Published in final edited form as:

Acta Otolaryngol. 2017 August ; 137(8): 829–836. doi:10.1080/00016489.2017.1292050.

Effect of spatial separation and noise type on sentence recognition by Mandarin-speaking cochlear implant users

Yang-Wenyi Liu^{a,b,*}, Duo-Duo Tao^{c,*}, Ye Jiang^{a,b}, John J. Galvin III^d, Qian-Jie Fu^d, Ya-sheng Yuan^{a,b}, and Bing Chen^{a,b}

^aDepartment of Otology and Skull Base Surgery, Eye Ear Nose and Throat Hospital, Fudan University, Shanghai, China

^bKey Laboratory of Hearing Medicine, Ministry of Health, Eye Ear Nose and Throat Hospital, Fudan University, Shanghai, China

^cDepartment of Ear, Nose, and Throat, The First Affiliated Hospital of Soochow University, Suzhou, China

^dDepartment of Head and Neck Surgery, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

Abstract

Objectives—To investigate the effects of spatial separation and noise type on sentence recognition by unilateral Mandarin-speaking cochlear implant (CI) users and normal-hearing (NH) listeners.

Method—Twenty-two unilateral Mandarin-speaking CI users and six NH listeners participated in this study. Speech reception thresholds were measured for three noise types (steady state noise, speech babble, and music). Sentences from the Mandarin Speech Perception test were presented directly in front of the listener (0°). Noise was presented from one of the five speaker locations: −90°, −45°, 0°, +45°, and +90°.

Results—Overall, CI performance was significantly poorer than NH performance for all spatial separation and noise type conditions. NH listeners performed best with music and poorest with steady noise. CI users performed best with steady noise, and poorest with babble. Performance was significantly affected by noise location and noise type. There was no significant difference in head shadow effects among the different noise types for CI users.

Conclusions—Performance was much poorer in CI than in NH listeners for all noise types and spatial separations. Noise type differently affected unilateral CI users and NH listeners. The limited spectral resolution in CI users did not appear to affect head shadow.

Keywords

Cochlear implant; speech perception; head shadow effect; noise

Correspondence: Bing Chen, bingchen@fudan.edu.cn, Department of Otology and Skull Base Surgery, Eye Ear Nose and Throat Hospital, Fudan University, 83 Fenyang Road, Shanghai 200031, China.

*These authors have contributed equally to this work.

Introduction

Normal-hearing (NH) individuals can take advantage of gaps in fluctuating or modulated maskers by “listening in the dips” in spectral and/or temporal envelopes to perceive a speech target. NH listeners regularly experience a release from masking with dynamic maskers relative to steady noise [1, 2]. Different from NH listeners, cochlear implant (CI) users have greater difficulty with dynamic than with steady noise [1]. The limited electric dynamic range (DR), the coarse spectral resolution, the broad analysis filter banks, and interactions among the implanted electrodes are all thought to contribute to CI users’ difficulty with dynamic noise. Given that most sounds are dynamic in nature, CI users experience greater interference for the most common listening conditions in daily life (multiple talkers, music backgrounds, etc.) than do NH listeners. The type of masker has also been shown to affect speech intelligibility [3]. While many previous studies have compared speech intelligibility between steady noise and interfering speech or modulated noise, relatively few have evaluated the effects of music maskers on speech understanding. Gfeller et al. [4] found that speech reception thresholds for CI listeners in background music was affected by different types of music and hearing status. Eskridge et al. [2] found that CI speech understanding worsened as the spectro-temporal complexity increased within the maskers.

Mandarin is one of the most widely spoken languages in the world, with over one billion native speakers. There are a growing number of Mandarin-speaking CI users. Mandarin language is different from English in a number of ways. First, the structure of a Mandarin syllable typically consists of two components: an initial consonant and a final vowel. Mandarin syllables have more voiceless consonants and fewer voiced consonants compared to English. Voiceless consonants are more easily masked due to less energy, suggesting that Mandarin speech may be more susceptible to energetic masking than English. Second, Mandarin is a tonal language, in which meaning is conveyed via lexical tones. Fundamental frequency (F0) is the primary cue for lexical tones, although duration and amplitude cues co-vary with F0 [5]. Current CI devices do not provide sufficient F0 cues due to the coarse spectral resolution. Consequently, CI users have difficulty with listening tasks that depend on F0 cues, such as vocal emotion recognition, voice gender discrimination, prosody perception, talker identification, and music perception. When duration and amplitude cues are not available, CI users’ tone recognition worsens due to poor perception of F0 [6] and mainly depends on the similarity between amplitude contours and F0 [7]. The amplitude contour may be more severely affected in dynamic maskers than in steady noise. Furthermore, because pitch contour of the vowel referring to lexical tone is phonemic, changes in pitch contours are likely to involve activity not only in auditory but also in the language pathways.

