

Improving Speech Recognition in Bilateral Cochlear Implant Users by Listening With the Better Ear

Trends in Hearing
Volume 22: 1–11
© The Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216518772963
journals.sagepub.com/home/tia



Alan Kan¹

Abstract

For patients with bilateral cochlear implants (BiCIs), understanding a target talker in a noisy situation can be difficult. Current efforts for improving speech-in-noise understanding have focused on improving signal-to-noise ratio by using multiple microphones or signal processing, with only moderate improvements in speech understanding performance. However, BiCI users typically report having a *better ear* for listening which can lead to an asymmetry in speech unmasking performance. This work proposes a novel listening strategy for improving speech-in-noise understanding by combining (a) a priori knowledge of a *better ear* and having a BiCI user selectively attend to a target talker in that ear with (b) signal processing that delivers the target talker to the *better ear* and the noisy background to the opposite ear. This strategy is different from traditional noise reduction strategies because it maintains situational awareness (background sounds are delivered to the ear contralateral to the *better ear*) while improving speech understanding. Speech recognition performance was evaluated with and without the *better ear* strategy in a speech-in-noise listening test using a virtual auditory space created from individualized head-related transfer functions. Listeners showed an average improvement of 4.4 dB signal-to-noise ratio in their speech reception threshold when using the *better ear* strategy with no listener showing a decrement in performance. This implies that the strategy has the potential to boost speech-in-noise recognition in BiCI users and may be useful in other hearing assistance devices such as hearing aids.

Keywords

bilateral cochlear implants, speech-in-noise understanding, signal processing strategy, Wiener filter, noise reduction

Date received: 16 September 2017; revised: 19 March 2018; accepted: 30 March 2018

Introduction

Cochlear implants (CIs) are being provided at an increasing rate to those with a severe-to-profound hearing loss in order to restore hearing. CIs restore hearing by taking an acoustic sound signal and converting it into electrical stimulation by separating the incoming sound into a number of channels with different center frequencies. Then, in each channel, the slow-varying envelope of the signal is extracted and used to modulate the amplitude of electrical pulses. These pulses stimulate different parts of the cochlea, taking advantage of its tonotopic organization. Through this mode of stimulation, the profoundly deaf have been able to recover a remarkable amount of speech understanding, especially in quiet conditions, where performance on understanding sentences can be as high as 100% in some users (Firszt et al., 2004; Loizou, Mani, & Dorman, 2003;

Wilson & Dorman, 2007). Recently, bilateral implantations have become more common, with demonstrated benefits in sound localization and understanding speech-in-noisy situations when compared with single CI use (Litovsky, Parkinson, Arcaroli, & Sammeth, 2006; Litovsky et al., 2004, Litovsky, Parkinson, & Arcaroli, 2009; van Hoesel & Tyler, 2003). However, even with two implants, CI users typically do not perform as well as normal hearing (NH) listeners in the same tasks, especially in the presence of noise (Kerber & Seeber, 2012; Loizou et al., 2009; Misurelli & Litovsky, 2012).

¹Waisman Center, University of Wisconsin–Madison, WI, USA

Corresponding author:

Alan Kan, Waisman Center, University of Wisconsin–Madison, 1500 Highland Avenue, Madison, WI 53705, USA.
Email: ahkan@waisman.wisc.edu



In particular, speech reception thresholds (SRTs), that is, the signal-to-noise ratio (SNR) needed to achieve 50% correct word recognition, can be 15 to 20 dB higher in CI users compared with NH listeners (Loizou et al., 2009).

Much of the current research in improving speech-in-noise understanding in bilateral CIs has focused on two areas. One area of focus has been to improve the spatial hearing abilities of bilateral CI users. This includes improving sound localization performance and speech-in-noise understanding, both of which are much better in NH listeners. A measurable spatial hearing benefit is spatial release from masking (SRM), which is the improvement in the SRT due to spatial separation of target and masker talkers. Using current sound processing strategies, bilateral CI users have shown very little benefit of SRM (2–5 dB), and much of this gain was due to monaural head shadow effects (Loizou et al., 2009). In contrast, NH listeners typically show SRM as large as 10 to 15 dB under similar conditions (Hawley, Litovsky, & Culling, 2004; Jones & Litovsky, 2011). The reason for small SRM in bilateral CI users may be due to poor sensitivity to interaural time differences (ITDs) when listening with clinical processors (Aronoff et al., 2010; Grantham, Ashmead, Ricketts, Labadie, & Haynes, 2007). It is speculated that ITDs cannot be used by CI users because they are not presented to the auditory system with fidelity with current clinical processors. This is because bilateral CIs are not synchronized and act as monaural systems when analyzing the incoming acoustic signals at each ear (Litovsky et al., 2012; van Hoesel, 2004). In addition, most processors operate at a pulse rate that is too high to provide useful ITD cues in the electrical pulses (Laback, Egger, & Majdak, 2015). This lack of access to ITDs may be limiting a bilateral CI user's ability to take advantage of binaural unmasking mechanisms that are enjoyed by NH listeners. There are several models that describe how ITDs might be important in binaural unmasking for speech-in-noise understanding (see Colburn & Durlach, 1978 for an overview). One such model, the equalization-cancellation model (Durlach, 1972), assumes that spatial separation of the target and masker allows the auditory system to apply internal delays to compensate for the interaural configuration of the noise. Then the noise can be canceled from the binaural signal by subtraction after equalization, thereby improving the SNR. While ITD sensitivity has been poor with clinical processors, the use of synchronized research processors in the laboratory has shown that some bilateral CI users do have sensitivity to ITDs, particularly at low rates of stimulation (for a review of the literature, see Kan & Litovsky, 2015; Laback et al., 2015). This is promising for the possibility of restoring binaural unmasking benefits to bilateral CI users, but technical challenges still exist. It is likely that changes in clinical mapping practices, coordinated signal

processing between processors, and new speech coding strategies will need to be developed before significant improvement in speech-in-noise understanding, and sound localization can be seen in CI users.

