisy environments are common to all CI users, future studies should evaluate performance with variable speech in noise.

Summary and Conclusion

In this study, sentence recognition of Mandarin sentences produced in different speaking styles was measured in adult and pediatric Chinese CI users. Major findings are as follows:

1. There was a large and significant deficit in recognition of whispered speech for both adult and pediatric CI subjects, relative to voiced speech, demonstrating the importance of pitch cues to speech understanding for Mandarin-speaking CI users.
2. Pediatric CI subjects were more robust to changes in speaking rate and speaking style than were adult CI subjects.
3. Chinese CI subjects, especially pediatric subjects, had greater difficulty with the MHINT than the MSP speech materials, most likely due to the faster speaking rate and the greater number of words per sentence, suggesting that auditory memory should be considered when selecting speech materials to evaluate CI performance.

Acknowledgements

The authors thank all the subjects who participated in this study.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partially supported by NIH grant DC004993 and Beijing Municipal Administration of Hospitals XMLX201409.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636.

Chang, Y. P., & Fu, Q. J. (2006). Effects of talker variability on vowel recognition in cochlear implants. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1331–1341.

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, X., Liu, B., Liu, S., Mo, L., Li, Y., Kong, Y., . . . Han, D. (2013). Cochlear implants with fine structure processing improve speech and tone perception in Mandarin-speaking adults. *Acta Otolaryngologica*, *133*(7), 733–738.

Chen, F., Wong, L. L., & Wong, E. Y. (2013). Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility. *Journal of the Acoustical Society of America*, *134*(2), EL178–EL184.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595.

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. *Journal of the Acoustical Society of America*, *107*(5), 2704–2710.

Eskenazi, M. (1993). Trends in speaking styles research. *Proceedings of EUROSPEECH '93, Berlin*, 501–509.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*(6), 1497–1510.

Freyman, R. L., Griffin, A. M., & Oxenham, A. J. (2012). Intelligibility of whispered speech in stationary and modulated noise maskers. *Journal of the Acoustical Society of America*, *132*(4), 2514–2523.

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

Fu Q. J. & Zeng F. G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific journal of speech, language, and hearing*, *5*, 45–57.

Fu, Q. J., Zhu, M., & Wang, X. (2011). Development and validation of the Mandarin speech perception test. *Journal of the Acoustical Society of America*, *129*(6), EL267–EL273.

Gao, N., Xu, X. D., Chi, F. L., Zeng, F. G., Fu, Q. J., Jia, X. H., . . . Jiang, Y. (2016). Objective and subjective evaluations of the Nurotron Venus cochlear implant system via animal experiments and clinical trials. *Acta Otolaryngologica*, *136*(1), 68–77.

Ito, T., Takeda, K., & Itakura, F. (2005). Analysis and recognition of whispered speech. *Speech Communication*, *45*(2), 139–152.

Ji, C., Galvin, J. J., Chang, Y. P., Xu, A., & Fu, Q. J. (2014). Perception of speech produced by native and nonnative talkers by listeners with normal hearing and listeners with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *57*(2), 532–554.

Ji, C., Galvin, J. J. 3rd, Xu, A., & Fu, Q. J. (2013). Effect of speaking rate on recognition of synthetic and natural speech by normal-hearing and cochlear implant listeners. *Ear and Hearing*, *34*(3), 313–323.

Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1395–1405.

Li, Y., Zhang, G., Kang, H. Y., Liu, S., Han, D., Fu, Q. J. (2011). Effects of speaking style on speech intelligibility for Mandarin-speaking cochlear implant users. *Journal of the Acoustical Society of America*, *129*(6), EL242–EL247.

Liang, Z. A. (1963). The auditory perception of Mandarin tones. *Acta Physica Sinica*, *26*, 85–91.

Lin, M. C. (1988). The acoustic characteristics and perceptual cues of tones in Standard Chinese. *Chinese Yuwen*, *204*, 182–193.

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F. G. (2004). Clear speech perception in acoustic and electric hearing. *Journal of the Acoustical Society of America*, *116*(4), 2374–2383.

Liu, C., Galvin, J. J. 3rd, Fu, Q.-J., & Narayanan, S. S. (2008). Effect of spectral normalization on different talker speech recognition by cochlear implant users. *Journal of the Acoustical Society of America*, *123*(5), 2836–2847.

Luo, X., Fu, Q. J., & Galvin, J. J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*, *11*(4), 301–315.

Ma, X., McPherson, B., & Ma, L. (2013). Chinese speech audiometry material: Past, present, future. *Hearing, Balance and Communication*, *11*(2), 52–56.

Meng, Q., Zheng, N., & Li, X. (2016). Mandarin speech-in-noise and tone recognition using vocoder simulations of the temporal limits encoder for cochlear implants. *Journal of the Acoustical Society of America*, *139*(1), 301–310.

