Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech

Soha N. Garadat,[a] Ruth Y. Litovsky,[b] and Gongqiang Yu[c]
Waisman Center, University of Wisconsin, 1500 Highland Avenue, Madison, Wisconsin 53705

Fan-Gang Zeng
University of California-Irvine, 364 Med Surge II, Irvine, California 92697

A cochlear implant vocoder was used to evaluate relative contributions of spectral and binaural temporal fine-structure cues to speech intelligibility. In Study I, stimuli were vocoded, and then convolved through head related transfer functions (HRTFs) to remove speech temporal fine structure but preserve the binaural temporal fine-structure cues. In Study II, the order of processing was reversed to remove both speech and binaural temporal fine-structure cues. Speech reception thresholds (SRTs) were measured adaptively in quiet, and with interfering speech, for unprocessed and vocoded speech (16, 8, and 4 frequency bands), under binaural or monaural (right-ear) conditions. Under binaural conditions, as the number of bands decreased, SRTs increased. With decreasing number of frequency bands, greater benefit from spatial separation of target and interferer was observed, especially in the 8-band condition. The present results demonstrate a strong role of the binaural cues in spectrally degraded speech, when the target and interfering speech are more likely to be confused. The nearly normal binaural benefits under present simulation conditions and the lack of order of processing effect further suggest that preservation of binaural cues is likely to improve performance in bilaterally implanted recipients.
© 2009 Acoustical Society of America. [DOI: 10.1121/1.3238242]

I. INTRODUCTION

Cochlear implants (CIs) have been highly successful at providing hearing to profoundly deaf individuals. As a result of continual progress made in this advanced technology, auditory perception in recipients has improved significantly in the past few decades. Today, most CI users are able to perform well in quiet listening situations. However, their performance deteriorates considerably in the presence of background noise and competing speech (Skinner *et al.*, 1994; Muller-Deile *et al.*, 1995; Battmer *et al.*, 1997; Stickney *et al.*, 2004). Numerous studies that focus on performance in unilateral CI users have attempted to identify some of the factors that can account for this deterioration, including the role of speech coding strategies and number of frequency bands (e.g., Gantz *et al.*, 1988; Waltzman *et al.*, 1992; Dorman and Loizou, 1997; Friesen *et al.*, 2001; Stickney *et al.*, 2004). In an alternative approach, bilateral CIs have been provided to a growing number of recipients, with the hope that stimulation of both ears will lead to improved performance in difficult listening situations. Results to date suggest that many bilateral CI users perform better at understanding speech in adverse listening conditions when using two CIs compared with a single CI (e.g., Schleich *et al.*, 2004; Iwaki *et al.*, 2004; Litovsky *et al.*, 2004; 2006; 2009; Tyler *et al.*,

2002). However, despite this improved performance, bilateral CI users are still considerably challenged in dynamic listening situations. In addition, there remains a gap in performance between bilateral CI users and normal hearing listeners (NHLs). The reasons for the gap remain to be understood.

When addressing the deficit in speech intelligibility that is experienced by CI users in presence of noise or competing speech, the complexity of the everyday auditory scene should be considered alongside the possibility that performance is limited by signal processing in the prosthetic devices. In real-world listening, the signal and unwanted "competing" sounds may overlap spectrally and temporally, as well as spatially. Often, there also exists acoustic variability in the auditory environments that may result in increased similarity between a target sound and a competing source, rendering extraction of the target signal rather difficult. The difficulty associated with source segregation under such conditions is often attributed to informational masking (Neff, 1995; Brungart, 2001; Kidd *et al.*, 2002). Although the effects of informational masking can be decreased by introducing dissimilarity between the target and interferer (Durlach *et al.*, 2003), this approach might not be realistically feasible in many listening situations due to the unpredictability of the auditory environments.

Overcoming informational masking can be achieved if listeners have access to a variety of other auditory cues. For example, NHLs exploit spectral (Assmann and Summerfield, 1990, 1994; Bird and Darwin, 1998; Vliegen and Oxenham, 1999) as well as temporal (Tyler *et al.*, 1982; Buss and Flo-

---

rentine, 1985; Bacon *et al.*, 1998; Summers and Molis, 2004) cues to segregate overlapping and competing auditory streams. It is also well known that NHLs can take advantage of spatial cues to segregate speech from competing sounds. This is manifested in as much as a 12 dB improvement in speech reception thresholds (SRTs) when target speech and competitors are spatially separated compared with situations in which they are co-located. This benefit is known as spatial release from masking (SRM), and is an effect that has been studied extensively in NHLs (Freyman *et al.*, 1999; Arbogast *et al.*, 2005; Hawley *et al.*, 1999; 2004; Drullman and Bronkhorst, 2000; Litovsky, 2005).

Spatial cues appear to become more prominent under conditions in which informational masking is relatively large (Kidd *et al.*, 1998; Arbogast *et al.*, 2002). This suggests that spatial hearing plays a crucial role in helping listeners to overcome informational masking. Given the growing number of bilateral CI users, the extent to which spatial cues can be made available to these listeners is a timely question with regard to addressing the gap in performance noted above. The contribution of spatial cues can be explored in these individuals by controlling the inputs to the two ears and comparing performance under bilateral vs unilateral listening modes. A recent study by Loizou *et al.* (2009) has shown that, compared with NHLs, bilateral users are less capable of taking advantage of binaural cues for source segregation, in particular, under conditions of informational masking. This may be due to the fact that CIs have limited spectral resolution (Freisen *et al.*, 2001) and ineffective encoding of F0 information (Stickney *et al.*, 2007). The novelty of the study of Loizou *et al.* (2009) lies in the tighter stimulus control utilized by presenting binaural stimuli directly to the CI users' processors, with spatially appropriate stimuli that were convolved with head related transfer functions (HRTFs).

Limitations in performance of participants in the study of Loizou *et al.* (2009) are of interest here, as they may have arisen from two factors that are highly difficult to control in CI users. One factor is the lack of obligatory coordination between specific pairs of electrodes across the two ears, which would have reduced the extent to which binaural cues could be preserved with fidelity upon reaching the brainstem. A second issue arises when participants whose auditory system has undergone periods of auditory deprivation are tested. Disruptions in the neural processing mechanisms are likely to be present and to contribute to variability in performance within the population of CI users, leading to difficulty in identifying and understanding mechanisms involved in the processes of binaural cues under complex listening conditions.

CI vocoders can offer a powerful tool for investigating effects of CI signal processing independently of other confounds inherent in cochlear implantation. In the current study, a CI vocoder was utilized to investigate whether limitations in performance on spatial auditory tasks that are observed in bilateral CI users are due to the signal processing itself. One of the main issues addressed in the present study is whether CI users are susceptible to informational masking that is borne out of crude signal processing in their prosthetic devices. This issue was investigated by using testing conditions that represent simple but realistic everyday listening situations, yet at the same time in which informational masking in the non-CI conditions may be small. Spondaic target words were presented in the presence of sentences, a combination of target and interferer that deliberately creates relatively easy testing conditions. This approach enabled a systematic examination of a number of critical factors related to speech intelligibility in adverse listening conditions, akin to those that occur with CI processors when a limited number of frequency bands are available. Specifically, speech intelligibility and SRM were evaluated using spectrally degraded stimuli, under binaural and monaural listening conditions. Of a particular interest in this study was the extent to which CI signal processing might impact the role of binaural hearing in providing benefits on measures of speech intelligibility and SRM.

Listening conditions in this study utilized "virtual space" techniques (e.g., Hawley *et al.*, 2004; Loizou *et al.*, 2009) such that all acoustic stimuli were convolved with HRTFs[1] to introduce more realistic, perceptually spatialized and separated target and competing stimuli. By controlling the stage at which stimuli were convolved with HRTFs, effects of signal processing and CI vocoding can be examined independent of the potential loss of binaural cues. Given that one of the future goals in bilateral CIs is to design and provide systems that capture and mimic the way that acoustic information is transmitted in NHLs, the present study could shed light on factors that could potentially enhance vs impair outcomes for effects due to binaural squelch, binaural summation, and the head-shadow effect.

II. STUDY I

In this study, conditions that are more idealized relative to true CI listening were examined by first processing the speech stimuli through the vocoders and subsequently convolving the output through the HRTFs. This approach is akin to a situation in which a NHL is presented with spectrally degraded stimuli through loudspeakers in a room, an approach that has previously been used to investigate effects of spectral degradation on speech perception but without considering effects of binaural hearing and/or spatial cues (e.g., Shannon *et al.*, 2002; Başkent and Shannon, 2007). The current study was designed to preserve as many cues as possible that would be naturally available to listeners for SRM. These include cues that are known to be available to bilateral CI users to some extent, such as head shadow and envelope interaural time differences (van Hoesel, 2004). In addition, we could preserve cues that contribute to spatialized percepts through temporal fine structure, an important binaural cue that is lost in CI processing. While the original speech fine structure in any band has been replaced with a tone, with the idealized order of processing applied here, the new fine structure is filtered through the HRTFs and thus contains the acoustic cues that are used for spatialization.

