Original Article

Benefits of Beamforming With Local Spatial-Cue Preservation for Speech Localization and Segregation

Trends in Hearing
Volume 24: 1–11

DOI: 10.1177/2331216519896908
journals.sagepub.com/home/tia
SAGE

Le Wang[1], Virginia Best[2], and Barbara G. Shinn-Cunningham[3]

Abstract

A study was conducted to examine the benefits afforded by a signal-processing strategy that imposes the binaural cues present in a natural signal, calculated locally in time and frequency, on the output of a beamforming microphone array. Such a strategy has the potential to combine the signal-to-noise ratio advantage of beamforming with the perceptual benefit of spatialization to enhance performance in multitalker mixtures. Participants with normal hearing and with hearing loss were tested on both speech localization and speech-on-speech masking tasks. Performance for the spatialized beamformer was compared with that for three other conditions: a reference condition with no processing, a beamformer with no spatialization, and a hybrid beamformer that operates only in the high frequencies to preserve natural binaural cues in the low frequencies. Beamforming with full-bandwidth spatialization supported speech localization and produced better speech reception thresholds than the other conditions.

Keywords

hearing aids, hearing loss, binaural hearing, cocktail party

Received 6 August 2019; revised 21 November 2019; accepted 1 December 2019

Introduction

In many realistic communication settings, a fundamental task of the listener is to perceptually segregate the various sources of sound, select one source upon which to focus attention, and then receive and process the information coming from the chosen source. The ability of listeners to succeed in this task varies widely and depends not only on the properties of the acoustic environment and types of competing sound sources that are present but also on a range of factors specific to individual listeners. For the task of understanding speech in the presence of competing talkers, some factors that have been shown to adversely influence performance include advanced age and hearing loss (e.g., Gallun, Diedesch, Kampel, & Jakien, 2013; Glyde, Cameron, Dillon, Hickson, & Seeto, 2013; Marrone, Mason, & Kidd, 2008) although even young normal-hearing (NH) listeners may vary widely in their abilities (e.g., Kidd et al., 2016; Ruggles & Shinn-Cunningham, 2011).

Many front-end signal-processing approaches have been proposed to improve speech recognition in noisy situations. The most successful of these approaches use directionality to improve the signal-to-noise ratio (SNR) and are a common feature of commercial assistive listening devices such as hearing aids (Launer, Zakis, & Moore, 2016) and cochlear implants (Loizou, 2006). Directional systems make use of multiple microphones and beamforming to emphasize sound sources from one direction and attenuate sound sources from other directions (see reviews in Doclo, Gannot, Moonen, & Spriet, 2010; Greenberg & Zurek, 2001). Such systems can make use of the microphones available in current hearing aids (i.e., two in a single hearing aid or four across a pair of hearing aids) or an array of microphones that may be

[1]Department of Biomedical Engineering, Boston University, Boston, MA, USA
[2]Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, USA
[3]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA

Corresponding Author:
Virginia Best, Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA 02215, USA.
Email: ginbest@bu.edu

mounted on the head or on eyeglasses (e.g., Anderson et al., 2018; Greenberg, Desloge, & Zurek, 2003; Kidd, 2017). The spatial tuning can be extremely narrow in systems based on a large number of microphones, which can dramatically improve the SNR for a sound located in the focus of the beamformer. Under relatively simple conditions with a frontal speech source and one or more spatially separated noise sources, reported improvements in speech reception thresholds (SRTs) for beamformers relative to omnidirectional microphones range from around 5 to 12 dB (e.g., Luts, Maj, Soede, & Wouters, 2004; Saunders & Kates, 1997; Soede, Bilsen, & Berkhout, 1993).

A drawback of many beamforming strategies is that they combine the microphone signals to produce a single-channel output that conveys no binaural information. This obviously compromises the ability to localize sounds and may impede the segregation of competing sounds based on differences in spatial position as well as the ability to selectively attend to (or suppress) different sounds. To mitigate this problem, a variety of strategies have been proposed to preserve or restore spatial cues in beamformer systems (see reviews in Doclo et al., 2010; Kollmeier & Kiessling, 2016). Most of these strategies involve the combination of processed with unprocessed signals or the selective application of beamforming to some parts of the signal. Generally, beamformer systems designed for hearing-aid applications try to reach a balance between SNR improvement and spatial-cue preservation (Van den Bogaert, Doclo, Wouters, & Moonen, 2008, 2009) and thus may not provide speech-in-noise benefits as large as those that are theoretically possible. Indeed, recent studies that evaluated beamforming hearing aids under relatively complex listening situations found rather modest improvements in speech intelligibility relative to standard directional microphones (e.g., Picou, Aspell, & Ricketts, 2014; Best, Mejia, Freeston, van Hoesel, & Dillon, 2015; Picou & Ricketts, 2019; Völker, Warzybok, & Ernst, 2015; Wu et al., 2019).

By using experimental systems, it is possible to explore the trade-off between SNR improvement and spatial-cue preservation systematically while bypassing some of the constraints associated with hearing-aid applications. We have previously used such a system to evaluate a spatial-cue preservation strategy in which the beamforming is restricted to frequencies above a specific cutoff, and natural binaural signals are allowed to pass through below that cutoff (Desloge, Rabinowitz, & Zurek, 1997). This hybrid approach improves the SNR while maintaining the perceived spatial separation of and the ability to localize sources in the scene. A drawback of this approach, in addition to the loss of any SNR advantage in the low frequencies, is that potentially useful spatial information in the higher frequencies is discarded. Moreover, in this scheme, spatial cues in the high and low frequencies are inconsistent (with natural location information at low frequencies, but spatial information consistent with a source in the median plane for high frequencies). Despite these issues, our experimental results show that both NH and hearing-impaired (HI) listeners benefit from the preservation of the low-frequency spatial cues in situations involving speech-on-speech masking, where differences in the perceived locations of the competing sounds are thought to be critical for their segregation (Best, Roverud, Mason, & Kidd, 2017). Notably, under these conditions, if no binaural cues are preserved, some listeners perform more poorly with beamforming than without, indicating that the improvement in SNR is more than counteracted by the loss of spatial information (Best et al., 2017; Kidd, Mason, Best, & Swaminathan, 2015; see also Neher, Wagener, & Latzel, 2017).

