



Published in final edited form as:

Ear Hear. 2016 ; 37(3): 271–281. doi:10.1097/AUD.0000000000000265.

Mandarin Tone and Vowel Recognition in Cochlear Implant Users: Effects of Talker Variability and Bimodal Hearing

Yi-ping Chang^a, Ronald Y. Chang^b, Chun-Yi Lin^a, and Xin Luo^c

^aSpeech and Hearing Science Research Institute, Children's Hearing Foundation, Taipei, Taiwan

^bResearch Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

^cDepartment of Speech, Language, and Hearing Sciences, Purdue University, 715 Clinic Drive, West Lafayette, IN 47907, USA

Abstract

Objectives—For cochlear implant (CI) users with residual low-frequency acoustic hearing in the non-implanted ear, bimodal hearing combining the use of a CI and a contralateral hearing aid (HA) may provide more salient talker voice cues than CI alone to handle the variability of talker identity across trials. This study tested the effects of talker variability, bimodal hearing, and their interaction on response accuracy and time of CI users' Mandarin tone, vowel, and syllable recognition (i.e., combined Mandarin tone and vowel recognition in this study).

Design—Fifteen prelingually deafened native Mandarin-speaking CI users (at age 20 or lower) participated in this study. Four talkers each produced six Mandarin single-vowel syllables in four lexical tones. The stimuli were presented in quiet via a single loudspeaker. To study the effects of talker variability, Mandarin tone, vowel, and syllable recognition was tested in two presentation conditions: with stimuli blocked according to talker (blocked-talkers condition) or mixed across talkers from trial to trial (mixed-talkers condition). To explore the effects of bimodal hearing, two processor conditions were tested: CI alone or CI+HA. The cumulative response time was recorded as an indirect indicator of the cognitive load or listening effort in each condition. The correlations were computed between demographic/hearing factors (e.g., hearing thresholds in the non-implanted ear) and bimodal performance/benefits (where bimodal benefits refer to the performance differences between CI alone and CI+HA).

Results—Mandarin tone recognition with both CI alone and CI+HA was significantly poorer in the mixed-talkers condition than in the blocked-talkers condition, while vowel recognition was comparable in the two presentation conditions. Bimodal hearing significantly improved Mandarin tone recognition but not vowel recognition. Mandarin syllable recognition was significantly affected by both talker variability and bimodal hearing. The cumulative response time significantly reduced with CI+HA compared with CI alone, but remained invariant with respect to talker variability. There was no interaction between talker variability and bimodal hearing for any performance measure adopted in this study. Correlation analyses revealed that the bimodal performance and benefits in Mandarin tone, vowel, and syllable recognition could not be predicted

by the hearing thresholds in the non-implanted ear or by the demographic factors of the participants.

Conclusions—Talker variability from trial to trial significantly degraded Mandarin tone and syllable recognition performance in both the CI alone and CI+HA conditions. While bimodal hearing did not reduce the talker variability effects on Mandarin tone and syllable recognition, generally better Mandarin tone and syllable recognition performance with shorter response time (an indicator of less listening effort) was observed when a contralateral HA was used in conjunction with the CI. On the other hand, vowel recognition was not significantly affected by either talker variability or bimodal hearing, because ceiling effects could not be counted out of the vowel recognition results.

Keywords

cochlear implant; bimodal hearing; talker variability; Mandarin tone recognition; vowel recognition; listening effort

INTRODUCTION

Human listeners recognize different phonemes based on acoustic features such as voice onset time, formant frequencies, and formant transitions. However, these acoustic cues vary across different talkers with different speaking rates and vocal tract lengths. The cross-talker acoustic differences may be perceptually confused with cross-phoneme acoustic differences. Much speech perception research has focused on understanding a listener's ability to reliably extract the phonological information from the variable speech signals produced by different talkers. Researchers have theorized a "talker normalization" process (e.g., Krulee et al., 1983; Klatt, 1986; Pisoni, 1993), which enables listeners to normalize the cross-talker acoustic variations and achieve perceptual constancy. For example, when listening to a single talker, listeners may gain talker-specific cues and adapt to the voice characteristics of the particular talker (Nygaard and Pisoni, 1998). When the talker voice varies from trial to trial in the mixed-talker condition, normal-hearing (NH) listeners' speech recognition performance in noise degrades as compared to when the talker voice remains invariant in the blocked-talker condition, a phenomenon called "talker variability effects" (Creelman, 1957; Mullennix et al., 1989; Ryalls and Pisoni, 1997). The adverse talker variability effects on speech recognition reveal the greater challenge of talker normalization in the mixed-talker condition, since more frequent talker normalization process is needed as compared to the blocked-talker condition.

Talker variability is pertinent to the recognition of not only phonemic cues such as vowels/consonants (e.g., Creelman, 1957; Mullennix et al., 1989; Ryalls and Pisoni, 1997), but also prosodic cues such as lexical tones (e.g., Wong and Diehl, 2003; Lee et al., 2010). Mandarin is a tonal language with four lexical tones that carry lexical meanings. For example, the syllable /ma/ can mean *mother* in Tone 1, *hemp* in Tone 2, *horse* in Tone 3, and *scold* in Tone 4. Mandarin tone recognition is thus an essential part of Mandarin speech understanding. The Mandarin tones are primarily characterized by pitch levels and pitch contours (Tone 1: high-flat, Tone 2: mid-rising, Tone 3: low-falling-rising, and Tone 4: high-falling; e.g., Chao, 1948). Because the main acoustic correlate of pitch (i.e., fundamental

frequency or F0) varies across different talkers with different glottal pulse rates, listeners need to recognize Mandarin tones with reference to the voice pitch range of the talker. For example, the same Mandarin tone may be perceived as a low lexical tone for a talker with a high-pitched voice, or as a high lexical tone for a talker with a low-pitched voice. Information about the voice pitch range of the talker is particularly important for the recognition of Cantonese level tones differing only in pitch levels. Thus, NH listeners' lexical tone recognition in both Mandarin (Lee et al., 2010) and Cantonese (Wong and Diehl, 2003) deteriorated when the talker voice varied across trials as compared to when the talker voice remained invariant.

In studies of talker variability effects on speech recognition (e.g., Mullennix et al., 1989; Lee et al., 2010), response time was also longer in the mixed-talker condition than in the blocked-talker condition. Although there was no direct evidence, longer response time in the mixed-talker condition suggests that talker variability may have increased the cognitive load and occupied the cognitive resources for speech recognition, resulting in lower recognition scores (Mullennix et al., 1989). Longer response time in the mixed-talker condition may also indicate greater listening effort with talker variability, because listening effort, although often loosely defined, includes cognitive demands and other factors. Actually, response time has been used as a measure of listening effort (e.g., Gatehouse and Gordon, 1990; Baer et al., 1993; Houben et al., 2013) and has yielded results consistent with other measures of listening effort such as pupil dilation (Zekveld et al., 2006, 2010). Together, talker variability may increase cognitive load or listening effort while reducing speech recognition scores.

Cochlear implants (CIs) electrically stimulate auditory neurons via a small number of implanted electrodes to help profoundly deaf people regain hearing. Studies have shown that crude CI signals can support good speech recognition in quiet but contain only limited talker voice information such as the F0 and formants reflecting the glottal pulse rate and vocal tract length, respectively (e.g., Cleary and Pisoni, 2002; Vongphoe and Zeng, 2005). For example, voice pitch cues are mainly encoded by the amplitude modulation (AM) rates of pulse trains on individual electrodes in CIs (Geurts and Wouters, 2001), because the F0 and harmonics are unresolved with the small number of implanted electrodes (12–22). However, CI users can only discern temporal pitch changes for frequencies up to around 300 Hz (e.g., Luo et al., 2010). Thus, both spectral and temporal resolutions are greatly decreased for CI users.

