The effects of binaural spectral resolution mismatch on Mandarin speech perception in simulated electric hearing

Fei Chen[a)] and Lena L. N. Wong
Division of Speech and Hearing Sciences, The University of Hong Kong, Prince Philip Dental Hospital, 34 Hospital Road, Hong Kong
feichen1@hku.hk, llnwong@hku.hk

Qudsia Tahmina, Behnam Azimi, and Yi Hu[a)]
Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, 3200 N. Cramer Street, Milwaukee, Wisconsin 53211
qtahmina@uwm.edu, bazimi@uwm.edu, huy@uwm.edu

Abstract: This study assessed the effects of binaural spectral resolution mismatch on the intelligibility of Mandarin speech in noise using bilateral cochlear implant simulations. Noise-vocoded Mandarin speech, corrupted by speech-shaped noise at 0 and 5 dB signal-to-noise ratios, were presented unilaterally or bilaterally to normal-hearing listeners with mismatched spectral resolution between ears. Significant binaural benefits for Mandarin speech recognition were observed only with matched spectral resolution between ears. In addition, the performance of tone identification was more robust to noise than that of sentence recognition, suggesting factors other than tone identification might account more for the degraded sentence recognition in noise.

1. Introduction

The full benefits of bilateral cochlear implant (CI) stimulation may significantly depend on unilateral CI performance, which involves factors such as the hearing loss duration of each ear, the processing ability of the CI user with the device, and the listening environment (Litovsky *et al.*, 2006). An earlier study conducted by Litovsky *et al.* (2006) reported that bilateral CI users who had a large difference in unilateral CI performance showed no significant binaural benefits, and, in contrast, those who had similar unilateral performance showed significant binaural benefits. In a more recent study, Yoon *et al.* (2011a) evaluated bilateral and unilateral speech recognition performance for English phonemes and sentences presented in quiet and in noise, and found that the binaural benefit was greater when unilateral performance was similar across both ears. As the difference in unilateral performance between ears increased, the binaural advantage decreased. This functional relationship was observed across different speech materials at various signal-to-noise ratio (SNR) levels, even though there was substantial intra- and inter-subject variability.

Although results from these studies indicate that subjects who have very different speech recognition performance between implanted ears in general show poor binaural benefit, there are only a few studies that focus on the likely sources of the performance mismatch between ears. For instance, Yoon *et al.* (2011b) showed that the CI insertion-depth difference between ears might contribute to mismatched unilateral performance, thereby decreasing binaural benefits. Garadat *et al.* (2010) studied the

---

[a)]Authors to whom correspondence should be addressed.

effects of "dead regions" for the performance difference between two implanted ears and suggested that customized programming for bilateral CI processors based on knowledge about dead regions could enhance performance in adverse listening situations. In our current study, we used simulated electric hearing to investigate the impact of mismatched spectral resolution between implanted ears on the binaural benefits for recognition of Mandarin tones and sentences. The rationale for the present work is three-fold. First, it is common for bilateral CI users to have mismatched spectral resolution between implanted ears, which might be a major source for mismatched unilateral performance. Second, it is not clear how mismatched spectral resolution affects binaural benefits for perception of Mandarin speech in noise. Third, in the context of binaural benefits in noise, the relationship between performance of tone recognition and that of sentence recognition has not been examined.

Many studies have shown binaural benefits in English-speaking CI listeners, and it is natural to foresee that binaural benefits could also be obtained from CI patients speaking tonal languages such as Mandarin Chinese. Nevertheless, considering the difference in linguistic cues used by native speakers of non-tonal languages vs tonal languages, present CI coding strategies do not result in satisfactory tone identification in unilateral Mandarin CI users (Wei *et al.*, 2004). Therefore, it is worthwhile to assess and verify whether binaural stimulation could improve tone identification, and whether the performance improvement in Mandarin tone identification (if any) could account for the performance improvement of Mandarin sentence understanding in noise.