Mitchell, R. L. (2007). fMRI delineation of working memory for emotional prosody in the brain: Commonalities with the lexico-semantic emotion network. *Neuroimage*, *36*(3), 1015–1025.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, *93*(2), 1097–1108.

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 Neurootology*, *14*(5), 327–337.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing. I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*(1), 96–103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*(4), 434–446.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1989). Speaking clearly for the hard of hearing. III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, *32*(3), 600–603.

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.

Shannon, R. V., Fu, Q. J., & Galvin, J. J. III (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngologica Supplement*, *552*, 50–54.

Smiljanić, R., & Bradlow, A. R. (2008). Temporal organization of English clear and conversational speech. *Journal of the Acoustical Society of America*, *124*(5), 3171–3182.

Sommers, M., Nygaard, L., & Pisoni, D. (1992). *Stimulus variability and the perception of spoken words: Effects of variation in speaking rate and overall amplitude*. The Second International Conference on Spoken Language Processing, ICSLP 1992, Banff, Alberta, Canada, October 13–16, 1992, 217–220.

Song, P. L., Li, H. J., & Wang, N. Y. (2011). Benefits of spatial hearing to speech recognition in young people with normal hearing. *Chinese Medical Journal (English)*, *124*(24), 4269–4274.

Soto, J. A., & Levenson, R. W. (2009). Emotion recognition across cultures: The influence of ethnicity on empathic accuracy and physiological linkage. *Emotion*, *9*(6), 874–884.

Stuart, A., Zhang, J., & Swink, S. (2010). Reception thresholds for sentences in quiet and noise for monolingual English and bilingual Mandarin-English listeners. *Journal of the American Academy of Audiology*, *21*(4), 239–248.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*, 455–462.

Tao, D., Deng, R., Jiang, Y., Galvin, J. J. 3rd, Fu, Q. J., Chen, B. (2014). Contribution of auditory working memory to speech understanding in mandarin-speaking cochlear implant users. *PLoS One*, *9*(6), e99096.

Trainor, L. J., Austin, C. M., & Desjardins, N. (2000). Is infant-directed speech a result of the vocal expression of emotion? *Psychological Science*, *11*, 188–195.

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America*, *107*, 3438–3451.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Disorders*, *39*, 494–509.

van Rijn, S., Aleman, A., van Diessen, E., Berckmoes, C., Vingerhoets, G., Kahn, R. S. (2005). What is said or how it is said makes a difference: Role of the right fronto-parietal operculum in emotional prosody as revealed by repetitive TMS. *European Journal of Neuroscience*, *21*(11), 3195–3200.

Volkova, A., Trehub, S. E., Schellenberg, E. G., Papsin, B. C., & Gordon, K. A. (2014). Children's identification of familiar songs from pitch and timing cues. *Frontiers in Psychology*, *5*, 863. doi:10.3389/fpsyg.2014.00863.

Wong, L. L., Soli, S. D., Liu, S., Han, N., & Huang, M. W. (2007). Development of the Mandarin Hearing in Noise Test (MHINT). *Ear and Hearing*, *28*(2 Suppl): 70S–74S.

Xu, L., Tsai, Y., & Pfingst, B. (2002). Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *Journal of the Acoustical Society of America*, *112*(1), 247–258.

Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Busso, C., Deng, Z., . . . Narayanan, S. (2004). An acoustic study of emotions expressed in speech. In: *Proceedings International Conference on Spoken Language Processing (ICSLP '04)* (pp. 2193–2196).

Zhang, C., & Hansen, J. H. L. (2007). Analysis and classification of speech mode: Whispered through shouted," in *Proceedings from Interspeech 2007*, Antwerp, Belgium, pp. 2289–2292.

Zhang, C., & Hansen, J. H. L. (2011). Whisper-island detection based on unsupervised segmentation with entropy-based speech feature processing. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(4), 883–894.

Zhang, N., Liu, S., Xu, J., Liu, B., Qi, B., Yang, Y., . . . Han, D. (2010). Development and applications of alternative methods of segmentation for Mandarin hearing in noise test in normal-hearing listeners and cochlear implant users. *Acta Otolaryngologica*, *130*(7), 831–837.

Zhang, J., Xie, L., Li, Y., Chatterjee, M., & Ding, N. (2014). How noise and language proficiency influence speech recognition by individual non-native listeners. *PLoS One*, *9*(11), e113386.

Zhu, M., Fu, Q. J., Galvin, J. J. 3rd, Jiang, Y., Xu, J., Xu, C., Tao, D., . . . Chen, B. (2011). Mandarin Chinese speech recognition by pediatric cochlear implant users. *International Journal of Pediatric Otorhinolaryngology*, *75*(6), 793–800.