A second area of focus has been to improve speech-in-noise understanding by using directional or multi-microphone techniques and sometimes combined with adaptive beamforming or speech enhancement algorithms. Research in this area has focused on improving the SNR. Commercially, Advanced Bionics users have the option of using a T-MicTM adaptor which sits in front of the ears to take advantage of the natural acoustic filtering of the pinna to attenuate sounds coming from behind a listener. Using the T-Mic adaptor, it has been shown that SRTs can be improved by 4 dB over the behind-the-ear microphone, though an average of +10 dB SNR is needed for 50% correct speech understanding (Gifford & Revit, 2010). In the Cochlear Nucleus Freedom processor, the use of BEAMTM (a two-microphone adaptive beamforming algorithm) has been shown to improve SRTs by about 6 dB over a single directional microphone. However, an average SNR of +6.2 dB is needed to achieve 50% correct speech understanding (Spriet et al., 2007). The Cochlear Nucleus 5 processor has a dual-microphone preprocessing scheme (marketed as Zoom) that provided some improvement over BEAM. Using Zoom, an average of +2 dB SNR is needed to achieve 50% correct speech understanding (Wolfe et al., 2012). More recently, multi-microphone techniques have been implemented to that take advantage of the additional microphones made available through bilateral implantation. For example, Advanced Bionics has implemented StereoZoom in their commercial processors, which is a four-microphone adaptive beamformer that takes advantage of all the microphones in the processors across the ears. This binaural beamforming algorithm has been shown to provide approximately 7.1 dB improvement compared with listening with an omnidirectional microphone (Buechner, Dyballa, Hehrmann, Fredelake, & Lenarz, 2014). In the laboratory, where there are less computational and power limitations compared with existing processors, more sophisticated algorithms have been evaluated. Many of these approaches have either applied an ideal binary mask that fully eliminate time-frequency bins that have poor SNR (e.g., Hu & Loizou, 2008; Koning, Madhu, & Wouters, 2015), or more sophisticated filtering algorithms that reduce the noise by estimating the noise power (e.g., Goldsworthy, Delhorne, Desloge, & Braid, 2014; Hersbach, Grayden, Fallon, & McDermott, 2013; Koning et al., 2015), or enhancing the target signal (e.g., Healy, Delfarah, Vasko, Carter, & Wang, 2017; Kokkinakis, Azimi, Hu, & Friedland, 2012). Baumgärtel, Hu et al. (2015) provides a thorough evaluation of some of the most promising algorithms and

showed that these algorithms can provide up to 7 dB improvement in CI users in demanding noisy conditions. Despite these advances in signal processing, NH listeners can still perform far better than CI listeners at adverse SNRs.

In this work, a novel listening strategy is proposed for improving speech-in-noise understanding in CI users. The premise of this work is to show proof-of-concept that speech-in-noise listening can be better than just signal processing alone, if we try to take advantage of the inherent speech unmasking asymmetries of bilateral CI users. This work was motivated by recent results reported in Goupell, Kan, and Litovsky (2016). In one of the listening conditions in Goupell et al.'s study, a female target talker was presented to one ear while a male masker talker was presented to the contralateral ear, simultaneously. The talkers were presented through the auxiliary port of clinical processors to remove cross-talk between the two ears. In this condition, bilateral CI users were asked to selectively attend to the target talker, while ignoring the masker talker in the contralateral ear. An average SRT of -23 dB SNR was reported, which was a significant improvement to a second condition tested in the same listeners, where the target and masker were presented in both ears, simultaneously. In the later condition, an average SRT of -8 dB SNR was needed to achieve similar performance. This implies that by presenting target and maskers in separate ears, and instructing the listener to selectively attend to the ear with the target, an average improvement of 15 dB can be obtained when compared with the nonseparated condition. For many listeners, different SRTs were achieved depending on which ear was attended to, suggesting a *better ear* for attending to speech in the presence of a masker. The phenomenon of a *better ear* has also been reported in a number of other studies (e.g., Baumgärtel, Hu, et al., 2015; Litovsky et al., 2006; van Hoesel & Tyler, 2003). There are a number of reasons why CI users have a *better* or preferred ear for listening. This ear may be their first implanted ear, or the ear they always use the telephone on, or the ear that had better hearing prior to implantation, and so on. However, there may also be more objective reasons for a *better ear*, which include differences in the two ears in terms of duration and etiology of deafness, neural survival, and quality of electrode array implantation. These factors may affect the quality of the speech signal being presented to the brain. The existence of a *better ear* suggests that much better speech-in-noise performance may be achieved by exploiting this inherent asymmetry with additional signal processing. Hence, a *better ear* listening strategy is proposed that combines (a) *a priori* knowledge of a *better ear* with (b) a signal processing algorithm that separates the target talker from a noisy background and delivers the target talker to the *better ear* while the

remaining sound scene is presented to the contralateral ear (Figure 1). One reason for delivering the remaining background sounds to the contralateral ear is to maintain situational awareness. While traditional beamforming and noise reduction algorithms aim to remove as much of the background noise as possible to improve SNR, this may not be desirable for CI users who want to be aware of the situation around them. The *better ear* strategy can help maintain situational awareness because the listener can switch attention between their ears to attend to either the target talker or surrounding background talkers. In this work, a signal processing algorithm based on Wiener filtering principles (Kan, 2017; Kan, Jin, & Van Schaik, 2008) was used to implement the *better ear* listening strategy and evaluated with bilateral CI users as a proof-of-concept. This algorithm is by no-means the state-of-the-art in noise reduction signal processing but was useful in this work because of the simplicity in its implementation for separating a target talker from a noisy background.

Methods

Participants

Eleven (7 women, 4 men) adult bilateral CI patients, aged between 21 and 81 (mean: 56.5) years old participated in this study. All listeners were implanted as adults with CIs manufactured by Cochlear Ltd (Sydney, Australia) and used either Freedom or Nucleus 5 sound processors. All listeners had at least 2 years of bilateral CI experience. Listeners traveled to the University of Wisconsin–Madison for testing and received a stipend for their participation. All testing procedures were approved by the University of Wisconsin's human subjects institutional review board.

Stimuli

The *better ear* strategy was evaluated using virtual auditory space (VAS) techniques (Carlile, 1996). The VAS used in this study simulates a situation where a target talker is in front of the listener, and masking talkers are toward the left and right. The head-related transfer functions (HRTFs) used for creating the VAS were measured individually on each participant in a $2.90 \times 2.74 \times 2.44$ m single-walled sound proof booth using a blocked-ear canal technique (see Kan [2017] for a description of the HRTF recording method). HRTFs were measured for three loudspeaker positions located in front and on the left and right of the listener. The recording system transfer function was deconvolved from the measured HRTFs using a pseudoinverse technique (Epain et al., 2010). The HRTFs were then used to filter the speech stimuli for testing.

