Effects of spectral resolution and frequency mismatch on speech understanding and spatial release from masking in simulated bilateral cochlear implants

Kevin Xu, **Shelby Willis**, **Quinton Gopen**, **Qian-Jie Fu**

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

Abstract

Objectives—Due to inter-aural frequency mismatch, bilateral cochlear-implant (CI) users may be less able to take advantage of binaural cues that normal-hearing (NH) listeners use for spatial hearing, such as inter-aural time differences and inter-aural level differences. As such, bilateral CI users have difficulty segregating competing speech even when the target and competing talkers are spatially separated. The goal of this study was to evaluate the effects of spectral resolution, tonotopic mismatch (the frequency mismatch between the acoustic center frequency assigned to CI electrode within an implanted ear relative to the expected spiral ganglion characteristic frequency), and inter-aural mismatch (differences in the degree of tonotopic mismatch in each ear) on speech understanding and spatial release from masking (SRM) in the presence of competing talkers in NH subjects listening to bilateral vocoder simulations.

Design—During testing, both target and masker speech were presented in five-word sentences that had the same syntax but were not necessarily meaningful. The sentences were composed of five categories in fixed order (Name, Verb, Number, Color, and Clothes), each of which had 10 items, such that multiple sentences could be generated by randomly selecting a word from each category. Speech reception thresholds (SRTs) for the target sentence presented in competing speech maskers were measured. The target speech was delivered to both ears and the two speech maskers were delivered to: 1) both ears (diotic masker), or 2) different ears (dichotic masker: one delivered to the left ear and the other delivered to the right ear). Stimuli included the unprocessed speech and four 16-channel sine-vocoder simulations with different inter-aural mismatch (0, 1, and 2 mm). SRM was calculated as the difference between the diotic and dichotic listening conditions.

Results—With unprocessed speech, SRTs were 0.3 and −18.0 dB for the diotic and dichotic maskers, respectively. For the spectrally degraded speech with mild tonotopic mismatch and no inter-aural mismatch, SRTs were 5.6 and −2.0 dB for the diotic and dichotic maskers, respectively. When the tonotopic mismatch increased in both ears, SRTs worsened to 8.9 and 2.4 dB for the diotic and dichotic maskers, respectively. When the two ears had different tonotopic mismatch (e.g., there was inter-aural mismatch), the performance drop in SRTs was much larger for the dichotic than for the diotic masker. The largest SRM was observed with unprocessed speech (18.3 dB). With the CI simulations, SRM was significantly reduced to 7.6 dB even with mild tonotopic

Corresponding author: Qian-Jie Fu, Department of Head and Neck Surgery, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA, qfu@mednet.ucla.edu.

mismatch but no inter-aural mismatch; SRM was further reduced with increasing inter-aural mismatch.

Conclusions—The results demonstrate that frequency resolution, tonotopic mismatch, and inter-aural mismatch have differential effects on speech understanding and SRM in simulation of bilateral CIs. Minimizing inter-aural mismatch may be critical to optimize binaural benefits and improve CI performance for competing speech, a typical listening environment. SRM (the difference in SRTs between diotic and dichotic maskers) may be a useful clinical tool to assess inter-aural frequency mismatch in bilateral CI users and to evaluate the benefits of optimization methods that minimize inter-aural mismatch.

Keywords

Frequency resolution; tonotopic mismatch; inter-aural mismatch; spatial release from masking; competing speech; speech recognition threshold

1. Introduction

Due to interactions among the acoustic-to-electric frequency allocation, the limited extent/ insertion of the electrode array, and patterns of nerve survival, cochlear-implant (CI) patients may experience tonotopic mismatch, which is defined as the frequency mismatch between the acoustic center frequency assigned to CI electrode within an implanted ear relative to the expected spiral ganglion characteristic frequency based on the Greenwood's (1990) frequency-place formula as the insertion angles for the most apical and most basal electrode vary significantly among CI users (Landsberger et al., 2015). Non-uniform nerve survival may increase tonotopic mismatch, as information mapped onto electrodes in "dead" regions is processed by spatially remote healthy neurons (Shannon et al., 2002). Tonotopic mismatch has been shown to limit speech performance in CI users (Fu et al., 2002; Baskent and Shannon, 2004) and in normal-hearing (NH) subjects listening to vocoder simulations (Baskent and Shannon, 2003, 2007).

Bilateral CI, bimodal CI, and CI users with single-sided deafness (SSD) may also experience "inter-aural mismatch" due to differences in the degree of tonotopic mismatch in each ear. Inter-aural mismatch may significantly limit binaural performance and benefits for bilateral CI users (Long et al., 2003; Kan et al., 2013, 2015; Svirsky et al., 2015) and in NH subjects listening to bilateral CI and SSD vocoder simulations (Yoon et al., 2011ab, 2013; Zhou et al., 2016). Yoon et al. (2011b) found maximum binaural-summation benefit for speech in quiet and in noise when the inter-aural mismatch was 1 mm. Inter-aural mismatch has also been shown to limit perception of inter-aural time and level differences (ITDs and ILDs; Blanks et al., 2008; Oxenham et al., 2014; Kan et al., 2015; Bernstein et al., 2018; Francart et al., 2018), which in turn can limit sound localization (Suneel et al., 2017) and speech understanding when speech and noise are spatially separated (Yoon et al., 2013; Ma et al., 2016; Wess et al. 2017; Goupell et al., 2018; Zhou et al., 2018).

Recently, Goupell et al. (2018) measured speech understanding and spatial release from masking (SRM) for NH subjects listening to bilateral vocoder simulations. In their study, female target speech was presented from the front location while two male competing talkers

were presented from the front location (co-located) or from the right (spatially separated). Stimuli were then processed by an eight-channel tone vocoder that simulated different degrees of frequency mismatch within and across ears in bilateral CIs. SRM was calculated as the performance difference between the co-located and spatially separated conditions. As the degree of tonotopic mismatch was increased in both ears (i.e., no inter-aural mismatch), performance similarly declined for the co-located and spatially separated conditions; as such, SRM was not affected by tonotopic mismatch. As the degree of inter-aural mismatch was increased, performance declined for the co-located and spatially separated conditions; however, SRM also decreased with increasing inter-aural mismatch. These data suggest that while overall performance is likely affected by frequency mismatch within and across ears, SRM is primarily determined by inter-aural mismatch.

Many previous studies have shown that speech-recognition performance is strongly affected by tonotopic mismatch in CI users (Fu et al., 2002; Baskent and Shannon, 2004) and in NH subjects listening to vocoder simulations (Baskent and Shannon, 2003, 2007). Different degrees of tonotopic mismatch in each ear might result in different speech-recognition performance, resulting in performance asymmetry across ears. The degree of performance asymmetry may limit binaural benefits (e.g., binaural summation). Yoon et al. (2011a) found that binaural benefits in bilateral CI users were reduced as the degree of performance asymmetry increased. In a follow-up study using CI simulations in NH listeners, Yoon et al. (2011b) found the maximum binaural-summation benefit when the inter-aural mismatch was 1 mm. Besides binaural summation, performance asymmetry may also affect SRM. Zhou et al. (2018) evaluated the effects of simulated insertion depth on spatial speech perception in noise for SSD vocoder simulations. Speech recognition in noise was measured for three spatial conditions: (1) speech and noise from the front (S0N0), (2) speech from the front and noise from the right (simulated CI ear) (S0Nci), and (3) speech from the front and noise from the left (S0Nnh). SRM with binaural hearing was calculated as the difference in speech reception thresholds (SRTs) between spatially separated (S0Nci or S0Nnh) and co-located speech and noise (S0N0), similar to Goupell et al. (2018). SRM was −1.4 dB when noise was presented to the NH ear (S0N0 - S0Nnh), and 5.3 dB when noise was presented to the CI ear (S0N0 - S0Nci). These results suggest that SRM may depend on the ear to which noise was presented, especially for conditions with large performance asymmetry (e,g, SSD CI users).