Here, we describe an approach that extracts full-bandwidth spatial information in an acoustic mixture prior to beamformer processing and then reapplies this spatial information to the output of the beamformer. In contrast to the hybrid approach, the final stimulus includes SNR improvements provided by beamforming as well as full-bandwidth spatial information gleaned from the original sound mixture. Like the hybrid approach (but unlike other spatial-cue preservation approaches), the scheme presented here has the very useful characteristic that the spatial cues are applied to all sources with no need for an estimate of what is the target and what is the noise. The approach is based on the observation that in a mixture of spectrotemporally sparse, competing sounds (such as speech), different sources tend to dominate at any given time and frequency; as a result, spatial information in the original mixture at each time and frequency largely reflects the spatial attributes of the sound source dominating that time–frequency "tile." Indeed, it has been shown that the spatial cues of the nondominant source in a tile contribute very little to the perception and intelligibility of the mixture (Schoenmaker, Brand, & van de Par, 2016). This observation suggests the possibility of resynthesizing a binaural sound from the single-channel beamformer output signal by (re)imposing the binaural differences present in the original sound mixture on the appropriate points in time and frequency of the resynthesized sound. In those time–frequency tiles dominated by the target sound, the spatial cues will be consistent with the target's location; in the tiles dominated by a distracting sound, the spatial cues will be consistent with the spatial location of that sound. As a result, distractor energy that is not perfectly suppressed by the beamformer should still be perceptually separable from target energy, which may improve the efficacy of neural suppression of the distractor through spatial selective attention.

The following study was designed to compare this full-bandwidth spatialization approach with the hybrid approach and to a full-bandwidth beamformer with no spatialization. These three strategies were compared with a reference condition in which the listener received natural, unprocessed binaural signals. The main task of interest was a speech-on-speech masking task like that we have used previously. A localization experiment was included to confirm that natural binaural cues were conveyed by the full-bandwidth spatialization approach. For each task, both broadband and high-pass speech conditions were included to confirm that useful spatial cues were provided at high frequencies as well as at low frequencies. The hypothesis was that full-bandwidth spatialization would support accurate horizontal localization, which would in turn support the ability of listeners to focus spatial attention effectively on the target, and thus maximize the advantage of beamforming under speech-on-speech masking conditions.

Methods

Participants

Fourteen adults participated in the study, seven with normal hearing (aged 18–40 years, mean age 23 years) and seven with bilateral sensorineural hearing impairment (aged 20–56 years, mean age 36 years). There was no significant age difference between the NH and HI groups, $t(12) = 2.02$, $p = .07$. The NH participants had pure-tone averages (PTAs; mean threshold across both ears at 0.5, 1 and 2 kHz) that ranged from 0 to 6.7 dB hearing level (HL; mean 3.8 dB HL). The HI participants had a range of losses with PTAs from 2.5 to 73.3 dB HL (mean 35.8 dB HL). The losses were relatively symmetric, with a PTA difference between the ears of no more than 10 dB. Participants were paid for their participation, gave informed consent, and all procedures were approved by the Boston University Institutional Review Board. Total testing time for the localization and speech intelligibility experiments was approximately 2.5 hr. We note that one NH participant was unable to complete the high-pass condition for both localization and speech intelligibility experiments, and another was unable to complete the broadband speech intelligibility experiment. Results are based on only six NH participants in these cases.

Beamforming and Spatialization

The different listening conditions were tested using a headphone simulation. Impulse responses were measured on an acoustic manikin (KEMAR) seated in a large sound-treated booth (IAC Acoustics). The inner dimensions of the booth were approximately $3.75\,\text{m} \times 4\,\text{m} \times 2.25$ (Length × Width × Height). The manikin was seated halfway along one of the walls with a distance of about 0.6 m between its back and the wall. An array of loudspeakers (Acoustic Research 215PS) was arranged in front of the manikin at ear-height, at a distance of about 1.5 m. Loudspeakers were positioned between $-90°$ and $+90°$ azimuth at 7.5° intervals for the impulse response recordings, although only a subset of five positions was used for this study (see later). The manikin was fitted with a flexible headband that ran from ear-to-ear across the top of the head. The headband housed a microphone array (Sensimetrics Corporation), which consisted of 16 omnidirectional microphones arranged in four front-back-oriented rows. The rows were evenly spaced with a separation of 66.67 mm, for a total array length of 200 mm. More details about the array, including images of the microphone layout, can be found elsewhere (Kidd, 2017, Figure 6; Roverud, Best, Mason, Streeter, & Kidd, 2018, Figure 2).

Figure 1 provides an overview of the processing steps used to create stimuli for each condition. Two sets of impulse responses were recorded. One set of impulse responses, which captured the signals received by the manikin's in-ear microphones, were used to simulate a natural binaural listening situation ("KEMAR" condition). The other set of impulse responses captured the 16-channel output of the microphone array for each

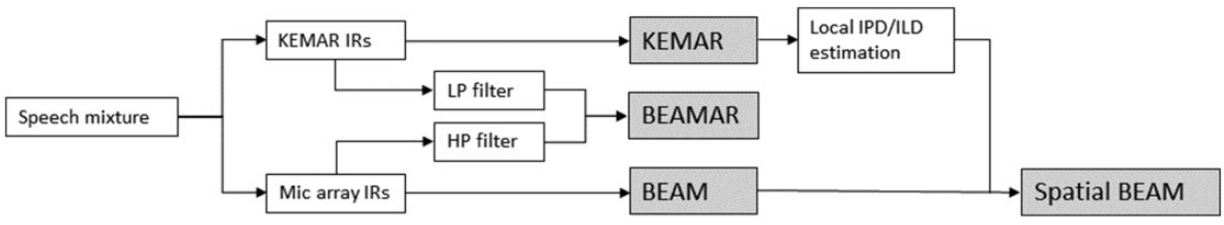

Figure 1. Overview of the steps used to create stimuli for each condition. The speech mixture was filtered with IRs measured in KEMAR's ear canals create the KEMAR stimuli and with IRs obtained from the microphone array to create the BEAM stimuli. IRs for the BEAMAR condition were created by combining LP-filtered KEMAR IRs and HP-filtered microphone array IRs. For spatial-BEAM stimuli, IPDs and ILDs were extracted from the KEMAR signals and applied to the BEAM signal. LP = low-pass; HP = high-pass; IPD = interaural phase difference; ILD = interaural level difference; IR = impulse response.

source location. These outputs were weighted and combined according to the optimal-directivity algorithm of Stadler and Rabinowitz (1993) for a look direction of 0° azimuth. The single-channel output for this condition ("BEAM") was presented diotically. The hybrid condition described earlier was simulated by combining low-pass-filtered binaural KEMAR impulse responses with high-pass-filtered BEAM impulse responses and is referred to as "BEAMAR." The crossover frequency was chosen to be 800 Hz, which was shown previously to be optimal (Best et al., 2017; Desloge et al., 1997).