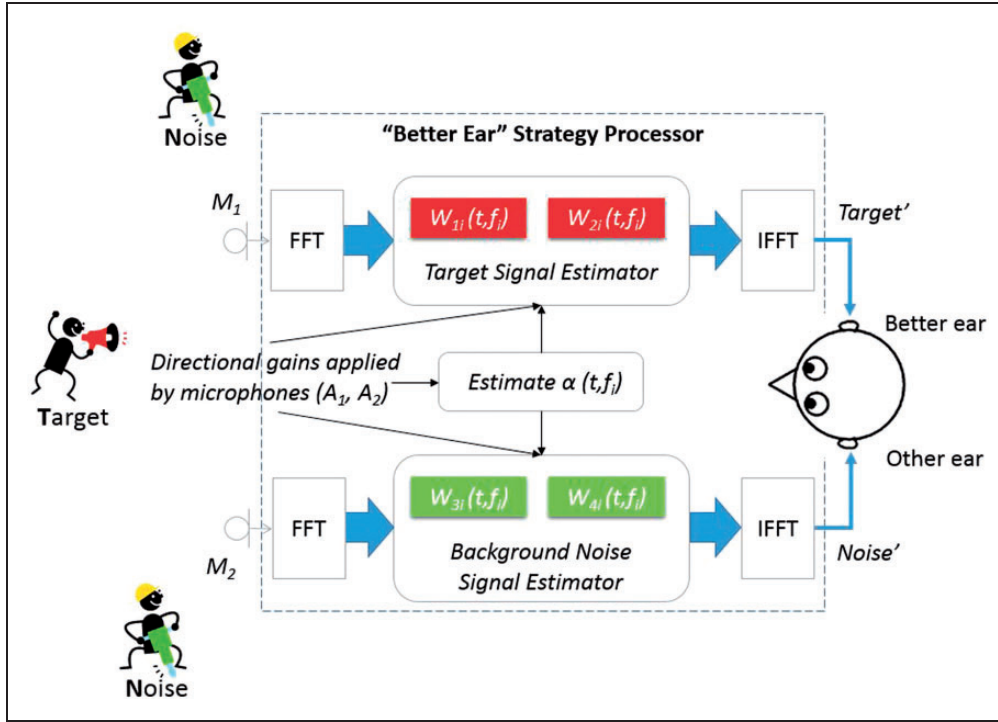


Figure 1. The proposed *better ear* listening strategy takes the microphone signals from the left and right ears (M_1 and M_2 , respectively) and makes estimates of the target and background noise. The estimated target and background noise signals are sent to the better ear and contralateral ear, respectively. In the current implementation of the *better ear* strategy processor, the Wiener filtering algorithm described in Kan (2017) was used to calculate weights by assuming that the location of the target talker was known, and that the target and background noise signals were not correlated. Here, W_{11} and W_{21} are the weights applied to M_1 , and W_{31} and W_{41} are the weights applied to M_2 in equations (2) and (5).

The speech stimuli used for testing was from Kidd, Best, and Mason (2008). The target and two interfering talkers were different male voices, speaking a five-word sentence of the form: name, verb, number, adjective, and noun. The target talker was filtered with the HRTFs corresponding to the front location, while the two interfering talkers were filtered with HRTFs corresponding to the left and right locations. All talkers began speaking their sentences simultaneously. The filtered speech stimuli were then combined at different SNRs to generate the VAS. An implementation of the *better ear* strategy was created by using the algorithm described in Kan (2017). Figure 1 illustrates how the algorithm was used to implement the *better ear* strategy. In this algorithm, there are two microphone signals, M_1 and M_2 . At a particular time frame, t , and frequency bin, f , the signals captured by the two microphones can be modeled by:

$$\begin{aligned} M_1(t, f) &= A_1(t, f)T(t, f) + B_1(t, f)N(t, f) \\ M_2(t, f) &= A_2(t, f)T(t, f) + B_2(t, f)N(t, f) \end{aligned} \quad (1)$$

where A_1 and A_2 are the direction-dependent gains applied by the microphones to the target signal T (for now, A_1 and A_2 are assumed to be known), and B_1 and B_2 are the unknown directional gains applied to all noise

sources N in the scene. For brevity of notation, the time and frequency indices are omitted in the following derivation. By rearranging Equation (1), and letting $B_2 = 1$ and $\alpha = B_1/B_2$,¹

$$T' = \frac{1}{A_1 - \alpha A_2} M_1 - \frac{\alpha}{A_1 - \alpha A_2} M_2 \quad (2)$$

That is, an estimate of the target, T' , can be found at a particular time frame, t , and frequency bin, f , if α can be estimated. The value α can be optimally estimated in the least-mean square error sense from the auto- and cross-correlations of M_1 and M_2 , which can be written as:

$$\begin{aligned} E\{M_1^2\} &= A_1^2 E\{T^2\} + B_1^2 E\{N^2\} + 2A_1 B_1 E\{TN\} \\ E\{M_2^2\} &= A_2^2 E\{T^2\} + B_2^2 E\{N^2\} + 2A_2 B_2 E\{TN\} \\ E\{M_1 M_2\} &= A_1 A_2 E\{T^2\} + B_1 B_2 E\{N^2\} \\ &\quad + (A_1 B_2 + A_2 B_1) E\{TN\} \end{aligned} \quad (3)$$

where $E\{X^2\}$ and $E\{XY\}$ denote the auto-correlation of X and cross-correlation of X and Y , respectively. If the target signal is assumed to be uncorrelated with the noise (i.e., $E\{TN\} = 0$), it can be shown that α can be estimated

by substituting $B_2=1$ and $\alpha=B_1/B_2$ into Equation (3) and rearranging:

$$\alpha = \frac{A_2 E\{M_1^2\} - A_1 E\{M_1 M_2\}}{A_2 E\{M_1 M_2\} - A_1 E\{M_2^2\}} \quad (4)$$

Hence, by solving Equations (4) and (2), an estimate of the target signal for a particular time-frequency bin can be obtained. In a similar fashion, the background noise can be estimated by rearranging Equation (1) to obtain:

$$N' = \frac{A_2}{\alpha A_2 - A_1} M_1 - \frac{A_1}{\alpha A_2 - A_1} M_2 \quad (5)$$