To better understand the effects of frequency mismatch within and across ears on speech understanding and SRM, performance asymmetry should be carefully considered. As reported in the previous studies (Zhou et al., 2016; Goupell et al., 2018), the amount of SRM was significantly higher when noise (or a speech masker) was presented to the poorer ear. Such differences are likely driven by the favorable signal-to-noise ratio (SNR) in the better ear due to head shadow effects. Using spatially symmetrically placed maskers (SSMs) for a frontal target (i.e., one speech masker presented to the left ear and another speech masker presented to the right ear) would provide similar long-term SNR in both ears, thus reducing the effect of SNR and performance asymmetry. SSMs have been used to measure binaural benefits (or SRM) in NH listener and bilateral CI users (Hu et al., 2018). SSMs are similar to those used in the Listening in Spatialized Noise - Sentence test (LiSN-S; Cameron and Dillon, 2007). Previous studies have shown significant SRM using the LiSN-S test (e.g.,

Brown et al., 2010). Hu et al. (2018) also found that SRM was significantly poorer for bilateral CI simulations and for real bilateral CI users than for NH subjects listening to unprocessed speech. However, both the simulated and real bilateral CI listeners benefitted greatly from ideal monaural better-ear masker (IMBM) processing. This suggests that the binaural benefit for SRM in bilateral CI users may be partly due to better-ear glimpsing to select time-frequency segments with favorable SNR in either ear.

These results suggest that SSMs might be a better alternative than asymmetrically placed maskers to measure SRM due to potential better ear effects. The present study evaluated the effects of frequency resolution, as well as the degree of frequency mismatch within and across ears on perception of competing speech and SRM using SSMs; NH subjects were tested while listening to unprocessed speech or 16-channel vocoders that simulated different degrees of frequency mismatch within and across ears. The target speech was delivered to both ears and the two speech maskers were delivered to: 1) both ears (diotic maskers), or 2) different ears (dichotic maskers), with one speech masker delivered to the left ear and the other to the right ear. SRM was calculated as the performance difference in SRTs between the diotic and dichotic masker conditions. The diotic masker was similar to the co-located target-masker conditions in Zhou et al. (2016) and Goupell et al. (2018). However, the present dichotic masker condition was slightly different from the spatially separated conditions in these previous studies, which used head-related transfer functions (HRTFs). For an HRTF-based spatial configuration, the left channel contains target sentence from the front, stronger "ipsilateral" masking from the left and weaker "contralateral" masking from the right (and vice versa). In this study, HRTFs were not used. The dichotic condition in the present study was similar to the infinite inter-aural level difference (ILDinf) conditions from Hu et al. (2018), in which the acoustic crosstalk from the contralateral interferer introduced by HRTF was artificially removed. Hu et al. (2018) found that SRM was approximately 5 dB or more for ILDinf than for HTRF-based processing, for both real bilateral CI listening and for bilateral CI simulations. Therefore, the present approach allows for better stimulus control, allowing for the largest possible differences between the diotic and dichotic performance and potentially larger SRM.

2. Methods

2.1 Subjects

Fifteen NH adults (7 males and 8 females; mean age = 24.7 yrs) participated in the study. All NH subjects had pure tone thresholds < 25 dB HL at all audiometric frequencies between 250 and 8000 Hz in both ears. 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 Institutional Review Board in The University of California, Los Angeles (UCLA).

2.2 Test Materials

SRTs, defined here as the SNR that produces 50% recognition of keywords in competing speech, were measured using a closed-set, matrix-styled test paradigm (Crew et al., 2015, 2016). The test materials consisted of five-word sentences, constructed for each test trial by

randomly selecting one of ten words in each of five categories (Name, Verb, Number, Color, and Clothing); the sentences were syntactically correct but not necessarily semantically coherent. The same test materials were used to construct the target sentence (produced by a male talker) and the masker sentences (produced by two male talkers that were different from each other and different from the target talker).

For each test trial, the first word in the target sentence was always the Name "**John**," followed by randomly selected words from the remaining categories. Thus, the target sentence could be "**John** moves *Six Gold* pants" or "**John** needs *Two Green* shoes," etc. (Name to cue target talker in **bold**; keywords in ***bold italic***). Note that only words from the Number and Color categories in the target sentence were regarded as the keywords (Tao et al., 2017, 2018). The two competing sentences were also for each test trial; words were randomly selected from each category. The words used in the target sentence were excluded, and the words were different between the two masker sentences. Thus, the target and masker sentences all had different words in each category. For example, the target sentence could be "**John** moves *Six Gold* pants" and the masking sentences could be "Bob Finds *Two Blue* coats," and "Greg loans *Five Grey* Jeans".

2.3 Signal Processing

Stereo stimuli for the experiment were generated in real-time and delivered to headphones (Sennheiser HAD 200) via audio interface (Edirol UA-25) connected to a mixer (Mackie 402). The target sentence was always presented diotically while the two competing sentences were presented either diotically or dichotically. For the dichotic masker condition, one competing sentence was presented to the left channel alone and the other was presented to the right channel alone. The dichotic masker condition was similar to the "spatially separated" condition in Cameron and Dillon (2007) except that no HRTFs were used in this study.

In each test trial, target and masker sentences were mixed at the specified SNR. Note that for the diotic masker condition, the "noise" contained two separate sentences with the same root-mean-square (RMS) level in each channel (left or right). Both channels contained three sentences (one target and two maskers) and SNR was defined as the ratio between the target and the total summed energy of the two maskers. For the dichotic masker condition, the "noise" only contained one masking sentence in each channel but each masker was presented at the same RMS level as in each ear. As such, each channel contained two sentences (one target and one masker; note that the masker sentences were different between the left and right channels). The SNR was defined as the ratio between the target sentence and the masker sentence in each channel. The mixed target and masker sentences in each channel were further processed by a 16-channel sine-wave vocoder as in Fu et al. (2004). First, the signal was processed through a high-pass pre-emphasis filter with a cutoff of 1200 Hz and a slope of −6 dB/octave. The input frequency range (200–8000 Hz) was then divided into 16 frequency bands, using 4th order Butterworth filters distributed according to Greenwood's (1990) frequency-place formula. The temporal envelope from each band was extracted using half-wave rectification and low-pass filtering with a cutoff frequency of 160 Hz. The extracted envelopes were then used to modulate the amplitude of sinewave carriers.

The distribution of the carrier sinewaves assumed a 20 mm electrode array and three different insertion depths (IDs; 27, 26, and 25 mm, relative to the base) were simulated. These conditions represented a mild, mild-moderate, and moderate tonotopic mismatch between the acoustic input frequency and the expected spiral ganglion characteristic frequency of the CI electrode position. The corresponding output frequency ranges were 354–7771, 428–8944, and 513–10290 Hz for 27-, 26-, and 25-mm IDs, respectively, according to the Greenwood's (1990) frequency-place formula. Table 1 lists detailed parameters for the 16-channel sine-wave vocoders.

Stimuli in the left and right channels were processed similarly or differently depending on the degree of inter-aural mismatch. The five listening conditions were:

1. *Unprocessed:* Unprocessed speech with no frequency mismatch within or across ears.

2. *ID27–27:* 16-channel bilateral CI simulation with mild frequency mismatch within each ear but no inter-aural mismatch. In both channels, a 27-mm ID was simulated; the input frequency range was 200–8000 Hz and the output range was 354–7771 Hz.

3. *ID27–26:* 16-channel bilateral CI simulation with mild-to-moderate frequency mismatch in each ear and a 1-mm inter-aural mismatch. In the left channel, a 27-mm ID was simulated; the input frequency range was 200–8000 Hz and the output range was 354–7771 Hz. In the right channel, a 26-mm ID was simulated; the input frequency range was 200–8000 Hz and the output range was 428–8944 Hz.

4. *ID27–25:* 16-channel bilateral CI simulation with mild-to-moderate frequency mismatch in each ear and a 2-mm inter-aural mismatch. In the left channel, a 27-mm ID was simulated; the input frequency range was 200–8000 Hz and the output range was 354–7771 Hz. In the right channel, a 25-mm ID was simulated; the input frequency range was 200–8000 Hz and the output range was 513–10290 Hz.