By preserving the fine-structure cues, it was assumed that there should be sufficient spatial information to acquire the classic release from masking for spectrally degraded

TABLE I. List of cutoff frequencies.

Band	16-band			8-band			4-band		
	Lf	Cf	Hf	Lf	Cf	Hf	Lf	Cf	Hf
1	300	350	400	300	411	521	300	574	848
2	400	460.5	521	521	686	848	848	1445	2042
3	521	595	669	848	1089	1330	2042	3341	4640
4	669	758.5	848	1330	1686	2042	4640	7470	10300
5	848	957	1066	2042	2566	3091			
6	1066	1198	1330	3091	3866	4640			
7	1330	1490.5	1651	4640	5784	6927			
8	1651	1845.5	2042	6927	8613	10300			
9	2042	2279	2516						
10	2516	2803.5	3091						
11	3091	3441	3791						
12	3791	4215.5	4640						
13	4640	5156.5	5673						
14	5673	6300	6927						
15	6927	7688.5	8450						
16	8450	9375	10300						

stimuli; hence, informational masking that is created by signal processing can be evaluated with limited confounds.

A. Material and methods

1. Listeners

Nine NHLs (three male, six female; age range 19–25 years) participated. All subjects were native speakers of English and had pure tone thresholds better than 15 dB hearing loss at octave frequencies ranging from 250–8000 Hz. Participants signed a consent form approved by University of Wisconsin-Madison Institutional Review Board and were paid for their participation. Testing was conducted in five two-hour sessions.

2. Signal processing

Speech signals with a bandwidth between 300 and 10300 Hz[2] were bandpass filtered into 4, or 8, or 16 contiguous frequency bands (see Table I) by sixth-order Butterworth filters using a MATLAB software simulation of CI signal processing strategies (e.g., Shannon *et al.*, 1995). Briefly, the envelope was extracted from each band by full-wave rectification and low-pass filtering at 50 Hz with a second order Butterworth filter. The extracted envelope was used to amplitude modulate a sinusoidal carrier at the band's central frequency followed by the same bandpass filter as the analysis filter to remove spectral splatter. All bands were summed and then convolved with HRTFs (Gardner and Martin, 1994) to create perceptually spatialized and virtually separated target and interferers. For each stimulus (target or interferer), the carrier tones in the right and left ears were in phase. The phase relationship between the carrier tones for target and interferer waveforms was arbitrary. Target and interfering stimuli were then summed and presented to the listeners through headphones (Sennheiser HD 580) under binaural and monaural (right-ear) conditions. In the vocoded speech conditions, target and interfering sentences were processed in the same manner.

3. Stimuli materials and virtual spatial configuration

Target stimuli consisted of a closed set of 25 spondees recorded in our laboratory with a male-talker and presented in quiet as well as in the presence of competing speech. The interferer stimuli were sentences from the Harvard IEEE corpus (Rothauser *et al.*, 1969) recorded with a different male talker than the target. Thirty sentences were strung together, and segments were randomly chosen and played for 6 s per trial. The target words began approximately 1.5 s after the onset of the competing sentence. On each trial the 25 spondees were visually presented to the subjects on a computer monitor. Subjects were instructed to respond by using a mouse button to select the appropriate target word. Feedback was provided following each response by flanking the correct stimuli on the computer screen in front of the listener.

Given that all stimuli were convolved through HRTFs to enable virtual spatial separation of target and interfering speech, data were collected for each subject using the following location combinations: (1) quiet: target at 0° azimuth and no interferer, (2) front: target and interferer both at 0° azimuth, (3) right: target at 0° azimuth and interferer at 90° azimuth, and (4) left: target at 0° azimuth and interferer at −90° azimuth.

4. Stimulus levels and threshold estimation

All stimuli were calibrated using an artificial ear coupler (AEC101 IEC 318, Larson Davis). Calibration was conducted after stimuli were convolved through the HRTFs. Stimulus levels were set based on calibration for token sentences from the speech corpus presented from the simulated front condition. The level of the interferer was fixed at 60 dB sound pressure level (SPL); thus for the front condition, interferer levels were set to 60 dB SPL in each ear. For nonfront conditions, interferer levels were 60 dB SPL for the ear ipsilateral to the interferers, and change in signal-to-noise ratio (SNR) represents the change in target level relative to interferer level at the ipsilateral ear. The level varied natu-
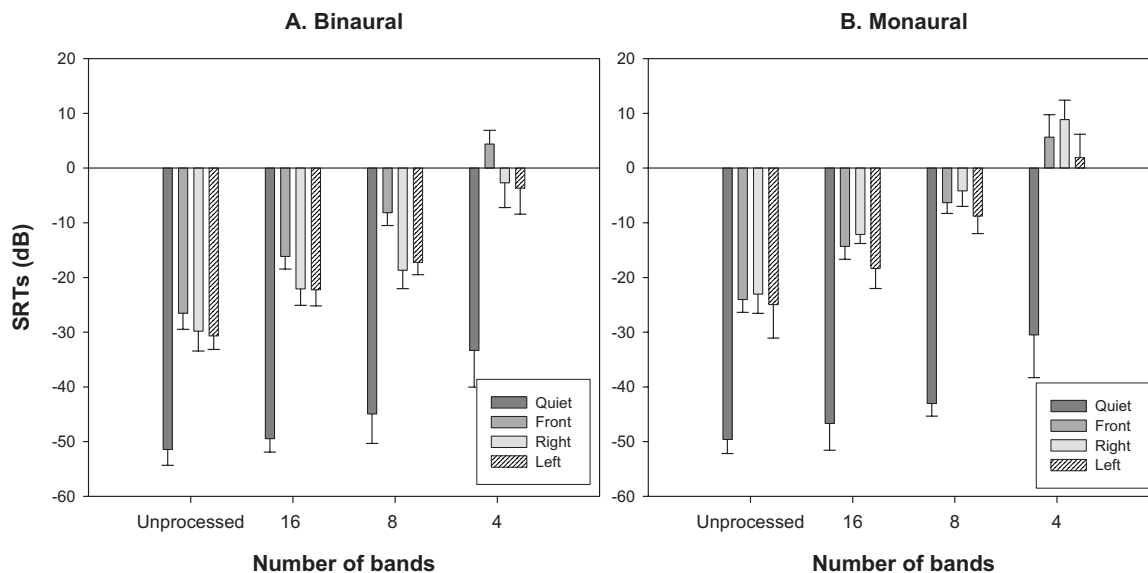
FIG. 1. Average SRTs (+1 SD) in dB are shown for all spatial conditions tested relative to interferer level (60 dB). Left panel represent binaural data and right panel represents monaural (right-ear) data. Within each panel, SRTs are plotted for the different interferer conditions as a function of frequency band conditions.

rally with the HRTF at the contralateral side to create a head-shadow effect. The level of the target was varied adaptively using an algorithm that targets the 79.4 point on the psycho-metric function (Levitt, 1971). The target level was initially 65 dB SPL and was decremented by 8 dB following each correct response. After the first incorrect response, a modified adaptive 3-down/1-up algorithm was used in which the step size was halved after each reversal, with the minimum step size set to 2 dB. If the same step size was used twice in a row in the same direction, the next step size was doubled in value. Testing was terminated following eight reversals.

SRTs were estimated from the adaptive tracks by using a constrained maximum-likelihood method of parameter estimation (MLE), which has been described by Wichmann and Hill (2001a, 2001b). Based on this method, data from each experimental run for each participant were fitted to a logistic function and thresholds were calculated by taking the level of the target at a specific probability level. This approach has been shown to yield comparable results to the well-known approach in which SRT is defined as the average of levels at which reversals occur. However, the MLE approach has the advantage, with this stimulus corpus and adaptive tracking method, of producing smaller group variance (Litovsky, 2005).

5. Procedure and training

Data collection was conducted in blocks with the number of frequency bands (unprocessed, 16, 8, and 4) fixed. To ensure familiarity with the task, participants completed the unprocessed conditions first. Subsequently, all other blocks (16, 8, and 4) were presented in a random order generated with a different seed for each subject. Within a block, all other conditions were randomized. Prior to each testing block with vocoded stimuli, subjects received additional listening exposures to familiarize them with the quality of the speech they were about to hear in the upcoming blocked condition. During these vocoded exposure periods, four

SRTs (two in quiet, and two with front interferer) were collected; these SRTs were excluded from the main analyses. After data completion, all conditions were re-randomized and a second set of data was collected based on the assumption that with more exposure to vocoded speech, listeners' performance would be more stable. This second set of data was used in the analyses and reported in this paper. However, statistical analyses comparing the two sets of data revealed that learning effects occurred only in the 4-band conditions.

