Masking Effects in the Perception of Multiple Simultaneous Talkers in Normal-Hearing and Cochlear Implant Listeners

Biao Chen[1] ⓘ, Ying Shi[1], Lifang Zhang[1], Zhiming Sun[1],
Yongxin Li[1] ⓘ, Quinton Gopen[2], and Qian-Jie Fu[2] ⓘ

Abstract

For normal-hearing (NH) listeners, monaural factors, such as voice pitch cues, may play an important role in the segregation of speech signals in multitalker environments. However, cochlear implant (CI) users experience difficulties in segregating speech signals in multitalker environments in part due to the coarse spectral resolution. The present study examined how the vocal characteristics of the target and masking talkers influence listeners' ability to extract information from a target phrase in a multitalker environment. Speech recognition thresholds (SRTs) were measured with one, two, or four masker talkers for different combinations of target-masker vocal characteristics in 10 adult Mandarin-speaking NH listeners and 12 adult Mandarin-speaking CI users. The results showed that CI users performed significantly poorer than NH listeners in the presence of competing talkers. As the number of masker talkers increased, the mean SRTs significantly worsened from −22.0 dB to −5.2 dB for NH listeners but significantly improved from 5.9 dB to 2.8 dB for CI users. The results suggest that the flattened peaks and valleys with increased numbers of competing talkers may reduce NH listeners' ability to use dips in the spectral and temporal envelopes that allow for "glimpses" of the target speech. However, the flattened temporal envelope of the resultant masker signals may be less disruptive to the amplitude contour of the target speech, which is important for Mandarin-speaking CI users' lexical tone recognition. The amount of masking release was further estimated by comparing SRTs between the same-sex maskers and the different-sex maskers. There was a large amount of masking release in NH adults (12 dB) and a small but significant amount of masking release in CI adults (2 dB). These results suggest that adult CI users may significantly benefit from voice pitch differences between target and masker speech.

Keywords

cochlear implant, vocal characteristics, speech recognition threshold, multitalker environment

Received 10 June 2019; revised 4 March 2020; accepted 9 March 2020

Everyday speech communication often requires listeners to understand the messages from a specific target (e.g., a specific talker or a talker from specific spatial location) that are masked by one or more competing talkers from the same or different spatial locations. When the target and masker talkers originate from the same direction relative to the listener, or when the target and masker talkers are presented monaurally, the spatial acoustic cues are not available for talker segregation. In this case, listeners must rely on monaural cues to segregate the competing messages. Many acoustic cues can be used to segregate the competing speech, including vocal

[1]Department of Otolaryngology, Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Ministry of Education of China
[2]Department of Head and Neck Surgery, David Geffen School of Medicine, University of California

Corresponding Authors:
Yongxin Li, Department of Otolaryngology, Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing 100730, China.
Email: entlyx@sina.com
Qian-Jie Fu, Department of Head and Neck Surgery, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.
Email: qfu@mednet.ucla.edu

characteristics (e.g., vocal tract, fundamental frequency, voice pitch, etc.; Başkent & Gaudrain, 2016; Brokx & Nooteboom, 1982; Brungart, 2001; Brungart et al., 2001; Darwin et al., 2003; Darwin & Hukin, 2000; Drullman & Bronkhorst, 2004; El Boghdady et al., 2019; Vestergaard et al., 2009), prosodic features (e.g., Darwin & Hukin, 2000), and overall speech levels (e.g., Bregman, 1994).

The effects of vocal characteristics and differing speakers on the ability to segregate target and masker talkers have been well documented in normal-hearing (NH) listeners (e.g., Brungart 2001; Brungart et al. 2001). Brungart (2001) measured the intelligibility of a target talker masked by a single competing talker as a function of the signal-to-noise ratio (SNR). The competing talker was the same as the target talker, was the same sex as the target talker ("same-sex masker"), or was a different sex from the target talker ("different-sex masker"). The results implied that the amount of masking strongly depended on the similarity between the target and masker talkers in terms of vocal characteristics. Performance was best with the different-sex masker and worst when the masker talker was the same as the target. Similar results were also reported by Darwin et al. (2003), where the vocal characteristics were directly manipulated on the same target/masker speaker. In a follow-up study, Brungart et al. (2001) further examined the effects of the vocal characteristics of competing talkers on listeners' ability to recognize target speech in the presence of three or four competing talkers. Similar to when there was only one competing talker, recognition performance decreased when the target and masker talkers had similar vocal characteristics. Similarly, Cullington and Zeng (2008) measured speech recognition with varying number of competing talkers in NH listeners, finding a significant advantage with fewer competing talkers and significant masking release with up to three competing talkers. NH listeners also benefited from voice pitch differences between the target and maskers, with more masking produced by the same-sex maskers than by the different-sex maskers. They argued that NH listeners may have attended to favorable SNRs in the temporal and/or spectral gaps, thereby obtaining masking release with one or two competing talkers. Several studies examined the energetic masking (EM) component of speech-on-speech masking (Anzalone et al., 2006; Brungart et al., 2006), finding that EM plays a relatively small role when speech is masked by interfering speech but a much greater role when speech is masked by interfering noise. Note that EM was defined as the loss of detectable target information due to the temporal and spectral overlap between the target and maskers. However, adding more maskers (three or more) likely fills in these temporal and spectral gaps, resulting in increased EM as the target and maskers produce overlapping excitation patterns in the auditory

nerve. As the number of masker talkers increase, NH listeners may be able to hear some words in the competing speech, but they cannot decide whether the words were spoken by the target or the maskers. Such uncertainty will also increase informational masking (IM; Brungart et al., 2001; Durlach et al., 2003; Kidd et al., 2016). Note that in this study, IM refers to listening situations where the target and masker signals are clearly audible, but the listener cannot segregate the target from similar-sounding distracters. Kidd et al. (2016) found a large masking release with different-sex maskers when compared with the same-sex maskers, primarily due to the reduction in IM. These results imply that voice pitch cues play an important role in the segregation of speech signals in multitalker environments. Increasing the number of competing talkers may limit NH listeners' ability to use voice pitch cues to help segregate target and competing speech.

For cochlear implant (CI) users, the vocal characteristics of target and masker talkers may not be discriminable due to the lack of fine spectrotemporal information (Gaudrain & Başkent, 2018). As such, CI users are less able to take advantage of differences in the spectrotemporal properties to segregate competing talkers. Previous studies have shown reduced sensitivity to voice pitch (related to fundamental frequency, or F0) and vocal-tract length (VTL; related to the height of the speaker and formant frequencies) in CI users (Gaudrain & Başkent, 2018). El Boghdady et al. (2019) also found that the sensitivity to both F0 and VTL was correlated with the intelligibility of speech masked by a background talker. Cullington and Zeng (2008) also measured speech recognition with varying number of competing talkers in CI users. In contrast to the NH data, there was no significant advantage with fewer competing talkers in CI users. However, CI users did benefit from voice pitch differences between the target and maskers, with the different-sex masker providing significantly less masking than the same-sex masker. Similar results were also reported by a few other studies (Meister et al., 2020; Visram & McKay, 2012). These results suggest that voice pitch cues also play an important role in CI users' ability to segregate targets from competing speech.