5. *ID25–25:* 16-channel bilateral CI simulation with moderate frequency mismatch within each ear but no inter-aural mismatch. In both channels, a 25-mm ID was simulated; the input frequency range was 200–8000 Hz and the output range was 513–10290 Hz.

2.4 Test Procedure

Testing was completed in a sound-attenuating booth. SRTs were measured using an adaptive procedure (1-up/1-down; Levitt, 1971) that produced 50% correct sentence recognition. During each test trial, a sentence was presented at a designated SNR; the initial SNR was 10 dB. The subject clicked on one of the 10 responses for the Number and Color categories; no selections could be made from the remaining categories, which were greyed out. If the subject correctly identified both keywords, the SNR was reduced by 4 dB (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 SNRs for the last 6 reversals. If there were fewer than 6 reversals within 20 trials, the test run was discarded and another run was measured. Due to the expected wide range in SRTs, the overall presentation level was fixed (instead of fixing the target or masker levels) avoid loud sounds at low SNRs. Once target and masking sentences were combined at the specified SNR, the overall presentation level was further adjusted to maintain the same RMS value for all presentations in all experimental conditions.

Two masker conditions (diotic and dichotic) and five listening conditions (Unprocessed, ID27–27, ID27–26, ID27–25, and ID25–25) were tested, resulting in a total of 10 experimental conditions. All experimental conditions were grouped into one block; a minimum of two blocks were tested and the order of test conditions was randomized across blocks. For each experimental conditions, if SRTs for the first two runs differed by more than 3 dB, then a third run was tested. The SRT was averaged across all runs. Subjects were not given any practice prior to testing, and all testing was completed in one session. No preview, no feedback and no training were provided. Previous studies showed little improvement for recognition of spectrally degraded and shifted speech if subjects received no preview, no feedback, and no training (Fu et al., 2005). Learning effects related to listening to vocoded speech were likely minimal given the short testing period (~2 hours) even though these were naïve listeners. The randomization of conditions across test blocks, as well as conducting a third test run if SRTs differed by more than 3 dB between the first two runs, further reduced potential learning effects.

3. Results

Fig. 1 shows boxplots of SRTs for the five listening conditions and two masker conditions. For the diotic masker conditions, the best SRT was for unprocessed speech (0.3 dB) and the worst SRT was for moderately shifted speech with no inter-aural mismatch (ID25–25; 8.9 dB). For the dichotic masker conditions, the best SRT was for unprocessed speech (−18.0 dB) and the worst SRT was for speech with mild-to-moderate tonotopic mismatch and 2 mm of inter-aural mismatch (ID27–25; 5.5 dB). A two-way repeated-measures analysis of variance (RM ANOVA) was performed with masker condition (diotic, dichotic) and listening condition (Unprocessed, ID27–27, ID27–26, ID27–25, ID25–25) as the factors. Results showed significant effects for the masker [$F(1,56)=211.899$, $p<0.001$; $\eta^2=0.938$] and listening conditions [$F(4,56)=192.954$, $p<0.001$; $\eta^2=0.932$]; there was a significant interaction [$F(4,56)=50.912$, $p<0.001$; $\eta^2=0.784$]. For the diotic masker conditions, post-hoc Bonferroni pairwise comparisons showed significantly better SRTs for unprocessed speech than those for 16-channel vocoded speech ($p<0.001$). SRTs for the ID25–25 condition were significantly worse than SRTs for the ID27–27 ($p=0.005$) and ID27–26 conditions ($p=0.02$). No other significant differences were observed. For the dichotic masker conditions, SRTs were significantly better for unprocessed speech than those for 16-channel vocoded speech ($p<0.001$). SRTs for the ID27–27 condition were significantly better than SRTs for the ID27–26, ID27–25, and ID25–25 conditions ($p<0.001$). SRTs for the ID27–25 condition were worse than SRTs for the ID25–25 ($p=0.008$) and ID27–26 conditions ($p=0.009$).

SRM was calculated as the performance difference between diotic and dichotic masker conditions. Note we use SRM to describe benefits related to spatially separated targets and

maskers even though HRTFs were not used and performance was measured using headphones rather than in free field. A beneficial SRM refers a statistically significant difference (p<0.05) between the diotic and dichotic conditions. Fig. 2 shows boxplots of SRM for the five listening conditions. The largest SRM (18.3 dB) was observed for unprocessed speech. With spectrally degraded speech, SRM was significantly reduced to 7.6 dB with mild tonotopic mismatch but no inter-aural mismatch (ID27–27); SRM was further reduced with increasing inter-aural mismatch. A one-way RM ANOVA was performed with listening conditions as the factor. Results showed a significant effect for listening condition on SRM [$F(4,56)=50.912$, $p<0.001$; $\eta^2=0.784$]. Post-hoc Bonferroni pairwise comparisons showed significantly larger SRM with unprocessed speech than with spectrally degraded speech ($p<0.001$). For spectrally degraded speech, SRM was significantly larger for ID27–27 than for ID27–26 ($p=0.024$) or ID27–25 ($p<0.001$); SRM was significantly larger for ID25–25 than for ID27–25 ($p=0.005$).

4. Discussion

The data from the present study demonstrate that frequency resolution, tonotopic mismatch, and inter-aural mismatch have differential effects on speech understanding and SRM for simulated bilateral CI listening. NH subjects had better SRTs and larger SRM when listening to unprocessed speech. Speech understanding was more significantly affected by inter-aural mismatch in the dichotic than the diotic masker condition. The loss of spectral detail resulted in smaller SRM and SRM was further reduced as the degree of inter-aural mismatch increased. Tonotopic mismatch affected overall speech performance, but not SRM. Below, we discuss the results in greater detail.

Effects of Frequency Resolution on Overall Performance

The effects of frequency resolution on speech recognition in co-located noise have been extensively studied (e.g., Dorman et al., 1998; Shannon et al., 2004; Fu and Nogaki, 2005; Croghan et al., 2017; Berg et al., 2019). In general, speech recognition performance improves with the number of spectral channels (i.e., better frequency resolution). As speech materials become more difficult, more spectral channels of information are required. For example, Fu and Nogaki (2005) showed that the SRTs in speech-shaped steady noise (SSN) worsened from −4.6 dB with unprocessed speech to 0.8 dB with a 16-channel vocoder, a difference of 5.4 dB. An additional 3–4 dB drop was also observed when the spectral resolution was further reduced to 8 channels. In the present study, SRT for the diotic masker condition was reduced from 0.3 dB with unprocessed speech to 5.6 dB with the ID27–27 vocoder simulation (mild tonotopic mismatch but no inter-aural mismatch). Regardless of differences in masker type (SSN vs. competing speech), the performance difference between unprocessed speech and 16-channel vocoded speech was similar between the present study and Fu and Nogaki (2005). When the degree of tonotopic mismatch was further increased (ID25–25), SRT was further reduced to 8.9 dB, comparable to that with further reduction of frequency resolution in Fu and Nogaki (2005).

Effects of Inter-aural Mismatch on Overall Performance

In addition to tonotopic mismatch, speech understanding was also affected by inter-aural mismatch. With a slight inter-aural mismatch (1 mm in the ID27–26 condition), the SRT for the ID27–26 condition (6.0 dB) was comparable to that for the matched ID27–27 condition (5.6 dB). However, when the inter-aural mismatch was increased to 2 mm, the SRT for the ID27–25 condition was further reduced to 7.3 dB. The performance drop with increasing inter-aural mismatch was consistent with Yoon et al. (2013), who showed that increased inter-aural mismatch provided poorer binaural benefit. These results suggest that the loss of spectral detail results in a large performance drop, which is compounded by tonotopic mismatch and, further, by inter-aural mismatch. The patterns are generally similar regardless of the masker type (SSN in previous studies vs. speech maskers in the present study).