The full-bandwidth spatialization strategy, referred to as "spatial BEAM," combined the spatial information from the KEMAR condition and the noise suppression from the BEAM condition across all frequencies. Specifically, for each spatial configuration, two spatial cues (interaural phase difference [IPD] and interaural level difference [ILD]) in each time–frequency tile were extracted from the binaural KEMAR signals. To do this, the KEMAR signal was broken down into time frames using a 92.9-ms hamming window (4,096 samples) that shifted by 23.2 ms (1,024 samples) for each frame. Within each frame, the spectrum of the left signal and the right signal was computed (frequency resolution = 10.8 Hz). The IPD and ILD were then calculated as the phase difference and the magnitude difference between the left and right signal for each frequency bin in the spectrum. These frequency-dependent IPDs and ILDs were then imposed on the corresponding time slice in the BEAM signal, creating a left and right signal per time frame. To do this, half of the IPD and ILD values were applied to the BEAM signal to create a left-ear signal; half of the IPD and ILD values inverted in polarity were applied to another copy of the BEAM signal to create a right-ear signal. The resynthesized binaural signals in each time frame were then summed to create a continuous output without additional temporal smoothing.

Stimuli

Target stimuli were taken from a 40-word corpus containing eight monosyllabic words in each of five distinct word categories (Kidd, Best, & Mason, 2008). Eight female voices were used in this study. Two speech bandwidth conditions were tested: broadband speech and high-pass speech. The broadband speech condition used the full spectrum speech signals in each processing condition, while the high-pass speech condition removed the frequency content below 800 Hz. The high-pass speech condition served as a control condition to test whether the *high-frequency* spatial information preserved in the spatial-BEAM condition offers useful information for localization and speech understanding, or whether benefits can only be obtained from the salient *low-frequency* spatial information. In the high-pass speech condition, the BEAMAR condition became identical to the BEAM condition because the low-frequency KEMAR portion of the BEAMAR signal was filtered out.

Stimuli were generated using MATLAB software (MathWorks Inc.) and presented via a 24-bit soundcard (RME HDSP 9632) through a pair of circumaural headphones (Sennheiser HD280 Pro). The participant was seated in a small sound-treated booth fitted with a computer monitor and mouse. For HI participants, individualized linear amplification according to the National Acoustic Laboratories' Revised, Profound (NAL-RP) prescription (Dillon, 2012) was applied to each stimulus just prior to presentation. A linear prescription was chosen to avoid potentially complicating interactions between nonlinear amplification and the different processing strategies.

Localization Experiment

A localization experiment was conducted to confirm that the spatial BEAM provided appropriate spatial information and produced lateral percepts in line with those produced by natural binaural stimuli (KEMAR). The stimuli were single words drawn at random from the speech corpus, presented at random from one of the five locations: −60°, −30°, 0°, +30°, and +60° azimuth. The nominal level of each word was 55 dB sound pressure level. Each word was processed according to one of the four conditions described earlier (KEMAR, BEAM, BEAMAR, and spatial BEAM), with the look direction of the beamformer always fixed at 0°. Trials were organized into blocks of 100 (five repetitions for each combination of processing condition and location), presented in a different random order for each participant. One block was completed for each of the two speech bandwidth conditions. Participants reported the perceived location of each stimulus by clicking on a graphical user interface showing a continuous arc representing the azimuthal plane from −90° to +90°. Before the experiment, each participant was given a training demo containing example trials from the KEMAR condition until they were familiar with the procedure.

Speech Intelligibility Experiment

The second experiment tested speech intelligibility in the presence of spatially separated competing talkers. Target stimuli were five-word sentences created by concatenating words from the speech corpus (with no added gaps between words). Each sentence had the form name-verb-number-adjective-noun (e.g., "Sue bought two red shoes"). The target sentence was spoken by one voice (chosen randomly on each trial from the set of eight)

and was identified on the basis of the first word (which was always "Sue"). The target was presented simultaneously with four speech maskers. The speech maskers were also five-word sentences, assembled in the same manner as the target sentence. The five presented sentences were spoken by different talkers and had no words in common. The target was located at 0° azimuth, and the four maskers were located at −60°, −30°, +30°, and +60° azimuth. Each masker was presented at 55 dB sound pressure level and the level of the target was varied to set the target-to-masker ratio (TMR; note that this is relative to *each* masker not the sum) to −20, −15, −10, −5, or 0 dB.

Participants completed three blocks of trials in each of the eight conditions (two bandwidths × four processing conditions) for a total of 24 blocks. Within each block, each condition was tested 5 times at each of the five TMRs (25 trials total). Responses were given by clicking on a graphical user interface containing a grid of the 40 possible key words. Psychometric functions were generated for each participant in each condition by plotting percentage correct (calculated across all four key words in all trials) as a function of TMR and fitting a logistic function. SRTs corresponding to the TMR at 50% correct were extracted from each function. Participants were given a training demo of the speech intelligibility experiment, which contained example trials from the KEMAR condition. The TMR in the example trials was set at 0 dB to make sure all participants were able to perform the task and follow the instructions before they moved on to the main experiment.

Acoustic Analyses

Before examining the results of the behavioral experiments, acoustic analyses are presented to describe the performance of the spatial-BEAM algorithm for the speech materials used in this study.