Different from English, Mandarin Chinese is a tonal language in which lexical tones convey linguistic meaning (Liang, 1963). While the F0 is the primary cue for lexical tones, listeners may also make use of duration and amplitude cues that covary with F0 to recognize lexical tones (Liang, 1963). Due to the lack of fine spectrotemporal information, voice pitch is not well perceived by CI users, which limits the recognition of lexical tones for Mandarin-speaking CI users (Fu & Zeng, 2005; Luo et al., 2009). Mandarin-speaking CI users depend more strongly on the covarying amplitude contour to recognize lexical tones (Fu & Zeng, 2005).

The amplitude contour of the target is likely more disrupted by competing speech than by speech-shaped steady noise (SSN), thus negatively affecting Mandarin-speaking CI users' ability to correctly perceive lexical tones with competing talkers. Luo et al. (2009) measured Mandarin-speaking CI users' ability to recognize concurrent vowels, tones, and syllables. They found that concurrent vowel and syllable recognition were not significantly different between the same- and different-talker conditions. However, concurrent tone recognition was significantly better with the same-talker condition, consistent with the eight-channel CI simulation results in Luo and Fu (2009). Such unexpected results were likely because lexical tone recognition primarily depends on the amplitude contour in CI users and CI simulations, as pitch contours could not be reliably detected due to the lack of fine spectrotemporal information (Fu & Zeng, 2005). The interference between concurrent tones may undermine the benefits of voice pitch cues in Mandarin-speaking CI users with competing-talker backgrounds.

Tao et al. (2018) measured speech recognition thresholds (SRTs) in SSN or in the presence of a masker talker with the same or different sex as the target in pediatric Mandarin-speaking CI and NH listeners. Similar to previous studies (e.g., Leibold et al., 2018), Mandarin-speaking NH children were able to greatly benefit from voice pitch differences between the target and masker talker. Mandarin-speaking CI children performed significantly better with SSN than with the competing talker. In contrast to Cullington and Zeng (2008), there was no significant performance difference in SRTs between the different-sex masker and the same-sex masker in Mandarin-speaking CI children. Poor CI performance in competing speech may be due to IM or some kind of modulation interference, as CI users appear unable to take advantage of temporal glimpsing when there is only one competing talker. As the number of competing talkers increases, the temporal gaps may be filled up and the amplitude contour of the resultant masker signals may be flattened. In Cullington and Zeng (2008), varying the number of competing maskers had little effect on English-speaking CI users who were unable to take advantage of temporal glimpsing. The flattened temporal envelope of the multitalker masker may be less disruptive to the amplitude contour of the target. Due to the importance of the amplitude contour for lexical tone recognition in Mandarin-speaking CI users (Fu & Zeng, 2005), it is possible that Mandarin-speaking CI users' recognition of target speech may improve as the number of competing talkers increases.

In the present study, we measured recognition of target speech in the presence of one, two, or four competing talkers in 10 adult Mandarin-speaking NH listeners and 12 adult Mandarin-speaking CI users. A number of different combinations of target-masker vocal characteristics were tested. We hypothesized that for Mandarin-speaking CI users, recognition of the target speech would significantly improve as the number of masker talkers was increased, due to the resultant flattening of the masker amplitude contour. We further hypothesized that Mandarin-speaking CI users would not benefit from voice pitch differences between the target and masker, as the interference between concurrent tones may undermine the benefits of voice pitch cues with competing-talker backgrounds (Luo et al., 2009).

Methods

Subjects

Twelve adult Mandarin-speaking Chinese CI users participated in the study (eight males and four females). The mean age at testing was 29.6 years (range = 18–47 years), the mean duration of auditory deprivation was 11.2 years (range = 0.5–29.7 years), and the mean CI experience was 3.4 years (range = 0.4–16.5 years). CI subject demographic information is shown in Table 1. Ten NH adults (5 males and 5 females; mean age = 25.7 years, range = 23–30 years) served as experimental controls for the CI users. All NH subjects had pure tone thresholds <20 dB HL at all audiometric frequencies between 125 and 8000 Hz. All CI and NH subjects were native speakers of Mandarin. In compliance with ethical standards for human subjects, written informed consent was obtained from all participants before proceeding with any of the study procedures. This study was approved by the institutional review board in Beijing Tongren Hospital, Capital Medical University.

Test Materials

The Closed-set Mandarin Speech (CMS; Tao et al., 2017) test materials were used to test speech understanding in the presence of one or more competing talkers. The CMS test materials consist of familiar words selected to represent the natural distribution of vowels, consonants, and lexical tones found in Mandarin Chinese. Ten keywords in each of five categories (Name, Verb, Number, Color, and Fruit) were produced by a native Mandarin talker, resulting in a total of 50 words that can be combined to produce 100,000 unique sentences. The CMS test materials produced by three male and two female native Mandarin talkers were used in the present study. One of the three male talkers was selected as the target talker (mean F0 across all 50 words: 139 Hz). The other two male talkers were used as the competing talkers (mean F0s: 143 and 178 Hz). The two female talkers were also used as the competing talkers (mean F0s: 208 and 248 Hz).

SRTs, defined as the SNR that produced 50% correct word recognition, were adaptively measured using a modified coordinate response matrix (CRM) test

Table 1. Cochlear Implant (CI) Subject Demographic Information.

Subject	CI ear	Sex	Age at testing (years)	Duration of auditory deprivation (years)	Duration of CI use (years)	Hearing age (years)	Device
S1	Left	Male	27	24.4	0.6	2.6	MED-EL
S2	Left	Male	18	1.5	16.5	16.5	Cochlear
S3	Right	Female	18	1.6	16.4	16.4	MED-EL
S4	Right	Female	24	11.0	0.6	13.0	Nurotron
S5	Right	Male	47	1.1	0.5	45.9	MED-EL
S6	Left	Male	26	6.9	0.6	19.1	Cochlear
S7	Right	Male	45	10.1	0.5	34.9	MED-EL
S8	Left	Male	22	7.4	0.4	14.6	Advanced Bionics
S9	Left	Female	29	0.8	0.9	28.2	Advanced Bionics
S10	Left	Male	36	29.7	0.9	6.3	Nurotron
S11	Left	Female	36	20.1	1.5	15.9	Nurotron
S12	Both	Male	27	19.7	0.9	7.3	MED-EL

Note. The shaded cells indicate CI subjects with less than 16 years of hearing age; hearing age was defined as the difference between age at testing and duration of auditory deprivation.

