JASA ARTICLE

Enhancing the perceptual segregation and localization of sound sources with a triple beamformer

Gerald Kidd, Jr.,[a),b)] Todd R. Jennings, and Andrew J. Byrne
Department of Speech, Language and Hearing Sciences and Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215, USA

ABSTRACT:
A triple beamformer was developed to exploit the capabilities of the binaural auditory system. The goal was to enhance the perceptual segregation of spatially separated sound sources while preserving source localization. The triple beamformer comprised a variant of a standard single-channel beamformer that routes the primary beam output focused on the target source location to both ears. The triple beam algorithm adds two supplementary beams with the left-focused beam routed only to the left ear and the right-focused beam routed only to the right ear. The rationale for the approach is that the triple beam processing exploits sound source segregation in high informational masking (IM) conditions. Furthermore, the exaggerated interaural level differences produced by the triple beam are well-suited for categories of listeners (e.g., bilateral cochlear implant users) who receive limited benefit from interaural time differences. The performance with the triple beamformer was compared to normal binaural hearing (simulated using a Knowles Electronic Manikin for Auditory Research, G.R.A.S. Sound and Vibration, Holte, DK) and to that obtained from a single-channel beamformer. Source localization in azimuth and masked speech identification for multiple masker locations were measured for all three algorithms. Taking both localization and speech intelligibility into account, the triple beam algorithm was considered to be advantageous under high IM listening conditions.
© *2020 Acoustical Society of America*. https://doi.org/10.1121/10.0002779

(Received 9 July 2020; revised 4 November 2020; accepted 9 November 2020; published online 11 December 2020)

[Editor: Karen S. Helfer] Pages: 3598–3611

I. INTRODUCTION

The "cocktail party problem" (CPP) has received considerable attention in speech and hearing science literature for more than 70 years (see Middlebrooks *et al.*, 2017, for a series of recent reviews). One part of this broad and complex problem concerns the increased difficulty experienced by certain subgroups of the general population in typical multisource/multitalker listening situations and consideration of what can be done to enhance communication under those conditions. Even among young adult listeners with normal hearing (NH) and cognitive abilities, there exists a considerable range of performance in speech-on-speech (SOS) masking conditions (e.g., Swaminathan *et al.*, 2015; Clayton *et al.*, 2016). This variability across NH listeners in communication performance in SOS masking conditions stands in contrast to that usually found for performance in speech-on-noise masking conditions, which tend to be much less variable both within and across subjects (e.g., Kidd *et al.*, 2014). The reasons for the wide range of performance found in SOS masking for young NH listeners are complex, and this topic continues to be of great interest to speech communication researchers (e.g., see the review in Kidd and Colburn, 2017). Increasing age can adversely affect SOS masking performance even when hearing is normal or typical for a given age (e.g., Helfer and Freyman, 2008; Ellinger *et al.*, 2017). Sensorineural hearing loss is known to adversely affect SOS masking performance (e.g., Marrone *et al.*, 2008b; Best *et al.*, 2011; Best *et al.*, 2013) with higher masked thresholds and less benefit from masking release conditions often observed even when frequency specific amplification is provided to compensate for the hearing loss (e.g., Marrone *et al.*, 2008b; Kidd *et al.*, 2019). Certain non-peripheral disorders that affect selective attention may also manifest as poorer than normal performance in SOS or speech-on-noise masking (e.g., Hoover *et al.*, 2017; Villard and Kidd, 2019). Efforts to enhance the communication performance of those persons who exhibit particular difficulty in CPP listening situations have largely focused on assisting listeners with hearing loss. This enhancement is usually achieved by fitting a hearing aid on one or both ears in an attempt to compensate for reduced audibility. For the other subgroups of the population who have difficulty solving the CPP but possess NH sensitivity, the options are more limited and usually involve mitigation of the difficulties using strategies such as managing the acoustic environment, making better use of visual cues, etc., and, less frequently, mild-gain amplification (e.g., Roup *et al.*, 2018). The wide range of performance raises the possibility that some listeners with NH sensitivity may obtain a benefit from technology designed to enhance communication abilities in multiple-source listening situations.

[a)] Electronic mail: gkidd@bu.edu
[b)] Also at: Department of Otolaryngology - Head and Neck Surgery, Medical University of South Carolina, Charleston, South Carolina 29425, USA.

https://doi.org/10.1121/10.0002779

Improving audibility is usually the implicit goal of amplification. Directional amplification aims to improve the audibility of target sounds more than competing sounds (i.e., improving signal-to-"noise" ratio, S/N) by differentially amplifying sounds from a particular azimuth and attenuating sounds from other azimuths. However, although it is clear that improving the S/N is strongly correlated with improvements in masked speech intelligibility, there can be other factors that contribute significantly to solving the CPP as well. Specifically, the ability to perceptually segregate the sound sources may be the key to reducing masking when multiple talkers compete with a target talker and create a high level of "informational masking" (IM; see Kidd *et al.*, 2008a). For example, suppose the target talker is masked by two other nearby talkers, one on either side of the target. It has been shown that target speech reception performance is not governed simply by the acoustic S/N (where the "signal" in this case is the target speech and the noise is the masking speech[1]) and may depend on many nonacoustic factors. This is easily demonstrated in the laboratory simply by time reversing the speech of the masker talkers. Without altering the S/N, the intelligibility of the target speech may improve significantly (cf. Kidd *et al.*, 2016). Or, in a similar vein, if the speech from the competing maskers is spoken in a language that is unfamiliar to the listener and different from the familiar target speech, or is spoken with an unfamiliar accent, a significant improvement in target speech intelligibility can result (e.g., Freyman *et al.*, 2001; Calandruccio *et al.*, 2010; Brouwer *et al.*, 2012; Calandruccio *et al.*, 2014) compared to intelligible speech at the same level. Other similar examples related to linguistic or source segregation/selection variables have been reported. These factors generally fall under the broad category of IM (in contrast to energetic masking, EM, due to spectrotemporal overlap of sounds; cf. review in Culling and Stone, 2017). Although improving the S/N is usually (cf. Brungart *et al.*, 2001) correlated with improving speech recognition performance regardless of the nature of the masking environment, i.e., whether it is dominated by EM or IM, the logic underlying the emphasis on improving the S/N without regard to other factors (discussed below) is deeply rooted in the goal of reducing the EM (e.g., Kidd *et al.*, 2008a). However, in many cases, reducing the IM could yield a substantial performance advantage as well. This begs the question, though, as to how an amplification strategy might be devised to reduce the IM in realistic listening situations. In this study, we examined a means for improving the S/N in a multiple-talker sound field while also enhancing the perceptual segregation of sound sources that can lead to a reduction in the IM.