B. Results

1. SRTs

SRTs were obtained using the MLE procedure described above and were normalized relative to interferer level. These data are displayed in Fig. 1 as a function of number of frequency bands under binaural and monaural conditions. Two-way repeated measure analyses of variance (ANOVAS) on SRTs were conducted for listening mode (binaural and monaural) and number of frequency bands (unprocessed, 16, 8, and 4); these analyses were conducted separately for each interferer condition (quiet, front, right, and left). A significant main effect of listening mode was found such that binaural SRTs were lower than monaural SRTs in all conditions; quiet $[F(1,8)=9.874, p<0.05]$, front $[F(1,8)=14.752, p<0.01]$, right $[F(1,8)=77.763, p<0.0001]$, and left $[F(1,8)=42.205, p<0.0001]$.

Significant main effects of number of bands were also observed for all conditions; quiet $[F(3,8)=53.243, p<0.0001]$, front $[F(3,8)=571.718, p<0.0001]$, right $[F(3,8)=298.448, p<0.0001]$, and left $[F(3,8)=2364.374, p<0.0001]$ SRTs. The lack of interactions with listening mode suggests that the effect of number of bands applies to binaural and monaural listening modes. *Post-hoc* Scheffe's tests revealed that, in quiet, SRTs for the unprocessed condition were comparable to those in the 16-band condition but lower (better performance) than those in the 8- and 4-band
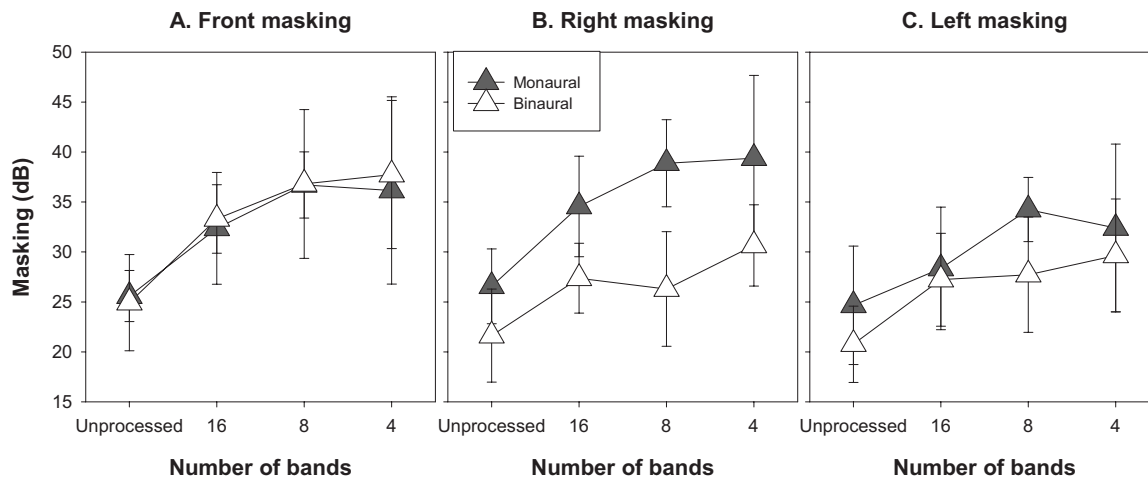
FIG. 2. Masking values (±1 SD) are plotted for the monaural (filled symbols) and binaural (unfilled symbols) as a function of the different frequency bands and shown for the front (panel A), right (panel B), and left (panel C) interferer conditions.

conditions ($p < 0.001$). However, in the presence of a speech interferer, the improvement in SRTs continued with further increases in number of frequency bands ($p < 0.0001$). Specifically, SRTs for the unprocessed conditions were lower than those in the 16-, 8-, and 4-band conditions. In addition, SRTs for the 16-band conditions were lower than those in the 8- and 4-band conditions, with lower SRTs in the 8-band than those in the 4-band conditions.

2. Masking

In this study, masking was defined as the absolute change in SRTs when interferer stimuli were present compared with the quiet condition. Masking values for the front, right, and left, respectively, were computed as ($SRT_{front} - SRT_{quiet}$), ($SRT_{right} - SRT_{quiet}$), and ($SRT_{left} - SRT_{quiet}$). These masking values, shown in Fig. 2, were subjected to two-way repeated measures ANOVAs for listening mode (binaural, monaural) and number of frequency bands (unprocessed, 16, 8, and 4) as described above for SRTs.

A main effect of listening mode was not found for front masking, indicating comparable amount of masking for the binaural and monaural conditions. However, a main effect of listening mode was obtained for right [$F(1,8) = 88.280$, $p < 0.0001$] and left [$F(1,8) = 13.346$, $p < 0.01$] maskings, such that the amount of masking was greater in the monaural than in the binaural conditions. These results suggest that binaural listening provides mechanisms for reduction in masking that are not available in the single-ear listening mode. In addition, a main effect of number of frequency bands was obtained for front [$F(3,8) = 20.502$, $p < 0.0001$], right [$F(3,8) = 14.511$, $p < 0.0001$], and left [$F(3,8) = 6.944$, $p < 0.005$] masking. Scheffe's *post-hoc* analyses revealed that the amount of masking was significantly smaller in the unprocessed condition than the 16-, 8-, and 4-band conditions ($p < 0.01$), suggesting that spectrally degraded speech is more susceptible to masking than natural speech. Finally, differences in masking were not statistically significant across the three spectrally degraded conditions; this finding occurred in all three spatial masker configurations: front, right, and left.

3. Spatial release from masking

Figure 3 summarizes the findings for SRM which was computed for two spatial configurations: right ($Masking_{front} - Masking_{right}$) and left ($Masking_{front} - Masking_{left}$). These data were subjected to two-way repeated measures ANOVAs for listening mode (binaural and monaural), and number of frequency bands (unprocessed, 16, 8, and 4); separate analyses were conducted for the right and left SRM values. A main effect of listening mode suggested that SRM was larger in the binaural than in the monaural conditions (right-ear) for both right [$F(1,8) = 51.317$, $p < 0.0001$] and left [$F(1,8) = 24.700$, $p < 0.005$] interferer configurations.

A main effect of number of frequency bands was not found for the right spatial configuration but was obtained for the left configuration [$F(1,8) = 3.424$, $p < 0.05$]. Scheffe's *post-hoc* analysis revealed that, in comparison with the unprocessed condition, the amount of SRM was greater for spectrally degraded conditions: 16-band ($p < 0.05$), 8-band ($p < 0.005$), and 4-band ($p < 0.001$). Differences in SRM were not statistically significant across the different spectrally degraded conditions.

A significant interaction of listening mode × number of frequency bands [$F(3,24) = 5.116$, $p < 0.01$] was found for the right spatial configuration. Scheffe's *post-hoc* analysis showed that, under monaural listening, SRM was statistically comparable for the different spectral conditions. In the binaural conditions, SRM for the unprocessed condition was smaller than that in the 8 ($p < 0.0001$) and 4 ($p < 0.005$) bands, and comparable to the 16-band condition. In addition, SRM was greater in the 8-band condition compared with 16-band ($p < 0.001$) and 4-band ($p < 0.05$) conditions. SRM for the 16- and 4-band conditions was comparable.

4. Bilateral effects

Further analyses were conducted in order to facilitate comparisons with studies in bilateral CI users. The variables of interest were head shadow, binaural squelch, and binaural summation (e.g., Muller *et al.*, 2002; Tyler *et al.*, 2003; Schleich *et al.*, 2004; Litovsky *et al.*, 2006). Head shadow in the monaural (right-ear) condition was defined as the advan-
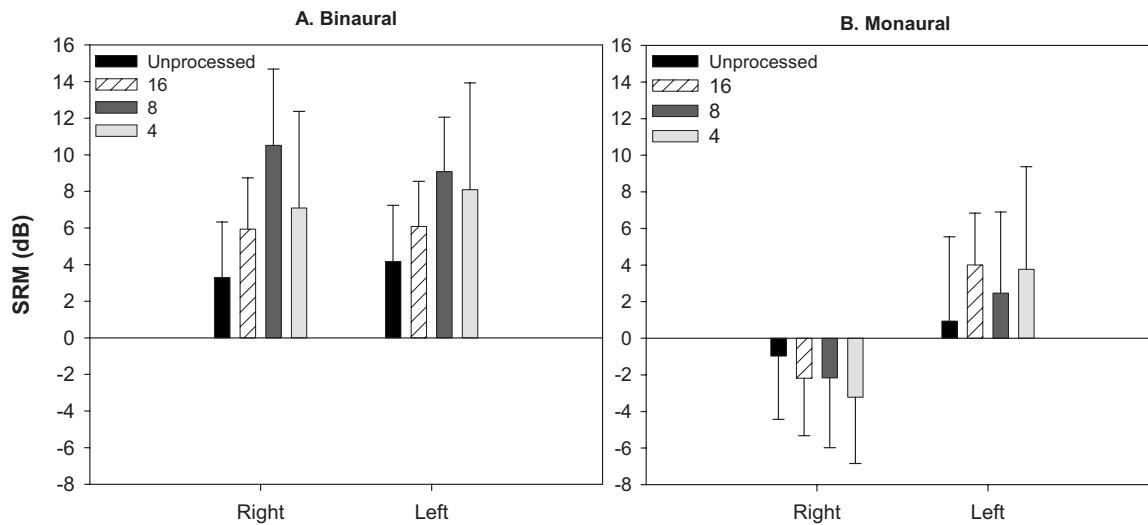
FIG. 3. Average amounts of SRM ($+1$ SD) are shown for the binaural (A) and monaural (B) listening modes. Within each panel, SRM is compared for the different frequency bands, as a function of interferer location.