One concern with the approach is that there may be a number of time–frequency tiles that are dominated by an interferer before beamforming, but that flip to being dominated by the target after beamforming. In these cases, binaural cues extracted from the original mixture may be inappropriately assigned to the target and lead to a distorted spatial representation. To assess how often this would have occurred for the stimuli in our speech intelligibility experiment, a simulation was conducted using 100 random stimuli configured in the same way as the experiment with a target at 0° and four maskers at ±60° and ±30°. The analysis was conducted for a range of TMRs from −20 to +20 dB. For each time–frequency tile, the SNR was compared before and after beamforming, and the number of tiles in which the SNR "flipped" from below 0 dB (masker-dominated) in the original signal to above 0 dB (target-dominated) in the BEAM signal was counted. Figure 2(a) shows this flip rate as a percentage of the total number of tiles. The flip rate ranged from around 4% at the lowest TMR to around 22% at a TMR of +10 dB (solid line) and covered the range 4% to 19% over the TMRs tested in the experiment. However, we speculated that many of these SNR sign flips did not lead to drastic changes in the value of the SNR but rather represented small absolute changes in SNR from slightly negative to slightly positive. Using a more conservative definition of what

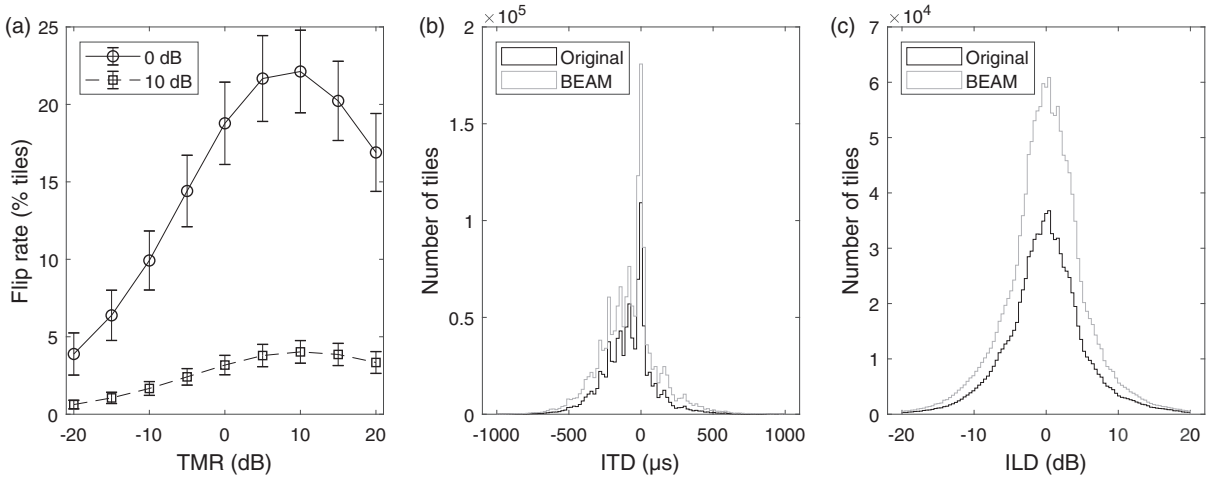

Figure 2. (a) Proportion of time–frequency tiles that were masker-dominated in the natural stimulus and flipped to being target-dominated after beamforming. Flip rates are shown for a wide range of nominal TMRs and for two criteria for defining flipped tiles: 0 dB (any tile whose local SNR changed from <0 dB to >0 dB) and 10 dB (any tile whose local SNR changed from <−10 dB to >+10 dB). (b) ITD histograms for target-dominated tiles before and after beamforming. (c) ILD histograms for target-dominated tiles before and after beamforming. ILD = interaural level difference; ITD = interaural time difference; TMR = target-to-masker ratio.

constitutes a flipped tile, by counting only those in which the SNR was <10 dB before and >10 dB after beamforming, the flip rate never exceeded 4% (dashed line). We can conclude from this analysis that the number of tiles that flip from being masker-dominated to target-dominated was relatively low but not negligible. Thus, we considered it important to gain some intuition about the effect of these flipped tiles on the spatial representation of the target.

To this end, another analysis was conducted to estimate the binaural cues associated with the target signal before and after the spatial-BEAM processing. First, time–frequency tiles were identified that were dominated by the target in the original mixture. For these tiles, IPDs and ILDs were calculated from the KEMAR stimuli, and IPDs were transformed to interaural time differences (ITDs). Histograms of ITDs and ILDs are plotted in Figure 2(b) and (c) (black lines) for a mixture with a TMR of 0 dB. As expected for a centrally located source, the histograms are centered on 0-μs ITD and 0-dB ILD (any asymmetries are due to asymmetries in the impulse responses, including minor effects of the room and the alignment of KEMAR within it, etc.). Second, time–frequency tiles were identified that were dominated by the target after beamforming. For these tiles, ITD and ILD distributions were again calculated from the KEMAR stimuli (gray lines in Figure 2(b) and (c)). These distributions capture the binaural cues that would be applied to the target in the spatial-BEAM condition. The first thing to notice is that there are more target-dominated tiles after beamforming. Moreover, these tiles continue to be centered on 0-μs ITD and 0-dB ILD, although some spread in the histograms can be observed. In general, we can conclude that the binaural cues associated with the target talker are not drastically distorted by the tiles that were previously masker-dominated. A similar pattern was found when the same analysis was applied to each of the four masker talkers, although in those cases fewer tiles (and not more) were available after beamforming.

Results

Localization Experiment

Figure 3 shows the group average localization responses as a function of the true location in the broadband speech condition (a and c) and the high-pass speech condition (b and d) for NH participants (a and b) and HI participants (c and d). The gray line indicates a slope of one, or perfect performance, where the perceived