Normally, there are many cues that can reduce masking in multitalker listening situations and may potentially be exploited by hearing aids. Of these various cues, those that arise from the different locations of competing sound sources have received particular attention in the literature, in part, because those factors were a focus of much of the original, seminal work on the CPP (e.g., Cherry, 1953; Pollack and Pickett, 1958; Schubert and Schultz, 1962; Carhart, 1969a,b). Differences in sound source location create differences in the time of arrival and level of sounds at the two ears so that the ability to exploit binaural processing is often considered to be a major determinant of success in multiple-source communication conditions. Of the various signal processing strategies for providing amplification, a single channel beamformer may be particularly effective in improving the S/N, especially if it can be rapidly directed toward the target source and if the focus of the beam (referred to as the "acoustic look direction") may be redirected when the location of the target source changes (e.g., during turn-taking in conversation, e.g., Best *et al.*, 2016; also, Doclo *et al.*, 2008; Kidd *et al.*, 2013; Adiloğlu *et al.*, 2015; Favre-Felix *et al.*, 2018; Hladek *et al.*, 2018).

The present study examines one particular approach to amplification that combines a standard beamformer—intended to improve the S/N when directed toward a target source—with a system incorporating two supplemental beamformers aimed at other directions. These side beams, which may actually reduce the S/N reaching the listener in some conditions, are intended to exploit binaural processing to enhance the perceived source segregation of the target and maskers. This triple beam approach, which was initially described by Jennings and Kidd (2018) and has subsequently been used in a study of listeners fitted with cochlear implants (Yun *et al.*, 2019), exaggerates interaural level differences (ILDs) at the two ears to provide both source segregation and the ability to localize the different sources. The basic idea of enhancing target separation through binaural beamforming has precedent in the literature (e.g., Bissmeyer and Goldsworthy, 2017) and has found some success both with users of hearing aids (e.g., Lotter and Vary, 2006; Rohdenburg *et al.*, 2007; Moore *et al.*, 2016) and those with cochlear implants (e.g., Baumgärtel *et al.*, 2015; Dieudonné and Francart, 2018; Williges *et al.*, 2018). The current investigation takes a parametric approach to characterizing this triple beam amplification scheme using listeners with NH in a laboratory setting. The tasks of the listener involved SOS masking as the competing talkers (maskers) were varied in azimuth surrounding the target talker and judging the location of individual sound sources in azimuth. The focus is on the relationship between the position of the masking sources relative to the target and the orientation of the two side beams. Understanding the functional consequences of the relationship between source location and side beam orientation—which underlie the interaural differences produced by the triple beam—is key to determining the conditions under which the triple beam approach may be beneficial. Finally, we consider an overall metric of a benefit that takes into account the performance observed in both SOS masking and source localization tasks.

II. GENERAL METHODS

Two separate experiments will be presented. The methods common to both experiments are described in this section while the procedures that were specific to each experiment are described under separate headings below.

A. Subjects

There were a total of six listeners, all female with NH as determined by a standard audiometric assessment (hearing thresholds at or below 20 dB Hearing Threshold Level (HTL) at octave frequencies from 250 to 8000 Hz). The subjects ranged in age from 18 to 26 years old [$M = 21.3$, standard deviation (SD) $= 2.7$]. All listeners reported U.S. English as their first and primary language. Five of the six subjects indicated on the self-report information form that they are musicians.

B. Stimuli

The speech materials were single words from a matrix identification test (Kidd *et al.*, 2008b) constructed in a manner that parallels other similar speech matrix tests (e.g., Hagerman, 1982; Bolia *et al.*, 2000). There were five syntactic categories (name, verb, number, adjective, and object) in the test with eight exemplars in each category. Each speech exemplar was recorded from 11 different female talkers. For the SOS masking experiment (experiment 1), each test stimulus was a syntactically correct five-word sentence comprising a random selection of words, one from each category (excluding the name designating the target sentence). An example would be "Sue found three big shoes." Two other five-word sentences constructed in a similar manner were presented concurrently with the target sentence. These masker sentences were independent, mutually exclusive selections from the same matrix of words also presented in syntactically correct order. Other than the name "Sue," which always designated the target sentence, all of the words were chosen at random without replacement on every trial. On each trial, 3 different talkers were selected at random from the set of 11 talkers to be the target and the 2 maskers. This random selection process—an independent, mutually exclusive selection of 3 of the 11 talkers—occurred for every trial. For the single-source localization experiment (experiment 2), the stimuli were single words randomly selected on each trial from the matrix corpus and the set of 11 talkers.

C. Apparatus

All stimuli were presented binaurally through Sennheiser HD280 Pro headphones (Sennheiser Electronic GmbH and Co. KG, Wedemark, Germany). The computer-generated waveforms were digital-to-analog converted at a rate of 44.1 k samples per second and played through an RME (Haimhausen, Germany) HDSP 9632 (ASIO) 24-bit sound card. Stimulus generation, response recording, and experimental control were implemented using MATLAB software (The MathWorks, Natick, MA). Listening took place in individual Industrial Acoustics Company (IAC; North Aurora, IL) double-walled booths, containing a keyboard and display monitor connected to a personal computer located outside of the booths. For experiment 1, the response alternatives were displayed on the monitor in matrix format and the response selection was registered by mouse-clicking a graphical user interface (GUI) showing the words in columns in syntactic order left to right, whereas for experiment 2, responses were recorded by clicking on a GUI showing a semicircle to illustrate the forward horizontal plane for localization.

D. Beamforming and control algorithms

1. Natural binaural cues

For the natural binaural listening control in this experiment, impulse responses from the Knowles Electronic Manikin for Auditory Research (KEMAR; G.R.A.S. Sound and Vibration, Holte, DK) as reported by Gardner and Martin (1994) were used. These impulse responses were convolved with the stimuli to position the sound sources in azimuth. This microphone condition is termed "KEMAR."

2. Beam

For both the beam and triple beam conditions, impulse responses were recorded through the microphone array mounted on the head of the KEMAR manikin and used to create the spatial filters that oriented the beam(s) toward the desired acoustic look directions. The impulse responses were recorded by Sensimetric Corporation (Malden, MA) from a loudspeaker located one meter from the KEMAR in a low-reverberation chamber and extended over a range of azimuths from −90° to +90° with a resolution of 2°. To obtain the recordings, the manikin was fitted with an array of 16 omnidirectional microelectromechanical systems (MEMS) microphones extending across the top of the manikin's head flush mounted on a band covering a flexible circuit board (see also Kidd, 2017, for an illustration and description). The output from each microphone in the array was weighted and combined to yield a matrix of gain values that optimized the response of the array to the specified direction (cf. Stadler and Rabinowitz, 1993; Desloge *et al.*, 1997). In the single beam case (referred to simply as "beam"), the stimulus was processed through a spatial filter having an acoustic look direction of 0° and presented diotically. This condition has been tested in several previous studies from our laboratory (e.g., Kidd *et al.*, 2013; Kidd *et al.*, 2015; Favrot *et al.*, 2013; Best *et al.*, 2017a; Best *et al.*, 2017b; Roverud *et al.*, 2018) and the spatial tuning properties of the single beam have been described in detail in those articles. Of particular importance for the current study is the frequency dependence of tuning of the array which affects the beam and triple beam algorithms equally because the filters are created using the same algorithm. As shown in Best *et al.* (2017a), the spatial selectivity of the beamformer increases with increasing center frequency. As a rough estimate, the −3 dB bandwidths for octave bands from the impulse responses are not measureable at 125 and 250 Hz center frequencies (i.e., broader than 180°) and then progressively decrease from about 80° at a center frequency of 500 Hz to less than 30° at 4000 Hz center frequency. Despite the cross-frequency differences in tuning, speech sources subjectively sound coherent as the azimuth is varied, although changes in the timbre are apparent for a broadband reference sound that is moved across the front hemifield as different parts of the spectrum are differentially attenuated (cf. Kidd *et al.*, 2015, for related discussion). The 16 microphones that provide the inputs to the beamforming algorithm are arranged in 4 front-to-back oriented rows along the flexible circuit board/headband. The total length of the array is 200 mm with a spacing of 67 mm between rows. Within each row, the microphones are arranged into two pairs with 10 mm spacing (15 mm spacing between the pairs). The outputs of the 16 microphones are combined to give a spatially tuned single-channel array output (e.g., Desloge *et al.*, 1997; Favrot *et al.*, 2013). For the triple beam case, this directional processing is implemented three times to provide three separate acoustic look directions (i.e., three separate single-channel beams).