Spatial cues can also be used to separate speech from maskers. For NH listeners, interaural time differences (ITDs) and interaural level differences (ILDs) can be used to segregate targets and maskers that originate from different spatial locations. There are a number of ways that binaural listening can help speech understanding in noise, including binaural summation, binaural squelch, and binaural spatial release from masking (SRM). Binaural summation describes the advantage of two ears over one ear when segregating co-located target and masker sounds. For bilateral or bimodal CI listeners, binaural summation has been

shown to improve SRTs by 1–2 dB [8]. Binaural squelch describes the advantage when adding the ear with the poorer signal-to-noise ratio (SNR) to the ear with the better SNR, when target and masker sounds are spatially separated. Binaural squelch benefits have been shown to be similarly small. SRM describes advantage with binaural listening when target and masker sounds are spatially separated rather than co-located. SRM has been shown to exhibit substantial binaural benefits in CI users, with 5–6 dB of improvement in SRT, most likely due to head shadow effects [8]. However, SRM in CI users remains smaller than that observed in NH listeners [9].

While bilateral implantation or bimodal listening may improve binaural perception, the vast majority of Mandarin-speaking CI users are unilateral CI users due to stringent implant criteria and financial considerations. However, when speech and noise sources are spatially separated, unilateral CI patients may still benefit from head shadow effects. It is unclear whether head shadow benefit may be affected by different types of noise and/or the limited spectral resolution in CIs. In this study, speech understanding in different types of maskers (steady noise, multi-talker babble, and music) was measured in Mandarin-speaking NH and CI users; the speech was always presented directly in front of the listener (0°), and the masker was presented from 1 of 5 locations (–90°, –45°, 0°, +45°, and +90°, where the minus and plus signs indicate that speaker was to the left or right of the listener, respectively). The motivation for this study was to better understand how tonal components of Mandarin might interact with different types of maskers, as well as how head shadow might interact with different types of maskers and/or the spectral resolution (i.e., the full resolution afforded by NH listeners versus the reduced resolution experienced by CI users).

Methods

Subjects

Twenty-two Mandarin-speaking CI users (11 males and 11 females) participated in this study. Eight CI subjects were adults (mean age at testing: 26.8 yrs; range: 18.8 – 54.1) and fourteen were children (mean age at testing: 12.3 yrs; range: 9.8 – 16.3). Eight subjects were post-lingually deafened and fourteen were pre-lingually deafened. All CI users were implanted at the Eye, Ear, Nose & Throat Hospital, Fudan University, and all had at least one year of experience with their device. All CI subjects were unilaterally implanted with Cochlear Corp. devices, and all used the Advanced Combination Encoder (ACE) strategy. CI subjects who wore hearing aids (HAs) for everyday listening were asked to remove the HA for testing. CI subjects were tested using their clinical speech processors and settings. Volume and sensitivity were set for comfortably loud conversation levels and were not changed during testing. Table 1 lists demographic information for the CI subjects. Across all CI subjects, the mean age at testing was 17.6 yrs (range: 9.8–54.1). The mean age for diagnosis of deafness onset was 6.9 yrs (range: 0–42). The mean duration of deafness was 2.2 yrs (range: 0.3–7.7). The mean age at implantation was 9.1 yrs (range: 1.0–43.8). The mean amount of CI experience was 8.5 yrs (range: 1.8–14.8).

Six NH listeners (3 males and 3 females) served as experimental controls. The mean NH age at testing was 21.5 yrs (range: 11–30). NH listeners all had pure-tone thresholds better than 20 dB HL for audiometric frequencies 250, 500, 1000, 2000, 4000, and 8000 Hz.

Stimuli

Word-in-sentence recognition was assessed using the Mandarin Speech Perception (MSP) materials, which consist of 5 lists with 20 sentences within each list [10]. Each sentence contains 7 monosyllabic words, resulting in a total of 140 monosyllabic words for each list. All test sentences are familiar and commonly used in daily life; sentences were phonetically balanced across lists. The MSP materials consist of high quality digital recordings of speech produced by an experienced female Mandarin talker speaking at a normal rate. The mean duration across all sentences was 1974 ms (± 129), the mean speaking rate was 3.55 words per second (± 0.08), and the mean fundamental frequency (F0) was 223 Hz (± 15).