(Brungart et al., 2001; Tao et al., 2017, 2018). Similar to CRM tests, two target keywords (randomly selected from the Number and Color categories) were embedded in a five-word carrier sentence uttered by the Mandarin-speaking male target talker. The first word in the target sentence was always the Name "Xiaowang," followed by randomly selected words from the remaining categories. Thus, the target sentence could be (in Mandarin) "**Xiaowang** sold *Three Red* strawberries," "**Xiaowang** chose *Four Brown* bananas," and so forth (Name to cue target talker in **bold**; keywords in ***bold italic***).

Recognition of the target keywords was measured in the presence of one or more competing talkers. The number of competing talkers ranged from one to four, and the competing talkers had a combination of different vocal characteristics. For the purpose of comparison, the acronym (i.e., TS, TD, etc.) was adopted from Brungart et al. (2001) to code the combination of different vocal characteristics. T represents the target (*male* talker), S indicates that the competing talker has the same voice gender as the target (i.e., male talker), and D indicates that the competing talker has a different voice gender as the target (i.e., female talker). Six different combinations based on the number and vocal characteristics of competing talkers were generated, including one female talker (TD; mean F0 across all words = 248 Hz), one male talker (TS; mean F0 across all words = 178 Hz), two female talkers (TDD; mean F0s: 208 and 248 Hz), one male and one female (TSD; mean F0s: 178 and 248 Hz), two male talkers (TSS; mean F0s: 143 and 178 Hz), or two male and two female talkers (TSSDD; mean F0s: 143, 178, 208, and 248 Hz). For the competing talkers, masker sentences were randomly generated for each test trial using the CMS materials; words were randomly selected from each category, excluding the words used in the target

and other masker sentences. Thus, the Chinese masker could be "Xiaozhang saw *Two Blue* kumquats," "Xiaodeng took *Eight Green* papayas," and so forth (competing keywords in italic). For conditions with multiple competing talkers, the target and all maskers have different words for each category. Figure 1 shows the waveforms mixed at 0 dB SNR for the target (red) and different maskers (TD, TS, TDD, TSD, TSS, and TSSDD). Note as the number of competing talkers increases, the spectrotemporal dips in speech maskers begin to weaken and the amplitude contour of the masker becomes flatter.

Test Protocols

Due to the expected wide range in SRTs, a fixed overall presentation level was used instead of a fixed target level or a fixed masker level to avoid overly loud sounds for some experimental conditions. Once target and competing sentences were combined according to a specific SNR, the overall speech level was further adjusted to have the same root mean square value in each presentation for all experimental conditions. All stimuli were presented in sound field at 65 dBA via a single loudspeaker; subjects were seated directly facing the loudspeaker at a 1-m distance. For CI users, SRTs were measured using the clinical settings for their device, which were not changed throughout the study. During each test trial, a sentence was presented at the designated SNR; the initial SNR was 10 dB. Subjects were instructed to listen to the target sentence (produced by the male target talker and beginning with the name "Xiaowang") and then click on 1 of the 10 response choices for each of the Number and Color categories; no selections could be made from the remaining categories, which were grayed out. If the subject correctly identified both keywords, the SNR was reduced by 4 dB
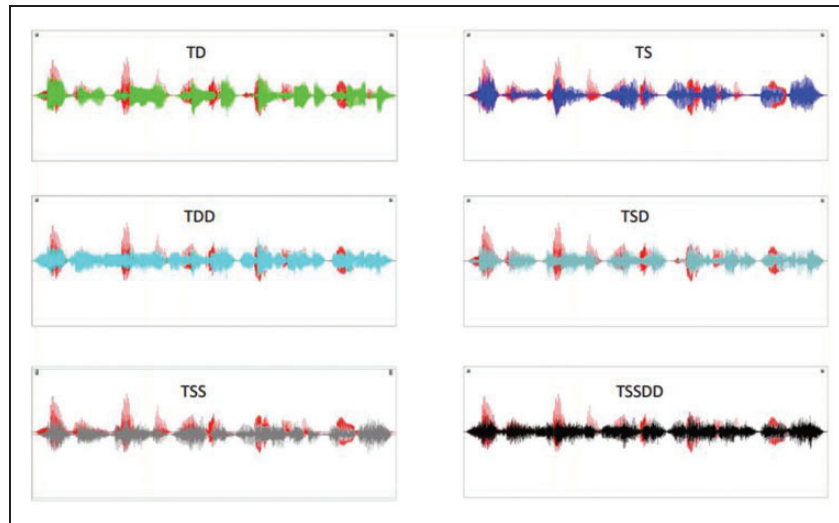
Figure 1. Target Sentence (Red) Combined With Different Masker Conditions at 0 dB SNR.
TD = one female talker; TS = one male talker; TDD = two female talkers; TSD = one male and one female talker; TSS = two male talkers; TSSDD = two male and two female talkers.

(initial step size); if the subject did not correctly identify both keywords, the SNR was increased by 4 dB. After two reversals, the step size was reduced to 2 dB. The SRT was calculated by averaging the last six reversals in SNR. If there were fewer than 6 reversals with 2-dB step size within 20 trials, the test run was discarded and another run was executed. Two test runs were completed for each condition, and the SRT was averaged across runs. The masker conditions were randomized within and across subjects.

Results

Effect of Number of Competing Talkers on SRTs

Figure 2 shows mean SRTs as a function of number of competing talkers for the NH and CI groups. Note that the one-talker masker data represent the mean scores averaged across the TS and TD conditions, the two-talker masker data represent the mean scores averaged across the TDD, TSD, and TSS conditions, and the four-talker masker data represent the mean score for the TSSDD condition. For NH listeners, mean SRTs gradually worsened from –22.0 dB to –5.2 dB as the number of competing talkers increased from 1 to 4. For CI users, mean SRTs gradually improved from +5.9 dB to +2.8 dB as the number of masker talkers increased from 1 to 4. A split plot analysis of variance (ANOVA) was performed on the data shown in Figure 2, with number of competing talkers (one, two, and four) as the within-subject factor and group (NH and CI) as the between-subject factor. Results showed significant effects for the number of competing talkers, $F(2, 40) = 94.36$, $p < .001$, and group, $F(1, 20) = 128.12$, $p < .001$; there was a
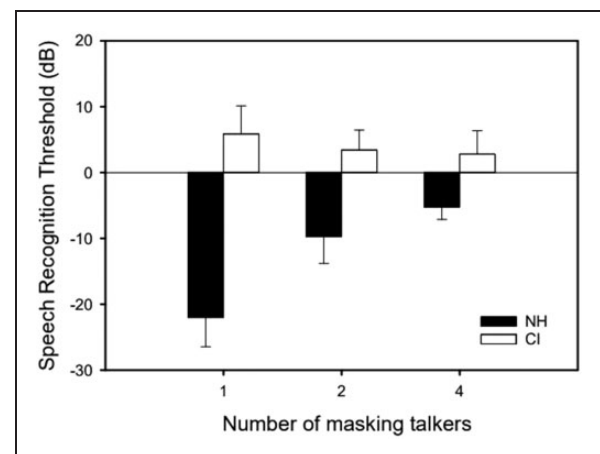