It is generally accepted that listeners need talker voice information and talker-specific acoustic-phonetic cues to achieve talker normalization. When these perceptual cues are not clearly transmitted, as in CIs, talker normalization may require more cognitive resources, and consequently, the talker variability effects on speech recognition may be more severe for CI users than for NH listeners. Previous results regarding the talker variability effects on CI users' speech recognition have been inconsistent. Kirk et al. (2000) found that pediatric CI users had better (rather than worse) word recognition in the mixed-talker condition than in the blocked-talker condition. In their study design, only one of the talkers in the mixed-talker condition was tested in the blocked-talker condition. This talker may be less intelligible to the participants than the other talkers and the possible difference in talker intelligibility may have complicated the testing of talker variability effects. Using a similar design, Kaiser et al. (2003) found that for adult CI users, single-talker word recognition was

better than multi-talker word recognition in audiovisual mode, but not in visual only and auditory only modes. Different from previous studies, Chang and Fu (2006) tested the same talkers and stimuli in both the blocked- and mixed-talker conditions and showed better vowel recognition for adult CI users when the talker voice remained invariant as compared to when the talker voice varied across trials. However, they repeated the stimuli eight times in the blocked-talker condition but only two times in the mixed-talker condition. This design rendered it unclear whether CI users' poorer vowel recognition in the mixed-talker condition was due to talker variability or fewer repeats of stimuli. While the talker variability effects on CI users' recognition of phonemic cues such as vowels and consonants remain a topic of debate (as summarized here), the talker variability effects on CI users' recognition of prosodic cues such as Mandarin tones have not been addressed.

For CI users with residual low-frequency acoustic hearing in the non-implanted ear, bimodal hearing with the combined use of a CI and a contralateral hearing aid (HA) may enhance the access to low-frequency acoustic cues (e.g., the F0, harmonics, and first formant). Compared to CI alone, CI+HA has been shown effective in improving word and sentence recognition, although evidence of bimodal benefits to phoneme and lexical tone recognition has been mixed (Dorman et al., 2008; Luo and Fu, 2006; Yuen et al., 2009). More talker voice information may be contained in aided residual acoustic hearing than in electric hearing, as indicated by better talker (especially gender) identification with CI+HA than with CI alone (Krull et al., 2010). We hypothesized that, since a contralateral HA may enhance talker voice information and talker-specific acoustic-phonetic cues, talker normalization may be more feasible and speech recognition may be more immune to talker variability for bimodal users. To our knowledge, the perspective of talker variability in conjunction with bimodal hearing has not been previously examined.

This study tested Mandarin tone, vowel, and syllable recognition with either CI alone or CI+HA in native Mandarin-speaking prelingually deafened CI users. The same talkers and stimuli were tested without repetition in both blocked- and mixed-talker conditions. The cumulative response time, as a primitive indicator of the cognitive load or listening effort (Mullennix et al., 1989; Gatehouse and Gordon, 1990; Houben et al., 2013), was recorded in each condition. This design allowed us to investigate the effects of talker variability, bimodal hearing, and their interactions on CI users' Mandarin tone, vowel, and syllable recognition. Based on the literature review, we hypothesized that Mandarin tone, vowel, and syllable recognition may improve and the response time may reduce in the blocked-talker condition than in the mixed-talker condition and in the CI+HA condition than in the CI alone condition. Furthermore, we hypothesized that the effects of talker variability on Mandarin tone, vowel, and syllable recognition (i.e., the performance differences between blocked- and mixed-talker conditions) may lessen in the CI+HA condition than in the CI alone condition.

METHODS

Participants

Eight male and seven female native Mandarin-speaking CI users aged between 10 and 20 years with an average age of 15 years took part in this study. Table 1 lists the demographic

information of participants. All participants were prelingually deafened and have used CI for more than one year. One group of participants (listed in the upper section of Table 1) were bimodal users tested in both CI alone and CI+HA conditions. Figure 1 shows the bimodal users' unaided and aided hearing thresholds in the non-implanted ear. The other group of participants (listed in the lower section of Table 1) were tested in the CI alone condition only. They did not wear a contralateral HA because they had no residual acoustic hearing in the non-implanted ear. All participants started learning oral communication shortly after the diagnosis of hearing loss by enrolling in the Auditory Verbal Therapy at the Children's Hearing Foundation in Taiwan. No participant had bilingual or extensive musical experience. All participants and parents of the participants who were younger than 20 years of age gave informed consent and received compensation for their participation in this study.

Test Materials

Mandarin tone, vowel, and syllable recognition was tested using a subset of the stimuli in Luo and Fu (2005). Two female (FT1 and FT2) and two male (MT1 and MT2) native Mandarin-speaking talkers each produced six single-vowel syllables (/a/, /o/, /e/, /i/, /u/, and /ü/ in Pinyin or [a], [o], [ɤ], [i], [u], and [y] in International Phonetic Alphabet) in four lexical tones (Tone 1: high-flat, Tone 2: mid-rising, Tone 3: low-falling-rising, and Tone 4: high-falling), creating a total of 96 lexically meaningful Mandarin syllable tokens. The first two formant frequencies of these stimuli can be found in Figure 1 of Luo and Fu (2005) and the F0 contours of the stimuli produced by FT1 and MT1 can be found in Figure 1 of Luo et al. (2009). Across the four lexical tones, FT1 had an F0 ranging from 76 to 317 Hz (mean: 213 Hz), FT2 had an F0 ranging from 75 to 315 Hz (mean: 214 Hz), MT1 had an F0 ranging from 76 to 211 Hz (mean: 131 Hz), and MT2 had an F0 ranging from 80 to 248 Hz (mean: 135 Hz). The stimuli were digitized at a 16-kHz sampling rate with 16-bit resolution and were normalized to have the same long-term root mean square energy.

Procedure

This study was reviewed and approved by the local institutional review board (IRB) committees. Testing was performed in a sound-treated room with the walls providing a 44-dB sound attenuation at the Children's Hearing Foundation in Taiwan. A single loudspeaker (Kinyo ps-285b) was placed half a meter in front of the participant to present the stimuli at the participant's most comfortable level (ranging from 60 to 70 dBA). For each participant, the presentation level was held consistent in different testing conditions. Participants used their own clinical CI and HA (if available) devices with audiologist-recommended programs and settings that remained unchanged during testing. Half of the bimodal users were tested in the order of first CI alone and then CI+HA, and the other half in the reverse order. P1 was only tested with CI+HA due to time limitations.

In each processor condition, Mandarin tone, vowel, and syllable recognition was tested in blocked- and mixed-talker presentation conditions. Half of the participants were tested in the order of first blocked-talker condition and then mixed-talker condition, and the other half in the reverse order. In both presentation conditions, the same 96 syllable tokens (six vowels × four lexical tones × four talkers) were tested only once. Each condition comprised of four blocks with a one-minute break between blocks. In the blocked-talker condition, each block

contained 24 syllable tokens (six vowels \times four lexical tones) from a particular talker. The testing orders of the talker blocks and of the syllable tokens within each talker block were both randomized. In the mixed-talker condition, each block contained 24 syllable tokens from all four talkers and the tokens were randomly presented without replacement. The whole testing took an hour over one day.

As in Luo and Fu (2009) and Luo et al. (2009), combined Mandarin tone and vowel recognition (or Mandarin single-vowel syllable recognition) was tested to simulate everyday listening situations and to save testing time. In each trial, a syllable token was selected from the stimulus set (without replacement) according to the aforementioned rules and presented to the participant. There were 24 response buttons (six columns \times four rows) on a computer screen. Each button was labeled with a letter showing the Mandarin vowel (/a/, /o/, /e/, /i/, /u/, and /ü/ for column one to six, respectively) followed by a number representing the Mandarin tone (1, 2, 3, and 4 for row one to four, respectively). Participants were asked to identify the Mandarin tone and vowel in the presented token by clicking on the corresponding response button. Each token was presented only once and no feedback was provided. Participants were told that they should take their best guesses and their responses will be timed. These instructions were given to each participant before testing using five randomly selected sample stimuli. As in Lee et al. (2010), the cumulative response time between the stimulus offset and the button click of individual trials for each testing condition was recorded. The responses were scored for the percentage of correct identification of Mandarin tones, vowels, and syllables (i.e., both tone and vowel were correctly identified). Chance level was 25%, 16.67%, and 4.17% correct for Mandarin tone, vowel, and syllable recognition, respectively.