Sentence recognition was measured in three types of maskers, as used in Eskridge et al. [2]: 1) speech-shaped steady state noise (SSN), 2) six-talker babble, and 3) an excerpt of music. The maskers differed in terms of spectro-temporal properties. The SSN was spectrally matched to average spectrum across all of the MSP sentences, and was thus expected to produce maximal energetic masking. The spectrum of the six-talker babble overlapped that of the MSP sentences, but also contained temporal fluctuations; as such, the babble produced both energetic and informational masking. The spectrum of the music excerpt (the “Power theme” sample used in Eskridge et al. [2]) partially overlapped that of the MSP sentences and contained temporal fluctuations according to the musical performance; as such, the music masker was expected to provide less energetic masking than the SSN and babble maskers and a different type of informational masking than the babble masker in terms of temporal fluctuations and content. Note that the “Power theme” sample was used in this study because it produced the largest interference with CI users’ speech understanding in Eskridge et al. [2].

Spatial test setup

The spatial test setup included 5 loudspeakers separated by 45° (−90°, −45°, 0°, +45°, and +90°, where the minus and plus signs indicate that speaker was to the left or right of the listener, respectively), positioned at ear level 1 meter away from the center of subject’s head. Figure 1 illustrates the setup with speaker labels for NH listeners (top) and CI users (bottom). Note the speakers are labeled for CI users with respect to the CI ear. Thus, when the noise was presented to the CI side (ipsilateral), the speakers are labeled “I-90” and “I-45.” When the noise was presented to non-implanted ear (contralateral), the speakers are labeled “C-90” and “C-45.” When the noise was presented directly in front of the CI user, the speaker is labeled “0°” (the same as for NH listeners). Speech and noise stimuli were delivered to the loudspeakers (Pyle audio) via multi-channel audio interface (MOTU 16a) connected to a multi-channel amplifier (Dayton Audio 1240a).

Procedure

Sentence recognition in noise was measured using an adaptive procedure in which the masker level was adjusted according to the correctness of response, ultimately converging on the SNR that produced 50% correct word-in-sentence recognition, defined as the speech reception threshold (SRT). Before testing in noise, sentence recognition was tested in quiet to ensure good performance; subjects were required to score > 80% correct to participate in

the study. This preliminary test also provided an opportunity to familiarize subjects with the test procedure.

One of five MSP lists (20 sentences in each list) were randomly selected for testing at each masker condition; the masker conditions and loudspeaker locations were randomly ordered within and across subjects. During testing, a sentence was randomly selected from the list and delivered to the 0° speaker location at 60 dBA. The masker was presented to 1 of the 5 speaker locations at the target SNR. Subjects were instructed to repeat the sentence as accurately as possible, and to guess if they were not sure. No trial-by-trial feedback was provided during the test. If the subject repeated 50% or more words correctly, the SNR was reduced by 2 dB. If the subject repeated less than 50% of the words correctly, the SNR was increased by 2 dB. From trial to trial, the target SNR was calculated according to the long-term RMS of the speech and masker. The SRT was calculated as the average of the last 6 reversals in SNR. For most CI children, a parent or companion was present during testing to reduce any anxiety. Subjects were tested in a single session and were allowed to take breaks at any time they felt fatigued. The total amount of testing time was 60–100 minutes. The same examiner administered all test sessions.

Statistical analyses

Performance was analyzed within subject groups using repeated-measures analyses of variance (RM ANOVAs). Post-hoc Bonferroni pair-wise comparisons were analyzed where appropriate. Performance was also compared across subject using ANOVAs. Performance was also compared to that of the English-speaking CI subjects in Eskridge et al. [2], as the masker stimuli were similar between the present and the previous study.