Figure 2. SRTs as a Function of the Number of Competing Talkers in NH Listeners (Black) and CI Users (White). The error bars show the standard deviation.
NH = normal-hearing; CI = cochlear implant.

significant interaction, $F(2, 40) = 2$ 00.01, $p < .001$. Due to a significant interaction, within-subject effects were tested independently for each subject group. For NH subjects, a one-way repeated measures (RM) ANOVA showed a significant effect for the number of competing talkers, $F(2, 18) = 162.05$, $p < .001$. Holm–Sidak pairwise comparisons showed that mean SRTs were significantly better with the one-talker masker than with the two-talker masker ($p < .001$) or the four-talker masker ($p < .001$) and significantly better with the two-talker masker than with the four-talker masker ($p < .001$). For CI users, a one-way RM ANOVA showed a significant effect for the number of competing talkers, $F(2, 22) = 21.72$, $p < .001$. Holm–Sidak pairwise comparisons

showed that mean SRTs were significantly *poorer* with the one-talker masker than with the two-talker masker ($p < .001$) or the four-talker masker ($p < .001$), with no significant difference between the two- and four-talker maskers ($p = .692$).

Interaction of Vocal Characteristics in the Multiple Competing Talkers

Figure 3 shows mean SRTs for the NH and CI groups for the six masker conditions. For NH listeners, the best SRT was observed for the TD condition and the worst for the TSS condition. For CI users, the best SRT was observed for the TDD condition and the worst SRT for the TS condition. A split plot ANOVA was performed on the data shown in Figure 3, with listening condition (TD, TS, TDD, TSD, TSS, and TSSDD) as the within-subject factor and group (NH and CI) as the between-subject factor. Results showed significant effects for listening condition, $F(5, 100) = 70.76$, $p < .001$, and group, $F(1, 20) = 175.52$, $p < .001$; there was a significant interaction, $F(5, 100) = 90.050$, $p < .001$. Due to the significant interaction, within-subject effects were tested independently for each subject group. For NH subjects, a one-way RM ANOVA showed a significant effect for listening condition, $F(5, 45) = 77.95$, $p < .001$. Holm–Sidak pairwise comparisons showed that SRTs were significantly better for the TD condition than for the remaining conditions ($p < .001$). SRTs were significantly poorer for the TSS condition than for the TD, TS, or TDD conditions ($p < .001$ in all cases) and significantly poorer for the TSSDD condition than for the TD, TS, or
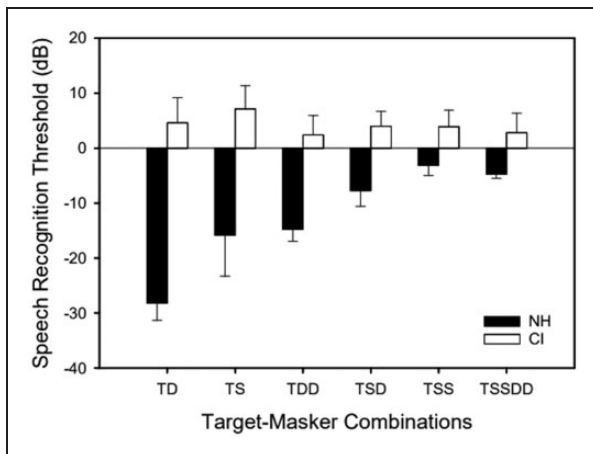
TDD conditions ($p < .001$ in all cases). SRTs were also significantly poorer for the TSD than for the TSS condition ($p = .014$). However, there was no significant difference between the TSD and TSSDD conditions ($p = .150$) or between the TSS and TSSDD conditions ($p = .484$). For CI users, a one-way RM ANOVA showed a significant effect for listening condition, $F(5, 55) = 17.34$, $p < .001$. Holm–Sidak pairwise comparisons showed that SRTs were significantly poorer for the TS condition than for the remaining conditions ($p < .001$ in all cases). SRTs were also significantly poorer for the TD than for the TDD ($p = .003$) and TSSDD conditions ($p = .023$). No significant differences were observed among the remaining conditions. Table 2 lists all pairwise multiple comparisons for the listening conditions.

Effects of the Number of Competing Talkers on Masking Release

In this context, masking release represents the benefit (in dB) between the different-sex and same-sex masker conditions. The amount of masking release was calculated for the one-talker (TS–TD) or two-talker masker conditions (TSS–TDD). Figure 4 shows the amount of masking release with the one- or two-talker maskers for NH and CI listeners. Note that positive values indicate masking release, and negative values indicate masker interference. Paired *t* tests (adjusted for multiple comparisons) showed significant masking release for NH listeners ($p < .001$) and CI listeners ($p < .005$) for both the one-talker and two-talker masker conditions. The amount of masking release was significantly larger for NH listeners than for CI users ($p < .001$) for both the one-talker and two-talker masker conditions.

Figure 3. SRTs as a Function of Target-Masker Combinations in NH Listeners (Black) and CI Users (White). The error bars show the standard deviation.
NH = normal-hearing; CI = cochlear implant; TD = one female talker; TS = one male talker; TDD = two female talkers; TSD = one male and one female talker; TSS = two male talkers; TSSDD = two male and two female talkers.

Table 2. All Pairwise Multiple Comparisons (Holm–Sidak Method) for the Different Masker Conditions.