tage (reduction in SRT) obtained when the interferer was contralateral versus ipsilateral to the functional ear. It was thus computed as $[\text{SRT(monaural)}_{\text{right}} - \text{SRT(monaural)}_{\text{left}}]$. Binaural squelch describes the advantage obtained as a result of spatial separation between target stimuli and interfering stimuli. These values were obtained for each subject as $[(\text{monaural})_{\text{left}} - (\text{binaural})_{\text{left}}]$. Binaural summation, an advantage that can result from listening to identical stimuli with two ears, was calculated for each subject in two ways; first, by comparing SRTs in the conditions with no interferer $[(\text{monaural})_{\text{quiet}} - (\text{binaural})_{\text{quiet}}]$, and second, by comparing SRTs in the conditions with interferer in the front $[(\text{monaural})_{\text{front}} - (\text{binaural})_{\text{front}}]$.

For each of the four effects listed above, a one-way repeated measures ANOVA was conducted in which the variable of interest was number of frequency bands, including the unprocessed conditions. There were no statistically significant findings for any of the analyses, suggesting that the effects were not dependent on spectral resolution. Data for the vocoded speech were pooled across frequency band conditions and plotted as group means ($+1$ SD) in Fig. 4. Average values were 5.3 dB for head shadow, 5.9 dB for squelch, and 2.5 and 1.6 dB for binaural summation in quiet and in the presence of the front interferer, respectively. In Fig. 4, the bilateral effects are also plotted for the unprocessed conditions for comparison purposes. Given the lack of a significant main effect of number of spectral bands, the unprocessed and processed conditions were grouped for each condition and were subjected to one-sample t-tests (e.g., Schleich et $al.$, 2004). Results revealed that head shadow, squelch, and summation in quiet and in the presence of front interferer were each significantly different than zero ($p < 0.0001$, $p < 0.0001$, $p < 0.01$, $p < 0.01$, respectively).

III. STUDY II

Given the increased robustness of SRM found in the first study when using vocoded speech, the next question addressed here was whether this effect can also be observed in a scenario that more realistically simulates true bilateral CI

listening. Therefore, speech stimuli were first convolved through the HRTFs, as would occur in a real world to a person using CIs in the free field; the resulting stimuli were subsequently processed through the vocoder. This study, with the reversed order of signal processing, enabled us to examine whether the directionally dependent cues that are available in the HRTFs are immune to, or distorted by, the CI signal processing in ways that affect benefits from binaural hearing for the spatially separated conditions. Testing was conducted with a second group of listeners, and data from the two studies will be henceforth for conditions that are thought to involve the use of binaural directional cues for source segregation.

A. Material and methods

1. Listeners

Nine NHLs (two male, five female; age range 19–25 years) participated. All subjects were native speakers of Eng-
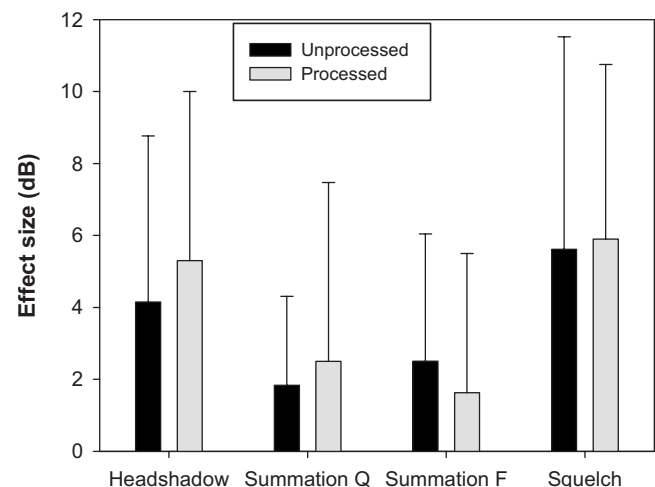

FIG. 4. Group means ($+1$ SD) are shown for head shadow, binaural squelch, and binaural summation as estimated from the quiet condition and binaural summation as estimated from the condition with interferer in front. Data are plotted for the unprocessed conditions (dark bars) and processed conditions (light bars).

J. Acoust. Soc. Am., Vol. 126, No. 5, November 2009

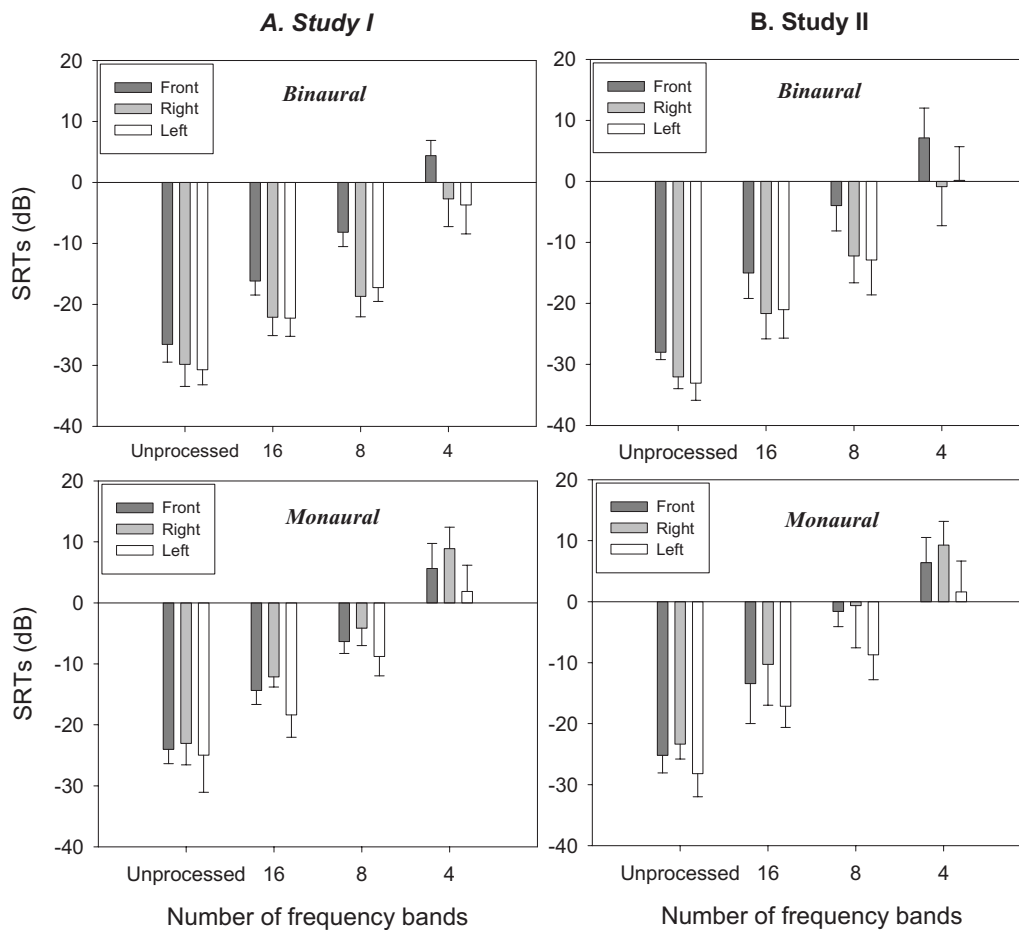Garadat *et al.*: Binaural vocoded speech   2527

FIG. 5. Average SRTs (+1 SD) in dB are shown for all spatial conditions tested relative to interferer level and displayed for Study I (left column) and Study II (right column). Top panels represent binaural data and bottom panels represent monaural data.

lish and had pure-tone thresholds better than 15 dB HL at octave frequencies ranging from 250–8000 Hz. Participants signed a consent form approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board and were paid for their participation.