Data Analyses

Each outcome measure (i.e., Mandarin tone, vowel, and syllable recognition scores, as well as the cumulative response time) was analyzed using a univariate general linear model at a significance level of 0.05 in SPSS version 20 (IBM, Armonk, NY, USA). In each model, the outcome measure was treated as the dependent variable, participant as the random factor, and presentation (mixed- or blocked-talker) and processor conditions (CI alone or CI+HA) as the fixed factors. The interaction between the fixed factors was also included in the model. Due to the concern of ceiling effects, all percent correct scores were arcsine-transformed (Studebaker, 1985) before analyses. In each model, the CI alone and CI+HA conditions were compared only for the nine participants tested in both processor conditions.

Mandarin tone recognition scores were compared across tones using a general linear model similar to the one described above, except that a fixed factor tone (Tones 1, 2, 3, and 4) and its interactions with the other factors (i.e., presentation and processor conditions) were added to the model. The effects of talker variability and bimodal hearing on the recognition scores of each Mandarin tone were further tested using the same general linear model as for the overall tone recognition scores. Mandarin tone recognition scores were also compared across talkers (FT1, FT2, MT1, and MT2) and acoustic features (e.g., durations and F0 ranges) of Mandarin tones were used to explain the talker differences. Vowel recognition scores for

individual vowels (/a/, /o/, /e/, /i/, /u/, and /ü/) or talkers (FT1, FT2, MT1, and MT2) were analyzed in similar manners as well.

Speech recognition scores in the CI alone and CI+HA conditions and the performance difference between them (i.e., bimodal benefits) were characterized by great inter-subject variability (e.g., Ching et al., 2004; Gifford et al., 2007; Yuen et al., 2009). The age at implantation and duration of CI use have been identified as two major demographic factors that may predict performance in pediatric CI users (e.g., Kirk et al., 2002; Zwolan et al., 2004; Han et al., 2009). Studies have also tested pure-tone thresholds among other psychophysical measures in the non-implanted ear as potential factors underlying individual differences in bimodal performance/benefits (e.g., Ching et al., 2004; Gifford et al., 2007; Yuen et al., 2009). Following these studies, we computed a series of Pearson correlations between the aided hearing thresholds in the non-implanted ear (at 250, 500, 1000 Hz, or averaged across these frequencies) and the bimodal performance/benefits in each listening task (i.e., Mandarin tone, vowel, and syllable recognition). Pearson correlations were also computed between the demographic factors of participants (i.e., the age at implantation, duration of CI use, duration of HA use, and age at testing, as listed or inferred in Table 1) and the CI alone (or CI+HA) performance (averaged across mixed- and blocked-talker conditions) in each listening task, to examine how these demographic factors may have contributed to the variable performance across participants in this study. In the multiple correlation analyses, a correlation was considered significant only when the probability level was below 0.005.

RESULTS

Overall Performance

Figure 2 shows the overall scores for Mandarin tone, vowel, and syllable recognition, as well as the cumulative response time in mixed- and blocked-talker conditions with either CI alone or CI+HA. In this and the following figures, the CI alone condition shows data from 14 participants except P1, while the CI+HA condition shows data from 10 participants except P3, P4, P6, P7, and P13 (see Methods for details). General linear models showed that in panel (a), both presentation ($F_{1,14} = 7.03$, $p = 0.02$) and processor conditions ($F_{1,8} = 6.57$, $p = 0.03$) significantly affected Mandarin tone recognition, and there was no interaction between the two factors ($F_{1,8} = 3.08$, $p = 0.12$). Vowel recognition in panel (b) was not significantly affected by either presentation ($F_{1,14} = 4.02$, $p = 0.06$) or processor condition ($F_{1,8} = 3.41$, $p = 0.10$). The two factors did not interact with each other ($F_{1,8} = 0.88$, $p = 0.38$). The percentage of correct syllable recognition (i.e., correct recognition of both Mandarin tone and vowel) was lower than that of correct Mandarin tone or vowel recognition. Similar to Mandarin tone recognition, syllable recognition in panel (c) was significantly affected by both presentation ($F_{1,14} = 5.23$, $p = 0.04$) and processor conditions ($F_{1,8} = 6.15$, $p = 0.04$), and no interaction was observed between the two factors ($F_{1,8} = 0.28$, $p = 0.61$). Overall, significantly better Mandarin tone and syllable recognition scores were observed in the blocked-talker condition than in the mixed-talker condition, showing talker variability effects. Mandarin tone and syllable recognition also significantly improved with CI+HA than with CI alone, showing bimodal benefits. In panel (d), the cumulative response

time was significantly affected by the processor condition ($F_{1,8} = 6.57, p = 0.03$) but not by the presentation condition ($F_{1,14} = 0.07, p = 0.80$). The two factors did not interact with each other ($F_{1,8} = 3.39, p = 0.10$). Participants had shorter response time with CI+HA than with CI alone, again showing bimodal benefits.

Detailed Mandarin Tone Recognition Performance

Figure 3 shows the Mandarin tone recognition scores for individual tones in different conditions. A general linear model revealed that the presentation condition ($F_{1,14} = 9.06, p = 0.01$), processor condition ($F_{1,8} = 7.89, p = 0.02$), and lexical tone ($F_{3,42} = 25.61, p < 0.001$) all had a significant effect on Mandarin tone recognition. The three factors did not interact with one another ($p > 0.25$). Post-hoc Bonferroni *t*-tests revealed significant differences in performance between any pair of Mandarin tones ($p < 0.001$). The Mandarin tones in the descending order of recognition scores were Tone 4, Tone 1, Tone 2, and Tone 3. Mandarin tone confusion matrices showed that participants were most often confused between Tone 3 and Tone 2, followed by between Tone 2 and Tone 1. Within each Mandarin tone pair, the asymmetric confusions generally reduced with CI+HA than with CI alone. For example, with CI alone, the percentage of misidentification of Tone 3 as Tone 2 was 48%, and that of Tone 2 as Tone 3 was 14%; with CI+HA, the percentage was 41% and 10%, respectively. There was a response bias towards Tone 2 when a pitch increase was perceived at the end of the vowel. The percentage of misidentification of high-flat Tone 1 as mid-rising Tone 2 (and low-falling-rising Tone 3 as high-flat Tone 1) was reduced by 5% (and 4%) in the blocked-talker condition than in the mixed-talker condition.

Table 2 shows the results of general linear models for the recognition scores of each Mandarin tone. The recognition of Tone 1 was significantly affected by both presentation and processor conditions, and the two factors did not interact with each other. The effects of both presentation and processor conditions and the interaction between the two factors were not significant for the recognition of Tone 2, Tone 3, and Tone 4.

Figure 4 shows the Mandarin tone recognition scores for individual talkers in different conditions. A general linear model revealed that Mandarin tone recognition was significantly affected by the presentation condition ($F_{1,14} = 8.30, p = 0.01$), processor condition ($F_{1,8} = 6.30, p = 0.03$), and talker ($F_{3,42} = 9.20, p < 0.001$). There was no interaction among the three factors ($p > 0.11$). Post-hoc Bonferroni *t*-tests showed that talkers FT1, MT1, and MT2 had significantly higher Mandarin tone recognition scores than talker FT2 ($p < 0.001$). Acoustic analyses showed that the F0 ranges for Mandarin contour tones (except high-flat Tone 1) of FT2 were on average smaller than those of FT1 but larger than those of MT1 and MT2. Also, FT2 produced all four Mandarin tones of shorter durations as compared to the other talkers. The shorter durations of the lexical tones rather than the F0 ranges may have accounted for the significantly lower Mandarin tone recognition scores for FT2.