	TD	TS	TDD	TSD	TSS	TSSDD
NH: $F(5, 45) = 77.95$, $p < .001$						
TD		**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
TS			.493	**<.001**	**<.001**	**<.001**
TDD				**<.001**	**<.001**	**<.001**
TSD					**.014**	.150
TSS						.484
CI: $F(5, 55) = 17.34$, $p < .001$						
TD		**<.001**	**.003**	.612	.603	**.023**
TS			**<.001**	**<.001**	**<.001**	**<.001**
TDD				.062	.084	.728
TSD					.869	.248
TSS						.291

Note. NH = normal-hearing; CI = cochlear implant; TD = one female talker; TS = one male talker; TDD = two female talkers; TSD = one male and one female talker; TSS = two male talkers; TSSDD = two male and two female talkers. When P < 0.05, the value is bold
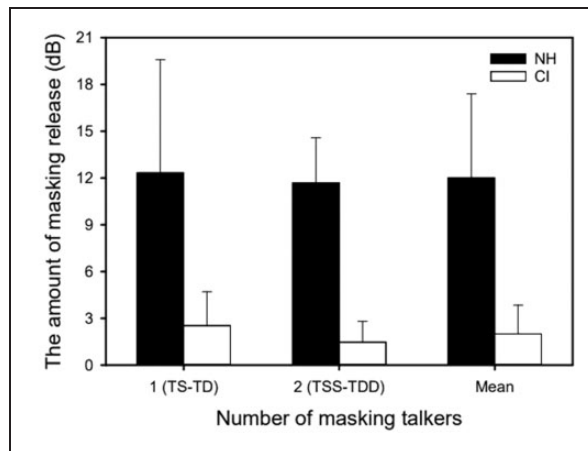
Figure 4. The Amount of Masking Release With One- or Two-Talker Maskers for NH Listeners (Black) and CI Users (White). The error bars show the standard deviation.
NH = normal-hearing; CI = cochlear implant; TS = one male talker; TD = one female talker; TSS = two male talkers; TDD = two female talkers.
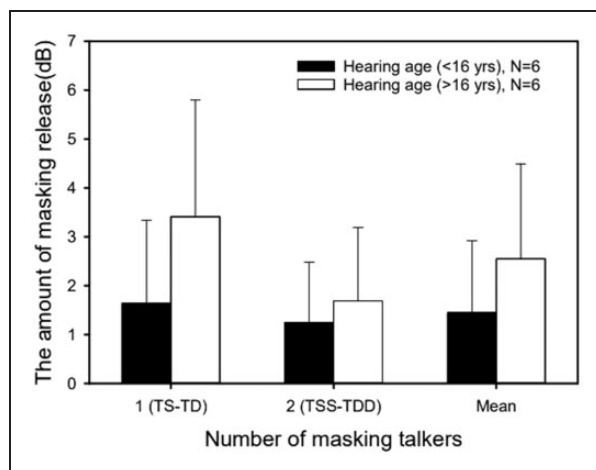

Figure 5. The Amount of Masking Release With One- or Two-Talker Maskers for CI Users With More Than 16 Years of Hearing Age (Black) or Less Than 16 Years of Hearing Age (White). Hearing age was estimated by the difference between age at testing and duration of auditory deprivation. The error bars show the standard deviation.

Effects of Hearing Age on Masking Release

In the present study, all CI users were at least 18 years old at testing. However, duration of auditory deprivation differed substantially among these CI users (range = 0.8–29.7 years). Hearing age is commonly used to indicate the period during which the listeners received auditory input with acoustic or electric hearing. For postlingually deafened CI users, hearing age can be estimated as the difference between age at testing and duration of auditory deprivation. Figure 5 shows the

amount of masking release with the one- or two-talker maskers for CI users with less than or more than 16 years of hearing age. A hearing age of 16 years was used as the threshold to divide the CI group, as previous studies have shown that SRTs reach adult-like levels at 16 years (Corbin et al., 2016), and the amount of masking release asymptotes at 16 or 17 years in NH listeners (Brown et al., 2010). The amount of masking release with the one-talker masker was markedly larger for CI users with more than 16 years of hearing age (3.61 dB vs. 1.64 dB). The amount of masking release with the two-talker maskers was slightly larger for CI users with more than 16 years of hearing age (1.69 dB vs. 1.25 dB). However, the difference in masking release was not significantly different between the two hearing age groups for the one-talker ($p = .17$) or two-talker masker conditions ($p = .58$).

Correlational Analyses

Demographic variables (age at testing, hearing age, duration of auditory deprivation, and duration of CI use) were compared with mean SRTs (averaged across all listening conditions) and the amount of masking release with the one- or two-talker maskers using Pearson correlations; Bonferroni correction for multiple comparisons was applied (adjusted $p = .004$). Moderately strong relationships were observed between mean SRTs and age at testing ($r = .628$, $p = .029$) and between mean SRTs and duration of auditory deprivation ($r = .559$, $p = .059$). However, the correlations were not significant after Bonferroni correction. No significant relationship was observed between the amount of masking release and any of the remaining demographic variables ($p > .1$).

Discussion

Consistent with many previous studies (Cullington & Zeng, 2008; El Boghdady et al., 2019; Fu & Nogaki, 2005; Iyer et al., 2010; Meister et al., 2020; Nelson et al., 2003; Stickney et al., 2004; Tao et al., 2018), CI users were much more susceptible to speech-on-speech maskers than were NH listeners, regardless of the number of competing talkers. While CI users performed significantly poorer in the presence of competing talkers, some interesting similarities and differences were observed between NH and CI listeners in terms of target-masker voice pitch effects.

Effect of Number of Competing Talkers on SRTs

For Mandarin-speaking NH listeners, mean SRTs worsened as the number of competing talkers increased. Such trends are generally consistent with the NH data reported in the literature (e.g., Freyman et al., 2007; Iyer et al.,

2010). The data can be well explained by EM and dip listening for NH listeners, who are likely able to take advantage of favorable SNRs in the spectrotemporal gaps to obtain masking release with one competing talker. As such, there was less EM with a one-talker masker than with SSN. Indeed, mean SRTs with the one-talker masker were as low as –22.0 dB, much lower than SRTs with SSN (–11.4 dB) reported by Tao et al. (2018) using the same test materials and protocols. When the number of competing talkers increased, due to the misaligned onsets and/or offsets for different competing talkers, the dips in the spectral and temporal envelopes were likely reduced, thereby smoothing the temporal envelopes in the multitalker masker. This likely resulted in fewer glimpses of the target speech (Figure 1). As noted previously, as the number of competing talkers increases, NH listeners may still be able to hear words in the competing speech but cannot decide whether the words were produced by target or the masker (i.e., increased IM). Indeed, mean SRTs were reduced to –9.7 dB with the two-talker masker and further reduced to –5.2 dB with the four-talker masker. Mean SRTs in the two- or four-talker masker conditions were generally worse than those with SSN in Tao et al. (2018). The present pattern of results is consistent with those in previous studies (e.g., Brungart et al., 2001; Cullington & Zeng, 2008), even though these studies differed in terms of testing materials, protocols, and language. Taken together, the results suggest that extracting information from target speech becomes more difficult when there is more than one interfering talker (the multimasker penalty described by Durlach, 2006).