It is interesting to analyze how the value of α affects the estimated signals. When the target and background noise are not correlated, that is, $E\{TN\}=0$, it can be seen that the signals are optimally estimated. When $\alpha=0$, there is no noise in the time-frequency tile of microphone M_1 , and Equation (2) gives a normalized estimate of the target using microphone signal M_1 , while Equation (5) returns the noise recorded by microphone M_2 . Conversely, if $\alpha=1$, there is no target, and Equation (2) returns a value of 0, while Equation (5) returns a scaled estimate of the noise. For $0 < \alpha < 1$, an appropriate amount of M_2 is subtracted from M_1 to estimate the target and noise from Equations (2) and (5), respectively. However, if $E\{TN\} \neq 0$, then the error in estimating α can be found by substituting Equation (3) into Equation (4) without assuming $E\{TN\}=0$, which yields:

$$\alpha = \frac{B_1 E\{N^2\} + A_1 E\{TN\}}{B_2 E\{N^2\} + A_2 E\{TN\}} \quad (6)$$

Substituting Equations (6) and (1) into Equation (2) yields:

$$T' = T - [E\{TN\}/E\{N^2\}]N \quad (7)$$

where the error in estimating the target is given by the second term in Equation (7). That is, if the target and background noise are correlated, the estimate of the target signal will be corrupted by some portion of the noise energy at that time-frequency bin which is related to the amount of correlation between target and background. This implies that the SNR at these time-frequency bins will be intermittently poorer. Depending on the time and frequencies where this occurs, the perceptual consequences on intelligibility may be situation dependent. To provide an objective measure for this algorithm, the intelligibility-weighted SNR (iSNR) (Greenberg, Peterson, & Zurek, 1993) was estimated using the method described in Baumgärtel, Krawczyk-Becker, et al. (2015). To calculate iSNR, two different

signals, $(T+N)$ and $(T-N)$, are processed using the algorithm. Assuming that both signals are processed the same way by the algorithm, the processed target and noise signals can be estimated by:

$$T_{processed} = \frac{1}{2} [(T+N)_{processed} + (T-N)_{processed}] \quad (8)$$

and

$$N_{processed} = \frac{1}{2} [(T+N)_{processed} - (T-N)_{processed}] \quad (9)$$

From these signals, the iSNR can be calculated as the difference between the weighted SNR before ($\text{SNR}_{W,in}$) and after ($\text{SNR}_{W,out}$) processing:

$$\text{iSNR} = \text{SNR}_{W,out} - \text{SNR}_{W,in} \quad (10)$$

where

$$\text{SNR}_W = \sum_{k=1}^K w_k \text{SNR}_k \quad (11)$$

and w_k is the band importance weight associated with the k th band (American National Standards Institute, 1997). For the stimuli used in this experiment, the mean iSNR was 2.02 dB.

In this work, the algorithm was applied to the VAS signals to implement the *better ear* strategy. Time-frequency analysis was conducted using 1,024-sample sine-windowed frames with a 64-sample shift per frame. A 1,024-point fast Fourier transform was applied to each frame. To ensure the smoothness and stability of the weights applied to M_1 and M_2 , the cross-correlation value, $E\{M_1 M_2\}$, was averaged over four frames, which is approximately equal to the maximum time difference between the two ears at a sampling rate of 44100 Hz. Further, the denominators of equations (2) and (5) were inverted with regularization, where the maximum permissible amplification at 0, 500, 1000, 8000, 12000, 16000, and 22050 Hz was 0, 0, 12, 12, 0, and 0 dB, respectively. All stimuli for each listener were made in advance of testing and stored as 32-bit two-channel wav files.

Procedure

Testing was completed in a double-walled, sound-proof booth. Stimuli were played using a TDT System3 with RP2.1, HB7, and PA5 units (Tucker-Davis Technologies, Alachua, FL) and delivered to the listener through the auxiliary port of their clinical processors with direct connect cables. A touchscreen monitor was used to present instructions to the listener in each trial

and to record their response. In order for listeners to be able to distinguish the target talker from the masker talkers, the target talker sentences always began with the same name and verb which was the phrase *Bob took*. Listeners responded with the number, adjective, and noun. For each of these categories, there were eight choices which were shown on the touchscreen. For the masker talkers, their sentences never began with the name *Bob*.

Since a reliable objective method for acquiring *a priori* knowledge of the better ear for each listener was not available, the proposed listening strategy was tested in both ears for each listener. Hence, there were three listening conditions tested: (a) VAS—no additional processing, (b) VAS with *better ear* listening strategy assuming the better ear was on the left, and (c) VAS with *better ear* listening strategy assuming the better ear was on the right. In each of these conditions, the listener was prompted prior to the beginning of the block of trials to pay attention to the target talker in front, left, or right, respectively. The three conditions were tested in interleaved blocks, which consisted of different SNRs presented in a pseudo-random order. In each block, each SNR was presented three times. All listeners were tested at SNRs of 3, 0, -3, -6, and -9 with additional SNRs added in 3 dB increments during the testing to obtain a well-fit psychometric function. Overall, each SNR was tested 18 times, and percent correct at each SNR was calculated by scoring the total number of keywords recalled correctly. A total of 54 words were scored per SNR. The psychometric function

was obtained by fitting the data with a logistic function using the psignifit software Version 2.5.6 (Wichmann & Hill, 2001), and the SRT calculated as the 50% correct point on the fitted function.

Results

Table 1 and Figure 2(a) show the SRTs obtained in each condition for each listener individually. SRTs when no strategy was applied (VAS only condition) ranged from -6.4 to 5.8 dB (median: 2.2 dB; mean: 1.8 dB). When the *better ear* strategy was applied, there was no ear that was consistently dominant across the group. However, all listeners showed improved SRTs when the applied signal processing sent the target talker to one of the ears and maskers to the contralateral ear. Across the group, SRTs for the better performing ear ranged from -9.3 to 4.8 dB (median: -4.1 dB; mean: -3.2 dB), and SRTs for the *poorer ear* ranged from -8.1 to 19.7 dB (median: -1.7 dB; mean: 0.2 dB). Using Friedman's test, a significant difference was found between the SRTs for the different listening conditions, $\chi^2(2, 11) = 17.35, p = .002$. Post hoc testing with Bonferroni correction revealed a significant improvement when listening with the better ear compared with the poorer ear and the VAS only condition.