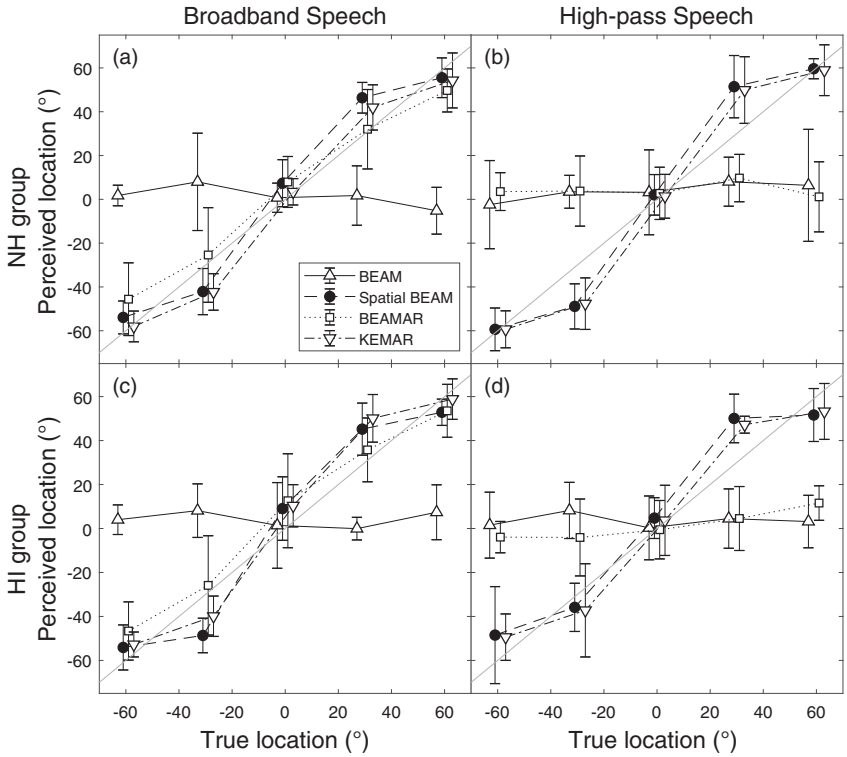

Figure 3. Mean perceived location plotted against true location for the broadband speech stimuli (a and c) and high-pass speech stimuli (b and d) for the NH group (a and b) and the HI group (c and d). Different symbols and line styles represent the different signal-processing schemes. The gray line represents the case when the perceived location is equal to the true location. Error bars show across-subject standard deviations. NH = normal-hearing; HI = hearing-impaired.

location is equal to the true location. Shallower slopes indicate a weaker spatial percept with a slope of zero indicating that the responses were not at all related to the true location.

When tested with broadband speech, both NH and HI groups were able to localize the stimulus with reasonable accuracy in all conditions except the BEAM condition. A mixed analysis of variance (ANOVA) model using true location and processing condition as within-subjects factors and group as a between-subjects factor found a significant main effect of true location, $F(4, 48) = 287.61$, $p < .001$, and a significant interaction between true location and processing condition, $F(12, 144) = 84.67$, $p < .001$. No main effect of group was identified, $F(1, 12) = 0.53$, $p = .48$, and none of the interactions involving group were significant. Follow-up ANOVAs comparing individual processing conditions indicated that the effect of true location in the spatial-BEAM condition was not significantly different from that in the KEMAR condition, $F(4, 52) = 0.68$, $p = .61$, while the effect of true location in the BEAMAR condition was significantly different than in the KEMAR condition, $F(4, 52) = 6.81$, $p < .001$. In terms of absolute localization error, when pooled across all trials for all locations for all participants, the mean values were 10° (KEMAR), 12° (spatial BEAM), 13° (BEAMAR), and 40° (BEAM).

With high-pass speech, because the spatial information in the BEAMAR condition was absent, participants were able to perceive lateral sounds only in the KEMAR and spatial-BEAM conditions. Similar to the broadband speech condition, there was a significant main effect of true location, $F(4, 44) = 173.08$, $p < .001$, and a significant interaction between true location and processing condition, $F(12, 132) = 74.67$, $p < .001$. No group effect was found, $F(1, 11) = 0.06$, $p = .81$, and no interactions involving group were significant. An ANOVA comparing the spatial-BEAM and KEMAR conditions found no significant difference in the effect of true location, $F(4, 48) = 0.07$, $p = .99$. Mean absolute localization errors in the high-pass condition were 13° (KEMAR), 12° (spatial BEAM), 35° (BEAMAR), and 37° (BEAM).

Speech Intelligibility Experiment

Figure 4 (a) and (b) shows group average SRTs for the different listening conditions, and Figure 4 (c) and (d) shows the same data but after each participant's SRT

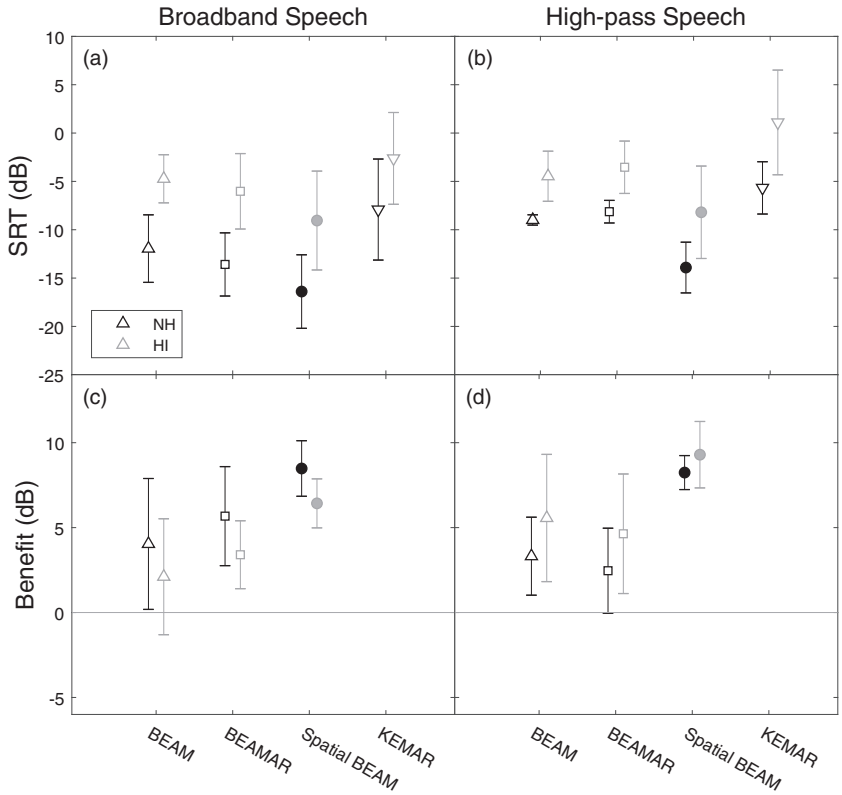

Figure 4. Mean SRTs (a and b) and benefits relative to KEMAR (c and d) for the broadband speech stimuli (a and c) and high-pass speech stimuli (b and d). The different processing schemes are shown along the abscissa, and NH and HI groups are shown in black and gray, respectively. Error bars show across-subject standard deviations. NH = normal-hearing; HI = hearing-impaired; SRT = speech reception threshold.

was normalized by their KEMAR SRT so that the ordinate represents the *benefit* provided by each beamforming condition relative to a natural binaural listening condition. As expected, the SRTs were generally higher in the HI group than in the NH group, for both the broadband speech (mean SRT −6 dB vs. −12 dB) and high-pass speech (mean SRT −4 dB vs. −9 dB).