2. Signal processing

Similar to Study I, all stimuli were bandpass filtered into 4, or 8, or 16 contiguous frequency bands by sixth-order Butterworth filters with equal bandwidths on a logarithmic scale from 300 to 10300 Hz. Target and interfering waveforms were convolved through HRTFs to create directionally dependent stimuli. Similar to that in Study I, the carrier tones for each of the target and interferer were in phase across the two ears but were not systematically in phase between target and interferer. Stimuli were then digitally mixed and subsequently passed through the CI simulation filters described in detail in Study I. A Tucker-Davis Technologies (TDT) RP2 array processor was used to attenuate the stimuli and to subsequently present them to listeners via TDT system III-hardware (HB7) and headphones (Sennheiser HD 580).

3. Stimuli and procedure

Speech stimuli and the testing apparatus were identical to those described in detail in Study I. Participants were tested on a total of 24 conditions consisting of all combina-

tions of three interferer locations (front, right and left), two listening modes (binaural and monaural/right-ear), and four spectral conditions (unprocessed, 16-, 8- and 4-band conditions). A similar training procedure to that described in the first study was used in this study in order to account for learning effects that might occur when listening to vocoded speech. Given that learning effects only occurred for the 4-band conditions in Study I, only that condition was repeated and included in the data analyses conducted in Study II. Data collection was completed in three two-hour sessions per participant.

B. Results

1. SRTs

SRTs obtained in the two studies are shown in Fig. 5, where results based on the two different orders of processing can be compared. Data were subjected to repeated measures ANOVAs with number of frequency bands (unprocessed, 16, 8, and 4) and interferer conditions (front, right, and left) as the within subject variables and processing order (Study I and Study II) as the between-subject variable. This analysis was conducted separately for binaural and monaural SRTs, in order to directly compare the processing order effects within each listening mode.

There was no significant main effect of processing order for either binaural or monaural SRTs. As was seen when data
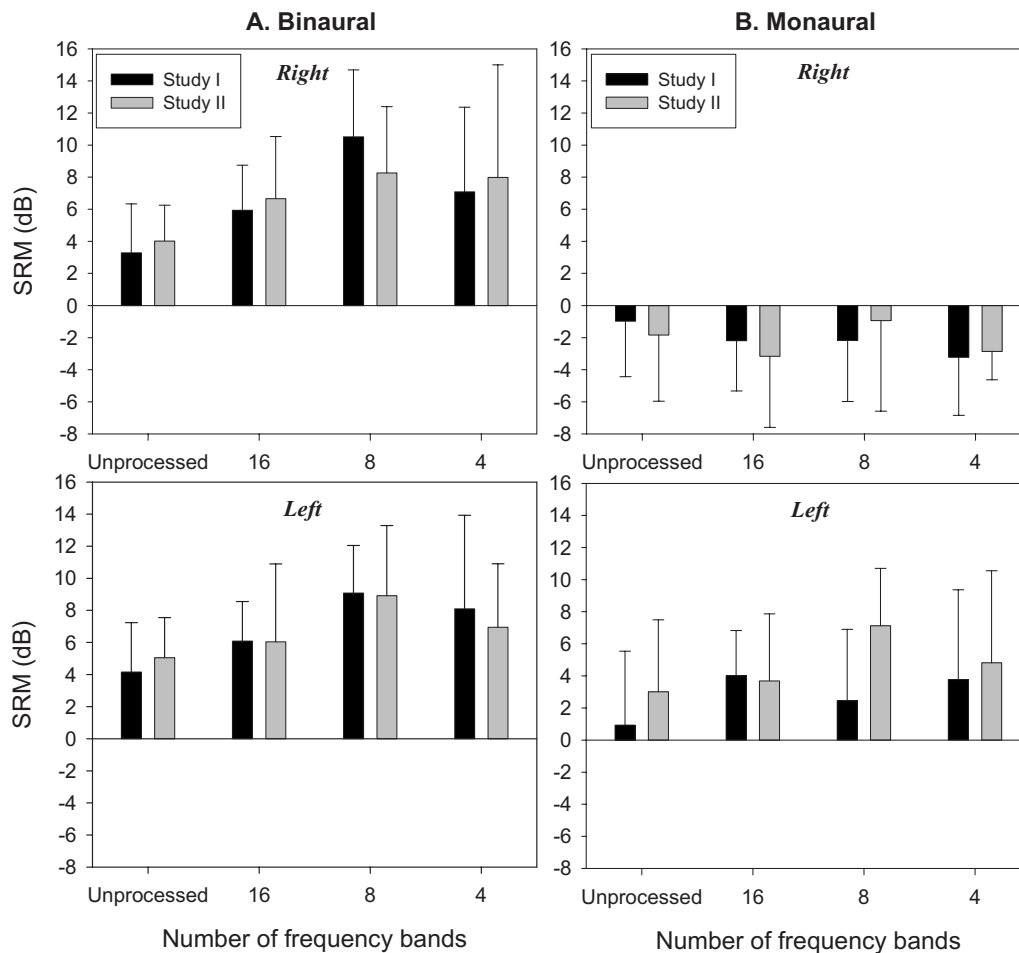
FIG. 6. Average amounts of SRM (+1 SD) are shown for the binaural (left column) and monaural (right column) listening modes. Within each panel, SRM is compared for the two studies across the different frequency bands. Each panel represents different interferer condition.

from Study I were analyzed, a significant main effect of number of frequency bands was found for both binaural [$F(3,16)=341.404$, $p<0.0001$] and monaural [$F(3,16)=346.285$, $p<0.0001$] conditions. *Post-hoc* tests were conducted for between-subject effects; all *t*-tests reported in these experiments were corrected for multiple comparisons using the Holm–Bonferroni procedure. Results from the *post-hoc* analyses (significance levels, $p<0.0001$) showed that SRTs for the unprocessed conditions were significantly lower than those in the 16-, 8- and 4-band conditions. In addition, SRTs with 16-bands were lower than those in the 8- and 4-band conditions; SRTs with 8-bands were lower than with 4-bands.

There was a significant main effect of interferer condition for both binaural [$F(2,16)=96.270$, $p<0.0001$] and monaural [$F(2,16)=86.737$, $p<0.0001$] SRTs. *Post-hoc* analyses (significance levels, $p<0.0001$) showed that binaural SRTs were higher when the interferer was placed in the front than that in the right or left configurations, the latter two conditions resulted in comparable SRTs. On the other hand, *post-hoc* analyses for the monaural right-ear conditions showed that when the interferer was located in the front, SRTs were higher than when the interferer was on the left but lower than when the interferer was on the right; monaural SRTs for the right were higher than SRTs for the left configuration. Note that the lower monaural SRTs when the in-

terferer was in the front relative to the right occurred despite the calibration procedure that equated the interferer SPLs in the two conditions. The difference in SRTs could have been due to differences in the interferer's spectral profile resulting from frequency-dependent head shadow effects.

A significant interaction was observed for number of frequency bands and processing order in the binaural conditions [$F(3,48)=4.504$, $p<0.01$] but not in the monaural right-ear conditions. *Post-hoc* tests revealed that for the 8-band conditions, binaural SRTs were lower in Study I compared with Study II. This suggests that when vocoding is conducted on stimuli that have already been filtered with HRTFs, there are likely to be adverse affects on performance.

2. SRM

The approach for deriving SRM values was similar to that used in Study I, and values from the two studies are compared in Fig. 6. Repeated measures ANOVAs, with number of frequency bands and interferer condition as the within-subject variables and processing order as a between-subject variable, were conducted separately for binaural and monaural conditions. There was no significant effect of processing order on SRM. A main effect of number of frequency bands was found for the binaural [$F(3,16)=7.493$, $p<0.0001$] but
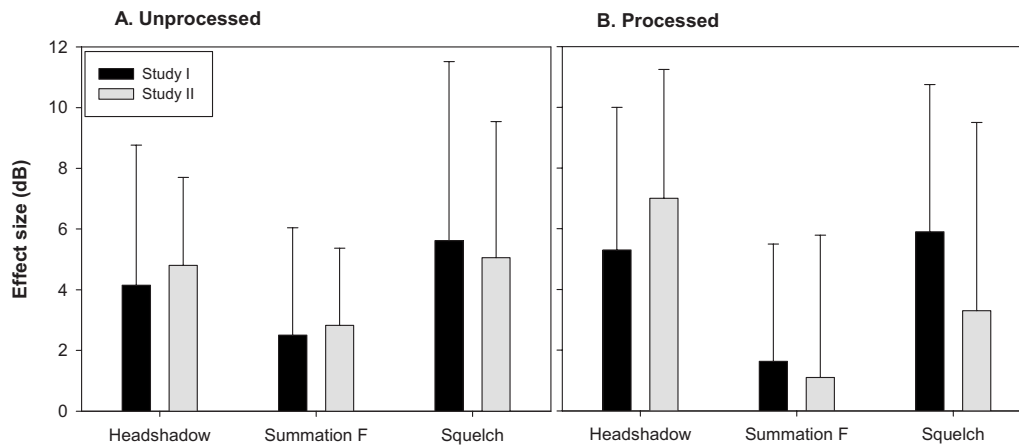
FIG. 7. Group means (+1 SD) are shown for head shadow, binaural summation in front, and binaural squelch. Results from Study I and Study II are compared for the unprocessed conditions (panel A) and the processed conditions (Panel B).

not for the monaural conditions. *Post-hoc t*-tests for this effect revealed that the amount of SRM for the unprocessed conditions was smaller than for the 16- ($p < 0.01$), 8- ($p < 0.0001$), and 4-band ($p < 0.005$) conditions. Additionally, greater SRM was found for the 8-band than the 16-band condition ($p < 0.0001$). SRM for the 4-band condition was comparable to SRM in the 16- and 8-band conditions.