Detailed Mandarin Vowel Recognition Performance

The vowel recognition scores for individual vowels in different conditions are shown in Figure 5. A general linear model revealed a significant effect of vowel on vowel recognition ($F_{5,70} = 20.00, p < 0.001$). No significant effect of the presentation ($F_{1,14} = 2.59, p = 0.13$)

or processor condition ($F_{1,8} = 2.77$, $p = 0.14$) was found and there was no interaction among the factors ($p > 0.17$). Post-hoc Bonferroni t -tests revealed significant differences in performance between any pair of vowels ($p < 0.01$), except between the least intelligible vowels /o/ and /ü/ ($p = 1.00$). Vowel confusion matrices showed that the most frequent vowel confusions were between /ü/ and /i/, followed by between /o/ and /e/. The confusions were asymmetric within each vowel pair. For example, the percentage of misidentification of /ü/ as /i/ (and /i/ as /ü/) was 40% (and 25%), and that of /o/ as /e/ (and /e/ as /o/) was 24% (and 13%).

Table 3 shows the results of general linear models for the recognition scores of each vowel. The effects of both presentation and processor conditions and the interaction between the two factors were mostly not significant, except the effect of processor condition for the recognition of /o/ and the effect of presentation condition for the recognition of /e/.

The vowel recognition scores for individual talkers in different conditions are shown in Figure 6. A general linear model showed no significant effect of the presentation condition ($F_{1,14} = 3.63$, $p = 0.08$), processor condition ($F_{1,8} = 3.92$, $p = 0.08$), or talker ($F_{3,42} = 2.81$, $p = 0.05$) on vowel recognition. The three factors did not interact with one another ($p > 0.26$).

Correlation Results

There was a positive correlation between the 1000-Hz threshold in the non-implanted ear and the Mandarin tone recognition performance with CI+HA ($r = 0.86$, $p = 0.001$). This correlation was unexpected because 1000 Hz was above the voice pitch range and bimodal users with higher 1000-Hz thresholds in the non-implanted ear had better (rather than worse) Mandarin tone recognition scores with CI+HA. No significant correlations were observed between any other aided thresholds and the bimodal performance/benefits ($p > 0.05$).

None of the demographic factors listed or inferred in Table 1 was correlated with any CI alone or CI+HA performance ($p > 0.05$). In large-scale studies of pediatric CI population (e.g., Kirk et al., 2002; Zwolan et al., 2004), the age at implantation has been shown to be negatively correlated with English speech recognition performance. Similar correlations have been observed for Mandarin tone recognition performance in Han et al. (2009), but not in Peng et al. (2004) and Zhou et al. (2013). In addition, Zhou et al. (2013) showed that Mandarin tone recognition performance was positively correlated with the duration of CI use. Unlike the previous studies, most participants in this study had an age at implantation of 10 years or older and were long-term HA users before implantation (see Table 1). This may explain why the age at implantation and the duration of CI use were not predictive of the CI alone or CI+HA performance in this study.

DISCUSSION

Talker Variability Effects

Significant talker variability effects on Mandarin tone recognition were observed for prelingually deafened CI users. Mandarin tone recognition with both CI alone and CI+HA significantly deteriorated when the talker identity varied rather than fixed from trial to trial. Similar talker variability effects on Mandarin tone recognition have been reported for NH

listeners in Lee et al. (2010), where background noise was added to avoid ceiling effects in NH performance and 3% better performance was observed in the blocked-talker condition than the mixed-talker condition. In contrast, greater talker variability effects on Cantonese level tone recognition have been reported for NH listeners in Wong and Diehl (2003), where performance in quiet was 30% better in the blocked-talker condition than the mixed-talker condition. Compared to Cantonese level tones, Mandarin contour tones (Tones 2, 3, and 4) contain more pitch cues within syllables, leading to reduced impact of pitch variations across talkers in the mixed-talker condition. In this study, significant talker variability effects were only observed for the recognition of high-flat Tone 1, which, similar to Cantonese level tones, may rely more on the talker voice pitch range for recognition than the other Mandarin tones.

Contrary to one of our original hypotheses, the talker variability effects on Mandarin tone recognition were not reduced with CI+HA than with CI alone, as indicated by the lack of interaction between the presentation and processor conditions. Although Mandarin tone recognition of CI users in both mixed- and blocked-talker conditions significantly improved with bimodal hearing, the performance difference between the two presentation conditions did not decrease with bimodal hearing. The enhanced talker voice information (especially the more salient voice pitch cues) from aided residual low-frequency acoustic hearing (e.g., Krull et al., 2012) did not seem sufficient for CI users to fully normalize the talker differences in Mandarin tone productions of isolated syllables. However, Luo et al. (2014) tested the same group of CI users in this study and found that, when target syllables were presented in the context of a preceding sentence (as in daily conversations), the voice pitch cues in the context can assist CI users' Mandarin tone normalization in the CI+HA condition but not the CI alone condition. These results suggested that even with CI+HA, CI users needed a certain amount of exposure to the talker voice information (either from the context of a preceding sentence as in Luo et al., 2014 or from the previous test stimuli in the blocked-talker condition as in this study) to achieve Mandarin tone normalization. The critical context duration for CI users' Mandarin tone normalization is expected to be longer than that for NH listeners (250 ms corresponding to a single syllable in a normal speaking rate; Luo and Ashmore, 2014), although further verification is needed.

Unlike Chang and Fu (2006), this study did not find significant talker variability effects on vowel recognition with both CI alone and CI+HA. There were several notable differences between the two studies. First, the English vowel recognition test in Chang and Fu (2006) had a greater number of response choices (i.e., 12) than the Mandarin vowel recognition test in this study (i.e., 6), and the participants in Chang and Fu (2006) were postlingually deafened adult CI users using relatively old CI technologies. Despite the different study designs, the overall performance of vowel recognition was similar in the two studies and thus could not account for the different talker variability effects on vowel recognition. Second, as mentioned in the Introduction, Chang and Fu (2006) presented each stimulus from a particular talker eight times in the blocked-talker condition but only two times in the mixed-talker condition. The greater number of stimulus repeats may have led to greater learning effects for vowel recognition in the blocked-talker condition. In this study, each stimulus from a particular talker was presented only once in both mixed- and blocked-talker

conditions, and the vowel recognition performance was not significantly different in the two presentation conditions.

An intriguing finding in this study was that the talker variability effects were significant on Mandarin tone (and hence syllable) recognition but not on vowel recognition. Acoustic analyses showed that the formant frequencies of the six vowels varied across the four talkers but did not strongly overlap across the vowels, except between /i/ and /ü/ and between /e/ and /o/. This formant distribution may have led to the near-ceiling vowel recognition scores (especially for vowels /a/ and /u/), and thus reduced the talker variability effects on vowel recognition. Detailed analyses showed that significant talker variability effects were observed for the recognition of one of the poorly recognized vowels (/e/). In future studies, lowering the overall vowel recognition scores by adding noise may help avoid the ceiling effects, and thus facilitate the examination of talker variability effects on vowel recognition.

For NH listeners in Lee et al. (2010), the additional cognitive load or listening effort of processing mixed-talker stimuli in Mandarin tone recognition significantly increased the response time while reducing the response accuracy. However, the talker variability effects on response time were not significant in both CI alone and CI+HA conditions in this study. Specifically, the response time with CI alone decreased (while that with CI+HA increased) in the blocked-talker condition than in the mixed-talker condition. The inconsistency suggested that the response time may not be a reliable measure of the listening effort with respect to the talker variability effects in the case of combined Mandarin tone and vowel recognition. Instead, some refined measures such as the dual-task paradigm (e.g., Pals et al., 2013) and pupilometry (e.g., Zekveld et al., 2010) may be used to better examine the talker variability effects on the listening effort with either CI alone or CI+HA.

Bimodal Benefits

The present results showed that aided residual acoustic hearing in the non-implanted ear significantly improved Mandarin tone recognition for CI users. The mean Mandarin tone recognition score increased by 8% with CI+HA than with CI alone, similar to the results in Yuen et al. (2009). Improved Mandarin tone recognition directly led to improved Mandarin syllable recognition in the CI+HA condition.