3. Triple beam

The processing implemented by the triple beam algorithm is identical to that which occurs for the beam algorithm but includes processing for two additional beams that have acoustic look directions to the left and right of center. The triple beam produces a two-channel output by combining the center beam and each side beam as shown schematically in Fig. 1 (adapted from Jennings and Kidd, 2018).

First of all, for reference in Fig. 1, the beam processing condition described in Sec. II D 2 is simply the center beam without the left and right beams. This single channel output is routed to both ears forming a diotic signal (it could also be routed only to one ear forming a monotic signal; that condition was not included in this study). This beam condition has been characterized in detail in past work from our group (e.g., Kidd *et al.*, 2013; Favrot *et al.*, 2013; Kidd *et al.*, 2015; Best *et al.*, 2017a; Best *et al.*, 2017b; Roverud *et al.*, 2018). Second, the triple beam comprises the single diotic beam *and* each of the two side beams—left and right—routed exclusively to the left and right ears, respectively. The algorithm for combining and weighting of microphone outputs to create the side beams is identical to that which produces the center beam except that the side beams have different acoustic look directions. As indicated in the diagram, the center beam output is added to each side beam output. In the current study, the target source was always directly ahead of the listener so that the target signal was the same in both ears. The side beams were focused to the left and right of center (0°) spaced symmetrically in azimuth at ±15°, ±30°, ±45°, ±60°, ±75°, or ±90° under the different test conditions, labeled as "triple 15°," "triple 30°," etc., in the experiments that follow. For experiment 1, the masker talkers were located symmetrically to the left and right of the target so that the left ear received the sum of the center beam and the left beam while the right ear received the sum of the center beam and the right beam. The masker levels reaching each ear depended on the acoustic look direction of each side beam and the attenuation characteristics of the beam filters.

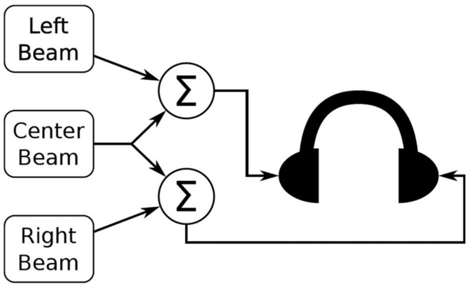

FIG. 1. A schematic illustration of the triple beam algorithm (adapted from Jennings and Kidd, 2018).

The acoustic results of the triple beam processing are illustrated in Fig. 2. The upper row of panels contains the attenuation characteristics of the three beams, left, center, and right, superimposed along coordinates of azimuth in degrees (with 0° indicating directly in front of the listener) by attenuation in dB. The different panels display the attenuation characteristics of the three beams for different side beam angles increasing from ±15° to ±90°, reading left to right. Note that there is significant overlap of these broadband response plots even at the wider separations. This overlap is frequency dependent because the sharpness of tuning—which depends on the spacing of the microphones relative to the wavelengths of the sounds—increases with frequency across the range of frequencies of interest (cf. Best *et al.*, 2017a). The middle row of panels is also organized from the narrow to wide side beam spacings in the panels from left to right. The two curves in each panel show the stimulus level measured after processing for a single broadband sound (speech-spectrum-shaped noise derived from the set of female talkers used in the study) moved across the range of azimuths for each ear separately. Thus, when the sound source is in the right hemifield, the input to the right ear—the sum of the right beam and the center beam outputs—is higher than the input to the left ear, whereas the converse is true when the source moves to the left of center. The resulting S/Ns that occur monaurally are given in the Appendix, along with the corresponding speech identification thresholds measured monaurally for two subjects. The difference in input level to the two ears forms an "ILD", which is fundamental to the benefits observed for the triple beam algorithm found in certain conditions in this study. The lower panel contains plots of the differences in level between the inputs to the left and right ears, i.e., the resulting ILDs, from the values displayed in the center panel.

As noted above, the spatial selectivity of the beamformer is frequency dependent. The ILDs that result from the combination of center and side beams are as well. Figure 3 shows the ILDs for octave bands of noise for each of the triple beam side beam angles. Also shown are the ILDs that were measured for the KEMAR manikin. With respect to the ILDs observed for the KEMAR manikin, the values vary across frequency in a manner that is representative of the average human head (cf. Shaw, 1974) increasing as stimulus center frequency increases. Of particular note for the current study is that large ILDs—much larger than

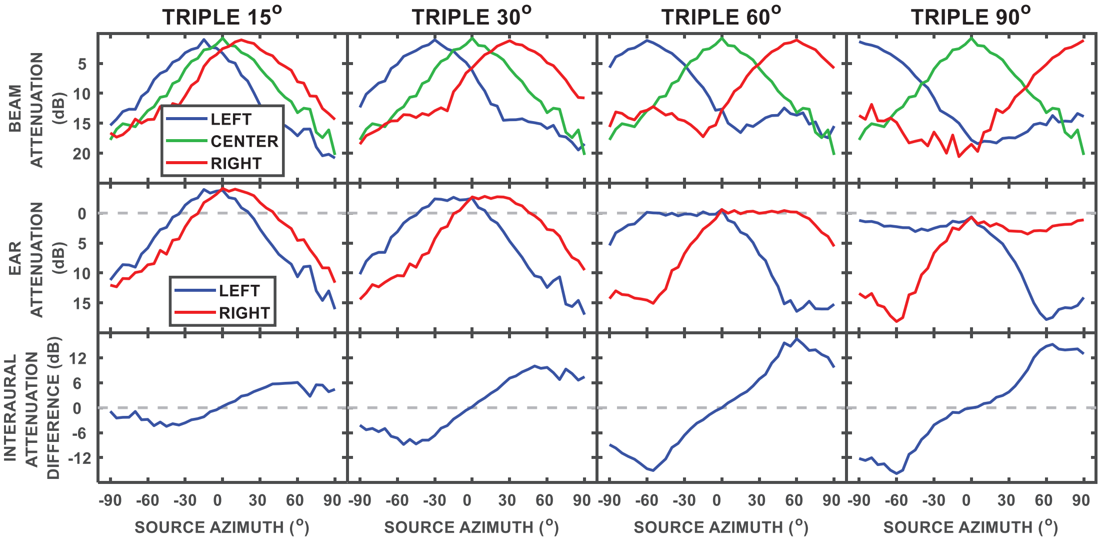

FIG. 2. (Color online) Each row of panels contains acoustic measurements for the triple beam algorithm for side beam angles of ±15°, ±30°, ±60°, and ±90°, reading left to right. Note the different range of values along the ordinate for each row. In the top row, the attenuation characteristics (spatial filters) are plotted in dB for each beam separately. In the middle row of panels, the input level to the right and left ears following processing is plotted for the broadband stimulus as it is moved across the horizontal plane. Note that the input to the left ear is the sum of the left and center beam outputs while the input to the right ear is the sum of the right and center beam outputs (refer to Fig. 1). The lower row of panels shows the difference between the right and left ear inputs from the middle panel, which is designated as the ILD. A negative value on the ordinate corresponds to a higher level in the left ear.