Results

Figure 2 shows the mean SRTs (across NH subjects) for the three masker types as a function of masker loudspeaker position; in all conditions, speech was presented directly in front of the listener (0°). Across all masker location conditions, the mean SRT was –11.3 dB for the SSN masker, –14.2 dB for the babble masker, and –16.0 dB for the music masker. Across all masker types, the mean SRT was –15.2 dB for the –90° location, –14.0 dB for the –45° location, –10.8 dB for the 0° location, –14.2 dB for the +45° location, and –15.0 dB for the +90° location. A two-way RM ANOVA, with masker type (SSN, babble, music) and masker location (–90°, –45°, 0°, +45°, and +90°) as factors, showed significant effects of masker type [$F(2, 40) = 43.5$; $p < 0.001$] and locations [$F(4, 40) = 139.8$, $p < 0.001$]; there was no significant interaction [$F(8, 40) = 0.89$; $p = 0.532$]. Post-hoc Bonferroni pairwise comparisons showed significant differences between most masker locations ($p < 0.05$), except for –45 versus +45 and –90 versus +90 ($p > 0.05$ in both cases). Post-hoc Bonferroni pairwise comparisons also showed significant differences among the masker types ($p < 0.05$ in all cases).

Figure 3 shows mean SRTs (across CI subjects) for the three masker types as a function of masker loudspeaker position; in all conditions, speech was presented directly in front of the listener (0°). Across all masker location conditions, the mean SRT was –1.1 dB for the SSN masker, +2.9 dB for the babble masker, and –0.7 dB for the music masker. Across all masker

types, the mean SRT was -3.8 dB for the C-90 location, -2.0 dB for the C-45° location, $+0.8$ dB for the 0° location, $+3.1$ dB for the I-45° location, and $+3.9$ dB for the I-90° location. A two-way RM ANOVA, with masker type (SSN, babble, music) and masker location (C-90, C-45, 0°, I-45, and I-90) as factors, showed significant effects of masker type [$F(2, 168) = 58.2$; $p < 0.001$] and masker location [$F(4, 168) = 293.1$; $p < 0.001$]; there was no significant interaction [$F(8, 168) = 0.81$; $p = 0.599$]. Post-hoc Bonferroni pairwise comparisons showed significant differences between most masker locations ($p < 0.05$), except for I-45 versus I-90 ($p > 0.05$). Post-hoc Bonferroni pairwise comparisons also showed significant difference between the babble and the SSN maskers ($p < 0.05$) and between the babble and the music maskers ($p < 0.05$), but not between the SSN and the music ($p > 0.05$).

Figure 4 shows mean head-shadow benefit for CI users. Head shadow was calculated as the difference in SRT when the masker was presented ipsilateral vs contralateral to the CI ear. In general, there was little difference in head shadow among the three maskers, but there was a stronger head shadow effect for the 90° location than for the 45° location. A two-way RM ANOVA, with masker type and location as factors, showed a significant head shadow effect for masker location [$F(1, 42) = 74.9$; $p < 0.001$], but not for masker type [$F(2, 42) = 0.8$; $p = 0.461$]; there were no significant interactions [$F(2, 42) = 0.1$; $p = 0.879$].

Because similar maskers were used, data from this study were compared to those from Eskridge et al. [2], in which SRTs were adaptively measured using HINT sentences [9]. In Eskridge et al. [2], SRTs were measured only for 0° azimuth; accordingly, only data for the 0° masker location were used for this study. Figure 5 shows boxplots for English sentences (from Eskridge et al. [2]) and Mandarin sentences (from the present study) as function of masker type, for NH and CI listeners. While the pattern of results is similar across studies, SRTs were markedly lower for the present Mandarin sentences. A split-plot RM ANOVA was performed on the NH data shown in Figure 5, with masker type as the within-subject factor and subject group (English, Mandarin) as the between-subject factor. Results showed significant effects for masker type [$F(2, 26) = 41.8$; $p < 0.001$] and subject group [$F(1, 13) = 847.8$; $p < 0.001$]; there was a significant interaction [$F(2, 26) = 4.8$; $p = 0.016$]. A split-plot RM ANOVA was also performed on the CI data shown in Figure 5, with masker type as the within-subject factor and subject group as the between-subject factor. Results showed significant effects for masker type [$F(2, 64) = 46.0$; $p < 0.001$] and subject group [$F(1, 32) = 32.3$; $p < 0.001$]; there was no significant interaction [$F(2, 64) = 3.5$; $p = 0.631$].