As indicated by the correlation results in this study as well as in Yuen et al. (2009), the bimodal benefits to Mandarin tone and syllable recognition could not be predicted by the aided hearing thresholds in the non-implanted ear. Ching et al. (2004) and Gifford et al. (2007) have also shown that the bimodal benefits to English speech recognition were similarly uncorrelated with the aided audiometric thresholds of residual acoustic hearing. It is possible that the pitch-ranking thresholds and frequency/temporal resolution in the non-implanted ear (e.g., Golub et al., 2012) may better indicate Mandarin tone and syllable recognition acuity with aided residual acoustic hearing than the pure tone detection thresholds considered in this study. This hypothesis needs to be verified in future studies of bimodal benefits to Mandarin speech recognition.

Detailed analyses in this study showed more accurate recognition of falling Tone 4 and flat Tone 1 than rising Tone 2 and falling-rising Tone 3, with either CI alone or CI+HA. Such

Mandarin tone recognition patterns were similar to those reported in Peng et al. (2004) and Luo et al. (2009) for CI users with CI alone, but different from those reported in Luo and Fu (2004, 2009) for NH listeners listening to CI simulations. In Luo and Fu (2004, 2009), NH listeners with limited access to pitch cues from 4–8 frequency channels recognized Tone 3 and Tone 4 better than Tone 2 and Tone 1, possibly due to the use of vowel duration cues (Tone 3 usually has the longest duration while Tone 4 usually has the shortest duration) and amplitude envelope cues that co-vary with pitch contours. In this study and in Luo et al. (2009), CI users were unable to take full advantage of these secondary tonal cues to aid the recognition of Tone 3, and they were confused between Tone 2 and Tone 3 that both end with a pitch increase. Adding a contralateral HA only significantly improved the recognition of Tone 1. This suggested that the bimodal benefits to Mandarin tone recognition may be due to generally enhanced pitch cues (e.g., the partially resolved F0 and low-frequency harmonics in aided residual acoustic hearing), rather than better use of the secondary tonal cues such as vowel durations and amplitude envelopes, which should have provided more benefits to the recognition of Tone 3 and Tone 4 than Tone 1 (e.g., Luo and Fu, 2004).

Different from Mandarin tone recognition, vowel recognition did not significantly (albeit slightly) improve with CI+HA than with CI alone. Again, this may be because the vowel recognition scores with CI alone (especially for vowels /a/ and /u/) were close to ceiling, leaving limited room for improvement. The bimodal benefits to vowel recognition were only significant for /o/, which had the lowest baseline performance with CI alone among the vowels. In acoustic simulations of bimodal hearing, Mandarin tone recognition significantly improved with low-frequency acoustic cues below 500 Hz, while vowel recognition significantly improved with those above 500 Hz (Luo and Fu, 2006). Note that the simulation results in Luo and Fu (2006) were obtained from NH listeners with a full capacity to perceive the spectral cues in the low-pass filtered speech (which simulated residual low-frequency acoustic hearing). Although the aided thresholds of our participants in the non-implanted ear (Figure 1) were mostly in the range of mild-to-moderate hearing loss (even for frequencies up to 2 kHz, which contain partial or full first formant cues depending on the vowels), the degraded frequency resolution associated with hearing loss (e.g., Moore, 1996) may have prevented participants from fully using the acoustic cues of vowel formants and thus from reaping the full bimodal benefits to vowel recognition. This hypothesis may be further tested by examining vowel recognition and frequency resolution with HA alone.

Not all vowels were equally well recognized in this study. Vowels /a/ and /u/ were best recognized because they are at different corners of the vowel space based on the first and second formant frequencies, and are distant from the other vowels in the vowel space (see Figure 1 of Luo and Fu, 2005). While /i/ is also a corner vowel, it could be confused with /ü/ at the same corner of the vowel space. Actually, when /ü/ was excluded from the vowel recognition test in Luo et al. (2009), /i/ became the best recognized vowel for CI users. Vowels /e/ and /o/ at the center of the vowel space were confused with each other due to their similar formant frequencies. For each of the two most confused vowel pairs /i/-/ü/ and /e/-/o/, participants tended to choose the vowel with higher second formant frequencies (i.e., /i/ over /ü/ and /e/ over /o/), for reasons that need further investigation.

An important finding in this study was that bimodal hearing not only significantly improved Mandarin tone (and hence syllable) recognition but also significantly reduced the response time of syllable recognition for CI users. In light of the possible link between the response time and listening effort (e.g., Gatehouse and Gordon, 1990; Baer et al., 1993; Houben et al., 2013), the shorter response time suggested that the participants' listening effort in the syllable recognition test may have decreased with the additional acoustic cues provided by the contralateral HA. Despite its inconsistent changes in the two presentation conditions, the response time significantly reduced with CI+HA than with CI alone, possibly reflecting the faster processing speed and less cognitive demands with the aid of a contralateral HA. To the best of our knowledge, there has been no previous report on the bimodal benefits to listening effort. Considering the limitations of response time as a measure of listening effort, future studies may use more refined measures such as the dual-task paradigm (e.g., Pals et al., 2013) and pupillometry (e.g., Zekveld et al., 2010) to verify the bimodal benefits to listening effort. Nevertheless, the present response time results provided supporting evidence from a previously unexamined perspective for the clinical recommendation of a contralateral HA for unilateral CI users with residual acoustic hearing in the non-implanted ear. Reduced listening fatigue with bimodal hearing may enhance children's classroom learning. Less listening effort with bimodal hearing may also provide bimodal users with more cognitive resources such as attention and working memory to perform talker normalization (Luo et al., 2014) and handle noisy and/or reverberant listening environments.

Acknowledgments

Conflicts of Interest and Source of Funding: This research was supported in part by NIH Grant R21-DC-011844.

The authors gratefully acknowledge the subjects who participated in this study and the support from the NIH (Grant R21-DC-011844).

References

- Baer T, Moore BCJ, Gatehouse S. Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times. *J Rehabil Res Dev.* 1993; 30:49–72. [PubMed: 8263829]
- Chang Y, Fu QJ. Effects of talker variability on vowel recognition in cochlear implants. *J Speech Lang Hear Res.* 2006; 49:1331–1341. [PubMed: 17197499]
- Chao, YR. Mandarin primer. Cambridge, MA: Harvard University Press; 1948.
- Ching TYC, Incerti P, Hill M. Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear Hear.* 2004; 25:9–21. [PubMed: 14770014]
- Cleary M, Pisoni DB. Talker discrimination by prelingually deaf children with cochlear implants: preliminary results. *Ann Otol Rhinol Laryngol Suppl.* 2002; 189:113–118. [PubMed: 12018337]
- Creelman CD. Case of the unknown talker. *J Acoust Soc Am.* 1957; 29:655.
- Dorman MF, Gifford RH, Spahr AJ, et al. The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiol Neurotol.* 2008; 13:105–112.
- Gatehouse S, Gordon J. Response times to speech stimuli as measures of benefit from amplification. *Brit J Audiol.* 1990; 24:63–68. [PubMed: 2317602]
- Geurts L, Wouters J. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am.* 2001; 109:713–726. [PubMed: 11248975]
- Gifford RH, Dorman MF, McKarns SA, et al. Combined electric and contralateral acoustic hearing: word and sentence recognition with bimodal hearing. *J Speech Lang Hear Res.* 2007; 50:835–843. [PubMed: 17675589]