those which occur naturally for human listeners as reflected in the KEMAR responses—are present in the low frequencies in many cases. For example, at the lowest octave frequency band shown—centered at 250 Hz—the ILD for the KEMAR condition varied less than 3 dB across the entire range of source angles. In contrast, ILDs greater than 15 dB were apparent for the 250-Hz octave band for the two wider triple beam conditions of ±60° and ±90°. At the higher frequencies, the KEMAR ILDs increase up to about 15 dB. The ILDs for the triple beam are roughly of the same magnitude as the KEMAR ILDs at the higher frequencies but tend to fluctuate more as the side beam angles change, producing an irregular function relating ILD to the source azimuth for some side beam angles.

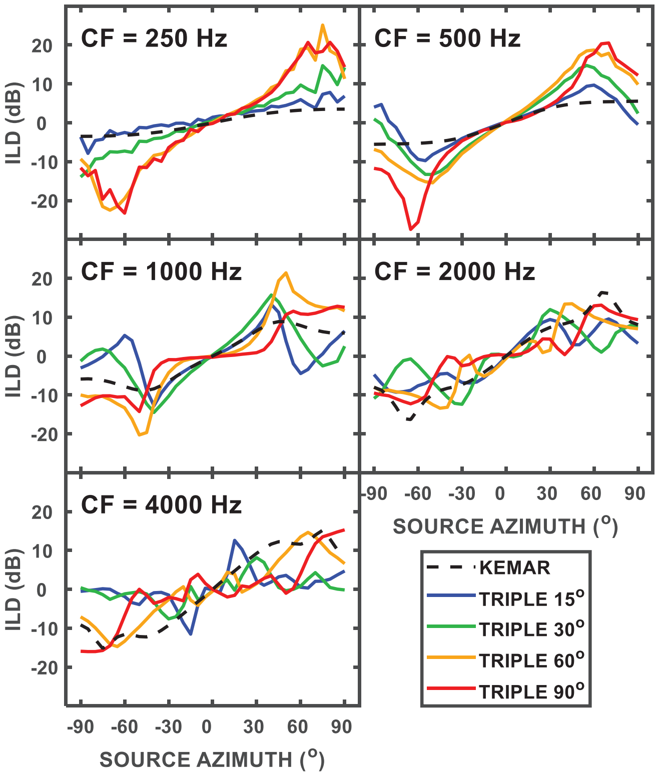

FIG. 3. (Color online) A plot of the ILDs for octave bands at different center frequencies measured for the different side beam angles (see the key). The ILDs for KEMAR (dashed line) are also shown for reference. The center frequencies of the octave bands increase in the panels from left to right and top to bottom.

III. EXPERIMENT 1: SOS MASKING EXPERIMENT

Because this study aimed to examine the effects of varying the parameters of the triple beam algorithm on performance and compare the triple beam with the KEMAR and beam algorithms, the focus of the experimental design was on spatially separated conditions. Also, previous work has found that the triple beam algorithm produced thresholds for SOS masking conditions that were roughly equivalent to those found for the KEMAR and beam algorithms for colocated target and masker (Jennings and Kidd, 2018; Yun *et al.*, 2019). However, we did not measure colocated thresholds and, thus, cannot make direct estimates of the spatial release from masking. The main concern here is on the masked thresholds obtained across the types of microphone/signal processing conditions: the KEMAR approximation of the natural listener interaural differences, the beam and triple beam algorithms. The SOS masking task we used is one that has been employed in several past studies of spatial hearing and masking release (e.g., Best *et al.*, 2011; Best *et al.*, 2013; Best *et al.*, 2017a; Best *et al.*, 2017b; Swaminathan *et al.*, 2015; Clayton *et al.*, 2016; Kidd *et al.*, 2016; Kidd *et al.*, 2019). It requires the listener to identify the words of a target talker denoted by a specific call sign (the first word "Sue" of the target sentence) while two other similar independent masker sentences are presented. The voice of the target talker is indicated when the listener hears the word "Sue" and then follows that voice throughout the remaining words of the sentence. The assumption is that the listener uses the vocal characteristics of the target talker to distinguish her speech from the other concurrent masker talkers. Furthermore, the target talker is the only voice originating from 0° azimuth so the source location also may serve as a means for designating the target talker.

A. Procedures

On each trial, the listener heard a target sentence from the matrix corpus described above that began with the designated name "Sue," along with two masker sentences also presented simultaneously, that each began with different names. Subject responses were registered on a GUI and displayed on a monitor, which showed the speech matrix arranged in columns in syntactic order. The responses were registered in order from left to right on the GUI. Response feedback was provided after every trial by indicating the target words and highlighting correct response items. The four words following the designated name in the target sentence were scored, and a response was counted as correct only if at least three of the four words were identified correctly. All conditions were repeated once before the next repetition, but the condition order was randomized in each repetition.

The stimulus level for the target sentence was fixed at 55 dB sound pressure level (SPL), specified prior to convolution with the microphone impulse response for a given test condition, while the level for each masker sentence was adapted during each block of trials according to a one-up one-down procedure that estimates the 50% correct point on the psychometric function. A T/M of 0 dB means that the level of each individual masker was equal to the level of the target. Each block began with a T/M of 10 dB and an initial adaptive step size of 6 dB, which was then reduced to 3 dB after the first three reversals were obtained. Each adaptive block continued until at least 20 trials and at least 9 reversals were completed, and the average of the last 6 reversals was defined as the T/M at threshold for that block of trials. The average number of trials required to obtain an adaptive threshold estimate was 22.

B. Results

Figure 4 shows the group mean masked speech reception thresholds as a function of the symmetric masker locations specified in azimuth. The parameter of the graph is the microphone condition which, for the triple beam, specifies the acoustic look directions of the side beams and also includes the KEMAR and beam results.