The benefit when listening with the proposed strategy can be more clearly seen in Figure 2(b). No listener showed a decrement in performance when listening with the better ear, and three listeners obtained improvements of ~9 dB. Median and mean improvement with

Table 1. Listener Data.

ID	Speech reception thresholds (dB)				First implanted ear	Years between implants	Etiology
	No strategy	Better ear strategy—left	Better ear strategy—right	Kan (2017) algorithm only			
IAZ	5.7	5.4	2.9		Left	1	Adult onset, hereditary
IBF	-6.4	-8.1	-9.3	-5.7	Right	1	Adult onset, hereditary
IBK	0.9	-4.1	-0.7	1.9	Left	6	Adult onset, noise-induced, possibly hereditary
IBO	0.6	-6.1	-8.7	-3.3	Right	3	Adult onset, Otosclerosis
IBR	4.8	9.3	4.8		Right	5	Adult onset, progressive
IBY	2.2	-4.4	-7.2	-0.1	Left	4	Adult onset, unknown
ICA	5.6	-3.3	-1.7		Right	7	Childhood onset, progressive, possibly from fever
ICB	-2.3	-6.6	-7	-0.3	Right	3	Childhood onset, hereditary
ICI	-0.5	-6.3	-6.4	-0.7	Left	1	Adult onset, unknown
ICP	5.8	4	19.7	7.8	Left	3	Childhood onset, nerve damage
ICV	3	-0.7	1.9		Simultaneous	0	Adult onset, traumatic injury

Note. Lowest speech reception thresholds across the different listening conditions are shown in bold.

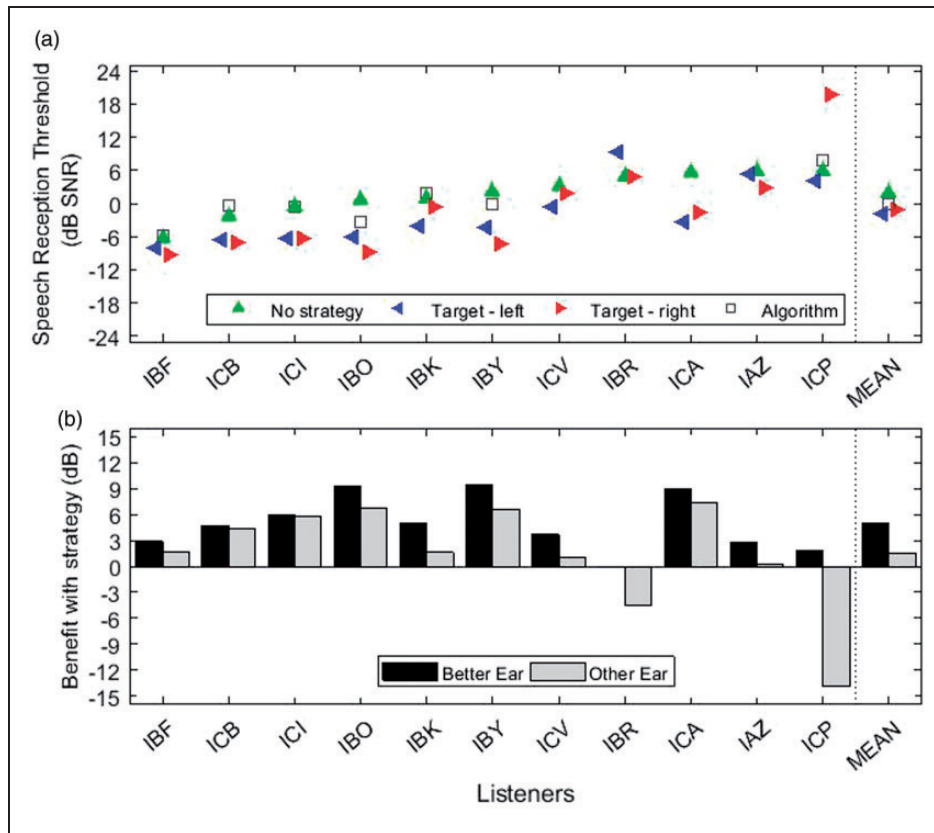


Figure 2. (a) The SRTs obtained in each listening configuration. (b) The improvement in SRTs when listening with the proposed strategy.

the better ear strategy was 4.7 and 4.9 dB, respectively. When listening with the poorer ear, most listeners still showed some improvement suggesting that the applied algorithm could provide some improvement in listening. Median and mean performance when listening with the poorer ear was 1.7 and 1.5 dB, respectively.

Discussion

In this work, a *better ear* listening strategy was proposed to take advantage of an asymmetry in speech understanding performance inherent in bilateral CI users to improve speech-in-noise understanding by applying some signal processing. Using a Wiener filtering algorithm to implement the proposed strategy, most listeners gained a significant benefit in speech recognition scores, and no listeners had a decrement in speech recognition with the proposed strategy. However, during testing, it was realized that it was necessary to determine whether listening to the better ear provided additional benefit over just using the Kan (2017) algorithm for simple noise reduction in both ears. Hence, seven listeners were tested during a return visit to the laboratory in a configuration where the Kan (2017) algorithm was used as a noise reduction algorithm in both ears. Their SRTs

are shown in Table 1 and Figure 2. Mean SRT was -0.1 dB SNR which was significantly poorer than listening with the better ear, $\chi^2(1, 7) = 7, p = .008$. This suggests that much of the gain in performance is from listening to the *better ear* rather than the algorithm itself.

Using the current implementation of the *better ear* listening strategy, the average SRT was -3.2 dB SNR. Compared with existing noise-reduction methods, where positive SNRs are needed to achieve comparable performance (e.g., 50% SRTs for Zoom requires $+2$ dB SNR; Wolfe et al., 2012), this is a promising result, though it should be noted that this was a closed-set task which may have artificially enhanced performance. Further, both ears were tested using the *better ear* strategy which inherently introduced a bias toward observing a larger effect. However, the results can be considered a best case scenario for determining the better ear, and provides a proof of concept that bilateral CI users can take advantage of a *better ear* listening strategy.

Taken together, the results suggest that bilateral CI users who have a noticeable asymmetry in speech understanding performance may be able to take advantage of their *better ear* even without the special signal processing suggested in this article. This can be achieved by simply having the target talker on the side of their better ear and

directing their attention to that ear. One could speculate that further improvements could be obtained in this configuration by applying directional beamforming in the *better ear* only, so that the target talker's speech is enhanced. This means that the *better ear* listening strategy can be employed with currently available technology and some clinician-guided advice.