CI data was analyzed in terms of demographic factors. For the 0° location (standard measure for speech understanding in noise), the mean SRT across all maskers was 2.2 and 0.0 dB for post- and pre-lingually deafened CI subjects, respectively. However, one-way ANOVA, with subject group (pre- or post-lingual) as factor, showed no significant difference between subject groups [$F(1, 21) = 3.9$; $p = 0.062$]. Note that most CI recipients in China are pre-lingually deafened and implanted at an early age, while many post-lingually deafened patients have a long duration of deafness. CI speech performance data were compared to demographic factors shown in Table 1. Because many of these factors are inter-related, a principal components analysis (PCA) was performed on the demographic data. Results showed that age at testing and age at diagnosis of deafness explained 65.1% and 23.9% of

the variance, respectively. Linear regression analyses were performed between these two demographic factors and SRTs at the 0° location, and head shadow at C-45 and C-90°; the significance level was adjusted to 0.017 to correct for family-wise error due to multiple comparisons (Bonferroni adjustment). Results showed significant correlations between the SRT at the 0° location and age at testing ($p = 0.005$) and age at diagnosis of deafness ($p = 0.007$). Results showed no significant correlation between either of the demographic factors and head shadow for the C-45 and C-90 locations ($p > 0.2$ in all cases).

Discussion

The data from the present study show that Mandarin-speaking CI users are highly susceptible to masking, but are also able to take advantage of head shadow to separate speech from a variety of maskers. Below, we discuss the study in greater detail.

NH versus CI listeners

In general, CI users were much more susceptible to masking than were NH listeners. At the 0° masker location, the mean NH SRT with the SSN masker was 8.0 dB better than that of CI users, consistent with previous studies with English-speaking NH and CI listeners [1, 2]. With the babble masker, the mean NH SRT was 15.0 dB better than that of CI users, reflecting both the masking release for NH listeners and the increased susceptibility for CI users to dynamic maskers. With the music masker, the mean NH SRT was 12.6 dB better than that of CI users, consistent with Eskridge et al. [2]. Thus, differences between target language did not appear to interact with the deficit in speech performance experienced by CI users, relative to NH listeners.

Speech performance for the present Mandarin-speaking CI listeners was significantly better than that of English-speaking CI listeners in Eskridge et al. [2]. Both studies used SSN that was spectrally matched to the talker, and both used the same babble and music excerpt. Performance differences between the studies may be due to differences between the English HINT and the Mandarin MSP sentences. Both databases contain approximately 7 syllables per sentence. Both studies used Rule 3 [11]. However, scoring for the HINT materials was word-based, and the words contain as many as 3 syllables; if the word was not correctly identified (even if some syllables were), the word was not scored. In the present study, the MSP sentences all contained 7 monosyllable words. As such, performance may have been better in this study as each correctly identified syllable was also a correctly identified word. Note that the M-HINT Mandarin sentence materials [12] contain 10 monosyllables, compared to the 7 monosyllables in this study, perhaps indicating that the present materials were relatively easy. The present SRTs with the SSN masker were also better than observed for English-speaking listeners [1]. In this study, the mean SRTs with SSN for the 0° masker location were -8.1 dB and -0.9 dB for NH and CI listeners, respectively. In Fu & Nogaki [1], the mean SRT with SSN was -4.6 dB and +10.0 dB for NH and CI listeners, respectively. Again, differences in speech materials and scoring may explain differences in SRTs between studies. It is also possible that age at diagnosis of deafness, which was much higher for Fu & Nogaki [1] (mean = 46.6 yrs) than for the present study (mean = 6.9 yrs). However, age at diagnosis of deafness is difficult to compare, as all subjects were post-

lingually deafened in Fu and Nogaki [1] with extended acoustic hearing experience before cochlear implantation. The mean age at implantation for the present Mandarin-speaking CI subjects was 9.1 yrs.

CI performance

Although there was no significant difference between pre- and post-lingual CI subjects, linear regression analyses showed that CI performance was negatively correlated with age at testing and age at diagnosis of deafness. However, each factor explained less than 35% of the variability in CI performance. Note that for the Mandarin-speaking CI population in general, most subjects were implanted at an early age, and thus with an early age at diagnosis of deafness; these tend to be pre-lingually deafened pediatric patients who develop speech with their CI. Late implantation with a prolonged duration of deafness can occur, and there are relatively few post-lingually deafened adult CI users in China. In this study, all subjects were required to score better than 80% in quiet, which may have reduced some of the variability that is observed in the overall CI population.