The pattern of results were different for the adult Mandarin-speaking CI users, who performed worst with the one-talker masker (+5.9 dB) and best with the four-talker masker (+2.8 dB). Such trends were consistent with our hypothesis but were in contrast to previous findings that showed no effect of the number of competing talkers in English-speaking CI users (Cullington & Zeng, 2008). With a one-talker masker, the mean SRT of +6.2 dB reported by Cullington and Zeng (2008) was comparable with the mean SRT of +5.9 dB in the present study. Previous studies have shown that mean SRTs with a one-talker masker were significantly poorer than those with SSN in English-speaking CI users (+6.2 dB with one-talker masker vs. +2.5 dB with SSN in Cullington and Zeng, 2008) and Mandarin-speaking CI users (+3.5 dB with one-talker masker vs. –3.6 dB with SSN in Tao et al., 2018). Given the lack of fine spectrotemporal information and other signal processing components (e.g., compressive amplitude mapping, automatic gain control, channel interaction, etc.), CI users may have greater difficulty segregating competing talkers using target and masker vocal characteristics (e.g., F0 and/or VTL, as in Fuller et al., 2014). As

such, CI users may be unable to take advantage of temporal glimpsing. CI users' inability to use temporal dips has also been reported for nonspeech dynamic maskers (e.g., Fu & Nogaki, 2005).

Relative to a one-talker masker, Cullington and Zeng (2008) found that mean SRTs for English-speaking CI users worsened by 2 dB. While they found similar trends in NH and CI listeners, there was no significant effect of the number of competing talkers in CI users. As mentioned earlier, Cullington and Zeng (2008) found that CI users performed more poorly with the one-talker masker than with SSN, suggesting the presence of IM with a single interfering talker. Increasing the number of competing talkers may not increase the amount of IM and may have little effect on SRTs in English-speaking CI users. Mandarin-speaking CI users are also unlikely able to use temporal glimpsing to recognize target speech in the presence of competing talkers (e.g., the poorer SRTs with the one-talker masker than with SSN in Tao et al., 2018). However, different from English-speaking CI users, mean SRTs for the present Mandarin-speaking CI users significantly improved by 3.1 dB as the number of competing talkers increased from 1 to 4. The flattened multitalker masker temporal envelope may be less disruptive to the amplitude contour of the target speech, which is an important cue for lexical tone recognition in Mandarin-speaking CI users (Fu & Zeng, 2005).

Interaction of Vocal Characteristics in the Multiple Competing Talkers

The different combinations of the target and masker vocal characteristics also revealed some interesting observations. For Mandarin-speaking NH listeners, the best performance was observed for the TD condition (–28.2 dB), in which the male target speech was masked by one female talker; the poorest performance was observed for the TSS condition (–3.1 dB), in which the male target speech was masked by two different male maskers. These data were consistent with previous findings in English-speaking NH listeners (Cullington & Zeng, 2008).

Iyer et al. (2010) examined the effects of vocal characteristics of competing talkers on the multimasker penalty (Durlach, 2006). They found that the multimasker penalty was greatest when one of the maskers contained contextually relevant information relative to the target; under such circumstances, adding maskers with no contextual relevance has nearly no effect. Similarly, Calandruccio et al. (2017) found that effects of a two-talker masker were largely driven by the masker that was most similar to the target. For the present two-talker maskers, mean SRTs were –14.81, –7.71, and –3.12 dB for the TDD, TSD, and TSS conditions, respectively.

This supports the notion that the two-talker masker effects were more driven by the masker that was most perceptually similar to the target in terms of vocal characteristics, consistent with the data from Calandruccio et al. (2017). However, the vocal characteristics of the second masker also had a small but significant effect on SRTs, as a significant difference was observed between the TSD and TSS conditions. Similar findings were also observed by Cullington and Zeng (2008). Note that it is difficult to ascertain the contextual similarity in the present CRM-like task, as it was a closed-set procedure with equally plausible target and masker words.

Several interesting discrepancies in terms of the interaction of vocal characteristics were observed between CI users and NH listeners and between Mandarin-speaking and English-speaking CI users. First, there was no multi-masker penalty in CI users. For English-speaking CI users, the number of competing talkers had little effect on SRTs, while for Mandarin-speaking CI users, mean SRTs actually improved with increasing number of competing talkers. One of the conditions for multimasker penalty in Iyer et al. (2010) is that SNR should be less than 0 dB. For CI users, SRTs were typically greater than 0 dB with competing talkers. With the two-talker maskers, the lowest SRT was observed for the TDD condition (+2.4 dB), with no significant difference between the TSD (+4.0 dB) and TSS conditions (+3.9 dB). The relatively high SRTs observed with the present Mandarin-speaking CI users may have precluded a mulitmasker penalty, according to the 0 dB SNR threshold put forth by Iyer et al. (2010).

Note that mean SRTs for the one-talker masker were averaged across the TS and TD conditions, and the mean SRTs for the two-talker masker were averaged across the TDD, TSS, and TSD conditions. As such, the effects of masker vocal characteristics may have not been fully considered when analyzing the mulit-masker penalty. When the TD and TDD conditions were excluded, mean SRTs were significantly improved from +7.1 dB with the one-talker masker (TS) to +3.9 dB with the two-talker masker (averaged across TSD and TSS) to +2.8 dB with the four-talker masker (TSSDD; $p < .05$ for all comparisons).

Effects of the Number of Competing Talkers and Hearing Age on Masking Release

Previous studies have shown that speech performance is most difficult when the target and interfering maskers are colocated, intelligible, and similar in terms of vocal characteristics. Any additional cues, such as spatial separation (spatial cues; Freyman et al., 1999, 2001; Kidd et al., 2016), degree of masker intelligibility (e.g., reversed speech maskers in Kidd et al., 2016), and differences in vocal characteristics (e.g., different-sex maskers

in Kidd et al., 2016), could be used to better segregate the target from maskers, resulting in masking release. Kidd et al. (2016) found that the large masking release with these cues was primarily due to a reduction in IM. In the present study, the amount of masking release was estimated between the same-sex masker and different-sex masker with the one-talker masker (TS–TD) or the two-talker masker (TSS–TDD). As shown in Figure 4, the amount of masking release was comparable between the one-talker (12.3 dB) and two-talker maskers (11.7 dB) in NH listeners. Large amounts of masking release were reported by Cullington and Zeng (2008) for one-talker (10.2 dB) and two-talker maskers (8.4 dB). Brown et al. (2010) also reported a similar amount of masking release (9.0 dB) with a two-talker masker using the listening in spatialized noise—sentence test when voice pitch cues were available.