Vocoded speech has long been used in cochlear implant simulations (e.g., Shannon *et al.*, 1995; Zhou and Xu, 2008; Yoon *et al.*, 2011b) for studying how various factors could influence speech recognition for unilateral and bilateral CIs, without a host of confounding factors (e.g., insertion-depth, neural surviving pattern). Therefore, in this study we assessed the effects of spectral resolution mismatch by using vocoded speech simulating unilateral and bilateral CIs. The binaural listening experiments were designed in the present study to answer the following questions: (1) How would spectral resolution mismatch affect the intelligibility of Mandarin speech in binaural electric hearing? (2) Does spectral resolution mismatch have the same impact on perception of Mandarin sentence as that of Mandarin tone? (3) Since binaural hearing in cochlear implants substantially improves sentence perception in noise, could this improvement be attributed to performance improvement in Mandarin tone identification? Based on the results from previous studies on the perception of English speech materials (e.g., Litovsky *et al.*, 2006; Yoon *et al.*, 2011a), we hypothesize that matched spectral resolution across the two implanted ears will yield more binaural benefits in Mandarin speech perception. As to the binaural Mandarin tone recognition and sentence perception, we hypothesize that binaural spectral resolution mismatch may have less effect on tone recognition than on sentence recognition. Our rationale is that, in simulating the unilateral CI performance for Mandarin speech perception in quiet, Fu *et al.* (1998) have shown that vowel, consonant, and tone recognition contributed equally to Chinese sentence recognition, and they also suggested that lexical tone recognition strongly relied on temporal envelope cues and was more robust to degraded spectral resolutions than vowel and consonant recognition.

2. Methods

2.1 Subjects and materials

Ten (four male and six female) normal-hearing native Mandarin-Chinese listeners participated in the experiment. The subjects' ages ranged from 23 to 35 years, and the majority of subjects were graduate students at The University of Hong Kong. All subjects were paid for their participation.

The sentence material consisted of sentences taken from the Mandarin Hearing in Noise Test database (Wong *et al.*, 2007). All the sentences were produced by a male speaker, with fundamental frequency ranging from 75 to 180 Hz. One adult male

and one adult female native Mandarin-Chinese speaker produced the following six single-vowel syllables in each of the four Mandarin tones (/a/, /o/, /e/, /i/, /u/, /ü/) in a sound-treated booth, resulting in a total of 48 vowel tokens (= two speakers × six vowels × four tones) for Mandarin tone identification. Both Mandarin sentences and vowels were recorded at a sampling rate of 16 kHz, and their waveforms were then adjusted to have the same root-mean-square (rms) values (Zhou and Xu, 2008). The duration of the vowel tokens was normalized (Fu and Zeng, 2000). A steady-state, speech-shaped noise was used to corrupt the target speech materials at 0 and 5 dB SNR levels before vocoding processing. The SNR levels were chosen to avoid the ceiling effect.

2.2 Signal processing

In implementing the eight-channel noise-vocoder, the corrupted Mandarin speech materials were first processed through a pre-emphasis high-pass filter (2000 Hz cutoff) with a 3 dB/octave roll-off, and then band-passed filtered into N=8 frequency bands between 80 and 6000 Hz using sixth-order Butterworth filters, where N can be viewed as the spectral resolution equivalently. The equivalent rectangular bandwidth scale (Glasberg and Moore, 1990) was used to allocate the eight channels with the specific bandwidth (see Table 1). The envelope of signal was extracted by full-wave rectification and low-pass filtering using a second-order Butterworth filter with a cutoff frequency of 160 Hz. The envelope of each frequency band was used to modulate white noise, followed by band-limiting using the same Butterworth band-pass filters. The envelope-modulated noises of each band were finally summed up and the level of the synthesized speech was adjusted to yield the same rms value as the original speech.

　　　　To simulate spectral-resolution mismatch across the two implanted ears, the right ear was presented with the eight-channel noise-vocoded stimuli, while the left ear was presented with the two-, four-, or eight-channel noise-vocoded stimuli. Our rationale to use two, four, and eight as the number of spectral channels is that the results from Friesen *et al.* (2001) showed that, despite the relatively larger number (16–22) of electrodes available, most CI users only receive a limited number (i.e., up to eight) of channels of spectral information, and those poor performers might receive spectral information of fewer than eight channels. To evaluate the effect of the number of spectral channels on speech intelligibility in CIs, it is common that many prior studies varied the channel number between 1 to a larger number such as six or eight in vocoder simulations (e.g., Shannon *et al.*, 1995). The cut-off frequencies used in the two- and four-channel noise vocoder are listed in Table 1. The three processing conditions described above are referred to as (1) R8_L2, (2) R8_L4, and (3) R8_L8, respectively, where R and L denote right and left ears, respectively. For comparison, we also implemented the condition R8_L0, which included only unilateral electric stimulation to the right ear.