Head shadow effect

The head shadow effect is elicited by the attenuation of a sound that occurs when it travels around the listeners' head to the opposite ear. For NH listeners, head shadow ranges from 9 to 11 dB [13]. Having a present and equal head shadow effect across ears is one of the most robust and early benefits of bilateral cochlear implantation [14]. However, several studies have reported marked improvement in head shadow effects when comparing unilaterally versus bilaterally CI users [15, 16]. This is primarily because unilateral CI patients will have head shadow for just one ear, provided that noise originates from side of the non-implanted ear. Head shadow effects have been shown to be similar in pediatric and adult bilateral CI listeners [17]. In the present study, head shadow was estimated as the difference in SRT between ipsilateral and contralateral masking. The mean head shadow effect (across three noise maskers) was 5.1 dB and 7.7 dB for the C-45 and C-90 conditions, respectively, somewhat less than observed in NH listeners [13]. Still, the present data suggest that unilateral CI users can make good use of head shadow effects. This is not very surprising, as head shadow reflects advantageous SNRs due to the physical locations of speech, masker, and CI ear.

Masker types

In this study, the best NH performance in terms of SRT was observed with the music masker while the poorest performance was observed with SSN, consistent with previous studies [1, 2]. Most likely, NH subjects were able to listen in the dips of the babble and music maskers where the energetic masking is momentarily reduced [12]. The music maskers also contained pitch and timbre information that was markedly different from target speech. In contrast to NH performance, CI users performed most poorly with the babble masker. The poorer functional spectral resolution most likely increased the perceptual similarity between the babble and the target speech, making them difficult to segregate [18]. Given the poor spectral resolution, the babble masker may have also produced more informational masking [19].

As shown in Figure 5, the effect of noise type on SRTs was similar for English- and Mandarin-speaking CI listeners, albeit with better absolute performance for Mandarin-speaking CI listeners, which was most likely due to differences in speech materials. While masker type affected absolute performance in this study, there were no significant differences among masker types in terms of head shadow benefit for NH or CI listeners. This finding was not consistent with our hypothesis that the tonal nature of Mandarin speech might interact with masker type for head shadow benefit. This finding is also not consistent with previous studies [3, 20]. Hawley et al. [3] found significant effects of and interactions between masker type (speech from the same talker, time reversed speech, speech-shaped noise, and speech-modulated noise) and spatial configurations in NH listeners. Maximum informational masking would be for the speech maskers, but NH listeners could listen in the spectro-temporal dips to hear the target speech. In this study, the maximum informational masking was with the multi-talker babble. While the present NH subjects were able to listen in the dips of the babble, as in Hawley et al. [3], CI subjects did not experience release from masking due to the limited spectro-temporal resolution. Arbogast et al. [20] found much greater SRM with a competing speech masker than with steady noise using 15-channel acoustic CI simulations with NH listeners. The advantage with the speech masker (presumably because of dip-listening) observed in Arbogast et al. [20] may be due to the number of channels (15) used in the CI simulations, which may have overestimated the functional resolution of real CI users, which is approximately 6–8 spectral channels [21]. The better resolution in the CI simulations in Arbogast et al. [20] may have allowed for some release from masking with competing speech that would not be typically observed in the real CI case. Also, the multi-talker babble in this study most likely produced greater energetic masking than the single-talker masker in Hawley et al. [3] or the spectrally-shifted single-talker sentences used in Arbogast et al. [20].

Implications for Mandarin-speaking CI users

Because of the limited number of bilateral and bimodal CI patients currently available in China, we were only able to test perception of spatially separated speech and noise in unilateral CI users. Also, we used a single female talker as the target; performance may be affected by different target talkers and different meaningful speech maskers, which were not tested in this study. Because we observed a significant head shadow effect when only one ear was implanted, we expect that bilateral implantation would offer a much greater advantage for spatially separated sound sources, as optimal SNRs would be available to either ear. Bilateral implantation would also allow CI users to benefit from SRM, further increasing the binaural benefit. As such, China should accelerate bilateral implantation for bilaterally deafened individuals.

Conclusions

Speech understanding with different maskers (steady noise, multi-talker babble, and music) originating from different locations (-90° , -45° , 0° , $+45^\circ$, and $+90^\circ$) was evaluated in Mandarin-speaking CI and NH listeners. Major findings include:

1. Overall, CI performance was much poorer than NH performance for all masker types and locations.

2. SRTs significantly differed across masker types, and the pattern of results was somewhat different for NH and CI listeners. For NH listeners, performance was poorest with steady noise and best with music, reflecting some release from masking with dynamic masking. For CI users, performance was poorest with the babble and best with steady noise, reflecting difficulties with dynamic maskers due to poor spectral resolution.
3. The effect of masker type was similar between the present Mandarin CI subjects and English-speaking CI subjects from Eskridge et al [2], suggesting that the tonal aspects of Mandarin did not interact differently with the spectro-temporal properties of the maskers, compared to non-tonal languages such as English.
4. Head shadow effects were greatest for the -90° and $+90^\circ$ masker locations. Head shadow effects did not significantly differ among masker types. Head shadow effects were not significantly different between CI and NH subjects, suggesting that spectral resolution may not limit head shadow effects.