- Golub JS, Won JH, Drennan WR, et al. Spectral and temporal measures in hybrid cochlear implant users: On the mechanism of electroacoustic hearing benefits. *Otol Neurotol.* 2012; 33:147–153. [PubMed: 22215451]
- Han D, Liu B, Zhou N, et al. Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users. *Ear Hear.* 2009; 30:169–177. [PubMed: 19194297]
- Houben R, van Doorn-Bierman M, Dreschler WA. Using response time to speech as a measure for listening effort. *Int J Audiol.* 2013; 52:753–761. [PubMed: 24053226]
- Kaiser AR, Kirk KI, Lachs L, et al. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J Speech Lang Hear Res.* 2003; 46:390–404. [PubMed: 14700380]
- Kirk KI, Hay-McCutcheon M, Sehgal ST, et al. Speech perception in children with cochlear implants: Effects of lexical difficulty, talker variability, and word length. *Ann Otol Rhinol Laryngol Suppl.* 2000; 185:79–81. [PubMed: 11141016]
- Kirk KI, Miyamoto RT, Lento CL, et al. Effects of age at implantation in young children. *Ann Otol Rhinol Laryngol Suppl.* 2002; 189:69–73. [PubMed: 12018353]
- Klatt, DH. The problem of variability in speech recognition and in models of speech perception. In: Perkell, JS.; Klatt, DH., editors. *Invariance and variability in speech processes.* Hillsdale, NJ: Erlbaum; 1986. p. 300-319.
- Krulse GK, Tondo DK, Wightman FL. Speech perception as a multilevel processing system. *J Psycholinguist Res.* 1983; 12:531–554.
- Krull V, Luo X, Kirk KI. Talker-identification training using simulations of binaurally combined electric and acoustic hearing: Generalization to speech and emotion recognition. *J Acoust Soc Am.* 2012; 131:3069–3078. [PubMed: 22501080]
- Lee CY, Tao L, Bond ZS. Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. *Speech Commun.* 2010; 52:900–910.
- Luo X, Ashmore KB. The effect of context duration on Mandarin listeners' tone normalization. *J Acoust Soc Am.* 2014; 136:EL109–EL115. [PubMed: 25096133]
- Luo X, Chang Y, Lin CY, et al. Contribution of bimodal hearing to lexical tone normalization in Mandarin-speaking cochlear implant users. *Hear Res.* 2014; 312:1–8. [PubMed: 24576834]
- Luo X, Fu QJ. Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. *J Acoust Soc Am.* 2004; 116:3659–3667. [PubMed: 15658716]
- Luo X, Fu QJ. Speaker normalization for Chinese vowel recognition in cochlear implants. *IEEE Trans Biomed Eng.* 2005; 52:1358–1361. [PubMed: 16042003]
- Luo X, Fu QJ. Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulations. *J Acoust Soc Am.* 2006; 120:2260–2266. [PubMed: 17069321]
- Luo X, Fu QJ. Concurrent-vowel and tone recognitions in acoustic and simulated electric hearing. *J Acoust Soc Am.* 2009; 125:3223–3233. [PubMed: 19425665]
- Luo X, Fu QJ, Wei CG, et al. Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. *Ear Hear.* 2008; 29:957–970. [PubMed: 18818548]
- Luo X, Fu QJ, Wu HP, et al. Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hear Res.* 2009; 256:75–84. [PubMed: 19595753]
- Luo X, Galvin JJ, Fu Q-J. Effects of stimulus duration on amplitude modulation processing with cochlear implants. *J Acoust Soc Am.* 2010; 127:EL23–EL29. [PubMed: 20136174]
- Moore BCJ. Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear Hear.* 1996; 17:133–161. [PubMed: 8698160]
- Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. *J Acoust Soc Am.* 1989; 85:365–378. [PubMed: 2921419]
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Percept Psychophys.* 1998; 60:355–376. [PubMed: 9599989]
- Pals C, Sarampalis A, Baskent D. Listening effort with cochlear implant simulations. *J Speech Lang Hear Res.* 2013; 56:1075–1084. [PubMed: 23275424]
- Peng SC, Tomblin JB, Cheung H, et al. Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear Hear.* 2004; 25:251–264. [PubMed: 15179116]

- Pisoni DB. Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning. *Speech Commun.* 1993; 4:75–95. [PubMed: 23226910]
- Ryalls BO, Pisoni DB. The effect of talker variability on word recognition in preschool children. *Dev Psychol.* 1997; 33:441–452. [PubMed: 9149923]
- Studebaker GA. A “rationalized” arcsine transform. *J Speech Lang Hear Res.* 1985; 28:455–462.
- Vongphoe M, Zeng FG. Speaker recognition with temporal cues in acoustic and electric hearing. *J Acoust Soc Am.* 2005; 118:1055–1061. [PubMed: 16158660]
- Wong PCM, Diehl RL. Perceptual normalization for inter- and intra-talker variation in Cantonese level tones. *J Speech Lang Hear Res.* 2003; 46:413–421. [PubMed: 14700382]
- Yuen KC, Cao KL, Wei CG, et al. Lexical tone and word recognition in noise of Mandarin-speaking children who use cochlear implants and hearing aids in opposite ears. *Cochlear Implants Int.* 2009; 10:120–129. [PubMed: 19195003]
- Zekveld AA, Heslenfeld DJ, Festen JM, et al. Top-down and bottom-up processes in speech comprehension. *Neuroimage.* 2006; 32:1826–1836. [PubMed: 16781167]
- Zekveld AA, Kramer SE, Festen JM. Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear.* 2010; 31:480–490. [PubMed: 20588118]
- Zhou N, Huang J, Chen X, et al. Relationship between tone perception and production in prelingually-deafened children with cochlear implants. *Otol Neurotol.* 2013; 34:499–506. [PubMed: 23442566]
- Zwolan TA, Ashbaugh CM, Alarfaj A, et al. Pediatric cochlear implant patient performance as a function of age at implantation. *Otol Neurotol.* 2004; 25:112–120. [PubMed: 15021769]

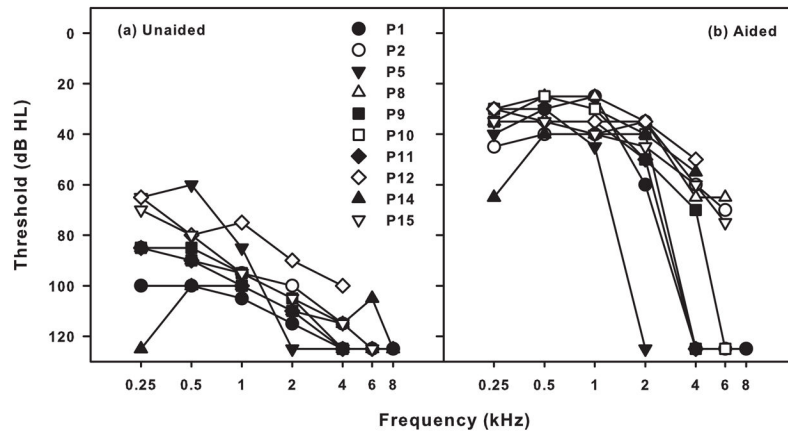


Fig. 1. Unaided and aided hearing thresholds in the non-implanted ear of the bimodal users in this study. Symbols at 125 dB HL represent no response at the maximum output of the audiometer.