2.3 Procedure

The experiment was performed in a sound-proof booth and stimuli were played to listeners through a Sennheiser HD 250 Linear II circumaural headphone at a comfortable listening level. Prior to the test, each subject participated in a 20 min training

Table 1. Filter cut-off (–3 dB) frequencies used for the noise-vocoder.

Band number, N	Filter cut-off frequency (Hz)
2	80, 1158, 6000
4	80, 426, 1158, 2710, 6000
8	80, 221, 426, 724, 1158, 1790, 2710, 4050, 6000

session and listened to the noise-vocoded speech with both matched and un-matched spectral resolution across the ears to become familiar with the vocoded stimuli and the testing procedure. Each subject participated in eight testing conditions ($=$ two SNR levels $\times$ four processing conditions) in Mandarin sentence and tone tests. The order of the eight testing conditions was randomized across subjects. For the Mandarin sentence test, a total of 20 Mandarin sentences were used per condition and none of the sentences were repeated across various testing conditions. The subjects were instructed to repeat the sentence as accurately as possible, and they were allowed to repeat the stimuli only once. For the Mandarin tone test, each condition consisted of two presentations of each vowel stimulus spoken by a male and a female talker, and each subject listened to 96 randomized vowel stimuli ($=$ two repetitions $\times$ two speakers $\times$ six vowels $\times$ four tones) per condition. The Mandarin tone responses were collected with custom software using a computer display of response alternatives and a mouse as a response key. The subjects were allowed to use a repeat key as many times as they wished to repeat the presentations of the test stimuli during the tone test. The percent correct score for each condition was calculated by dividing the number of words/tones correctly identified by the total number of words/vowel stimuli in a particular testing condition. Subjects were given a 5 min break every 30 min during the test.

3. Results and discussion

Figure 1(a) shows the mean scores of Mandarin sentence recognition for all conditions at different SNRs. Statistical significance was determined by using the percent correct score as the dependent variable, and SNR levels and processing conditions as the two within-subjects factors. Two-way analysis of variance (ANOVA) with repeated measures indicated a significant effect ($F[1, 9] = 159.7$, $p < 0.0005$) of SNR level, a significant effect ($F[3, 27] = 11.2$, $p < 0.0005$) of spectral resolution, and a non-significant interaction ($F[3, 27] = 2.15$, $p = 0.117$) between SNR level and spectral resolution. Significant improvement in intelligibility was observed when listeners had access to the binaural noise-vocoded stimuli with the same spectral resolution (i.e., condition R8_L8) across ears, compared with the binaural stimulation with un-matched spectral resolution (i.e., conditions R8_L4 and R8_L2). For instance, the improvement with the R8_L8 stimuli relative to the R8_L4 stimuli ranged from 11 percentage points at 5 dB SNR to 7 percentage points at 0 dB SNR. Furthermore, *post hoc* pairwise comparisons using the Sidak method revealed a significant binaural benefit only for the condition R8_L8 ($p < 0.05$), compared with the condition R8_L0, in which the noise-vocoded stimuli were presented unilaterally to the right ear. When the spectral resolutions are not matched between the ears (i.e., conditions R8_L4, and R8_L2), there is no performance improvement with binaural stimulation compared with condition R8_L0.
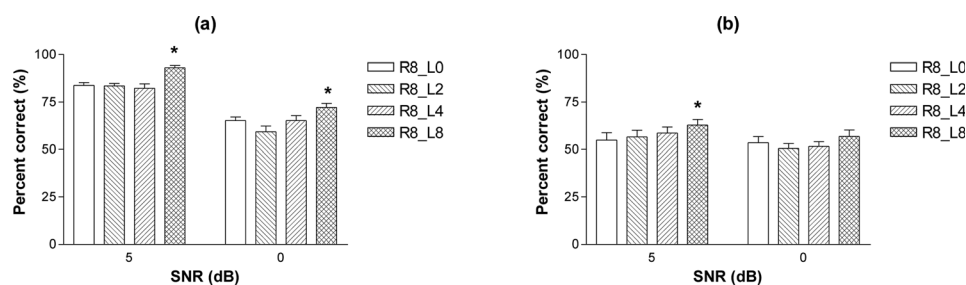

Fig. 1. Mean speech recognition scores (percent of Mandarin words/tones identified correctly) as a function of SNR level in recognizing (a) Mandarin sentences and (b) Mandarin tones. The error bars denote one standard error of mean (SEM). Asterisk indicates statistically significant ($p < 0.05$) difference between the binaural (i.e., R8_L8, R8_L4, or R8_L2) and unilateral (i.e., R8_L0) performances.