Effects of Frequency Resolution on SRM

SRM may be influenced by many factors, including the number of interfering talkers, spatial configuration of the sound sources, room acoustics, and similarity between the target and maskers (e.g., Zurek, 1993; Plomp and Mimpen, 1981; Marrone et al., 2008; Brungart et al., 2001). The data from the present study also suggest that SRM may be negatively affected by the loss of spectral detail, consistent with previous studies (Hu et al., 2018). SRM was 18.3 dB for the unprocessed speech and significantly reduced to 7.6 dB for the 16-channel spectrally degraded speech with mild tonotopic mismatch and no inter-aural mismatch (ID27–27), consistent with Hu et al. (2018), who measured SRM for both unprocessed speech and a 12-channel noise-vocoded bilateral CI simulation. In Hu et al. (2018), the target speech always originated from the front (0°) while the two uncorrelated speech maskers were either co-located with the target, or spatially separated at ±60° in the horizontal plane. The scenarios were simulated using HTRFs in virtual acoustics. SRM was 10.0 dB for unprocessed speech and 0.0 dB for the bilateral CI simulation. As mentioned earlier, the dichotic condition in the present study was similar to the infinite inter-aural level difference (ILDinf) condition in Hu et al. (2018), in which the acoustic crosstalk from the contralateral interferer introduced by HRTF was artificially removed. For the ILDinf condition, SRM was 19.0 dB for unprocessed speech and 6.4 dB for the bilateral CI simulation, which was highly consistent with the data from the present study.

The data from previous studies have shown that ILD cues may not suffice for restoring SRM in spectrally degraded speech and appropriate ITD cues are necessary for restoring SRM (Ihlefeld and Litovsky, 2012). However, the data from the present study suggested that reduced SRM for the spectrally degraded speech may not be explained in terms of ITD and/or ILD cues, as no ITD cues were available for all listening conditions and ILD cues were artificially enhanced (i.e., ILDinf) for both unprocessed and spectrally degraded speech. Suneel et al. (2017) found that localization performance was closely related to binaural fusion changes for spectrally degraded speech. In their study, two fusion measures (fused/unfused and punctate/diffuse), were used. For the punctate/diffuse measure, participants indicated what they heard, ranging from a punctate sound to a diffuse sound to separate sounds at each ear. Participants saw an image of a head with a small oval in the center. By turning a dial, they were able to make the oval, which represented the number and size of the auditory images in their head, larger or smaller. The dial created discrete steps

and the number of steps (i.e., 0=small oval in the center of the head; 8=large oval in the center of the head; see Fig. 1 in Suneel et al., 2017) was used as an ordinal scale of fusion. The smallest median value was 4 for all spectrally degraded speech, which indicated a moderately sized auditory image in the center of the head. The results from Suneel et al. (2017) suggest that reduced SRM for the spectrally degraded speech is likely caused by the partially overlapped auditory images of spatially separated target and masker, due to the diffused auditory images.

Effects of Inter-aural Mismatch on SRM

The present data also showed that SRM was negatively affected by the inter-aural mismatch but not tonotopic mismatch. SRM was 7.6 dB for the 16-channel spectrally degraded speech with mild tonotopic mismatch and no inter-aural mismatch (ID27–27). SRM was slightly reduced to 6.4 dB for the 16-channel spectrally degraded speech with moderate tonotopic mismatch and no inter-aural mismatch (ID25–25). However, the difference between ID27–27 and ID25–25 was not statistically significant (p=.26). With slight inter-aural mismatch (1 mm; ID27–26), SRM was significantly reduced to 3.6 dB, and further reduced to 1.7 dB with moderate inter-aural mismatch (2 mm; ID27–25). SRM for the ID27–25 condition was significantly smaller than SRM for the ID27–27 (p<0.001) or ID25–25 conditions (p=0.005). Similar results were also reported by Goupell et al. (2018) who used a sine-vocoder to evaluate the effect of simulated frequency mismatch on speech understanding and SRM. They found that tonotopic mismatch in one or both ears decreased speech understanding. SRM, however, was only affected in conditions with inter-aural mismatch.

Previous studies have shown that inter-aural mismatch may affect binaural fusion and localization abilities (Goupell et al., 2013; Kan et al., 2013). Recently, Aronoff et al. (2018) found a close relationship between changes in binaural fusion and changes in localization ability using a vocoder simulation that manipulated the degree of inter-aural mismatch. Reduced binaural fusion introduced by inter-aural mismatch is likely responsible for the difference in SRM for the different listening conditions in the present study. As shown in Figure 1, SRTs for the diotic listening condition were generally similar among all listening conditions and SRM was largely driven by SRTs in the dichotic listening condition. The limited binaural benefits for SRM with inter-aural mismatch were likely driven by binaural interaction (or binaural fusion) in the dichotic presentation. For the dichotic listening condition with no inter-aural mismatch (i.e., ID27–27), the target speech in both channels was likely fused and merged into a central image. However, the speech masker in the left channel provided a lateralized image in the left side while the speech masker in the right channel provided a lateralized image in the right side. The "virtual" spatial separation depended on the amount of fusion and likely responsible for the reasonable SRM in the matched condition (i.e., ID27–27, ID25–25). However, with inter-aural mismatch, target speech was likely less perceived as centralized image. Listeners reported two lateralized images with moderate inter-aural mismatch even when target speech was presented alone. The lateralized target speech was likely mixed with the lateralized speech maskers in either side, resulting in two lateralized images that each contained both one target and one masker. In this case, the listener likely listened to the better ear alone (i.e., the ear with less tonotopic mismatch). For example, the mean SRT was 5.5 dB in the dichotic condition for ID27–25,

which was nearly the same as the SRT with the diotic listening condition for ID27–27 (5.6 dB). These results suggested that inter-aural mismatch reduced binaural fusion, thus reducing binaural benefits such as SRM.

However, there is the possibility that SRM measured in this study reflects a monaural dip-listening benefit, instead of or in addition to binaural factors (e.g., binaural fusion). With two uncorrelated fluctuating maskers, the ear with the better short-term SNR will alternate between ears over time (i.e., "better-ear glimpsing"). In this scenario, the binaural advantage relies on combining monaural speech cues across ears, rather than relying on binaural processing per se (Best et al. 2015; Brungart and Iyer, 2012). Hu et al. (2018) also found that real bilateral CI users and NH subjects listening to bilateral CI simulation exhibited strongly reduced benefit from spatial separation compared to NH subjects listening to unprocessed speech. However, both groups greatly benefited from IMBM processing. These results suggest that binaural benefits may be at least partly due to better-ear glimpsing. Other studies have shown that better-ear glimpsing cannot fully account for SRM in conditions with high informational masking (Glyde et al., 2013) or in bilateral sensorineural hearing impairment (Best et al., 2015). If SRM depended on better-ear glimpsing, SRM for the ID27–25 condition would likely be larger than that for the ID25–25 condition, as ID27 would be expected to provide better performance than ID25 due to less tonotopic mismatch. Such predictions are inconsistent with the present results, suggesting that better-era glimpsing cannot fully account for SRM when there is an inter-aural mismatch.

The masker changes between the diotic and dichotic conditions may also affect SRM. In each ear, the masker goes from being a mixture of two voices (diotic condition) to one voice (dichotic condition). This will result in increased speech information in each ear. Hu et al. (2018) found that SRM was about 5 dB better for the ILDinf condition (similar to the present dichotic condition) than the HRTF condition, of which 3 dB could be attributed to the reduced level of the ILDinf interferer. The remaining 2-dB effect may be attributed to the masker change. However, this masker change would be expected to boost SRM for all listening conditions. The deficit in SRM due to inter-aural mismatch likely remains a factor.

Effects of Testing Materials

For the diotic masker condition, SRT for unprocessed speech was 0.3 dB, which was slightly worse than the low-cue SRTs (−1.7 dB) reported in the previous LiSN-S study (Brown et al., 2010). One possible difference was the test materials. In the present study, both target and competing speech used the matrix-styled materials. In the previous LiSN-S study, the target was simple Bamford-Kowal-Bench (BKB) sentences while maskers consisted of speech from two stories (Cameron and Dillon, 2007). The similar sentence structure between target and competing sentences in the present study may have increased difficulty in segregating the target and maskers. Nevertheless, the difference between studies was relatively small (approximately 2 dB).