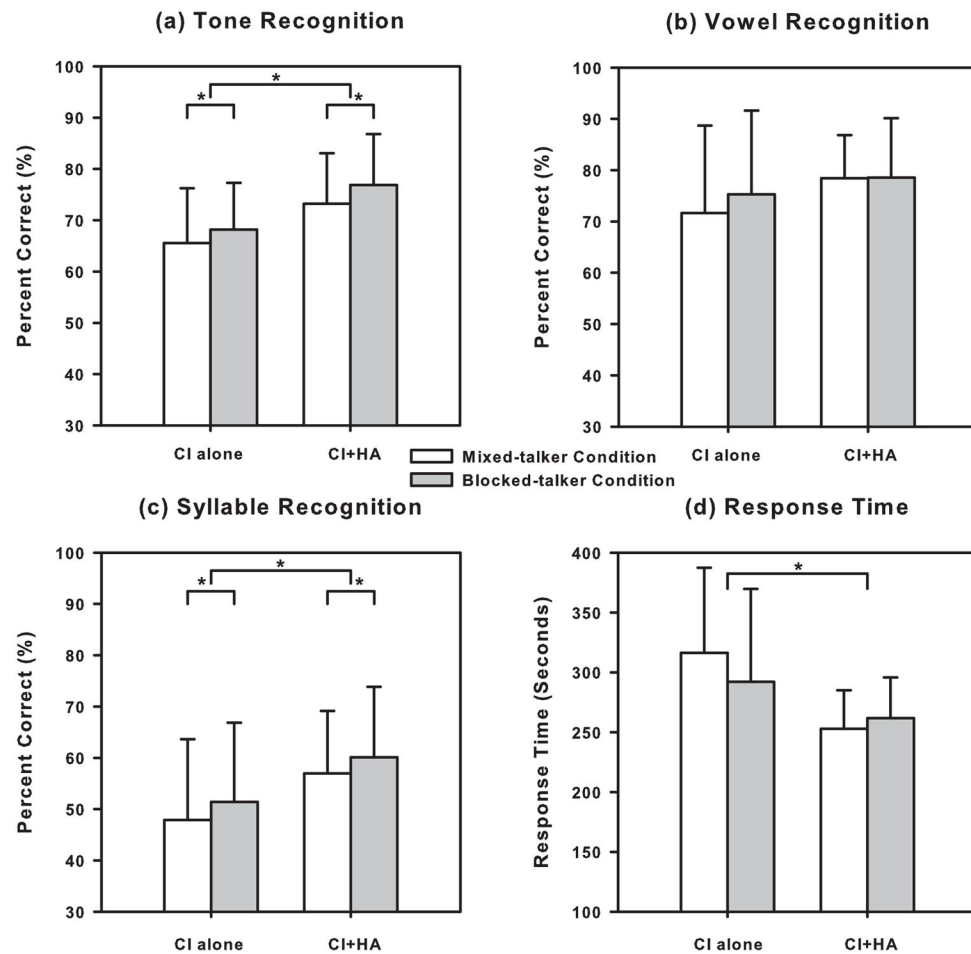


Fig. 2. Overall performance for Mandarin tone, vowel, and syllable recognition, as well as cumulative response time with either CI alone or CI+HA in mixed- and blocked-talker conditions. Vertical bars represent the mean while error bars represent the standard deviation across participants. Statistically significant differences ($p < 0.05$) between the processor or presentation conditions are indicated by asterisks.

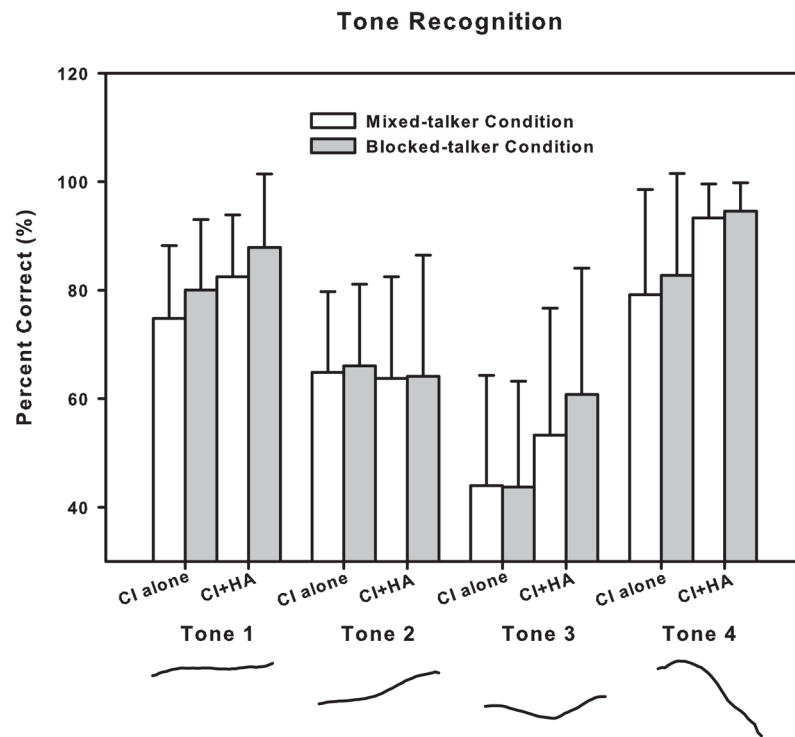


Fig. 3. Mandarin tone recognition scores for individual Mandarin tones with either CI alone or CI +HA in mixed- and blocked-talker conditions. Vertical bars represent the mean while error bars represent the standard deviation across participants. Example pitch contours of the four Mandarin tones are shown under the x-axis.

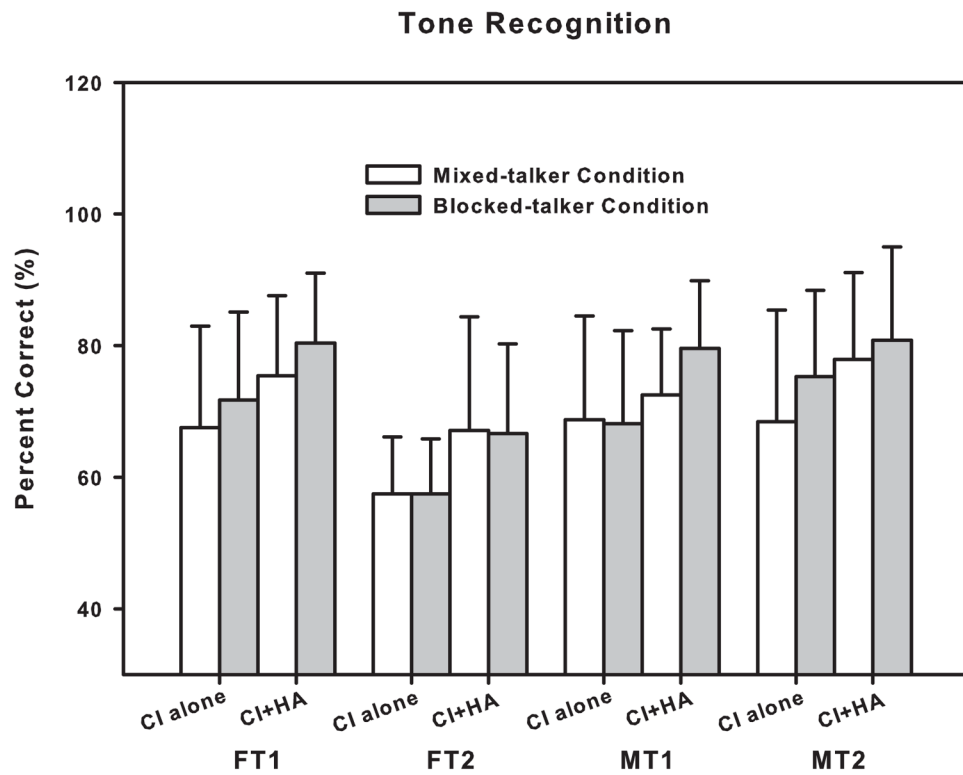


Fig. 4. Mandarin tone recognition scores for individual talkers with either CI alone or CI+HA in mixed- and blocked-talker conditions. Vertical bars represent the mean while error bars represent the standard deviation across participants.

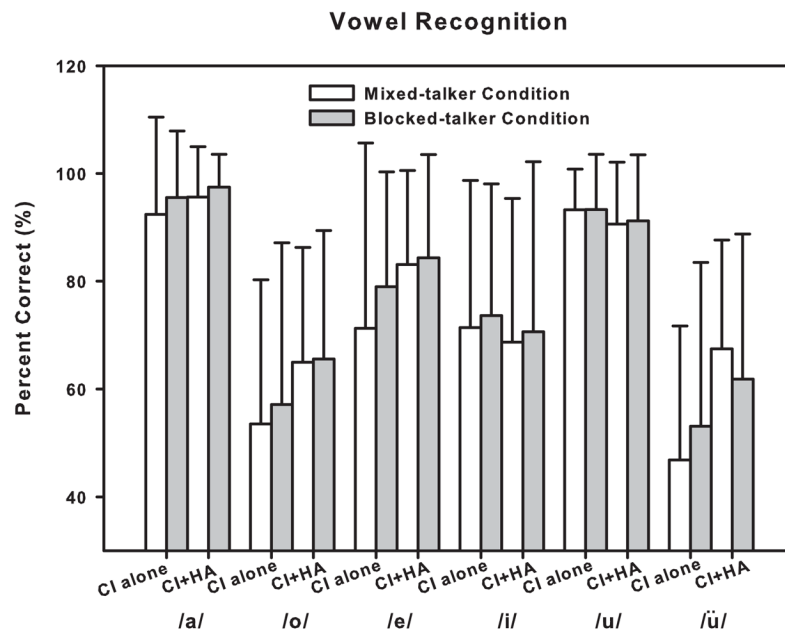


Fig. 5. Vowel recognition scores for individual vowels with either CI alone or CI+HA in mixed- and blocked-talker conditions. Vertical bars represent the mean while error bars represent the standard deviation across participants.