The 4.4 dB improvement in this study is much smaller than the improvement predicted from the results presented in Goupell et al. (2016). In that study, presenting target and maskers in separate ears provided a 15-dB improvement in SRTs compared with the condition where target and maskers were presented to both ears, simultaneously. One could consider the Goupell et al.'s (2016) result as the best-case scenario for the *better ear* listening strategy. The smaller improvement found in this study can probably be attributed to differences in the speech stimuli used in this experiment (Goupell et al. used a female target and one male masker, while in this study the target and two maskers were all male which would render the task much harder), interference from the background sounds from the contralateral ear, or artifacts that may have been introduced by the signal processing algorithm. The effectiveness of the signal processing applied in this study relies on the assumption that the target and background noise are not correlated. If this assumption does not hold, then the separation of the target from the background noise will be incomplete and residual background noise would be presented to the *better ear*, making the task harder. It is likely that with the use of more sophisticated signal processing, such as steered binaural beamformer (Adiloğlu et al., 2015), subspace methods (see Loizou, 2017 for review), blind source separation (see Kokkinakis & Loizou, 2010 for review), or machine learning algorithms (e.g., Healy et al., 2017), will lead to much greater improvements in performance. However, one benefit of the proposed algorithm lies in its computational simplicity and ease with which it can be added to the CI signal processing chain. While the current implementation of the algorithm was applied in the Fast Fourier transform domain, the algorithm can also be applied to the extracted signal envelopes in each channel, meaning that the number of computations necessary to calculate the necessary weights scale by the number of channels of the CI sound processor. It should be noted, however, that in the current evaluation, the location of the target was assumed to be known. This was necessary so that the gains applied by the microphone to the target signal could be accounted for. In practice, precise location tracking may be difficult in dynamic environments where either the listener, or the target, could be moving. However, given the smoothness of HRTF gains across nearby locations, and the fact that microphone gains have a scaling effect in Equations (2) and (5), it is anticipated that the

need for precise location tracking will not be necessary. What is unclear at the present moment is how different amounts of reverberation will affect the performance of the algorithm.

In this work, there was no *a priori* knowledge of the *better ear*. In initial pilot testing, the methodology of Goupell et al. (2016) was tried, whereby a target talker was played in one ear and a masker in the contralateral ear as a way of assessing which ear would be better. However, results using that method did not necessarily predict the ear that would provide the best results using the Kan (2017) algorithm implementation of the *better ear* listening strategy. It is presently unclear how to objectively predict the *better ear* because factors such as years of CI or acoustic hearing experience, quality of the surgery and implantation, and etiology and neural survival may all contribute to speech understanding outcomes. Table 1 shows some of these factors for the listeners in this study. For the majority of listeners who were sequentially implanted and had greater than 3 years between implants, their first implanted ear typically showed the greatest benefit when using the *better ear* listening strategy. However, there are two listeners (IBY and ICA) that do not follow this trend. Further work is needed to understand how to predict the *better ear*.

The *better ear* listening strategy offers unique advantages for improving speech-in-noise understanding over existing approaches. First, it takes advantage of an asymmetry in speech unmasking performance in order to increase intelligibility of a target talker. In bilateral CI patients, this is particularly useful because an inherent asymmetry exists due to the variable nature of candidacy and implantation (Goupell et al., 2016). A second unique advantage of the *better ear* strategy is that it allows the CI user to maintain situational awareness, because nontarget sounds in the environment are still transmitted. This is contrary to existing approaches which aim to increase the SNR by removing as much of the background noise as possible. However, there are situations where removal of background noise may not be desirable. For example, a child in a noisy classroom needs to be able to hear input from a target (teacher), as well as supplementary input from classmates. When using a traditional beamforming strategy, the speech from the teacher is transmitted clearly, but surrounding sounds from classmates are usually suppressed. However, by using the *better ear* strategy, a child listening with bilateral CIs can still be aware of other students by attending to the nontarget ear. An extra option would include the ability to choose different mixing ratios for the two ears to enhance audibility of either the target or the background noise, if so desired. A third benefit of this strategy is economy; the *better ear* strategy can be made a part of the sound processors firmware without

needing extra equipment, and hence it is cheaper to implement than existing radio frequency solutions such as FM and loop systems. This is cost efficient and makes the solution more easily accessible to CI users. Furthermore, reducing the number of physical components will simplify setup and likely reduce the load on teachers and other professionals who regularly need to learn about a student's various assistive devices. Lastly, the *better ear* strategy is not CI specific, but rather, is a general strategy that can be implemented in listeners with any form of hearing impairment and present with an asymmetry in speech unmasking abilities between the ears. These may include patients who use bilateral hearing aids or are bimodal.

As with all approaches to improving SNR, there are limitations with the *better ear* listening strategy. In this work, it was assumed that the target signal within the sound scene was known. This is an inherent problem with all noise reduction algorithms, and few solutions exist. In current practice, the assumption is that the target talker of interest is in front of the listener and beamforming algorithms in existing CIs and hearing aids are designed to maximize sounds coming from in front of the listener. The same assumption was also made in this experiment. For the *better ear* strategy to be useful in a multitarget situation, a method for selecting and extracting multiple targets from a sound scene is required. Proposed methods for obtaining this information include pointing, target selection through button press, and visual guidance (Hart et al., 2009; Kidd, Favrot, Desloge, Streeter, & Mason, 2013).

Another limitation of this approach is that spatial hearing abilities are lost, which is a significant problem of the proposed strategy. This problem is compounded by the fact that a target talker on the right of the listener would be presented in the left ear if the left ear was the *better ear*. However, there may still be a group of bilateral CI users who will benefit from the *better ear* strategy. For some bilateral CI users, it is likely that good spatial hearing may not be possible because of limited sensitivity to binaural cues, such as ITDs, especially those who have had little acoustic hearing experience (Ehlers, Goupell, Zheng, Godar, & Litovsky, 2017; Litovsky, Jones, Agrawal, & van Hoesel, 2010). Another group who may benefit from the *better ear* strategy may be those who have a large asymmetry in performance between the two ears. It is possible that the proposed strategy will give them much larger benefit for understanding speech-in-noise compared with that provided via binaural hearing benefits. While the proposed strategy may not be useful as an everyday strategy because of the loss of spatial hearing abilities, the *better ear* strategy will likely be a useful option that a bilateral CI listener can switch to when having a conversation in demanding noisy situations.