A mixed ANOVA on the SRTs in the broadband speech condition identified a significant main effect of processing condition, $F(3,33) = 38.5$, $p < .001$, and a significant main effect of group, $F(1, 11) = 10.44$, $p = .008$, but no significant interaction, $F(3, 33) = 1.14$, $p = .35$. Post hoc comparisons (paired t tests with Bonferroni correction) confirmed that SRTs in the spatial-BEAM condition were significantly lower than in all other conditions at −12 dB on average, while the SRTs in the KEMAR condition were the highest at −5 dB. In the high-pass speech condition, the effect of processing condition was again significant, $F(3, 33) = 55.75$, $p < .001$, as was the effect of group, $F(1, 11) = 11.00$, $p = .007$, and there was no significant interaction, $F(3, 33) = 1.21$, $p = .32$. Post hoc comparisons confirmed that SRTs in the spatial-BEAM condition were significantly lower than in all other conditions at −11 dB on average, while the SRTs in the KEMAR condition were the highest at −2 dB.

Consistent with these findings, the spatial BEAM provided larger benefits for speech understanding (8 dB on average) than both BEAM and BEAMAR (4 dB on average; Figure 4 (c) and (d)). When pooled across NH and HI participants, the SRT benefit in the spatial-BEAM condition was not significantly correlated with age or with PTA for either speech condition ($p > .05$). However, the spatial-BEAM benefit was positively correlated with the BEAM benefit (broadband: $r = .62$; high-pass: $r = .71$) and with the BEAMAR benefit (broadband: $r = .84$; high-pass: $r = .62$).

Overall Performance

Figure 5 illustrates the overall performance of individual participants, combining the results from both the localization and speech intelligibility experiments. The four panels show data for the broadband speech condition (a and c) and the high-pass speech condition (b and d) for NH participants (a and b) and HI participants (c and d). Within each panel, individuals with good performance in both experiments are represented as points in the lower left corner, corresponding to low localization errors and low SRTs. For broadband speech, all processing schemes except the BEAM show a small average localization error, and the spatial BEAM shows generally lower SRTs than KEMAR and BEAMAR. Thus, the

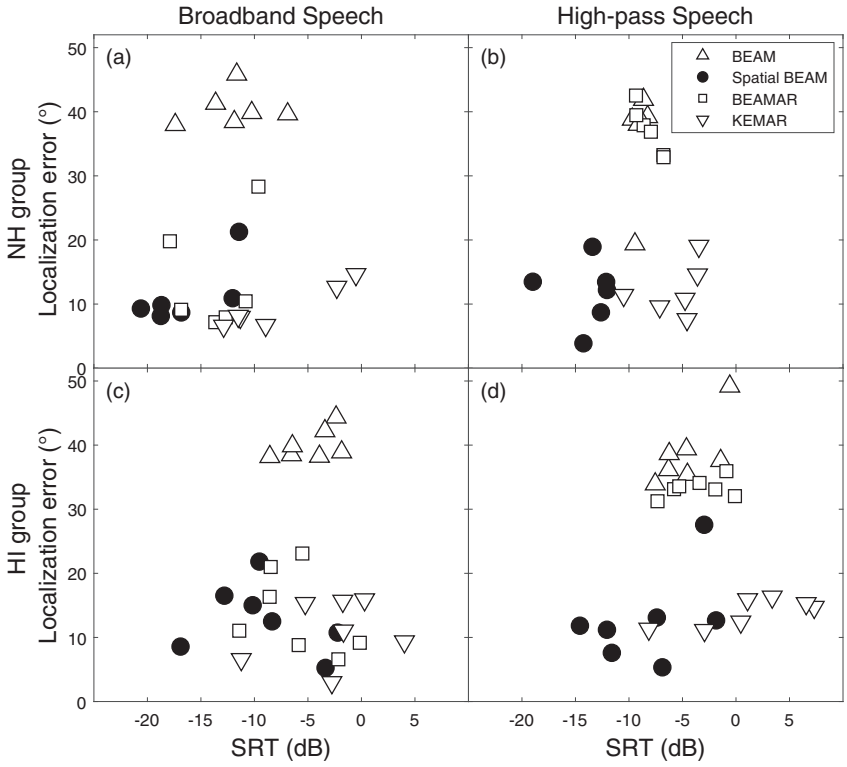

Figure 5. Individual localization errors plotted against individual SRTs for the broadband speech stimuli (a and c) and high-pass speech stimuli (b and d) for NH participants (a and b) and HI participants (c and d). Different symbols represent the different signal-processing schemes. NH = normal-hearing; HI = hearing-impaired; SRT = speech reception threshold.

spatial BEAM provided the most benefit in speech intelligibility among all conditions without degrading localization accuracy. For high-pass speech, results for the BEAMAR condition look essentially the same as the BEAM condition, confirming its limitation in situations when only high-frequency information is available. On the other hand, the spatial-BEAM condition preserves reliable spatial information at high frequencies and thus leads to low localization errors. Because the spatial BEAM also inherits the improved SNR from beamforming, overall performance was better than in the KEMAR condition.

Discussion

This study provided behavioral data to assess the benefits of a signal-processing strategy that reimposes natural, full-bandwidth, binaural information on the output of a highly directional beamformer. This spatial-BEAM strategy may be a promising option for assistive hearing devices because it has the potential to combine the SNR advantage of beamforming with the perceptual benefit of spatialization. Groups of participants with and without hearing loss were tested using this approach on both sound localization and speech intelligibility tasks. As anticipated, the spatial-BEAM strategy supported horizontal localization performance (as measured with single speech sources) that was equivalent to that observed in the natural binaural condition. Moreover, the spatial-BEAM strategy significantly improved speech understanding in the presence of competing speech relative to other implementations of the beamformer.