A significant main effect of interferer condition was found for the monaural mode [$F(1, 16) = 150.412$, $p < 0.0001$], where SRM for the left configuration was larger than that for the right configuration due to the presence of head shadow; there was no main effect of interferer condition for the binaural mode.

3. Bilateral effects

As in Study I, values for head shadow, binaural squelch, and binaural summation were calculated. A one-way repeated measures ANOVA was conducted on data from Study II for each effect, with the variable of interest being number of frequency bands (unprocessed, 16, 8, and 4). Similar to findings in Study I, none of the effects showed a dependence on number of frequency bands. However, to evaluate the effect of processing order on head shadow, summation, and squelch across the two studies, only data within the processed conditions were examined. Thus, data from each study within the processed conditions were pooled to yield an overall measure for each of the three effects. These pooled data are compared in Fig. 7 (panel B). A one-way ANOVA with processing order as the independent factor did not reveal statistically significant differences between the two studies. Differences in the squelch effect approached significance ($p = 0.06$). The effect sizes for the unprocessed conditions (common across the two studies) are shown in Fig. 7 (panel A). The differences between the two subject groups were small and not statistically significant.

IV. DISCUSSION

A. Effect of CI vocoding on speech intelligibility

The current work shows that the unprocessed speech yielded the lowest average thresholds, in particular, in the

conditions with interfering speech. This result is consistent with previous studies demonstrating that NHLs are less challenged by speech recognition in noise than subjects with hearing impairments (e.g., Chung and Mack, 1979; Dubno *et al.*, 1989; Pekkarinen *et al.*, 1990) or CI users (Nelson and Jin, 2004; Stickney *et al.*, 2004; Fu and Nogaki, 2005). This also confirms the well-known robustness of natural speech as a communication medium over spectrally degraded speech due to the listeners having access to spectral and temporal cues that have been shown to be important for speech recognition in noise (Assmann and Summerfield, 1990; Leek and Summers, 1993; Eisenberg *et al.*, 1995; Vliegen and Oxenham, 1999; Summers and Molis, 2004).

In the vocoder conditions speech recognition improved as the number of frequency bands was increased, which is consistent with previous reports using vocoded speech (Dorman and Loizou, 1997; Loizou *et al.*, 1999; Dorman *et al.*, 2000; Friesen *et al.*, 2001; Qin and Oxenham, 2003; Stickney *et al.*, 2004). In contrast with previous work (e.g., Dorman *et al.*, 1998), the present study demonstrated that 16 frequency bands are not sufficient to reach comparable performance to that with natural speech and that a larger number of bands is needed to support the dynamics of listening in complex auditory environments. In general, the elevated SRTs observed with a relatively large number of frequency bands underscore the importance of attempting to recapture and preserve information that is currently missing in today's clinical processors in order to enhance performance. Current CI processors encode temporal envelope cues to extract speech and discard fine-structure information, the latter perhaps being particularly important for listening to speech in the presence of interfering sounds (Rosen, 1992; Nelson *et al.*, 2003; Nie *et al.*, 2005; Fu and Nogaki, 2005).

B. Effect of CI vocoding on masking

Results showed that amount of masking increased considerably when spectrally degraded signals were used relative to the unprocessed conditions. However, amount of masking was comparable for the three frequency band conditions (16, 8, and 4) suggesting that once speech is spectrally degraded, the susceptibility to masking is relatively

high. In line with previous studies, current results underscore the limitation of CI vocoders in reproducing the fine structure information which is known to be important in speech recognition, particularly in the presence of temporally overlapping speech sounds (e.g., Smith *et al.*, 2002; Stickney *et al.*, 2005, 2007; Rubinstein and Hong, 2003; Wilson *et al.*, 2003, 2005). These results further suggest that when fine-structure information is reduced by vocoding, increasing the number of bands might not be the most constructive solution to the problem.

A further noteworthy finding is the relationship between masking and listening mode across the different spatial configurations. Overall, current results are consistent with reports in NHLs which showed that binaural hearing provides advantages over monaural hearing, in particular, in adverse listening environments (e.g., MacKeith and Coles, 1971; Bronkhorst and Plomp, 1988; Arsenault and Punch, 1999; Hawley *et al.*, 2004). Here we observed less masking under binaural than monaural conditions (see Fig. 2), with a dependence on the interferer location. When the interferer was placed in front, i.e., in absence of spatial separation between the target and interferer, binaural advantages were not observed.

C. Effect of CI vocoding on spatial release from masking

SRM increased as the number of frequency bands was reduced from unprocessed to 16- and then 8-bands. These results may be related to the finding in the normal-hearing literature that saliency of spatial cues increases as the listening environment becomes more difficult and complex. One such example is when the interferers carry linguistic content and context rather than consisting of noise or when the number of interfering talkers increases from 1 to 3 (e.g. Hawley *et al.*, 2004). Similarly, advantages of spatial cues become particularly robust when the target speech and interferers are comprised of identical or highly similar talkers, that is, when informational masking is present (e.g., Freyman *et al.*, 1999).

In the current studies, target-interferer similarity would have also been heightened when the number of frequency bands was reduced, leading to increase in informational masking. Thus, the increase in SRM in the vocoded conditions compared with the unprocessed conditions can be reasonably interpreted within the context of informational masking. What cannot be readily explained within that context is the fact that the increase in SRM was non-monotonic, declining for the 4-band conditions compared with the 8-band conditions. Perhaps, one interpretation is that the reduced SRM in the 4-band condition might have resulted from non-linearity in informational masking under conditions of degraded speech. As can be seen in Fig. 1, targets had to be presented at positive SNR values in order for listeners to achieve optimal performance, which reflects the difficulty of the task in that condition. It appears that when target recognition requires SNRs above 0 dB, listeners' susceptibility to informational masking is reduced, thereby minimizing SRM (Arbogast *et al.*, 2005; Freyman *et al.*, 2008).

The size of the bilateral effects measured with these simulations should be applied to actual CI users only with caution. The preservation and enhancement of SRM in the 8- and 16-band conditions was most likely due to availability of coordinated inputs across the two ears, such as identical frequency inputs at specific stimulation bands. As discussed below, simply smearing the spectral cues by processing HRTFs through CI simulation filters, as done in Study II, did not appear to have an effect on SRM. Thus, other aspects of signal processing in CIs need to be considered in order to understand how the gap between NHLs and CI users might be bridged.

D. Spatial effects related to the bilateral CI literature

Results from this study were analyzed in comparable ways to analyses that are typically conducted when bilateral CI users are tested. Three effects that are thought to reflect advantages stemming from having bilateral hearing were found to be significant: head shadow, binaural squelch, and binaural summation. The effects occurred regardless of the number of frequency bands in the signal, suggesting that benefits arising from bilateral hearing are not intimately dependent on frequency resolution.

In terms of benefits that are known to occur when two ears are activated, the head shadow is one of the largest. This effect occurs when the head of a listener acts as an acoustic barrier ("shadow") such that masker levels are attenuated at the ear contralateral to that of the masker location, improving SNRs at the contralateral ear. In a cocktail party environment, the benefit would arise when the target is near the "better" ear and the masker is contralateral to that ear (compared with ipsilateral to that ear). In the present study, when averaged across conditions, the head-shadow effect was approximately 5 dB, which is similar in magnitude to what has been found in persons with bilateral CIs (Gantz *et al.*, 2002; Muller *et al.*, 2002; Tyler *et al.*, 2002; Van Hoesel *et al.*, 2002; Schleich *et al.*, 2004; Litovsky *et al.*, 2006), and somewhat smaller than the 9–11 dB reported in NHLs (Arsenault and Punch, 1999; Bronkhorst and Plomp, 1988). The extent to which these differences are related to the choice of speech stimuli used in this study cannot be determined based on the current results; this topic should certainly be addressed in future studies. However, the similarity of the effect size between bilateral CI listeners and NHLs tested using binaural vocoded speech suggests that the vocoding reduced the effect size in a manner that effectively mimicked perceptual effects that arise when head-shadow cues are available to CI users.