As mentioned earlier, the testing materials and methods used in the present study were generally similar to the methods used in Brungart et al. (2001). Instead of measuring SRTs, Brungart et al. (2001) measured performance as a function of the target-to-masker ratio (TMR). When the voice gender of the target and the two masker talkers were the same, the

estimated SRT was approximately 3 dB TMR. Note that TMR in Brungart et al. (2001) refers to the ratio between the target and one masking talker. In the present study, the SNR refers to the ratio between the target and the summed masking signal in each ear. Thus, when the levels of all three talkers are the same in the diotic condition (i.e., same as 3-talker TSS condition in Brungart et al., 2001), the TMR of each masking talker would be 0 dB, but the overall SNR would be approximately −3 dB. Therefore, the estimated SRT of 3 dB TMR in Brungart et al. (2001) is theoretically equivalent to 0 dB SNR in the present study (which is comparable to 0.3-dB SNR observed in this study).

HRTF and Dichotic Listening for Spatial Representation

In previous studies, the spatial auditory listening environment was created by non-individualized HRTFs and the performance difference between co-located and spatially separated listening conditions was used to evaluate SRM (Goupell et al., 2018; Zhou et al., 2016) or spatial advantage (Cameron and Dillon, 2007). The present study used the dichotic masker presentation to create a "perceptually" spatially separated listening condition instead of using non-individualized HRTFs. As noted previously, the present dichotic condition was somewhat different from the spatially separated condition using HRTFs in the LiSN-S test, where stimuli in the left channel contained the target from the front, a stronger "ipsilateral" masker from the left, and a weaker "contralateral" masker form the right (and vice-versa). The mean RMS level difference between the "ipsilateral" and "contralateral" masker sentences was about 7.2 dB for the test stimuli used in the LiSN-S test.

Even though two interfering maskers had different levels in each ear, the combined two interfering maskers (even with different level) may reduce the temporal "dips" in both ears. In the present dichotic condition, only one masker was presented to each ear; as such, time-frequency segments with favorable SNRs in either ear were much more distinct, allowing for "better-ear" glimpsing. The present approach also allowed for more careful control of the stimuli to provide the largest possible differences between the diotic and dichotic conditions. With HRTFs, it would not be possible to fully isolate the signals in each ear. With unprocessed speech, SRM was approximately 12 dB using HRTFs (Brown et al., 2010) and more than 18 dB using the present ILDinf approach for the dichotic condition. Thus, the present approach may provide a large baseline SRM with which to compare the effects of frequency resolution, tonotopic mismatch, and inter-aural mismatch on perception of spatially separated competing speech.

Clinical Implications

The present data suggest that frequency resolution, tonotopic mismatch, and inter-aural mismatch may reduce listeners' ability to segregate target speech from spatially separated competing speech. It is critical to minimize inter-aural mismatch to restore binaural benefits in bilateral CI patients for spatially separated competing speech, a common daily realistic listening environment. Since SRM in this study were primarily affected by inter-aural mismatch, the present methodology may be a useful clinical tool with which to estimate inter-aural mismatch in bilateral CI users. Such tests may also be useful in evaluating the benefits of optimization methods that minimize the inter-aural mismatch (e.g., frequency allocations based on radiological imaging, inter-aural pitch matching, etc.). Radiological

imaging can be used to estimate intra-cochlear electrode position, which can then be used to generate tonotopically appropriate frequency-place allocation (e.g., Labadie et al., 2016). Noble et al. (2014) found that image-guided frequency allocation may significantly improve word and sentence recognition in both quiet and noise in pediatric CI users. Acoustic-electric (bimodal, SSD) or electric-electric (bilateral CI) pitch-matching is often used to assess inter-aural mismatch. Pitch-matching accommodates idiosyncrasies in the electrode neural interface and long-term adaptation that may not be captured by imaging. Pitch-matching data has also been used to generate frequency to-electrode allocations that improve localization performance in bilateral CI users (Kan et al., 2015).

In the present study, the performance difference was quite small with diotic maskers for the listening conditions with or without inter-aural mismatch. For example, the average performance difference between the ID27–27 and ID27–26 was only 0.4 dB, which is well within the range of testing variability. Even if the inter-aural mismatch could be perfectly corrected by appropriate frequency-place allocation, the improvement for diotic maskers may be marginal, making it difficult to validate the effectiveness of the optimization methods. With dichotic maskers, correcting for a mild inter-aural mismatch (1-mm) improved SRM by 4.4 dB. As such, the present method may be useful to validate whether optimization of the frequency allocation across ears may effectively reduce inter-aural mismatch.

Limitations of the Present Study

The present study used vocoder simulation to understand the potential effects of spectral resolution and frequency mismatch on speech understanding and SRM. This approach has several advantages. For example, testing vocoder simulation with NH listeners provides useful information regarding the effects of frequency resolution, tonotopic mismatch, and inter-aural mismatch, as simulations allow for explicit control of signal processing and place of stimulation that cannot be easily controlled in real CI users (Fitzgerald et al., 2017) and which likely contributed to the large inter-subject variability observed in CI users. While vocoders are highly imperfect simulations of CI listening (Dorman et al., 2017), there are many fundamental similarities between simulated and real CI performance. CI users do not have access to temporal fine structure (TFS) cues, and instead must rely on coarse spectral envelope cues and low-frequency temporal envelope cues (e.g., Shannon et al., 1995; Dorman et al., 1997). Vocoder simulations are able to capture CI users' limited functional spectral resolution (caused by channel interaction associated with current spread) by varying carrier bandwidth or mixing across channels (Fu and Nogaki, 2005; Crew et al., 2012; Grange et al., 2017). Acute and post-training effects of tonotopic mismatch for CI users (e.g., Fu and Shannon, 1999; Fu et al., 2002) can also be effectively simulated using NH listeners (e.g., Fu and Shannon, 1999; Fu and Galvin, 2003; Li et al., 2009).

Similarities in performance between real and simulated bilateral CI listening have been observed. Hu et al. (2018) found that real and simulated bilateral CI listening exhibited similar SRM across conditions, despite the considerably lower SRTs in the simulated bilateral CI group. This suggests that the bilateral vocoder simulations, which remove TFS and reduce the spectral resolution, might capture the major factors that limit the SRM in

bilateral CI users, relative to NH subjects listening to unprocessed speech. These main limitations persist even though the sound quality of CI simulations is quite different from that of real CIs (e.g., acoustic-electric quality comparison in SSD CI users by Dorman et al., 2017).

Another potential confounding factor may be the choice of carriers used in the vocoder simulations. Noise-band vocoders typically produce a noisy or whispered voice quality (Shannon et al., 1995); also, the temporal envelope is noisy due to noise carrier. Sinusoidal carriers with a frequency equal to the center frequency of a carrier band produce a tonal voice quality with a distinct flattening of the pitch contour (Dorman 1997; Goupell et al., 2018); the temporal envelope is less noisy than with noise-vocoders. The sound quality of noise- and sine-vocoders does not generally correspond to that of real CIs (Dorman et al., 2017), and the median rating is slightly better for sine-vocoders than for noise-vocoders.

In the present study, the use of sine-vocoders might very well have introduced binaural-fusion cues that real CI listeners may not be able to access. The NH listeners in the present study may have attended to correlated TFS cues available with the sine carriers when there was no inter-aural mismatch, allowing for better binaural fusion. Noise-vocoders that use uncorrelated noise in each ear do not preserve TFS cues, and therefore would limit potential binaural fusion benefits associated with sine-vocoders. Hu et al. (2018) used 12-channel noise-vocoders to simulate bilateral CIs. In that study, when there was no tonotopic mismatch or inter-aural mismatch, the mean SRM was 6.4 dB for the ILDinf condition, comparable to the 7.6 dB of SRM in the present study with a 16-channel, sine-vocoder that simulated a mild tonotopic mismatch and no inter-aural mismatch. As a point of further comparison, we measured SRM for the ID27–27 and ID27–25 conditions in three subjects listening to 16-channel noise-vocoders. SRM was 6.4 dB for the ID27–27 condition and −0.9 dB for the ID27–25 condition. While SRM might have been slightly smaller with noise- than sine-vocoders, the effects of inter-aural mismatch on SRM may not depend strongly on vocoder carrier type. Goupell et al. (2018) used sine-vocoders to evaluate the effects of simulated tonotopic mismatch on speech understanding and SRM. They found that while tonotopic mismatch in one or both ears reduced speech performance, SRM was only affected by the degree of inter-aural mismatch. Similarly, the present CI simulation results showed the relatively large effects of inter-aural mismatch on SRM, highlighting the importance of reducing inter-aural mismatch to maximize binaural benefits in bilateral CI users. However, future studies with real bilateral CI users are needed to confirm the effects of inter-aural mismatch on speech recognition and SRM.