Conclusions

Difficulty of understanding speech in noise is arguably the most common complaint of people with hearing loss. The *better ear* listening strategy has the potential to significantly improve speech understanding in noisy situations not just for bilateral CI users but for people with other hearing impairments as well.

Acknowledgments

I would like to thank all our listeners who traveled to Madison to participate in these experiments. I would also like to thank Ruth Litovsky for her support in letting me pursue my own research ideas while in her lab. Much thanks to Zachery Smith and Aaron Parkinson from Cochlear Americas for providing hardware and technical support. A portion of this work was presented at the 167th Meeting of the Acoustical Society of America.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by an Emerging Research Grant from the Hearing Health Foundation (12-00042), and in part by grants from the National Institute of Health (NIH)-National Institute on Deafness and other Communication Disorders (R03DC015321 to Alan Kan, R01DC003083 to Ruth Litovsky) and the NIH-Eunice Kennedy Shriver National Institute on Child Health and Human Development (P30 HD03352 to the Waisman Center).

Note

1. The choice of whether B_1 or B_2 is chosen to equal one is somewhat arbitrary. It is a necessary step because the number of variables that can be estimated using the auto- and cross-correlation functions is limited to three.

ORCID iD

Alan Kan  <http://orcid.org/0000-0001-6033-0958>

References

- Adiloğlu, K., Kayser, H., Baumgärtel, R. M., Rennebeck, S., Dietz, M., & Hohmann, V. (2015). A binaural steering beamformer system for enhancing a moving speech source. *Trends in Hearing, 19*, 2331216515618903. doi:10.1177/2331216515618903.
- American National Standards Institute. (1997). *ANSI S3.5-1997: Methods for calculation of the speech intelligibility index*. Washington, DC: Author.
- Aronoff, J. M., Yoon, Y., Freed, D. J., Vermiglio, A. J., Pal, I., & Soli, S. D. (2010). The use of interaural time and level

- difference cues by bilateral cochlear implant users. *The Journal of the Acoustical Society of America*, 127(3), EL87–EL92. doi:10.1121/1.3298451.
- Baumgärtel, R. M., Hu, H., Krawczyk-Becker, M., Marquardt, D., Herzke, T., Coleman, G., ... Dietz, M. (2015). Comparing binaural pre-processing strategies II. *Trends in Hearing*, 19, 2331216515617917. doi:10.1177/2331216515617917.
- Baumgärtel, R. M., Krawczyk-Becker, M., Marquardt, D., Völker, C., Hu, H., Herzke, T., ... Dietz, M. (2015). Comparing binaural pre-processing strategies I. *Trends in Hearing*, 19, 2331216515617916. doi:10.1177/2331216515617916.
- Buechner, A., Dyballa, K.-H., Hehrmann, P., Fredelake, S., & Lenarz, T. (2014). Advanced beamformers for cochlear implant users: Acute measurement of speech perception in challenging listening conditions. *PLoS One*, 9, e95542. doi:10.1371/journal.pone.0095542.
- Carlile, S. (1996). *Virtual auditory space: Generation and applications*. Neuroscience intelligence unit. Berlin/Heidelberg, Germany: Springer. doi:10.1007/978-3-662-22594-3.
- Colburn, H. S., & Durlach, N. I. (1978). Models of binaural interaction. In E. Carterette, & M. Friedman (Eds), *Handbook of perception* (pp. 365–466). New York, NY: Academic Press.
- Durlach, N. I. (1972). Binaural signal detection: Equalization and cancellation theory. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (pp. 369–462). New York, NY: Academic Press.
- Ehlers, E., Goupell, M. J., Zheng, Y., Godar, S. P., & Litovsky, R. Y. (2017). Binaural sensitivity in children who use bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 141, 4264–4277. doi:10.1121/1.4983824.
- Epain, N., Guillon, P., Kan, A., Kosobrodov, R., Sun, D., Jin, C., & van Schaik, A. (2010, August). Objective evaluation of a three-dimensional sound field reproduction system. In *Proceedings of the 20th International Congress on Acoustics*, Sydney, Australia.
- Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., ... Wackym, P. A. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear and Hearing*, 25(4), 375–387. doi:10.1097/01.AUD.0000134552.22205.EE.
- Gifford, R. H., & Revit, L. J. (2010). Speech perception for adult cochlear implant recipients in a realistic background noise: Effectiveness of preprocessing strategies and external options for improving speech recognition in noise. *Journal of the American Academy of Audiology*, 21(7), 441–451. doi:10.3766/jaaa.21.7.3.
- Goldsworthy, R. L., Delhorne, L. A., Desloge, J. G., & Braida, L. D. (2014). Two-microphone spatial filtering provides speech reception benefits for cochlear implant users in difficult acoustic environments. *The Journal of the Acoustical Society of America*, 136(2), 867–876. doi:10.1121/1.4887453.
- Goupell, M. J., Kan, A., & Litovsky, R. Y. (2016). Spatial attention in bilateral cochlear-implant users. *The Journal of the Acoustical Society of America*, 140(3), 1652. doi:10.1121/1.4962378.
- Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F., & Haynes, D. S. (2007). Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants. *Ear and Hearing*, 28(4), 524–541. doi:10.1097/AUD.0b013e31806dc21a.
- Greenberg, J. E., Peterson, P. M., & Zurek, P. M. (1993). Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *The Journal of the Acoustical Society of America*, 94(5), 3009–3010. doi:10.1121/1.407334.
- Hart, J., Onceanu, D., Sohn, C., Wightman, D., & Vertegaal, R. (2009). The attentive hearing aid: Eye selection of auditory sources for hearing impaired users. In T. Gross, J. Gulliksen, & P. Kotzé, et al. (Eds), *Human-computer interaction—INTERACT 2009: Proceedings of 12th IFIP TC 13 International Conference Part I* (pp. 19–35). Berlin/Heidelberg, Germany: Springer.
- Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America*, 115(2), 833–843. doi:10.1121/1.1639908.
- Healy, E. W., Delfarah, M., Vasko, J. L., Carter, B. L., & Wang, D. (2017). An algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker. *The Journal of the Acoustical Society of America*, 141(6), 4230–4239. doi:10.1121/1.4984271.
- Hersbach, A. A., Grayden, D. B., Fallon, J. B., & McDermott, H. J. (2013). A beamformer post-filter for cochlear implant noise reduction. *The Journal of the Acoustical Society of America*, 133(4), 2412–2420. doi:10.1121/1.4794391.
- Hu, Y., & Loizou, P. C. (2008). A new sound coding strategy for suppressing noise in cochlear implants. *The Journal of the Acoustical Society of America*, 124(1), 498–509. doi:10.1121/1.2924131.
- Jones, G. L., & Litovsky, R. Y. (2011). A cocktail party model of spatial release from masking by both noise and speech interferers. *The Journal of the Acoustical Society of America*, 130(3), 1463–1474. doi:10.1121/1.3613928.
- Kan, A. (2017, December). Improving speech intelligibility for bilateral cochlear implant users using Weiner filters and its impact on cognitive load. In *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Kuala Lumpur, Malaysia.
- Kan, A., Jin, C., & Van Schaik, A. (2008, July). Estimating a sound signal in a known direction from a Soundfield microphone recording. In *Proceedings of the International Conference on Audio, Language and Image Processing*, Shanghai, China.
- Kan, A., & Litovsky, R. Y. (2015). Binaural hearing with electrical stimulation. *Hearing Research*, 322, 127–137. doi:10.1016/j.heares.2014.08.005.
- Kerber, S., & Seeber, B. U. (2012). Sound localization in noise by normal-hearing listeners and cochlear implant users. *Ear and Hearing*, 33(4), 445–457. doi:10.1097/AUD.0b013e318257607b.
- Kidd, G. Jr, Best, V., & Mason, C. R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *The Journal of the Acoustical Society of America*, 124(6), 3793–3802. doi:10.1121/1.2998980.