For the KEMAR and beam conditions, the patterns of thresholds that were observed were usually consistent with past work (e.g., Kidd *et al.*, 2015; Wang *et al.*, 2020; Yun *et al.*, 2019) for the KEMAR and beam conditions. The thresholds obtained in the KEMAR condition fell in a relatively narrow range bounded by about −22.5 to −25 dB T/M. The relatively low T/Ms at threshold reflect, in part, the three of four correct criterion for counting a trial correct (cf. Swaminathan *et al.*, 2016) and for this set of subjects, indicates sharp spatial tuning that reaches a maximum by about ±30° (cf. Yun *et al.*, 2019). For the beam, thresholds for the wider masker spacings were lower than for the KEMAR condition extending down to about −32 dB. For the triple beam conditions, the masked thresholds generally improved as the masker spatial separation increased. For both the beam and triple beam, the acoustic attenuation provided by the beamformer tended to increase over the range of masker azimuths tested while the spatial filters internal to the listeners, as revealed by the KEMAR condition, reached a maximum attenuation by about ±30°. For each side beam angle, the lowest threshold for triple beam was found for the ±90° masker separation with the exception of the ±90° side beam angles where an upturn in threshold was apparent as the masker angles increased from ±60°. The upturn in threshold for the ±90° side beam angles and the flattening of the threshold curve for the ±60° side beam angles as the masker separation approached ±60° likely reflects an increase in the masker levels as the location of the masker equaled the orientation of the side beams. For the ±90° masker separation, the ordering of thresholds was negatively related to the degree of separation of the side beam angles: the lowest thresholds occurred for the narrowest side beam angles with progressively higher thresholds found as the side beam angles grew wider. These very low T/Ms at threshold for the narrowest side beam angles and widest masker separations were likely a combination of both the acoustic effect of the mixing of the center and side beams as well as the enhanced perceptual segregation of the different sources. The Appendix shows preliminary data from two subjects for a

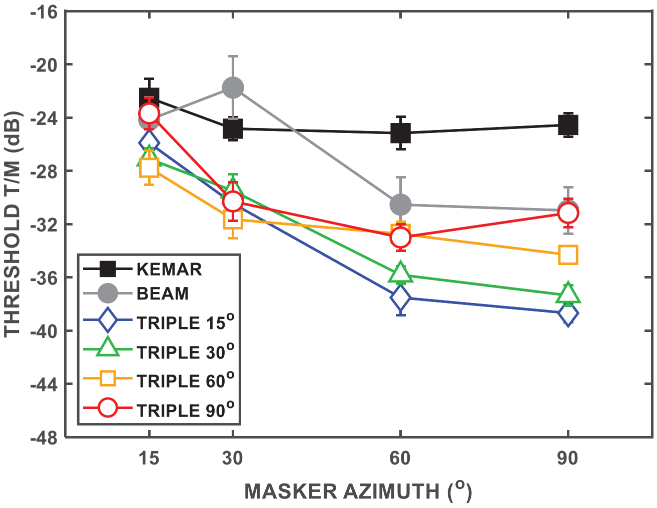

FIG. 4. (Color online) Group mean speech reception thresholds plotted as a function of (symmetric) masker location in degrees azimuth. The target speech was always presented at 0° azimuth. Error bars for this figure and all subsequent figures, unless stated otherwise, represent the standard error of the mean.

monaural control (side beam angles of ±30°), suggesting that for those limited conditions, monaural listening also yielded low thresholds that decreased as the masker separation increased. When compared to the group mean binaural data shown in Fig. 4, it appeared that adding the second ear further lowered thresholds by about 10 dB. However, these data await confirmation in a more extensive/rigorous experimental study. A further question concerns the magnitude of spatial release from masking as traditionally referenced to colocated thresholds; as noted above, we did not measure that condition here but note that very similar conditions yielded thresholds closer to 0 dB T/M in the study by Yun *et al*. (2019) for NH control subjects.

The group mean results from the different microphone conditions averaged across all masker separations are displayed in Fig. 5. Here, the average T/Ms at threshold (ordinate) are plotted for each microphone condition, including the KEMAR, beam and triple beam conditions with four side beam angles ranging from ±15° to ±90°. A repeated-measures analysis of variance (ANOVA) was performed on the data having the main factors of microphone condition and masker angle. The results revealed that both the microphone condition [$F(5,25) = 77.45$, $p < 0.001$] and masker azimuth [$F(3,15) = 112.20$, $p < 0.001$] were significant as was the interaction of the two [$F(15,75) = 5.68$, $p < 0.001$]. These effects were clear from the inspection of Fig. 4, including the interaction of the microphone and masker angle as the functions increasingly diverged as the masker separation increased.

IV. EXPERIMENT 2: SOURCE LOCALIZATION

The ability to locate sound sources in the environment is important in many social listening situations, and the extent to which various hearing aid algorithms—including those with beamforming as a primary component (e.g., Doclo *et al.*, 2008; Chalupper *et al.*, 2011; Bissmeyer and Goldsworthy, 2017; Hauth *et al.*, 2018; Wang *et al.*, 2020)—can preserve natural localization may be an important consideration in weighing the overall benefits of the different approaches. In experiment 2, localization judgments were obtained under each of the microphone conditions tested in experiment 1 for a stimulus consisting of randomized single words.

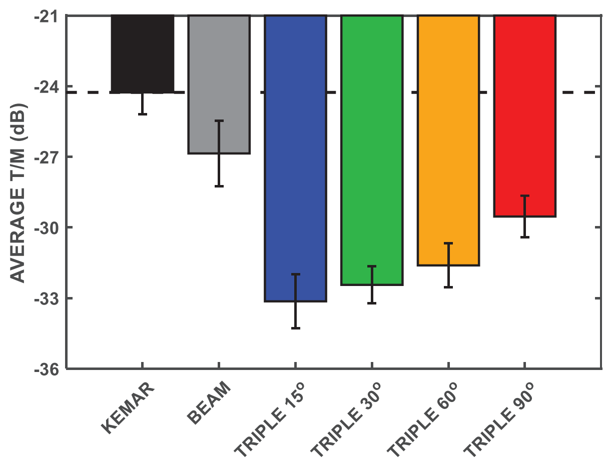

FIG. 5. (Color online) Group mean thresholds for the various microphone conditions. The graph shows the average threshold computed across all masker angles. The horizontal dashed line indicates the threshold for the KEMAR condition for reference.

A. Procedures

The same six listeners who participated in the SOS masking conditions in experiment 1 also participated in the single source localization conditions in experiment 2. The same corpus of words was used for localization with each trial consisting of a random selection with replacement of a single word from the entire matrix of words without regard to syntactic category. Also, each stimulus was randomly selected with replacement from the set of 11 female talkers used in experiment 1.

The stimulus level was fixed at 55 dB SPL specified prior to convolution with the microphone impulse response for a given test condition. The same microphone conditions as in experiment 1, KEMAR, beam, and triple beam, were tested in experiment 2. Also, for the triple beam condition, the same side beam acoustic look directions were tested, along with the addition of ±45° and ±75°, resulting in a total of eight microphone conditions. The target was presented at one location chosen at random with replacement from the range of −90° to 90° in 15° steps. Following the stimulus presentation, the listener registered a response using a GUI by mouse clicking the point on a semicircle corresponding to the perceived location of the source.

The experimental session consisted of 40-trial blocks with each of the 8 microphone conditions and 13 target azimuth locations ultimately tested a minimum of 10 times. One repetition of all of the conditions was obtained before proceeding to the next repetition of conditions, but the order was randomized within each repetition. Response feedback was not given during the task itself but during a familiarization period prior to data collection, the subject was presented with the full range of source azimuths paired with the visual representations of the location on the semicircular response GUI.