Another limitation to vocoder studies is the lack of long-term experience. Most CI users have ample opportunity to adapt over time to frequency mismatches within or across ears. Unilateral CI users are able to at least partly adapt to tonotopic mismatch (Rosen et al., 1999; Fu et al., 2002; Svirsky et al, 2004). Bilateral CI users have been shown to at least partly adapt to inter-aural mismatches (Svirsky et al., 2015); adjusting the frequency allocation in the CI ear with shallow insertion depth to reduce the inter-aural mismatch depth greatly improved binaural speech performance. These studies suggest that while acute measures may not fully reflect the effects of tonotopic and inter-aural mismatch, auditory plasticity may not fully compensate for these mismatches. Most of the present NH

participants were novice listeners. Previous vocoder simulation studies have shown that NH listeners can passively adapt to small amounts of frequency mismatch (Li et al., 2009); even greater adaptation is possible with explicit training (Rosen et al., 1999; Fu and Galvin, 2003; Fu et al., 2005). However, most of these previous studies focused on tonotopic mismatch. It is unclear whether training can improve speech performance and more importantly, SRM for inter-aurally mismatched speech. Future studies on training benefits for inter-aural mismatched speech on SRM will provide further evidence about the importance of minimizing inter-aural mismatch to maximize binaural benefit in bilateral CI users.

5. Summary and Conclusions

Speech understanding for a target sentence in the presence of competing speech was measured for NH listeners listening to unprocessed speech or to 16-channel sine-wave vocoded speech with and without inter-aural mismatch. Target speech was delivered to both ears and two speech maskers were delivered to: 1) both ears (diotic masker), or 2) different ears (dichotic masker: one delivered to the left ear and the other delivered to the right ear). SRM was calculated as the difference between the diotic and dichotic listening conditions. Major findings include:

1. Speech understanding and SRM were significantly worse for spectrally degraded speech in both diotic and dichotic masker conditions.

2. When there was no inter-aural mismatch, performance similarly worsened with increasing tonotopic mismatch for both the diotic and dichotic listening conditions. Inter-aural mismatch had a much larger effect on the dichotic than the diotic listening condition.

3. The degree of inter-aural mismatch was a major factor in SRM. While increased tonotopic mismatch increased the difficulty in segregating target from masking speech, the effect of tonotopic mismatch on SRM was smaller than the effect of inter-aural mismatch.

4. The present approach to measure SRM may provide a useful clinical tool with which to measure inter-aural mismatch in bilateral CI users and to evaluate the benefits of optimization methods that minimize inter-aural mismatch.

Acknowledgments

References

Aronoff JM, Shayman C, Prasad A, Suneel D, Stelmach J (2015). Unilateral spectral and temporal compression reduces binaural fusion for normal hearing listeners with cochlear implant simulations. Hearing research, 320, 24–29. [PubMed: 25549574]

Ba kent D, Shannon RV (2003). Speech recognition under conditions of frequency-place compression and expansion. J Acoust Soc Am. 2003 Apr;113(4 Pt 1):2064–76. [PubMed: 12703717]

Ba kent D, Shannon RV (2004). Frequency-place compression and expansion in cochlear implant listeners. J Acoust Soc Am. 116(5):3130–40. [PubMed: 15603158]

Ba kent D, Shannon RV (2007). Combined effects of frequency compression-expansion and shift on speech recognition. Ear Hear. 28(3):277–89. [PubMed: 17485977]

Berg KA, Noble JH, Dawant BM, Dwyer RT, Labadie RF, Gifford RH. (2019). Speech recognition as a function of the number of channels in perimodiolar electrode recipients. J Acoust Soc Am. 2019 3;145(3):1556. doi: 10.1121/1.5092350. [PubMed: 31067952]

Bernstein JG, Stakhovskaya OA, Schuchman GI, Jensen KK, Goupell MJ (2018). Interaural time-difference discrimination as a measure of place of stimulation for cochlear-implant users with single-sided deafness. Trends in hearing, 22, 2331216518765514. [PubMed: 29623771]

Best V, Mason CR, Kidd G Jr, Iyer N Brungart DS (2015). Better-ear glimpsing in hearing-impaired listeners. The Journal of the Acoustical Society of America, 137(2), EL213–EL219. [PubMed: 25698053]

Blanks DA, Buss E, Grose JH, Fitzpatrick DC, Hall JW 3rd. (2008). Interaural time discrimination of envelopes carried on high-frequency tones as a function of level and interaural carrier mismatch. Ear Hear. 2008 Oct;29(5):674–83.

Brown DK, Cameron S, Martin JS, 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. J Am Acad Audiol. 2010 Nov-Dec;21(10):629–41. [PubMed: 21376004]

Brungart D (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. 109, 1101–1109. [PubMed: 11303924]

Brungart DS, Iyer N. (2012). Better-ear glimpsing efficiency with symmetrically-placed interfering talkers. The Journal of the Acoustical Society of America, 132(4), 2545–2556. [PubMed: 23039448]

Brungart DS, Simpson BD, Ericson MA, Scott KR. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am. 2001 11;110(5 Pt 1):2527–38. [PubMed: 11757942]

Cameron S, Dillon H. (2007). Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). Ear Hear. 2007 4;28(2):196–211. [PubMed: 17496671]

Crew JD, Galvin JJ 3rd, Fu QJ (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. J Acoust Soc Am 132(5):EL429–35. PMCID: PMC3494451. [PubMed: 23145706]

Crew JD, Galvin JJ 3rd, Fu QJ. (2015). Melodic contour identification and sentence recognition using sung speech. J Acoust Soc Am. 2015 9;138(3):EL347–51. [PubMed: 26428838]

Crew JD, Galvin JJ 3rd, Fu QJ. (2016). Perception of Sung Speech in Bimodal Cochlear Implant Users. Trends Hear. 2016 11 11;20.

Croghan NBH, Duran SI, Smith ZM. (2017). Re-examining the relationship between number of cochlear implant channels and maximal speech intelligibility. J Acoust Soc Am. 2017 12;142(6):EL537. doi: 10.1121/1.5016044 Erratum in: J Acoust Soc Am. 2018

Dorman MF, Loizou PC, Rainey D (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. J Acoust Soc Am 102(5 Pt 1):2993–6. [PubMed: 9373986]

Dorman MF, Loizou PC, Fitzke J, Tu Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. J Acoust Soc Am. 1998 12;104(6):3583–5. [PubMed: 9857516]

Dorman MF, Natale SC, Butts AM, Zeitler DM, Carlson ML (2017). The sound quality of cochlear implants: Studies with single-sided deaf patients. Otol Neurotol. 38(8):e268–e273. [PubMed: 28806337]

Fitzgerald MB, Prosolovich K, Tan CT, Glassman EK, Svirsky MA (2017). Self-Selection of Frequency Tables with Bilateral Mismatches in an Acoustic Simulation of a Cochlear Implant. J Am Acad Audiol 28(5):385–394. doi: 10.3766/jaaa.15077. [PubMed: 28534729]

Francart T, Wiebe K, Wesarg T (2018). Interaural time difference perception with a cochlear implant and a normal ear. Journal of the Association for Research in Otolaryngology, 19(6), 703–715. [PubMed: 30264229]

Fu QJ, Shannon RV (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. J Acoust Soc Am 105(3):1889–900. [PubMed: 10089611]