The results from the tone identification shown in Fig. 1(b) are similar to those observed from sentence recognition in Fig. 1(a). Two-way ANOVA with repeated measures indicated a significant effect ($F[1, 9]=13.6$, $p<0.005$) of SNR level, a significant effect ($F[3, 27]=9.5$, $p<0.0005$) of spectral resolution, and a non-significant interaction ($F[3, 27]=0.98$, $p=0.41$) between SNR level and spectral resolution. Significant improvement in intelligibility was observed when listeners had access to the binaural noise-vocoded stimuli with the same spectral resolution (i.e., condition R8_L8) across ears compared with binaural stimulation with un-matched spectral resolution (i.e., conditions R8_L4 and R8_L2) at 5 dB SNR. The improvement with the R8_L8 stimuli relative to the R8_L2 stimuli was 6 percentage points at 5 and 0 dB SNR levels. Significant performance improvement ($p<0.05$) is, however, observed only at 5 dB SNR relative to the unilateral condition R8_L0. Again, when the spectral resolutions are not matched across ears (i.e., conditions R8_L4, and R8_L2), there is no significant binaural benefit compared with unilateral hearing (i.e., condition R8_L0) at both 5 and 0 dB SNR.

The above analyses indicate that spectral-resolution mismatch has a significant impact on the intelligibility of binaurally presented noise-vocoded stimuli. In other words, spectral-resolution difference may be a likely source for unilateral CI performance mismatch that leads to decreased binaural benefits. To some extent, this finding is consistent with previous results reported in Litovsky *et al.* (2006) and Yoon *et al.* (2011a). It is worthwhile to note that such benefits from binaural summation might not be affected by the "mismatch" of spectral resolutions between ears *per se*, but rather depend on the spectral resolution of the added left ear. When there were only two or four channels available in the left ear, the better spectral resolution in the right ear dominated the overall performance, yielding no benefits from binaural summation.

Another finding of the present study is the relative performance robustness of tone identification in either unilateral or bilateral listening. When the SNR level was decreased from 5 dB to 0 dB, the decrement of sentence recognition scores ranged from 17 percentage points (i.e., condition R8_L4) to 24 percentage points (i.e., condition R8_L2), while that of the tone identification scores ranged from one percentage point (condition R8_L0) to 7 percentage points (condition R8_L4). In other words, the percent correct of Mandarin tone identification is relatively stable and not notably affected by the noise interference in the SNR range [0–5] dB.

The relative robustness of tone identification performance has also been reported in previous studies. For instance, Fu *et al.* (1998) found that when the spectral resolution in noise-vocoder simulation was altered from N=1 to 4, the identification of Mandarin tones was almost consistent, i.e., around 60% and 80% with envelope filter frequencies at 50 and 500 Hz, respectively. When Zhou and Xu (2008) simulated the effect of mismatched spectral distribution of envelopes on Mandarin tone recognition using a noise-vocoder, they found that spectral shift might not pose a severe problem for tone recognition when the carrier bands were shifted basally relative to the analysis bands by 1–7 mm in the cochlear. This suggests that the resistance of tone recognition to spectral shift can be attributed to the overall amplitude contour cues that are independent from spectral manipulations.

The findings in the current study also shed light on investigating the relationship between Mandarin tone identification and Mandarin sentence recognition in maskers (at least in steady-state speech-shaped noise) by simulated electric hearing. Although many studies have examined Mandarin tone identification in quiet (e.g., Fu *et al.*, 1998, Wei *et al.*, 2004; Zhou and Xu, 2008), only a limited number of studies have investigated how electric hearing performs in tone-identification task with noise being present (e.g., Kong and Zeng, 2006), and whether in noise, performance in tone identification can predict performance in sentence recognition. Several clinically oriented studies have found that some Mandarin-speaking CI users perform poorly in Mandarin tone identification even in quiet (e.g., Wei *et al.*, 2004), and a relevant question could be raised on whether this poor performance in tone identification shown in

these CI users would predict decreased performance of their perception of Mandarin sentence in noise, which clearly presents a more clinically relevant situation.