Acronyms and abbreviations

ACE	Advanced Combination Encoder
CI	cochlear implant
DR	dynamic range
F0	fundamental frequency
HS	Head shadow effect
ILDs	interaural level differences
ITDs	interaural time differences
MSP	Mandarin Speech Perception
NH	normal hearing
PCA	principal components analysis
RM ANOVA	repeated-measures analysis of variance
RMS	root mean square
SNR	signal to noise ratio
SRM	spatial release from masking
SRTs	speech reception thresholds
SSN	steady-state noise
yrs	years

References

1. Fu QJ, Nogaki G. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J Assoc Res Otolaryngol.* 2005; 6:19–27. [PubMed: 15735937]
2. Eskridge EN, Galvin JJ 3rd, Aronoff JM, Li T, Fu QJ. Speech perception with music maskers by cochlear implant users and normal-hearing listeners. *J Speech Lang Hear Res.* 2012; 55:800–810. [PubMed: 22223890]
3. Hawley ML, Litovsky RY, Culling JF. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *J Acoust Soc Am.* 2004; 115:833–843. [PubMed: 15000195]
4. Gfeller K, Turner C, Oleson J, Kliethermes S, Driscoll V. Accuracy of cochlear implant recipients in speech reception in the presence of background music. *Ann Otol Rhinol Laryngol.* 2012; 121:782–791. [PubMed: 23342550]
5. Fu QJ, Shannon RV, Wang X. Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *J Acoust Soc Am.* 1998; 104:3586–3596. [PubMed: 9857517]
6. Fu QJ, Zeng FG. Identification of temporal envelope cues in Chinese tone recognition. *Asia Pac J Speech, Lang Hear.* 2005; 5:45–57.
7. 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]
8. Vermeire K, Van de Heyning P. Binaural Hearing after Cochlear Implantation in Subjects with Unilateral Sensorineural Deafness and Tinnitus. *Audiology and Neurotology.* 2009; 14:163–171. [PubMed: 19005250]
9. Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am.* 1994; 95:1085–1099. [PubMed: 8132902]
10. Fu QJ, Zhu M, Wang X. Development and validation of the Mandarin speech perception test. *J Acoust Soc Am.* 2011; 129:EL267–273. [PubMed: 21682363]
11. Chan JCY, Freed D, Vermiglio AJ, Soli SD. Evaluation of binaural functions in bilateral cochlear implant users. *Int J Audiology.* 2008; 47:296–310.
12. Li N, Loizou PC. Factors influencing glimpsing of speech in noise. *J Acoust Soc Am.* 2007; 122:1165–1172. [PubMed: 17672662]
13. Arsenault MD, Punch JL. Nonsense-syllable recognition in noise using monaural and binaural listening strategies. *J Acoust Soc Am.* 1999; 105:1821–1830. [PubMed: 10089605]
14. Koch DB, Soli SD, Downing M, Osberger MJ. Simultaneous bilateral cochlear implantation: prospective study in adults. *Cochlear Implants Int.* 2010; 11:84–99. [PubMed: 19247979]
15. Dunn CC, Tyler RS, Witt SA. Benefit of wearing a hearing aid on the unimplanted ear in adult users of a cochlear implant. *J Speech Lang Hear Res.* 2005; 48:668–680. [PubMed: 16197280]
16. Laszig R, Aschendorff A, Stecker M, Müller-Deile J, Maune S, Dillier N, et al. Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results. *Otol Neurotol.* 2004; 25:958–968. [PubMed: 15547426]
17. Van Deun L, van Wieringen A, Wouters J. Spatial speech perception benefits in young children with normal hearing and cochlear implants. *Ear Hear.* 2010; 31:702–713. [PubMed: 20548238]
18. Misurelli SM, Litovsky RY. Spatial release from masking in children with bilateral cochlear implants and with normal hearing: Effect of target-interferer similarity. *J Acoust Soc Am.* 2015; 138:319–331. [PubMed: 26233032]
19. Drullman R, Festen JM, Plomp R. Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am.* 1994; 95:1053–1064. [PubMed: 8132899]
20. Arbogast TL, Mason CR, Kidd G Jr. The effect of spatial separation on informational and energetic masking of speech. *J Acoust Soc Am.* 2002; 112:2086–2098. [PubMed: 12430820]
21. Friesen LM, Shannon RV, Baskent D, Wang XS. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am.* 2001; 110:1150–1163. [PubMed: 11519582]