Fu QJ, Shannon RV, Galvin JJ 3rd (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. J Acoust Soc Am. 112(4):1664–74. [PubMed: 12398471]

Fu QJ, Shannon RV, Galvin JJ 3rd (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. J Acoust Soc Am. 112(4):1664–74. [PubMed: 12398471]

Fu QJ, Galvin JJ 3rd (2003). The effects of short-term training for spectrally mismatched noise-band speech. J Acoust Soc Am. 113(2):1065–72. [PubMed: 12597199]

Fu QJ, Nogaki G, Galvin JJ 3rd (2005). Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. J Assoc Res Otolaryngol. 6(2):180–9. [PubMed: 15952053]

Goupell MJ, Stoelb CA, Kan A, Litovsky RY. (2018). The Effect of Simulated Interaural Frequency Mismatch on Speech Understanding and Spatial Release From Masking. Ear Hear. 2018 Sep/Oct;39(5):895–905. [PubMed: 29337763]

Grange JA, Culling JF, Harris NSL, Bergfeld S (2017). Cochlear implant simulator with independent representation of the full spiral ganglion. J Acoust Soc Am. 142(5):EL484. [PubMed: 29195445]

Greenwood DD (1990). A cochlear frequency-position function for several species - 29 years later. J Acoust Soc Am. 87(6):2592–605. [PubMed: 2373794]

Hu H, Dietz M, Williges B, and Ewert SD (2018). Better-ear glimpsing with symmetrically-placed interferers in bilateral cochlear implant users. J Acoust Soc Am. 87(6): 2128–2141.

Ihlefeld A, Litovsky RY (2012). Interaural level differences do not suffice for restoring spatial release from masking in simulated cochlear implant listening. PLoS One. 2012;7(9):e45296. doi: 10.1371/journal.pone.0045296. [PubMed: 23028914]

Kan A, Stoelb C, Litovsky RY, Goupell MJ (2013). Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users. J Acoust Soc Am. 134(4):2923–36. [PubMed: 24116428]

Kan A, Litovsky RY, Goupell MJ (2015). Effects of interaural pitch matching and auditory image centering on binaural sensitivity in cochlear implant users. Ear Hear. 2015a 36(3):e62–8. [PubMed: 25565660]

Labadie RF, Noble JH, Hedley-Williams AJ, Sunderhaus LW, Dawant BM, Gifford RH (2016). Results of postoperative, CT-based, electrode deactivation on hearing in prelingually deafened adult cochlear implant recipients. Otol Neurotol. 37(2):137–45. [PubMed: 26719955]

Landsberger DM, Svrakic M, Roland JT Jr, Svirsky M (2015). The Relationship Between Insertion Angles, Default Frequency Allocations, and Spiral Ganglion Place Pitch in Cochlear Implants. Ear Hear. 2015 Sep-Oct;36(5):e207–13. doi: 10.1097/AUD.0000000000000163. [PubMed: 25860624]

Long CJ, Eddington DK, Colburn HS, Rabinowitz WM (2003). Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user. J Acoust Soc Am. 114(3):1565–74. [PubMed: 14514210]

Levitt H (1971). Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, Supp 9(2), 467.

Li T, Galvin JJ 3rd, Fu QJ (2009). Interactions between unsupervised learning and the degree of spectral mismatch on short-term perceptual adaptation to spectrally shifted speech. Ear Hear 30(2):238–49. PMCID: PMC2889179. [PubMed: 19194293]

Ma N, Morris S, Kitterick PT (2016). Benefits to speech perception in noise from the binaural integration of electric and acoustic signals in simulated unilateral deafness. Ear and hearing, 37(3), 248. [PubMed: 27116049]

Marrone N, Mason CR, Kidd G Jr. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. J Acoust Soc Am. 2008 11;124(5):3064–75. doi: 10.1121/1.2980441. [PubMed: 19045792]

Noble JH, Gifford RH, Hedley-Williams AJ, Dawant BM, Labadie RF (2014). Clinical evaluation of an image-guided cochlear implant programming strategy. Audiol Neurootol. 19(6):400–11. [PubMed: 25402603]

Oxenham AJ, Kreft HA (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. Trends Hear. 13;18.

Plomp R and Mimpen AM (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception threshold for sentences. Acustica 48, 325–332.

Rosen S, Faulkner A, and Wilkinson L (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants, J Acoust Soc Am. 1999 12;106(6):3629–36. [PubMed: 10615701]

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995). Speech recognition with primarily temporal cues. Science 270(5234):303–4. [PubMed: 7569981]

Shannon RV, Galvin JJ 3rd, Baskent D (2002). Holes in hearing. J Assoc Res Otolaryngol. 3(2):185–99. [PubMed: 12162368]

Shannon RV, Fu QJ, Galvin J 3rd. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. Acta Otolaryngol Suppl 2004 5; (552):50–4.

Suneel D, Staisloff H, Shayman CS, Stelmach J, Aronoff JM (2017). Localization performance correlates with binaural fusion for interaurally mismatched vocoded speech. J Acoust Soc Am. 142(3):EL276. [PubMed: 28964063]

Svirsky MA, Silveira A, Neuburger H, Teoh SW, Suárez H (2004). Long-term auditory adaptation to a modified peripheral frequency map. Acta Otolaryngol. 124(4):381–6. [PubMed: 15224858]

Svirsky MA, Fitzgerald MB, Sagi E, Glassman EK (2015). Bilateral cochlear implants with large asymmetries in electrode insertion depth: implications for the study of auditory plasticity. Acta Otolaryngol. 135(4):354–63. [PubMed: 25719506]

Tao DD, Fu QJ, Galvin JJ, and Yu YF (2017). The development and validation of the Closed-set Mandarin Sentence (CMS) test. Speech Comm 92, 125–131.

Tao DD, Liu YW, Fei Y, Galvin JJ 3rd, Chen B, Fu QJ. (2018). Effects of age and duration of deafness on Mandarin speech understanding in competing speech by normal-hearing and cochlear implant children. J Acoust Soc Am. 2018 8;144(2):EL131. [PubMed: 30180674]

Wess JM, Brungart DS, Bernstein JG (2017). The effect of interaural mismatches on contralateral unmasking with single-sided vocoders. Ear and Hearing, 38(3), 374–386. [PubMed: 28002083]

Yoon YS, Li Y, Kang HY, Fu QJ (2011a). The relationship between binaural benefit and difference in unilateral speech recognition performance for bilateral cochlear implant users. Int J Audiol. 50(8):554–65. [PubMed: 21696329]

Yoon YS, Liu A, Fu QJ (2011b). Binaural benefit for speech recognition with spectral mismatch across ears in simulated electric hearing. J Acoust Soc Am. 130(2):EL94–100. [PubMed: 21877777]

Yoon YS, Shin YR, Fu QJ (2013). Binaural benefit with and without a bilateral spectral mismatch in acoustic simulations of cochlear implant processing. Ear Hear. 34(3):273–9. [PubMed: 22968427]

Zhou X, Li H, Galvin JJ III, Fu QJ, Yuan W (2016). Effects of insertion depth on spatial speech perception in noise for simulations of cochlear implants and single-sided deafness. Int J Audiol. 1:1–8.