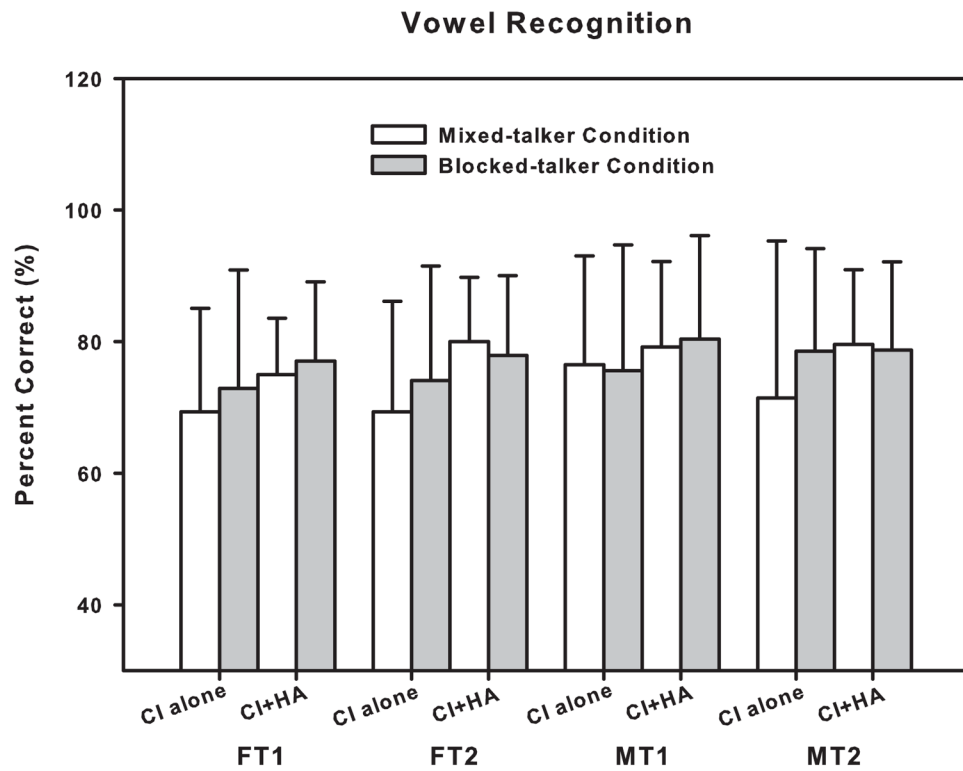


Fig. 6. Vowel recognition scores for individual talkers with either CI alone or CI+HA in mixed- and blocked-talker conditions. Vertical bars represent the mean while error bars represent the standard deviation across participants.

Table 1

Demographic information of the participants. Table reproduced from Luo et al. (2014) with permission from Elsevier.

Subject	Age (yrs)	Gender	Etiology	Implant/strategy (ear)	Age at diagnosis of hearing loss, enrollment in auditory verbal therapy, and implantation (yrs)	HA device (yrs of HA use before/after CI)	Test condition
							CI alone CI+HA
P1	13	F	Unknown	Freedom/ACE (R)	2, 2, 9	Phonak Naida III UP (7/4)	X X
P2	17	M	EVA	Freedom/ACE (L)	3, 3, 12	Phonak Supero 412 (10/5)	X X
P5	20	M	EVA	Nucleus 24M/ACE (R)	Unknown, Unknown, 9	C18+ (Unknown/Unknown)	X X
P8	16	M	Unknown	Freedom/ACE (L)	3, 4, 12	Phonak Naida III UP (9/4)	X X
P9	20	F	Unknown	Freedom/ACE (R)	5, 6, 17	Starkey Destiny 1200 P+ (11/3)	X X
P10	20	M	Unknown	Freedom/ACE (L)	Unknown, 4, 15	Phonak Naida III UP (11/5)	X X
P11	12	F	Hereditary	Freedom/ACE (L)	1, 2, 9	Targa 3 HP (7/3)	X X
P12	11	M	EVA	Freedom/ACE (L)	2, 2, 10	Widex B32 (8/1)	X X
P14	14	F	Unknown	Nucleus 24CS/ACE (L)	Unknown, 1, 4	Phonak Savia Art 411 (3/10)	X X
P15	11	M	High fever	Freedom/ACE (R)	1, 2, 5	Oticon Sumo DM (3/6)	X X
P3	17	M	EVA	Nucleus 24M/ACE (R)	Unknown, 1, 3	N/A	X X
P4	19	M	EVA	Freedom/ACE (R)	Unknown, 6, 16	N/A	X X
P6	16	F	Meningitis	Nucleus 24M/ACE (L)	1, 2, 1	N/A	X X
P7	13	F	EVA	Harmony/HiRes120 (R)	5, 5, 12	N/A	X X
P13	10	F	Unknown	Nucleus 24CS/ACE (L)	1, 2, 8	N/A	X X

EVA: Enlarged Vestibular Aqueduct; ACE: Advanced Combination Encoder. N/A: not available.

Table 2

Results of general linear models for the recognition scores of each Mandarin tone.

Mandarin Tone	Presentation Condition	Processor Condition	Interaction Between the Two Factors
Tone 1	$F_{1,14} = 15.82, p = 0.004$	$F_{1,8} = 11.18, p = 0.01$	$F_{1,8} = 0.01, p = 0.94$
Tone 2	$F_{1,14} = 0.07, p = 0.80$	$F_{1,8} = 0.01, p = 0.91$	$F_{1,8} = 0.08, p = 0.79$
Tone 3	$F_{1,14} = 3.70, p = 0.09$	$F_{1,8} = 1.58, p = 0.24$	$F_{1,8} = 2.81, p = 0.13$
Tone 4	$F_{1,14} = 0.05, p = 0.83$	$F_{1,8} = 3.31, p = 0.11$	$F_{1,8} = 0.62, p = 0.45$

Table 3

Results of general linear models for the recognition scores of each Mandarin vowel.

Mandarin Vowel	Presentation Condition	Processor Condition	Interaction Between the Two Factors
/a/	$F_{1,14} = 2.81, p = 0.13$	$F_{1,8} = 2.57, p = 0.15$	$F_{1,8} = 1.38, p = 0.27$
/o/	$F_{1,14} = 0.09, p = 0.77$	$F_{1,8} = 8.93, p = 0.02$	$F_{1,8} = 0.48, p = 0.51$
/e/	$F_{1,14} = 6.03, p = 0.04$	$F_{1,8} = 2.68, p = 0.14$	$F_{1,8} = 1.00, p = 0.35$
/i/	$F_{1,14} = 1.08, p = 0.33$	$F_{1,8} = 0.17, p = 0.69$	$F_{1,8} = 0.03, p = 0.86$
/u/	$F_{1,14} = 0.03, p = 0.86$	$F_{1,8} = 0.15, p = 0.71$	$F_{1,8} = 0.00003, p = 1.00$
/ü/	$F_{1,14} = 0.05, p = 0.82$	$F_{1,8} = 2.78, p = 0.13$	$F_{1,8} = 0.56, p = 0.48$