Another effect measured in this study is the squelch effect, quantified for each subject as $[(\text{monaural})_{\text{left}}-(\text{binaural})_{\text{left}}]$. This effect, thought to be helpful for source segregation when sounds are spatially separated, requires that the auditory system make use of the differences in signals arriving at the two ears. In the present studies participants showed a squelch effect averaging approximately 6 dB in Study I and 3.6 dB in Study II, which is highly similar to the range of effect sizes (3–7 dB) reported in NHLs with undegraded stimuli (Levitt and Rabiner, 1967; Arsenault and Punch, 1999; Bronkhorst and Plomp, 1988; Hawley *et al.*,

J. Acoust. Soc. Am., Vol. 126, No. 5, November 2009

Garadat *et al.*: Binaural vocoded speech   2531

2004). Our finding, that squelch occurs with either processing order, suggests that the smaller number of spectral bands available to CI users is not likely to be the limiting factor for eliciting the squelch effect.

If binaural temporal fine-structure cues are important for squelch, then one might have expected that the removal of those cues, as was done in Study II, would reduce or obliterate the effect. Statistical comparison of squelch between the two studies approached, but did not reach, significance, suggesting that the effect of processing order was weak or absent, or that variability in the data obscured the effect. However, the trend for squelch to be smaller in Study II, where the binaural cues in the stimuli were smeared by the vocoder, may help to understand the very small (Muller *et al.*, 2002; Schön *et al.*, 2002; Schleich *et al.*, 2004; Litovsky *et al.*, 2006) or absent (Gantz *et al.*, 2002; Van Hoesel and Tyler, 2003; Van Hoesel *et al.*, 2002) squelch seen in bilateral CI users. Clearly, lack of significant statistical effects here temper this conclusion and suggest that further work is needed in this area. Another, perhaps more likely, explanation for small squelch effects in bilateral CI users is the lack of coordinated inputs across the two ears and minimal or absent interaural timing cues. It has been reported that interaural level cues are the predominant cues for CI users (van Hoesel, 2004); this suggests that future advances in speech processors should include mechanism for restoring interaural timing cues.

In the third effect, known as binaural summation or redundancy, the signals reaching the two ears are very similar or identical, as the auditory stimulus is presented from 0° (front). The effect sizes in our studies were similar to those obtained in studies with hearing impaired listeners (Bronkhorst and Plomp, 1989) and were unaffected by order of processing. The summation effect is also an effect that is found in some but not all bilateral CI users (Schleich *et al.*, 2004; Litovsky *et al.*, 2006). In the study of Litovsky *et al.* (2006) 15/34 subjects (44%) demonstrated this effect when compared with either of the two ears alone, while 17/34 (50%) subjects had no effect, and 2/34 subjects (6%) showed a decrement in the bilateral condition rather than improved performance. Thus, like squelch, the summation effect might be a good example of a benefit that comes from having inputs present at both ears and depends on highly symmetrical (or identical) hearing integrity, and possibly also coordinated timing of inputs between the ears. These are factors that are known to be problematic in bilateral CI users, but that were clearly surmounted by the stimulation approaches used here.

E. Effect of simulation order on utility of spatial information

By varying the order of stimulus processing in the two studies, we were able to examine effects of two aspects of CI simulation; removal of speech fine-structure cues (Study I) and subsequently also removal of binaural temporal fine-structure cues (Study II). Because the stimulation to the two ears was coordinated, as occurs in NHLs, interaural timing and level differences in the envelope remained unperturbed,

which renders these good candidates for cues used by the listeners. A main effect of processing order was not found for SRTs. However, a significant interaction of number of frequency bands by processing order revealed that for the 8-band conditions binaural SRTs obtained from Study I were lower than those obtained in Study II (in which HRTFs were processed through the CI vocoder). In the 16-band condition, i.e., the condition with substantially richer spectral information, there was no effect of processing order. Together with the lack of interaction effects in the monaural conditions, these results suggest that under binaural conditions, spectral information that is available in the HRTFs might be useful for speech recognition in adverse listening environments. The underlying mechanisms responsible for this are likely to be ones that produce redundancy or summation of information that is required for speech perception.

The consequence of higher SNRs being required for listeners to achieve optimal performance is consistent with what is known about CI users and the challenges that they face in noisy situations. This finding has implications for true CI users who are shown to utilize up to eight independent channels of spectral information (Friesen *et al.*, 2001). Additionally, it is important to note that differences in SRTs across the two studies were not observed for the 4-band conditions. This is most likely due to the severe degradation of spectral information available in the 4-band conditions regardless of the presence of HRTF cues. On the other hand, the extent to which inter-subject variability precluded these differences cannot be ruled out. The lack of statistically significant effects in the monaural conditions suggests that directionally dependent cues in the HRTFs, which were available in Study I but were most likely eliminated in Study II by processing HRTFs through the vocoder, may not have served an important purpose for the effects studied here.

Of further interest is whether the amount of SRM is dependent on the preservation of directional cues that are available in the HRTFs. SRM was not significantly different across the two processing approaches. In Fig. 6, however, one can see a trend for smaller SRM in the right-masker configuration in Study II than in Study I. It is noteworthy that the data have an inherent level of inter-subject variability. Individual differences are a hallmark of some noted perceptual phenomena such as informational masking (e.g., Oh and Lutfi, 1998; Durlach *et al.*, 2003), which, as discussed above, seems to have arisen in the vocoded stimuli used in the current experiments. The choice of using a speech interferer in these experiments was based on the desire to utilize a stimulus paradigm that more realistically represents real-world listening situations encountered by CI users in everyday situations. The extent to which the inter-subject variability might have been so large as to obscure effects due to signal processing order or other manipulations conducted here might be investigated in future studies perhaps with fewer conditions but a much larger *N* size for participants. Alternatively, one might tackle this issue by using stimuli that are constructed so as to maximize energetic, rather than informational, masking.

Regardless of the variability, the spatial effects observed here were either comparable to or greater than those reported

in bilateral CI users. These results suggest that directional cues that exist at the output of the vocoder, even after the HRTFs have been processed, are sufficient for the occurrence of spatially dependent bilateral benefits. The present study demonstrated that, by preserving binaurally coordinated stimulation in the envelopes of the signals alone, benefits from bilateral CIs could be substantial, regardless of the amount of spectral degradation in the speech signal. This suggests that, while preservation of fine-structure in the signal may offer other benefits, envelope-based binaural differences are likely to offer a substantial portion of the advantage for listening in complex environments. The extent to which fine-structure vs envelope cues might each contribute to improved performance is obviously an important topic for further investigation.

V. SUMMARY AND CONCLUSION

The current study examined the effect of spectral resolution on speech intelligibility and SRM in binaural and monaural listening conditions in NHLs. The order of signal processing of the vocoded speech and the directionally dependent HRTFs had little effect on the results. The findings are consistent with the notion that increased spectral information is important for improved speech intelligibility. However, the benefit of spatial cues was most pronounced under conditions of spectral degradation of speech, when the target and interfering speech are more likely to be confused and thus when informational masking is likely to be larger. Benefits from binaural hearing that are rarely observed in true bilateral CI users were seen here. This suggests that for the effects studied in these experiments, preservation of binaural coordination between the two ears may be important to support bilateral implantation.

ACKNOWLEDGMENTS

[1]HRTFs from NHLs such as those used here are typically measured in the ear canal (Blauert, 1997), and contain high-frequency cues that are not preserved by the CI processors.

[2]In typical cochlear implant systems, the highest frequency is approximately 8000 which is lower than the 10,300 Hz cutoff used here. However, logarithmically this value is not that much higher than the values used in current CIs. There is good reason for providing higher frequencies because localization cues that result from directionally dependent filtering of sounds by the head and ears are greatest at frequencies between the range of 8–10 000 Hz than at the lower frequencies. Some of the MAPS that are provided by manufacturers do offer this range as an option (e.g., Table 9 in the cochlear system). Finally, the higher frequency cutoff used

here was selected for consistency and comparability with results from other studies being conducted by our group in which sound localization ability is investigated using the same stimuli.

Arbogast, T., Mason, C., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T., Mason, C., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal hearing and hearing impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Arsenault, M., and Punch, J. (**1999**). "Nonsense-syllable recognition in noise using monaural and binaural listening strategies," J. Acoust. Soc. Am. **105**, 1821–1830.

Assmann, P., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.

Bacon, S., Opie, J., and Montoya, D. (**1998**). "The effect of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.

Başkent, D., and Shannon, R. V. (**2007**). "Combined effects of frequency compression-expansion and shift on speech recognition," Ear Hear. **28**, 277–289.

Battmer, R., Feldmeier, I., Kohlenberg, A., and Lenarz, T. (**1997**). "Performance of the new Clarion speech processor 1.2 in quiet and in noise," Am. J. Otol. **18**, S144–S146.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London).

Blauert, J. (**1997**). *Spatial Hearing—Revised Edition: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Bronkhorst, A., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1374–1383.