Due to the loss of fine structure, CI users have much greater difficulty using voice pitch cues to segregate and competing speech. In the present study, mean SRTs improved from +7.1 dB in the TS condition to +4.6 dB in the TD condition, resulting in 2.5 dB of masking release due to voice pitch cues when there was one competing talker. This amount of masking release is consistent with previous studies (e.g., Cullington & Zeng, 2008; El Boghdady et al., 2019; Liu et al., 2019; Meister et al., 2020; Visram & McKay, 2012). For example, Visram and McKay (2012) reported an improvement of 1.3 dB in English-speaking CI users with the CI alone and 1.6 dB with bimodal listening (CI combined with contra-lateral hearing aid, or CI+HA) when voice pitch cues were available. Liu et al. (2019) also reported a similar amount of masking release (1.7 dB) in Mandarin-speaking CI children with bimodal listening. Cullington and Zeng (2008) found that mean SRTs improved from +7.6 dB in the same-sex masker condition to +2.1 dB in the different-sex masker condition, resulting in a relatively large improvement (5.5 dB) with one competing talker.

However, other studies have shown no significant masking release in CI users due to voice pitch cues (Liu et al., 2019; Stickney et al., 2004, 2007; Tao et al., 2018). There are several differences among the present and previous studies, including language (tonal vs. non-tonal), age at testing (children vs. adults), listening mode (CI-only vs. CI+HA), and test materials (CRM vs. sentences). Stickney et al. (2004, 2007) measured percent correct at fixed SNRs using Institute of Electrical and Electronics Engineers (IEEE) sentences (Rothauser et al., 1969), while other studies adaptively measured SRTs (Cullington & Zeng, 2008; Visram & McKay, 2012). Due to the difficulty of IEEE sentences, the overall performance was relatively low for all conditions, resulting in a potential floor effect. The present data also contrast previous studies that showed no masking

release in Mandarin-speaking CI children (Liu et al., 2019; Tao et al., 2018). The only difference between the present and these previous studies was age at testing, as all three studies used exactly same test materials and protocols in Mandarin-spearing CI users. These results suggest that age at testing may play an important role in masking release.

For NH listeners, age at testing represents the time period that listeners have been receiving the auditory input. However, for postlingually deafened CI adults, age at testing may be different from the time period that listeners have been receiving the auditory input due to the duration of auditory deprivation. In the present study, hearing age was defined as the difference between age at testing and duration of auditory deprivation. Brown et al. (2010) found that the amount of masking release increased from approximately 3 dB in 6-year-old children to approximately 9 dB in adults when there was two competing talkers. Because the amount of masking release typically asymptotes at 16 or 17 years in NH listeners (Brown et al., 2010), a 16-year threshold of hearing age was used to divide the CI group in the present study. The present data showed that CI users with >16 years hearing age had substantially larger masking release (3.4 dB) than CI users with <16 years hearing age (1.6 dB); however, this difference did not achieve statistical significance due to the small number of subjects. Aging effects on speech measures have also been reported in middle-aged and elderly adults (Başkent et al., 2014; Bergman et al., 1976; Gordon-Salant & Fitzgibbons, 1999; Tun et al., 2002). For example, Başkent et al. (2014) found that the difference in SRTs between young and middle-aged adults was 2.1 dB for competing speech and 0.8 dB for steady noise. Tun et al. (2002) suggested that attention may be an important factor in older adults' difficulty in segregating target and competing speech. These data suggest that aging effects are already evident in children and middle-aged or elderly adults when compared with young adults, especially when testing with competing speech.

In the present study, as the number of competing talker increased, the amount of masking release reduced and was more variable in CI users; this was not consistent with the NH data, which showed little difference between the one- and two-talker masker conditions. Mean CI SRTs improved from +3.9 dB in the TSS condition to +2.4 dB in the TDD condition, resulting in 1.5 dB of masking release with voice pitch cues. However, Cullington and Zeng (2008) found that mean SRTs worsened from 7.4 dB in the TSS condition to 8.7 dB in the TDD condition, resulting in a 1.3 dB deficit when voice pitch cues were available. Again, the different test materials and language across studies may have contributed to differences in outcomes.

Summary and Conclusions

Understanding of target Mandarin Chinese speech was measured in adult Mandarin-speaking NH listeners and CI users in the presence of one or multiple competing talkers using a modified CRM task. The number of competing talkers increased from one to four, and the competing talkers contained different combinations of vocal characteristics. Major findings include the following:

1. For NH listeners, mean SRTs worsened from –22.0 dB to –5.2 dB as the number of competing talkers increased. The flattened peaks and valleys with increasing numbers of competing talkers may limit NH listeners' ability to glimpse the target speech in the dips of the spectral and temporal envelopes.
2. For Mandarin-speaking CI users, mean SRTs slightly but significantly improved from 5.9 dB to 2.8 dB for CI users as the number of competing talkers increased. This finding is in contrast to some previous studies that showed no effect of increasing number of competing talkers on SRTs in English-speaking CI users. The flattened amplitude contour of the multi-talker masker may be less disruptive to the amplitude contour of the target, which is an important cue for Mandarin-speaking CI users.
3. Adult CI users significantly benefitted from voice pitch differences between target and masker speech, but the amount of masking release was much smaller (2 dB) for CI users than for NH listeners (12 dB). The present data are in contrast to some previous studies that did not show a benefit for target-masker voice pitch differences in CI children. This suggests that age at testing (or hearing age) may be important to consider when evaluating the benefits of voice pitch cues for CI users.

ORCID iDs

Biao Chen   https://orcid.org/0000-0002-2436-4621
Yongxin Li   https://orcid.org/0000-0001-6267-5730
Qian-Jie Fu   https://orcid.org/0000-0003-3494-7633

References

Anzalone, M. C., Calandruccio, L., Doherty, K. A., & Carney, L. H. (2006). Determination of the potential benefit of time-frequency gain manipulation. *Ear and Hearing, 27*, 480–492. https://doi.org/10.1097/01.aud.0000233891.86809.df

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America, 139*, EL51–EL56. https://doi.org/10.1121/1.4942628

Başkent, D., van Engelshoven, S., & Galvin, J. J., III. (2014). Susceptibility to interference by music and speech maskers in middle-aged adults. *The Journal of the Acoustical Society of America, 135*, EL147–EL153. https://doi.org/10.1121/1.4865261

Bergman, M., Blumenfeld, V. G., Cascardo, D., Dash, B., Levitt, H., & Margulies, M. K. (1976). Age-related decrement in hearing for speech: Sampling and longitudinal studies. *Journal of Gerontology, 31*, 533–538. https://doi.org/10.1093/geronj/31.5.533

Bregman, A. S. (1994). *Auditory scene analysis*. MIT Press.