While the HI group in our study had a poorer mean SRT than the NH group, we found no interaction between group and processing condition, suggesting that the benefit of spatialization was achieved by both groups. Indeed, all participants had lower SRTs for the spatial BEAM than for BEAM (or for BEAMAR) in both bandwidth conditions. As pointed out by Neher et al. (2017), however, a benefit of spatialization may only apply to listeners who are sufficiently sensitive to spatial information; for listeners with poor sensitivity, it may have no effect. We note though that as the spatial-BEAM strategy does not sacrifice beamforming for spatial-cue preservation, it is hard to imagine a case in which it would be *detrimental* to performance.

A particular strength of the spatial-BEAM strategy is that it provides appropriate binaural cues across all frequencies rather than in a restricted frequency range (cf., Best et al., 2017; Desloge et al., 1997). The full-bandwidth approach has several specific advantages. First, the consistency of cues across frequency provides a coherent spatial perception of sounds without the risk of "split images." Second, spatial cues are available across the spectrum, which increases the versatility of the strategy. For example, this strategy will be robust in situations where low-frequency spatial information is degraded or lost due to masking noise or reverberation. Moreover, the inclusion of both IPDs and ILDs increases the chances that some sense of spatialization will be preserved even if one of the cues is unavailable or inaccessible to a listener. This issue may be important for potential applications in listeners with poor sensitivity to IPDs (such as older listeners; Moore, 2014) or in bilateral cochlear-implant users who are almost entirely dependent on ILDs (e.g., van Hoesel, 2004).

It is worth noting that this strategy is ideally suited for situations containing spectrotemporally sparse targets and distractors, such as the speech mixtures tested here. However, it is likely to be less effective for conditions in which the component sounds overlap heavily in the spectrotemporal plane. For example, for a speech target in the presence of a continuous intense noise, there may be few tiles in which the target is very dominant and thus the location of the target may not be clearly represented. Moreover, there may be many tiles in which both the target and noise sources contribute significant and near-equal amounts of energy and the spatial cues are inconsistent with either source. Under such conditions, reconstituting the spatial information may be of little benefit. An important next step in this line of work will be to compare the spatial-BEAM approach with other kinds of beamforming strategies using a wide variety of stimuli and tasks.

Finally, while the experimental system we used here allowed us to explore the potential benefits of the spatial-BEAM strategy while retaining a good degree of experimental control, this approach comes with some clear limitations. First, we tested only a "fixed-head" situation in which the received signals were unaffected by head movements. It is possible that this scenario underestimates the performance that is possible in the BEAM condition and hence overestimates the benefit of spatialization. Second, the spatial-BEAM strategy relies on binaural cues contained in signals as they occur in the listener's ear canals, and these were captured in our system via the KEMAR impulse responses. It is likely, however, that equally useful binaural cues could be extracted from microphones in other locations (such as the outermost microphones of the microphone array). Third, it is not clear whether similarly robust effects of spatialization would be achieved within the constraints of a real hearing-aid system. For example, it may not be feasible to implement the spatial-BEAM processing in real time, especially the ideal version implemented in this study that had very fine spectral and temporal resolution. Systematic evaluations of lower resolution versions of the spatialization would be very informative in this case. It would also be critical to determine how well the preserved binaural cues are

transmitted to the wearer for different hearing-aid styles, ear pieces, vent sizes, and so on.

Acknowledgments

The authors would like to thank Gerald Kidd and Chris Mason for their helpful input and support.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by National Institutes of Health-National Institute on Deafness and Other Communication Disorders (R01 DC013286) and Air Force Office of Scientific Research (FA9950-16-10372).

ORCID iD

Virginia Best https://orcid.org/0000-0002-5535-5736

References

Anderson, M. H., Yazel, B. W., Stickle, M. P. F., Espinosa Inguez, F. D., Gutierrez, N. S., Slaney, M., . . . Miller, L. M. (2018). *Towards mobile gaze-directed beamforming: A novel neuro-technology for hearing loss*. Paper presented at the 40th Annual Conference of the IEEE Engineering in Medicine and Biology Society, Honolulu, HI. doi: 10.1109/EMBC.2018.8513566

Best, V., Mejia, J., Freeston, K., van Hoesel, R. J., & Dillon, H. (2015). An evaluation of the performance of two binaural beamformers in complex and dynamic multitalker environments. *International Journal of Audiology*, *54*(10), 727–735. doi: 10.3109/14992027.2015.1059502

Best, V., Roverud, E., Mason, C. R., & Kidd, G. (2017). Examination of a hybrid beamformer that preserves auditory spatial cues. *Journal of the Acoustical Society of America*, *142*(4), EL369–EL374. doi: 10.1121/1.5007279

Desloge, J. G., Rabinowitz, W. M., & Zurek, P. M. (1997). Microphone-array hearing aids with binaural output. I. Fixed-processing systems. *IEEE Transactions on Speech and Audio Processing*, *5*, 529–542. doi: 10.1109/89.641298

Dillon, H. (2012). *Hearing aids* (2nd ed.). Turramurra, Australia: Boomerang Press. doi: 10.1097/MAO.0b013e31827ca367

Doclo, S., Gannot, S., Moonen, M., & Spriet, A. (2010). Acoustic beamforming for hearing aid applications. In S. Haykin & K. J. Ray Liu (Eds.), *Handbook on array processing and sensor networks* (pp. 269–302). Hoboken, NJ: Wiley-IEEE Press. doi: 10.1002/9780470487068.ch9

Gallun, F. J., Diedesch, A. C., Kampel, S. D., & Jakien, K. M. (2013). Independent impacts of age and hearing loss on spatial release in a complex auditory environment. *Frontiers in Neuroscience*, *7*, 252. doi:10.3389/fnins.2013.00252

Glyde, H., Cameron, S., Dillon, H., Hickson, L., & Seeto, M. (2013). The effects of hearing impairment and aging on spatial processing. *Ear and Hearing*, *34*(1), 15–28. doi: 10.1097/AUD.0b013e3182617f94

Greenberg, J. E., Desloge, J. G., & Zurek, P. M. (2003). Evaluation of array-processing algorithms for a headband hearing aid. *Journal of the Acoustical Society of America*, *113*(3), 1646–1657. doi: 10.1121/1.1536624

Greenberg, J. E., & Zurek, P. M. (2001). Microphone-array hearing aids. In M. S. Brandstein & D. B. Ward (Eds.), *Microphone arrays: Techniques and applications*. New York, NY: Springer.