To this end, the present results show that, compared with the more pronounced response to noise level observed in Mandarin sentence recognition task, the performance of Mandarin tone identification is relatively robust and resistant to noise interference. The robustness of Mandarin tone recognition in noise might be partially due to two factors. First, acoustic cues on lexical tone identification (e.g., tonal envelope cue) mainly exist in vowel segments, and thus might be better recognized, as vowel segments usually have stronger energies and more favorable local SNRs than other segments (e.g., consonants). Second, there are only four choices for tones, and such a higher chance level may make tone recognition more robust in noise than consonant or vowel recognition. Note that further investigation in the future is needed to compare the Mandarin vowel and consonant recognition in noise with tone recognition in the experimental conditions.

The findings in this study indicate that factors other than tone identification account more for the degradation of Mandarin sentence understanding in noise. On the other hand, although the present study indicates that noise interference might have less of an impact on the performance of Mandarin tone identification, and the performance of Mandarin sentence recognition (e.g., when SNR was increased from 0 to 5 dB) might not result from correctly identifying more Mandarin tones, it does not mean tone identification does not play an important role in Mandarin sentence recognition. For instance, Fu *et al.* (1998) showed that in quiet, tones, vowels and consonants contributed equally to Chinese sentence recognition. More work still needs to be done to investigate the relationship between Mandarin sentence recognition and tone identification, particularly in noise.

4. Conclusions

The present study assessed the effects of spectral resolution mismatch for Mandarin speech (sentence and lexical tone) intelligibility in simulated binaural electric hearing. Consistent with previous findings on binaural benefit for English speech materials, the present results indicate that sentence perception performance of binaural CI listening in noise is affected by the difference of spectral resolution between ears, and binaural benefits are maximized with matched spectral resolution in the two implanted ears. In addition, the present study shows that tone identification in steady-state speech-shaped noise is more robust than sentence recognition, indicating that tone identification performance in noise might not predict sentence recognition performance in unilateral and bilateral CIs, and factors other than tone identification might significantly account for the performance of Mandarin sentence understanding in noise.

References and links

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q. J., and Zeng, F. G. (**2000**). "Identification of temporal envelope cues in Chinese tone recognition," Asia Pac. J. Speech, Language Hearing **5**, 45–57.

Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Garadat, S. N., Litovsky, R. Y., Yu, G. Q., and Zeng, F. G. (**2010**). "Effects of simulated spectral holes on speech intelligibility and spatial release from masking under binaural and monaural listening," J. Acoust. Soc. Am. **127**, 977–989.

Glasberg, B., and Moore, B. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Kong, Y. Y., and Zeng, F. G. (**2006**). "Temporal and spectral cues in Mandarin tone recognition," J. Acoust. Soc. Am. **120**, 2830–2840.

Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (**2006**). "Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study," Ear Hear. **27**, 714–731.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Wei, C. G., Cao, K. L., and Zeng, F. G. (**2004**). "Mandarin tone recognition in cochlear-implant subjects," Hear. Res. **197**, 87–95.

Wong, L. L., Soli, S. D., Liu, S., Han, N., and Huang, M. W. (**2007**). "Development of the Mandarin hearing in noise test (MHINT)," Ear Hear. **28**, 70S–74S.

Yoon, Y. S., Li, Y. X., Kang, H. Y., and Fu, Q. J. (**2011a**). "The relationship between binaural benefit and difference in unilateral speech recognition performance for bilateral cochlear implant users," Int. J. Audiol. **50**, 554–565.

Yoon, Y. S., Liu, A., and Fu, Q. J. (**2011b**). "Binaural benefit for speech recognition with spectral mismatch across ears in simulated electric hearing," J. Acoust. Soc. Am. **130**, EL94–EL100.

Zhou, N., and Xu, L. (**2008**). "Lexical tone recognition with spectrally mismatched envelopes," Hear. Res. **246**, 36–43.