Brokx, J., & Nooteboom, S. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10*, 23–36. https://doi.org/10.1016/s0095-4470(19)30909-x

Brown, D. K., Cameron, S., Martin, J. S., Watson, C., & Dillon, H. (2010). The North American listening in spatialized noise—sentences test (NA LiSN-S): Normative data and test-retest reliability studies for adolescents and young adults. *Journal of the American Academy of Audiology, 21*(10), 629–641. https://doi.org/10.3766/jaaa.21.10.3

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*, 1101–1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America, 120*(6), 4007–4018. https://doi.org/10.1121/1.2363929

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5 Pt 1), 2527–2538. https://doi.org/10.1121/1.1408946

Calandruccio, L., Bowdrie, K., & Buss, E. (2017). Effectiveness of two-talker maskers that differ in talker congruity and perceptual similarity to the target speech. *Trends in Hearing, 21*, 2331216517709385. https://doi.org/10.1177/2331216517709385

Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and Hearing, 37*(1), 55–63. https://doi.org/10.1097/AUD.0000000000000201

Cullington, H. E., & Zeng, F. G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *The Journal of the Acoustical Society of America, 123*(1), 450–461. https://doi.org/10.1121/1.2805617

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America, 114*, 2913–2922. https://doi.org/10.1121/1.1616924

Darwin, C. J., & Hukin, R. W. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *The Journal of the Acoustical Society of America, 107*(2), 970. https://doi.org/10.1121/1.428278

Drullman, R., & Bronkhorst, A. W. (2004). Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 116*, 3090–3098. https://doi.org/10.1121/1.1802535

Durlach, N. (2006). Auditory masking: Need for improved conceptual structure. *The Journal of the Acoustical Society of America, 120*, 1787–1790. https://doi.org/10.1121/1.2335426

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking. *The Journal of the Acoustical Society of America, 113*, 2984–2987. https://doi.org/10.1121/1.1570435

El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *The Journal of the Acoustical Society of America, 145*(1), 417. https://doi.org/10.1121/1.5087693

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America, 109*(5 Pt 1), 2112–2122. https://doi.org/10.1121/1.1354984

Freyman, R. L., Helfer, K. S., & Balakrishnan, U. (2007). Variability and uncertainty in masking by competing speech. *The Journal of the Acoustical Society of America, 121*(2), 1040–1046. https://doi.org/10.1121/1.2427117

Freyman, R. L., Helfer, K. S., Mccall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America, 106*(6), 3578. https://doi.org/10.1121/1.428211

Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology, 6*(1), 19–27. https://doi.org/10.1007/s10162-004-5024-3

Fu, Q.-J., & Zeng, F. G. (2005). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech Language and Hearing, 5*, 45–57. https://doi.org/10.1179/136132800807547582

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q. J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology, 15*(6), 1037–1048. https://doi.org/10.1007/s10162-014-0483-7

Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing, 39*(2), 226–237. https://doi.org/10.1097/AUD.0000000000000480

Gordon-Salant, S., & Fitzgibbons, P. J. (1999). Profile of audi-
tory temporal processing in older listeners. *Journal of Speech, Language, and Hearing Research*, 42, 300–311. https://doi.org/10.1044/jslhr.4202.300

Iyer, N., Brungart, D. S., & Simpson, B. D. (2010). Effects of target-masker contextual similarity on the multimasker pen-alty in a three-talker diotic listening task. *The Journal of the Acoustical Society of America*, 128(5), 2998–3010. https://doi.org/10.1121/1.3479547

Kidd, G., Mason, C. R., Swaminathan, J., Roverud, E., Clayton, K. K., & Best, V. (2016). Determining the energet-ic and informational components of speech-on-speech masking. *The Journal of the Acoustical Society of America*, 140, 132–144. https://doi.org/10.1121/1.4954748

Leibold, L. J., Buss, E., & Calandruccio, L. (2018). Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch. *Ear and Hearing*, 39(5), 935–945. https://doi.org/10.1097/AUD.0000000000000554

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

Liu, Y. W., Tao, D. D., Chen, B., Cheng, X., Shu, Y., Galvin, J. J., 3rd, & Fu, Q. J. (2019). Factors affecting bimodal benefit in pediatric Mandarin-speaking Chinese cochlear implant users. *Ear and Hearing*, 40(6), 1316–1327. https://doi.org/10.1097/AUD.0000000000000712

Luo, X., & Fu, Q.-J. (2009). Concurrent-vowel and tone rec-ognition in acoustic and simulated electric hearing. *The Journal of the Acoustical Society of America*, 125(5), 3223–3233. https://doi.org/10.1121/1.3106534

Luo, X., Fu, Q.-J., Wu, H. P., & Hsu, C. J. (2009). Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hearing Research*, 256, 75–84. https://doi.org/10.1016/j.heares.2009.07.001

Meister, H., Walger, M., Lang-Roth, R., & Müller, V. (2020). Voice fundamental frequency differences and speech recog-nition with noise and speech maskers in cochlear implant recipients. *The Journal of the Acoustical Society of America*, 147(1), EL19. https://doi.org/10.1121/10.0000499

Nelson, P. B., Jin, S. H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 113(2), 961. https://doi.org/10.1121/1.1531983

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., & Weinstock, M. (1969). I.E.E.E. recommended practice for speech quality measurements. *IEEE Transactions in Audio Electroacoustics*, 17, 227–246. https://doi.org/10.1109/tau.1969.1162058

Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F. G. (2007). Effects of cochlear implant processing and funda-mental frequency on the intelligibility of competing senten-ces. *The Journal of the Acoustical Society of America*, 122, 1069–1078. https://doi.org/10.1121/1.2750159

Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, 116(2), 1081. https://doi.org/10.1121/1.1772399

Tao, D. D., Fu, Q. J., Galvin, J. J., & Yu, Y. F. (2017). The development and validation of the closed-set mandarin sen-tence (CMS) test. *Speech Communication*, 92, 125–131. https://doi.org/10.1016/j.specom.2017.06.008

Tao, D. D., Liu, Y. W., Yei, F., Galvin, J. J., Chen, B., & Fu, Q. J. (2018). Effects of age and duration of deafness on Mandarin speech understanding in competing speech by normal-hearing and cochlear implant children. *The Journal of the Acoustical Society of America*, 144, EL131–EL137. https://doi.org/10.1121/1.5051051

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, 17, 453–467. https://doi.org/10.1037//0882-7974.17.3.453

Vestergaard, M. D., Fyson, N. R., & Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *The Journal of the Acoustical Society of America*, 125, 1114–1124. https://doi.org/10.1121/1.3050321

Visram, A. K. K., & Mckay, C. M. (2012). Voice gender differ-ences and separation of simultaneous talkers in cochlear implant users with residual hearing. *The Journal of the Acoustical Society of America*, 132(2), 135–141. https://doi.org/10.1121/1.4737137