B. Results

The group mean sound source localization judgments are plotted in Fig. 6. For the KEMAR condition, localization judgments varied in a manner that was roughly ogive shaped with compression at the extreme locations and exaggerated lateral judgments for sources around 0° azimuth. For example, a 15° change in the location from 75° to 90° yielded very little change in the perceived azimuth with both locations judged to be near 75°. In contrast, the change in the physical location from 0° to 15° yielded a much larger change in the perceived location with the judged location for 15° falling near 30°. This pattern of responses is likely a consequence of the generally poor ability to distinguish differences in the source azimuth for angles in the range

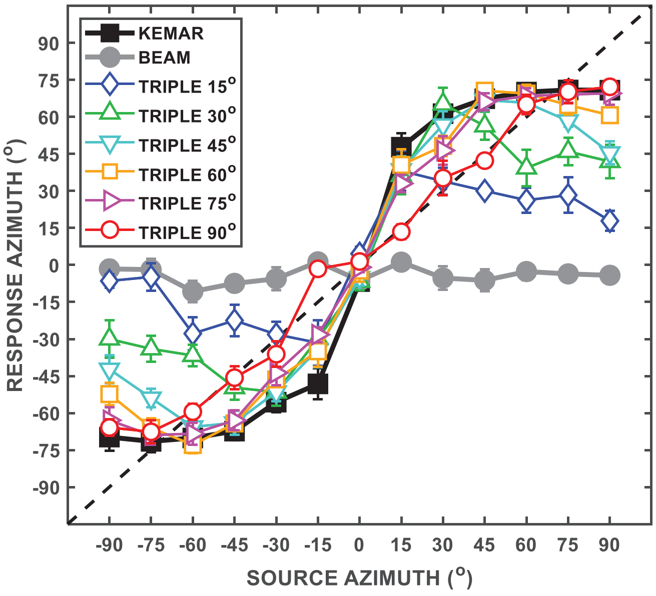

FIG. 6. (Color online) Group mean localization judgments for a single source ranging in location from −90° to +90° in 15° steps. The actual location of the target [using head-related impulse responses (HRIRs) to spatialize the stimuli] is given on the abscissa and the mean location judgment is given on the ordinate. The legend indicates the microphone condition for KEMAR, Beam, and six triple beam side beam angles.

tested (i.e., near 90°, e.g., the "cone of confusion," Mills, 1972) combined with a general tendency for observers to compress judgments toward the mean of a fixed range of values especially when response feedback is not provided. Furthermore, judgments of the source location that are obtained for stimuli convolved with head-related transfer functions (HRTF) and presented through headphones also tend to produce localization functions having this general form (cf. Wang *et al.*, 2020; Yun *et al.*, 2019), perhaps because of limited externalization of sound sources for HRTFs that are not accurately individualized (e.g., Best *et al.*, 2020). For this reason, the data are considered with respect to both the actual source location and the location as judged by listening through the KEMAR HRTFs. Of the various algorithms, the triple beam with side beam orientations at ±90° yielded the location judgments that were closest to the diagonal (actual location presented).

At the other extreme, the diotic beam condition provided no interaural differences, and the resulting perceived location function was essentially flat across the source azimuths. This result is also in line with past work (e.g., Best *et al.*, 2017a; Wang *et al.*, 2020; Yun *et al.*, 2019) for the triple beam. The localization judgments varied considerably with side beam orientation particularly with respect to the apparent location of the more extreme source angles. As a rough summary, the perceived locations of the sources presented from the wider angles (e.g., −90° or 90°) were compressed toward the center for the narrow side beam angles and progressively shifted toward the true location as the side beams were spaced farther apart.

The group mean localization error is displayed in two ways in Fig. 7. In the left two panels of the four-panel plot, the error is referenced to the intended source location as presented. In the right two panels, the error is computed referenced to the mean location judgments for the KEMAR condition with each subject referenced to her own average KEMAR judgments. In the upper row, the absolute value of the response error is plotted as a function of source location as presented. Thus, for example, in the upper left panel, the error for the beam is nearly equal to the distance of the source azimuth from 0° because the localization function (cf. Fig. 6) did not change appreciably as the source location was varied. The lower row of panels shows the group mean error averaged across all masker angles for each of the functions in the upper row in bar graph form. In the lower left panel, a dashed line indicates the average localization error for the KEMAR condition; the lower right panel references the error to the KEMAR function. As expected, the beam yielded the highest overall localization error followed by the two narrower triple beam side angle orientations. The two widest side beam angles, ±75° and ±90°, produced the lowest error overall for both references: the error was lower than the KEMAR condition in the lower left panel, and ±75° was the lowest of all of the conditions plotted in the lower right panel. A repeated-measures ANOVA conducted on the localization error results (the upper left panel of Fig. 7) revealed significant main effects of the microphone condition, [$F(7,35) = 60.43$, $p < 0.001$], and source azimuth, [$F(6,30) = 25.96$, $p < 0.001$], as well as a significant interaction of the microphone condition and source azimuth [$F(42,210) = 21.35$, $p < 0.001$].

V. DISCUSSION

A. Summary of findings

The results of the current study suggest that the triple beam algorithm can produce lower thresholds in SOS masking conditions for spatially separated sources than either natural binaural listening (simulated by the KEMAR HRTFs) or the single-channel diotic beamformer (beam). Furthermore, the triple beam algorithm yielded localization accuracy for a single speech source that was comparable to that which was found under the KEMAR conditions. The extremes of the different acoustic look directions for the triple beam side beams produced markedly different patterns of results on these two tasks: masked thresholds and localization. The narrowest side beam angles yielded the lowest (best) masked thresholds while producing the highest (worst) localization error. Conversely, the widest side beam angles yielded the highest masked thresholds but lowest localization error while the intermediate side beam angles generally produced an intermediate performance for these tasks. This variation in the advantage conferred by these different side beam angles relative to natural binaural listening and single-channel beamformer implies that should the triple beam algorithm be incorporated into the design of an assistive listening device, the option of adapting/adjusting