Brungart, D. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1119.

Buss, S., and Florentine, M. (**1985**). "Gap detection in normal and impaired listeners: The effect of level and frequency," in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, London), pp. 159–179.

Chung, D., and Mack, B. (**1979**). "The effect of masking by noise on word discrimination scores in listeners with normal hearing and with noise-induced hearing loss," Scand. Audiol. **8**, 139–143.

Dorman, M., Loizou, P., Fitzke, J., and 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. **104**, 3583–3585.

Dorman, M., and Loizou, P. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for normal hearing listeners and patients with cochlear implants," Am. J. Otol. **18**, S113–S114.

Dorman, M., Loizou, P., Kemp, L., and Kirk, K. (**2000**). "Word recognition by children listening to speech processed into small number of channels, data from normal-hearing children and children with cochlear implants," Ear Hear. **21**, 590–596.

Drullman, R., and Bronkhorst, A. (**2000**). "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," J. Acoust. Soc. Am. **107**, 2224–2235.

Dubno, J., Dirks, D., and Schaefer, A. (**1989**). "Stop-consonant recognition for normal hearing listeners and listeners with high frequency loss II: Articulation index predictions," J. Acoust. Soc. Am. **85**, 355–364.

Durlach, N., Mason, C., Shinn-Cunningham, B., Arbogast, T., Colburn, H., and Kidd, G., Jr. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Eisenberg, L., Dirks, D., and Bell, T. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.

Freyman, R., Balakrishnan, U., and Helfer, K. (**2008**). "Spatial release from

masking with noise-vocoded speech," J. Acoust. Soc. Am. **124**, 1627–1637.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Friesen, L., Shannon, R., 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 implant," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users, the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Gantz, B., Tyler, R., Knutson, J., Woodworth, G., Abbas, P., McCabe, B., Hinrichs, J., Tye-Murray, N., Lansing, C., Kuk, F., and Brown, C. (**1988**). "Evaluation of five different cochlear implant designs: Audiologic assessment and predictions of performance," Laryngoscope **98**, 1100–1106.

Gantz, B., Tyler, R., Rubinstein, J., Wolaver, A., Lowder, M., Abbas, P., Brwon, C., Hughes, M., and Preece, J. (**2002**). "Binaural cochlear implants placed during the same operation," Otol. Neurotol. **23**, 169–180.

Gardner, W., and Martin, K. (**1994**). "HRTF measurements of a KEMAR dummy-head microphone," The MIT Media Laboratory Machine Listening Group, http://sound.media.mit.edu/resources/KEMAR.html (Last viewed February 2009).

Hawley, M., Litovsky, R., and Colburn, H. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Hawley, M., Litovsky, R., and Culling, J. (**2004**). "The benefits of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Iwaki, T., Matsushiro, N., Mah, S., Sato, T., Yasuoka, E., Yamamoto, K., and Kubo, T. (**2004**). "Comparison of speech perception between monaural and binaural hearing in cochlear implant patients," Acta Oto-Laryngol. **124**, 358–362.

Kidd, G., Jr., Mason, C., Rohtla, T., and Deliwala, P. (**1998**). "Release from masking due to spatial separation of sources in the identification of non-speech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.

Kidd, G., Jr., Mason, C., and Arbogast, T. (**2002**). "Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **111**, 1367–1376.

Leek, M., and Summers, V. (**1993**). "The effect of temporal waveform shape on spectral discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **94**, 2074–2082.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Levitt, H., and Rabiner, L. (**1967**). "Binaural release from masking for speech and gain in intelligibility," J. Acoust. Soc. Am. **42**, 601–608.

Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.

Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (**2009**). "Spatial hearing and speech intelligibility in bilateral cochlear implant users," Ear Hear. **30**, 419–431.

Litovsky, R. Y., Parkinson, A., Arcaroli, J., and Sammath, C. (**2006**). "Clinical study of simultaneous bilateral cochlear implantation in adults: A multicenter study," Ear Hear. **27**, 714–731.

Litovsky, R., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., and Yu, G. (**2004**). "Bilateral cochlear implants in adults and children," Curr. Opin. Otolaryngol. Head Neck Surg. **130**, 648–655.

Loizou, P., Dorman, M., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097–3003.

Loizou, P., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. (**2009**). "Speech recognition by bilateral cochlear implant users in a cocktail-party setting," J. Acoust. Soc. Am. **125**, 372–383.

MacKeith, N., and Coles, R. (**1971**). "Binaural advantages in hearing of speech," J. Laryngol. Otol. **85**, 213–232.

Muller, J., Schon, F., and Helms, J. (**2002**). "Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system," Ear Hear. **23**, 198–206.

Muller-Deile, J., Schmidt, B., and Rudert, H. (**1995**). "Effects of noise on speech discrimination in cochlear implant patients," Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 303–306.

Neff, D. (**1995**). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.

Nelson, P., and Jin, S. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.

Nelson, P., Jin, S., Carney, A., and Nelson, D. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Nie, K., Stickney, G., and Zeng, F. G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Oh, E., and Lutfi, R. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3899.

Pekkarinen, E., Salmivalli, A., and Suonpaa, J. (**1990**). "Effect of noise on word discrimination by subjects with impaired hearing, compared with those with normal hearing," Scand. Audiol. **19**, 31–36.

Qin, M., and Oxenham, A. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Rosen, S. (**1992**). "Temporal information in speech acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Rothauser, E., Chapman, W., Guttman, N., Nordby, K., Silbigert, H., Urbanek, G., and Weinstock, M. (**1969**). "IEEE Recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Rubinstein, J., and Hong, R. (**2003**). "Signal coding in cochlear implants: Exploiting stochastic effects of electrical stimulation," Ann. Otol. Rhinol. Laryngol. Suppl. **191**, 14–19.

Schleich, P., Nopp, P., and D'Haese, P. (**2004**). "Head shadow, squelch, and summation effects in bilateral users of the Med-El COMBI 40/40+ Cochlear implant," Ear Hear. **25**, 197–204.

Schon, F., Muller, J., and Helms, J. (**2002**). "Speech reception thresholds obtained in a symmetrical four-loudspeaker arrangement from bilateral users of MED-EL cochlear implant," Otol. Neurotol. **3**, 710–714.

Shannon, R. V., Galvin, J. J. III, and Baskent, D. (**2002**). "Holes in hearing," J. Assoc. Res. Otolaryngol. **3**, 185–199.

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

Skinner, M., Clark, G., Whitford, L., Seligman, P., Staller, S., Shipp, D., Shallop, J., Everingham, C., Menapace, C., and Arndt, P. (**1994**). "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system," Am. J. Otol. **15**, 15–27.

Smith, Z., Delgutte, B., and Oxenham, J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Stickney, G., Assmann, P., Chang, J., and Zeng, F. G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.

Stickney, G., Nie, K., and Zeng, F. G. (**2005**). "Contribution of frequency modulation to speech recognition in noise," J. Acoust. Soc. Am. **118**, 2412–2420.

Summers, V., and Molis, M. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.

Tyler, R., Dunn, C., Witt, S., and Preece, J. (**2003**). "Update on bilateral cochlear implantation," Curr. Opin. Otolaryngol. Head Neck Surg. **11**, 388–393.

Tyler, R., Summerfield, Q., Wood, E., and Fernandes, M. (**1982**). "Psychoacoustic and phonetic temporal processing in normal and hearing impaired," J. Acoust. Soc. Am. **72**, 740–752.

Tyler, R., Gantz, B., Rubinstein, J., Wilson, B., Parkinson, A., Wolaver, A., Precce, J., Witt, S., and Lowder, M. (**2002**). "Three-month results with bilateral cochlear implants," Ear Hear. **23**, 80S–89S.

Van Hoesel, R. (**2004**). "Exploring the benefits of bilateral cochlear implants," J. Nurs. Meas **9**, 234–246.

Van Hoesel, R., and Tyler, R. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Van Hoesel, R., Ramsden, R., and O'Driscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination and speech intelligibility advantages in noise for a bilateral cochlear implant users," Ear Hear. **23**, 137–149.

Vliegen, J., and Oxenham, A. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. **105**, 339–346.

Waltzman, S. B., Cohen, N. L., and Fisher, S. (**1992**). "An experimental comparison of cochlear implant systems," Semin. Hear. **13**, 195–207.

Wichmann, F. A., and Hill, J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1290–1313.

Wichmann, F. A., and Hill, J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

Wilson, B., Lawson, D., and Muller, J. (**2003**). "Cochlear implant: Some likely next steps," Annu. Rev. Biomed. Eng. **5**, 207–249.

Wilson, B., Schatzer, R., Lopez-Poveda, E., Sun, X., Lawson, D., and Wolford, R. (**2005**). "Two new directions in speech processor design for cochlear implants," Ear Hear. **26**, 73S–81S.