Greenberg, J. E., & Zurek, P. M. (2001). Microphone-array hearing aids. In M. Brandstein & D. Ward (Eds.), *Microphone arrays. Signal processing techniques and applications* (pp. 229–253). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-662-04619-7_11

Kidd, G. (2017). Enhancing auditory selective attention using a visually guided hearing aid. *Journal of Speech, Language and Hearing Research*, *60*(10), 3027–3038. doi: 10.1044/2017_JSLHR-H-17-0071

Kidd, G., Best, V., & Mason, C. R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *Journal of the Acoustical Society of America*, *124*(6), 3793–3802. doi: 10.1121/1.2998980

Kidd, G., Mason, C. R., Best, V., & Swaminathan, J. (2015). Benefits of acoustic beamforming for solving the cocktail party problem. *Trends in Hearing*, *19*, 2331216515593385. doi:10.1177/2331216515593385

Kidd, G., Mason, C. R., Swaminathan, J., Roverud, E., Clayton, K. K., & Best, V. (2016). Determining the energetic and informational components of speech-on-speech masking. *Journal of the Acoustical Society of America*, *140*(1), 132–144. doi: 10.1121/1.4954748

Kollmeier, B., & Kiessling, J. (2016). Functionality of hearing aids: State-of-the-art and future model-based solutions. *International Journal of Audiology*, *57*(sup3), S3–S28. doi: 10.1080/14992027.2016.1256504

Launer, S., Zakis, J. A., & Moore, B. C. J. (2016). Hearing aid signal processing. In G. Popelka, B. Moore, R. Fay, & A. Popper (Eds.), *Hearing aids* (Vol. 56, pp. 93–130). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-33036-5_4

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. *Advances in Otorhinolaryngology*, *64*, 109–143. doi: 10.1159/000094648

Luts, H., Maj, J. B., Soede, W., & Wouters, J. (2004). Better speech perception in noise with an assistive multimicrophone array for hearing AIDS. *Ear and Hearing*, *25*(5), 411–420. doi: 10.1097/01.aud.0000145109.90767.ba

Marrone, N., Mason, C. R., & Kidd, G. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *Journal of the Acoustical Society of America*, *124*(5), 3064–3075. doi: 10.1121/1.2980441

Moore, B. C. J. (2014). *Auditory processing of temporal fine structure*. Singapore: World Scientific. doi: 10.1142/9064

Neher, T., Wagener, K. C., & Latzel, M. (2017). Speech reception with different bilateral directional processing schemes: Influence of binaural hearing, audiometric asymmetry, and

acoustic scenario. *Hearing Research*, *353*, 36–48. doi: 10.1016/j.heares.2017.07.014

Picou, E. M., Aspell, E., & Ricketts, T. A. (2014). Potential benefits and limitations of three types of directional processing in hearing aids. *Ear and Hearing*, *35*(3), 339–352. doi: 10.1097/AUD.0000000000000004

Picou, E. M., & Ricketts, T. A. (2019). An evaluation of hearing aid beamforming microphone arrays in a noisy laboratory setting. *Journal of the American Academy of Audiology*, *30*, 131–144. doi: 10.3766/jaaa.17090

Roverud, E., Best, V., Mason, C. R., Streeter, T., & Kidd, G. (2018). Evaluating the performance of a visually guided hearing aid using a dynamic auditory-visual word congruence task. *Ear and Hearing*, *39*(4), 756–769. doi: 10.1097/AUD.0000000000000532

Ruggles, D., & Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: Individual differences in normal-hearing listeners. *Journal of the Association for Research in Otolaryngology*, *12*(3), 395–405. doi: 10.1007/s10162-010-0254-z

Saunders, G. H., & Kates, J. M. (1997). Speech intelligibility enhancement using hearing-aid array processing. *Journal of the Acoustical Society of America*, *102*(3), 1827–1837. doi: 10.1121/1.420107

Schoenmaker, E., Brand, T., & van de Par, S. (2016). The multiple contributions of interaural differences to improved speech intelligibility in multitalker scenarios. *Journal of the Acoustical Society of America*, *139*(5), 2589–2603. doi: 10.1121/1.4948568

Soede, W., Bilsen, F. A., & Berkhout, A. J. (1993). Assessment of a directional microphone array for hearing-impaired listeners. *Journal of the Acoustical Society of America*, *94*(2 Pt 1), 799–808. doi: 10.1121/1.408181

Stadler, R. W., & Rabinowitz, W. M. (1993). On the potential of fixed arrays for hearing aids. *Journal of the Acoustical Society of America*, *94*(3 Pt 1), 1332–1342. doi: 10.1121/1.408161

Van den Bogaert, T., Doclo, S., Wouters, J., & Moonen, M. (2008). The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids. *Journal of the Acoustical Society of America*, *124*(1), 484–497. doi: 10.1121/1.2931962

Van den Bogaert, T., Doclo, S., Wouters, J., & Moonen, M. (2009). Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids. *Journal of the Acoustical Society of America*, *125*(1), 360–371. doi: 10.1121/1.3023069

van Hoesel, R. J. (2004). Exploring the benefits of bilateral cochlear implants. *Audiology and Neurootology*, *9*(4), 234–246. doi: 10.1159/000078393

Völker, C., Warzybok, A., & Ernst, S. M. A. (2015). Comparing binaural pre-processing strategies III: Speech intelligibility of normal-hearing and hearing-impaired listeners. *Trends in Hearing*, *19*, 2331216515618609. doi:10.1177/2331216515618609

Wu, Y. H., Stangl, E., Chipar, O., Hasan, S. S., DeVries, S., & Oleson, J. (2019). Efficacy and effectiveness of advanced hearing aid directional and noise reduction technologies for older adults with mild to moderate hearing loss. *Ear and Hearing*, *40*(4), 805–822. doi: 10.1097/AUD.0000000000000672