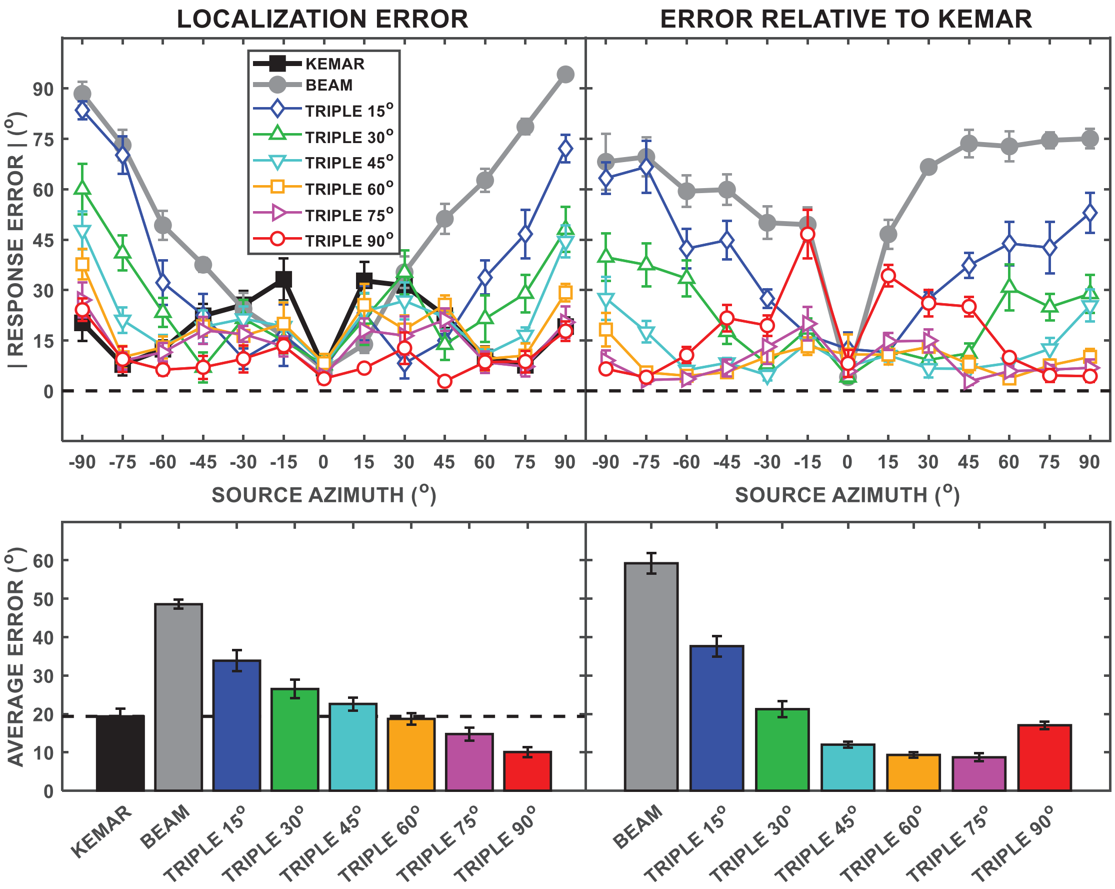

FIG. 7. (Color online) Group mean localization error plotted in four ways. In the upper left panel, the absolute value of the group mean error in degrees is plotted for each microphone condition (parameter of graph) as a function of the source azimuth. In the upper right panel, the same data are replotted with the error computed in reference to the KEMAR location judgments. In the lower left panel, the group mean average localization error is plotted as bars for each microphone condition averaged across all source angles. In the lower right panel, a similar bar graph of the group mean error is shown referenced to the judged location in the KEMAR microphone condition.

the properties of the device (e.g., side beam angles and gain) could be beneficial for optimizing amplification for different tasks and different listening environments and, perhaps, could be adjusted to account for the abilities of different listeners too. Furthermore, these trends serve to highlight the differences between the mechanisms underlying the enhancements in the performance across the microphone conditions observed here. For normal binaural hearing, it might be expected that improved localization accuracy would accompany enhanced source segregation likely resulting in lower masked thresholds. Finding the opposite trend for the beamforming conditions (e.g., the relationship between side beam angles for triple beam and the patterns of thresholds/localization noted above), in part, reflects the improvements in the S/N monaurally that occur from beamforming (as discussed in the Appendix) and not improved source segregation due to spatial separation *per se*. For example, the lowest masked thresholds occurred for the narrowest side beam angles, which provided monaural S/N improvements comparable to the beam condition (see Fig. 11), yet produced the smallest ILDs especially at the low frequencies (cf. Figs. 2 and 3). In contrast, the widest side beam angles produced the best localization performance with large ILDs at the wider masker azimuths observed across the different octave band center frequencies (cf. Fig. 3) with relatively small improvements observed in the monaural S/Ns (see Fig. 11). Thus, in that case, the exaggerated ILDs likely improved thresholds because of better segregation from the enhanced binaural cues. Section V B is a consideration of some of the factors that could influence the development of a composite performance metric for comparing across stimuli/conditions and gives an example of a simple equally weighted measure of combined performance for the tasks used here.

B. Factors governing overall performance

Consideration of the findings of experiments 1 and 2 leads to the question as to which approach provided the best performance overall. The greatest reduction in masking in the SOS tests—the lowest masked thresholds—in experiment 1 was produced by the triple beam algorithm. All of the triple beam masked thresholds were lower than the corresponding thresholds for the KEMAR and beam conditions with the lowest thresholds overall found for the ±15° and ±30° side beam angles. However, a different hierarchy of performance was observed for the single source localization in experiment 2. Here, the narrower side beam angles produced the greatest localization error (which was still less than the beam but greater than the KEMAR condition) while the widest side beam angles yielded the smallest error—even less than KEMAR.

An initial step in considering the relative merits of the different microphone algorithms for the two tasks used in this study is to visualize the joint performance along the relevant dimensions. Figure 8 is a plot of the group mean localization error averaged over all source locations (specified in degrees) plotted as a function of the group mean SOS masked thresholds (in dB T/M). The left panel shows the group mean performance on these two measures using the intended location (i.e., as presented via HRTFs) as the reference in the computing error while the right panel references the error to the group mean judgments obtained for the KEMAR condition.

SDs across subjects are shown in each dimension. The best joint performance would lie in the lower left corner of the plot, whereas the worst performance would lie in the upper right corner for both panels (cf. Wang *et al.*, 2020, for a similar plot of results from the related beamforming algorithms). In general, it is clear that the beam condition yielded the poorest joint performance while some triple beam conditions fared the best. There is overlap in the error bars that is apparent in some cases, especially along the abscissa (intelligibility).

A better descriptive approach to estimating the joint performance may be obtained by converting the group mean thresholds to standard normal deviates (i.e., *z*-scores) and taking the sum of the values observed for each task. It should be noted that taking the sum of the *z*-scores weights performance on the two tasks equally, which may not be a reasonable assumption in many cases (e.g., localization accuracy may be more important to the listener than an improvement in intelligibility in a given situation). This may be particularly true for some cases here where the T/Ms are much lower than are typically found in normal conversation (e.g., Weisser and Buchholz, 2019) and the threshold differences may not be as meaningful as would be the case for higher T/Ms. Figure 9 plots this simple sum of the *z*-scores metric.

For the SOS masking experiment, the positive *z*-scores reflected lower masked thresholds while for the single source localization task, positive *z*-scores were associated with lower errors. As in the localization error descriptions presented above, computations were made in reference both to the intended/presented location and the location judgments relative to KEMAR. For instance, the mean localization error values for all subjects and all conditions (plotted in the lower left panel of Fig. 7) were converted to *z*-score units by subtracting the mean of all values used for that plot and then dividing by the SD. The same computation was repeated for the localization error relative to the KEMAR condition. Note that the sum of *z*-scores in each plot equals zero. First, considering the data shown in the left panel of Fig. 9, the poorest joint performance was found for the beam algorithm due primarily to the large localization error in that condition. The other negative summed *z*-score was observed for the KEMAR condition, which did not yield masked thresholds as low as the beamformer-assisted conditions. All four triple beam conditions yielded positive values with the best overall performance apparent for the two widest side beam angles. When the reference for computing the localization error was the KEMAR condition, meaning that localization judgments were compared to the values obtained for that simulation of natural binaural cues, the