Figure 1. Speaker locations and labels for NH (top panel) and CI listeners (bottom panel). For CI listeners, the ipsilateral and contralateral labels are with respect to the CI ear (filled ear in Figure). I-45 and I-90 refer to maskers presented 45° and 90° ipsilateral to the CI ear, respectively. C-45 and C-90 refer to maskers presented 45° and 90° contralateral to the CI ear, respectively.

Figure 2.

Mean NH SRTs (across subjects) as a function masker location for the different masker types. Speech was always presented from the front (0°). The error bars show the standard error.

Figure 3. Mean CI SRTs (across subjects) as a function masker location for the different masker types. Speech was always presented from the front (0°). The error bars show the standard error.

Figure 4.

CI users' head shadow benefit as a function of masker type for the 45° and 90° masker locations. Head shadow benefit was calculated as the difference in SRTs between the ipsilateral and contralateral masker locations (i.e., I-45 – C-45, I-90 – C-90). The error bars show the standard error.

Figure 5.

Boxplots for NH (left panel) and CI listeners' SRTs as a function of masker type. English-speaking data are from Eskridge et al. (2012). Mandarin-speaking data are from the present study with the 0° masker location. In each box, the solid line shows the median, the error bars show the 10th and 90th percentiles, and the circles show outliers.

Table 1

CI subject demographics.

Subject	Gender	Age test yrs	Age deaf yrs	Dur deaf yrs	Age imp yrs	CI exp yrs	Pre/Post	L/R	Etiology	Music exp
S1	F	22.8	20.3	0.8	21.1	1.8	Post	L	Unknown	N
S2	F	26.1	16.0	7.7	23.7	2.4	Post	R	Drug	N
S3	F	23.3	1.5	2.3	8.5	14.8	Post	R	Unknown	N
S4	F	22.5	16.9	3.8	20.7	1.8	Post	L	Unknown	Y
S5	F	15.8	0.7	0.3	1.0	14.8	Pre	R	Genetic	Y
S6	F	13.0	0.3	1.8	2.1	10.9	Pre	R	Genetic	Y
S7	M	17.2	0.0	3.0	3.0	13.3	Pre	R	Genetic	Y
S8	M	20.9	11.8	2.7	14.4	6.5	Post	R	Unknown	Y
S9	F	54.1	42.0	1.8	43.8	10.3	Post	R	Unknown	N
S10	M	11.0	0.0	2.3	2.3	8.7	Pre	L	Genetic	N
S11	M	10.4	0.5	0.8	1.3	9.2	Pre	R	Genetic	N
S12	F	10.3	2.2	1.9	4.1	6.2	Pre	R	Genetic	N
S13	M	18.8	10.1	3.8	13.8	5.0	Post	R	Genetic	N
S14	M	9.8	0.7	0.3	1.0	8.8	Pre	R	Genetic	N
S15	F	10.2	0.4	3.5	3.9	6.3	Pre	R	Genetic	N
S16	M	26.0	20.6	2.0	22.6	3.4	Post	R	Unknown	Y
S17	M	13.4	0.6	0.6	1.2	12.3	Pre	R	Genetic	N
S18	M	11.9	0.7	0.9	1.6	10.3	Pre	R	Genetic	N
S19	F	12.7	2.3	1.9	4.2	6.5	Pre	L	Genetic	N
S20	M	13.3	0.0	1.3	1.3	12.1	Pre	R	Genetic	N
S21	F	13.1	0.3	1.2	1.5	11.4	Pre	R	Genetic	N
S22	M	12.8	0.0	3.3	3.3	9.5	Pre	R	Genetic	N
AVE		17.7	6.7	2.2	9.1	8.5				

F = female; M = male; Age Test = age at testing; yrs = years; Age Deaf = age at diagnosis of deafness; Dur Deaf = duration of deafness; Age Imp = age at implantation; CI Exp = CI experience; Pre = prelingual; Post = postlingual; L = left; R = right; Music exp = Music experience; N = no; Y = yes