Zurek PM (1993). Binaural advantages and directional effects in speech intelligibility, in Acoustical Factors Affecting Hearing Aid Performance, edited by Studebaker G and Hockberg I(Allyn and Bacon, Needham Heights, MA) pp. 255–276.
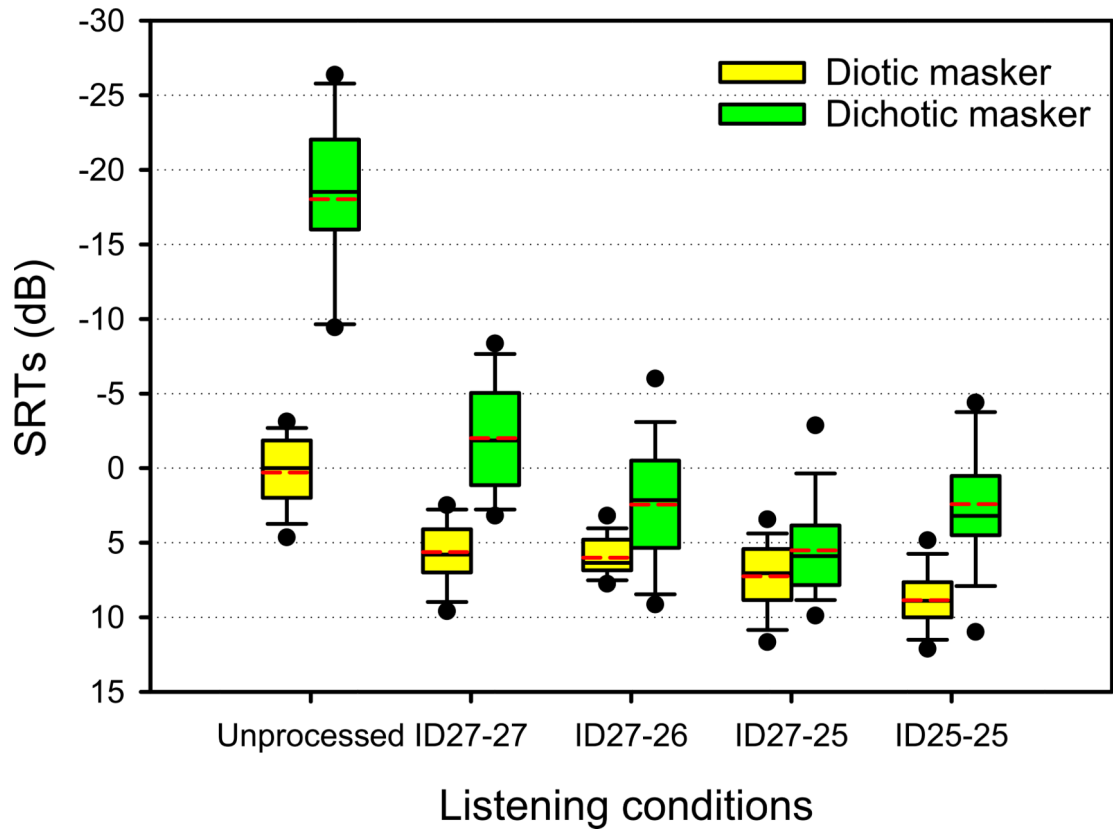
Figure 1.
Boxplots of SRTs with diotic and dichotic speech maskers for the different listening conditions. The boxes show the 25th and 75th percentiles, the error bars show the 5th and 95th percentiles, the symbols show outliers, the solid horizontal line shows the median and the dashed horizontal line shows the mean.

Figure 2.
Boxplots of SRM for the different listening conditions. The boxes show the 25th and 75th percentiles, the error bars show the 5th and 95th percentiles, the symbols show outliers, the solid horizontal line shows the median and the dashed horizontal line shows the mean.
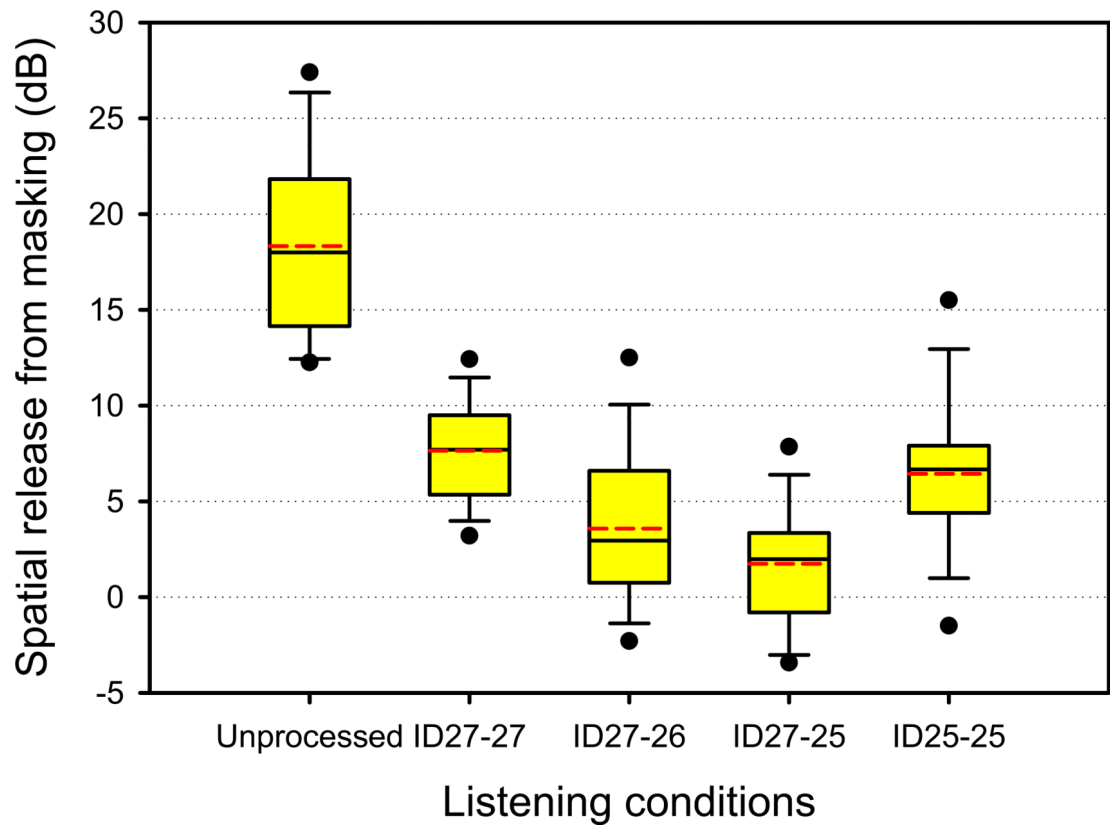
Table 1.

Detailed parameters for the 16-channel sine-wave vocoders, including the corner frequencies of each analysis band and carrier band as well as the center frequency (CF) of the carrier sinewaves for each analysis and carrier band for the simulated insertion depths at 27, 26, 25 mm.

	Input			Output	ID 27mm		Output	ID 26mm		Output	ID 25mm	
CH	Low	High	CF	Low	High	CF	Low	High	CF	Low	High	CF
1	200	275	**235**	354	448	**398**	428	536	**479**	513	637	**572**
2	275	367	**318**	448	560	**501**	536	665	**597**	637	785	**707**
3	367	479	**419**	560	693	**623**	665	817	**737**	785	960	**868**
4	479	616	**543**	693	851	**768**	817	999	**903**	960	1168	**1059**
5	616	782	**694**	851	1039	**940**	999	1215	**1102**	1168	1416	**1286**
6	782	985	**878**	1039	1262	**1145**	1215	1471	**1337**	1416	1710	**1556**
7	985	1231	**1101**	1262	1528	**1389**	1471	1776	**1616**	1710	2060	**1877**
8	1231	1532	**1373**	1528	1843	**1678**	1776	2138	**1949**	2060	2476	**2258**
9	1532	1899	**1706**	1843	2218	**2022**	2138	2568	**2343**	2476	2970	**2712**
10	1899	2345	**2110**	2218	2663	**2430**	2568	3080	**2812**	2970	3557	**3250**
11	2345	2889	**2603**	2663	3193	**2916**	3080	3688	**3370**	3557	4255	**3890**
12	2889	3551	**3203**	3193	3822	**3493**	3688	4410	**4033**	4255	5085	**4652**
13	3551	4358	**3934**	3822	4570	**4179**	4410	5269	**4820**	5085	6071	**5556**
14	4358	5342	**4825**	4570	5459	**4995**	5269	6289	**5756**	6071	7242	**6631**
15	5342	6540	**5911**	5459	6515	**5964**	6289	7502	**6869**	7242	8635	**7908**
16	6540	8000	**7233**	6515	7771	**7115**	7502	8944	**8191**	8635	10290	**9426**