- Kidd, G., Favrot, S., Desloge, J. G., Streeter, T. M., & Mason, C. R. (2013). Design and preliminary testing of a visually guided hearing aid. *The Journal of the Acoustical Society of America*, *133*(3), EL202–EL207. doi:10.1121/1.4791710.
- Kokkinakis, K., Azimi, B., Hu, Y., & Friedland, D. R. (2012). Single and multiple microphone noise reduction strategies in cochlear implants. *Trends in Amplification*, *16*(2), 102–116. doi:10.1177/1084713812456906.
- Kokkinakis, K., & Loizou, P. C. (2010). Advances in modern blind signal separation algorithms: Theory and applications. In A. Spanias (Ed.), *Synthesis lectures on algorithms and software in engineering* (pp. 1–100). San Rafael, CA: Morgan & Claypool Publishers. doi:10.2200/S00258ED1V01Y201003 ASE006.
- Koning, R., Madhu, N., & Wouters, J. (2015). Ideal time–frequency masking algorithms lead to different speech intelligibility and quality in normal-hearing and cochlear implant listeners. *IEEE Transactions on Biomedical Engineering*, *62*(1), 331–341. doi:10.1109/TBME.2014.2351854.
- Laback, B., Egger, K., & Majdak, P. (2015). Perception and coding of interaural time differences with bilateral cochlear implants. *Hearing Research*, *322*, 138–150. doi:10.1016/j.heares.2014.10.004.
- Litovsky, R., Parkinson, A., Arcaroli, J., & Sammeth, C. (2006). Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study. *Ear and Hearing*, *27*(6), 714–731. doi:10.1097/01.aud.0000246816.50820.42.
- Litovsky, R. Y., Goupell, M. J., Godar, S., Grieco-Calub, T., Jones, G. L., Garadat, S. N., . . . Misurelli, S. (2012). Studies on bilateral cochlear implants at the University of Wisconsin's Binaural Hearing and Speech Laboratory. *Journal of the American Academy of Audiology*, *23*(6), 476–494. doi:10.3766/jaaa.23.6.9.
- Litovsky, R. Y., Jones, G. L., Agrawal, S., & van Hoesel, R. (2010). Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans. *The Journal of the Acoustical Society of America*, *127*(1), 400–414. doi:10.1121/1.3257546.
- Litovsky, R. Y., Parkinson, A., & Arcaroli, J. (2009). Spatial hearing and speech intelligibility in bilateral cochlear implant users. *Ear and Hearing*, *30*(4), 419–431. doi:10.1097/AUD.0b013e3181a165be.
- Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., & Yu, G. (2004). Bilateral cochlear implants in adults and children. *Archives of Otolaryngology–Head & Neck Surgery*, *130*(5), 648. doi:10.1001/archotol.130.5.648.
- Loizou, P. C. (2017). *Speech enhancement theory and practice*. Boca Raton, FL: CRC Press, Taylor & Francis Group, LLC.
- Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., & Roland, P. (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *The Journal of the Acoustical Society of America*, *125*(1), 372–383. doi:10.1121/1.3036175.
- Loizou, P. C., Mani, A., & Dorman, M. F. (2003). Dichotic speech recognition in noise using reduced spectral cues. *The Journal of the Acoustical Society of America*, *114*(1), 475. doi:10.1121/1.1582861.
- Misurelli, S. M., & Litovsky, R. Y. (2012). Spatial release from masking in children with normal hearing and with bilateral cochlear implants: Effect of interferer asymmetry. *The Journal of the Acoustical Society of America*, *132*(1), 380–391. doi:10.1121/1.4725760.
- Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., . . . Wouters, J. (2007). Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom Cochlear Implant System. *Ear and Hearing*, *28*(1), 62–72. doi:10.1097/01.aud.0000252470.54246.54.
- van Hoesel, R. J. M. (2004). Exploring the benefits of bilateral cochlear implants. *Audiology and Neurotology*, *9*(4), 234–246. doi:10.1159/000078393.
- van Hoesel, R. J. M., & Tyler, R. S. (2003). Speech perception, localization, and lateralization with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, *113*(3), 1617–1630. doi:10.1121/1.1539520.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. doi:10.3758/BF03194544.
- Wilson, B. S., & Dorman, M. F. (2007). The surprising performance of present-day cochlear implants. *IEEE Transactions on Biomedical Engineering*, *54*(6), 969–972. doi:10.1109/TBME.2007.893505.
- Wolfe, J., Parkinson, A., Schafer, E. C., Gilden, J., Rehwinkel, K., Mansanares, J., . . . Gannaway, S. (2012). Benefit of a commercially available cochlear implant processor with dual-microphone beamforming. *Otology & Neurotology*, *33*(4), 553–560. doi:10.1097/MAO.0b013e31825367a5.