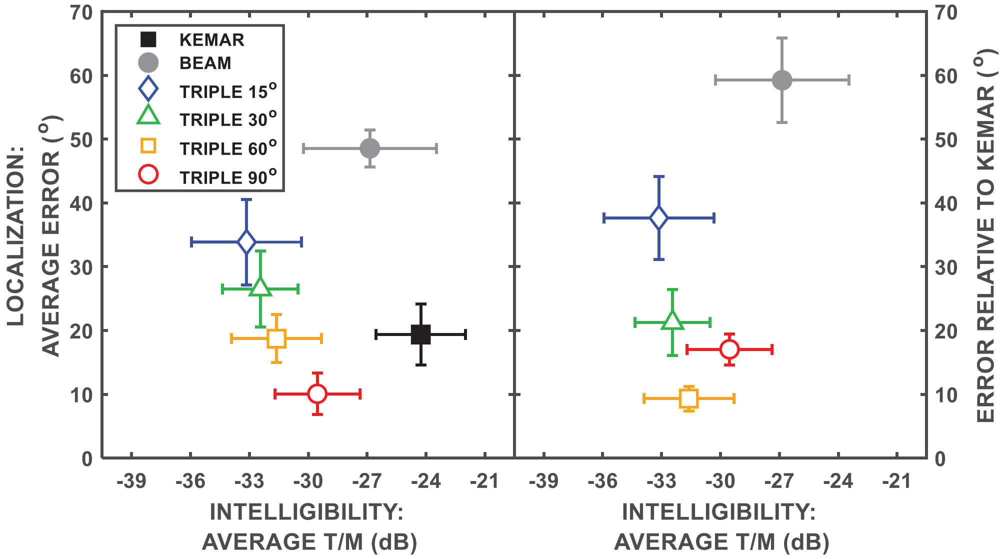

FIG. 8. (Color online) These panels show group means plotted to indicate joint performance on the SOS masking (intelligibility) task and single source localization task. The abscissa is the average target to masker ratio in dB for the SOS task while the ordinate is the average localization error in degrees. The error bars are plus and minus one SD of the mean in each dimension. The left panel contains the localization error computed relative to the actual source location while the right panel references the localization error to the localization judgments for the natural binaural simulation (KEMAR).

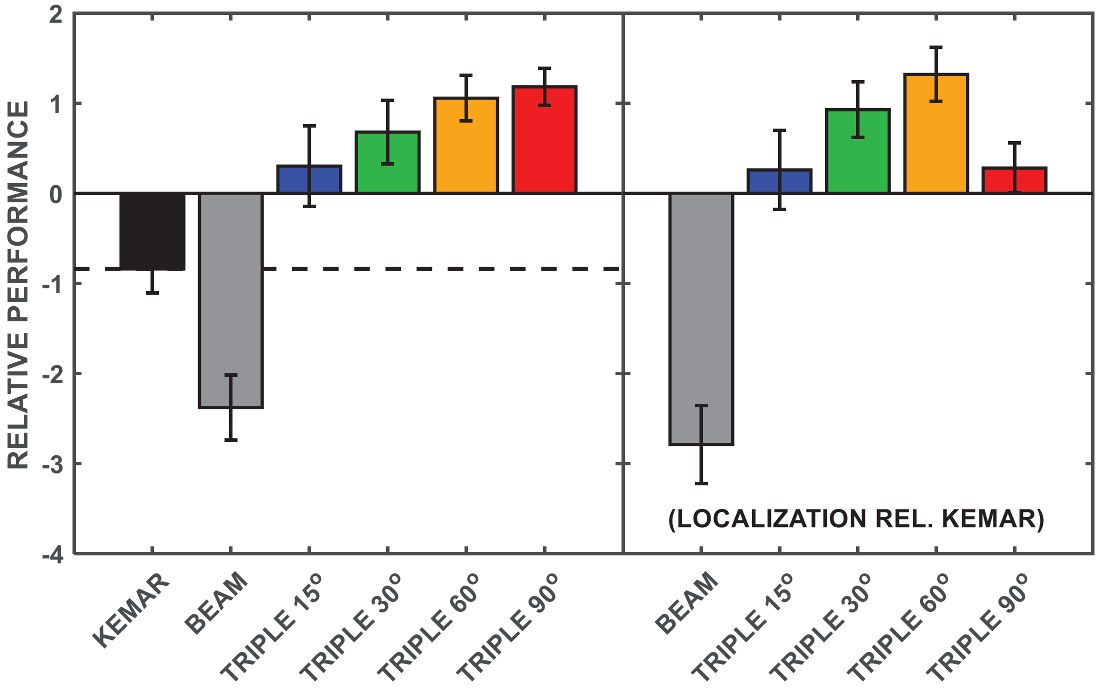

FIG. 9. (Color online) A plot of the joint performance based on the localization error and SOS masked thresholds. The abscissa shows the different microphone conditions while the ordinate is the summed z-scores for the two tasks. In the left panel, the mean error computation references the intended source location while in the right panel, the error is computed with respect to the group mean KEMAR judgments.

ranking of the combined z-scores changed somewhat with the best composite performance found for the triple beam ±60° side beam angles, followed by the ±30° side beam angles. It should be pointed out explicitly that the actual weighting of these two tasks—the importance of localization accuracy vs masked speech intelligibility—is likely highly context dependent. This would imply that the simple metric of joint performance used here (or any subsequent metric that takes more factors into account) ought to be weighted differentially when one task or the other is more important to the observer.

C. Limitations and future directions

The issues involved in fully answering the question about the best overall performance under realistic listening conditions are complex and were not addressed by the present study. One reason is that there are many listening tasks/stimuli that were not considered here and are likely important for typical communication situations. For example, the issue of a direct path acoustically to each ear that is overlaid on any processing from a device occluding the ear (e.g., an insert phone) alters the sound input and could affect the benefit of a beamformer. This issue has been addressed to some degree for the beam condition in past work from our group (e.g., "BEAMAR" algorithm; cf. Kidd *et al.*, 2015; Best *et al.*, 2017a) but has yet to be evaluated for the triple beam. A second reason is that the SOS masking task used here purposely emphasizes conditions high in IM (e.g., Kidd and Colburn, 2017) because the principle upon which the triple beam was designed was to foster improved sound source segregation, thereby reducing the IM. Conditions high in EM, such as Gaussian noise sources, may yield different patterns of results, especially for masker sources that are located at the acoustic look directions of the side beams or perhaps in listening environments that are diffuse/highly reverberant. A more complete understanding of the strengths and weaknesses of the triple beam algorithm in a range of listening situations awaits further investigation. Furthermore, the current approach does little to evaluate subjective dimensions of the signal processing, such as listening effort in speech masking (e.g., Rennies *et al.*, 2019) or the naturalness of speech quality. This latter dimension is also difficult to assess given the matrix style speech materials comprising individually recorded words concatenated to form syntactically correct sentences.

There could be practical implications for assessing the benefits/costs of the various microphone conditions. Recently, Yun *et al.* (2019) tested the triple beam algorithm for SOS masking and single source localization conditions for a group of bilateral cochlear implant subjects. In addition, they tested the beam and KEMAR conditions in a manner very similar to that tested here. They found that the triple beam algorithm yielded large improvements in T/M for SOS masking conditions relative to the KEMAR algorithm, although for the CI subjects, the improvements were greater for the beam than for the triple beam algorithms (triple beam thresholds were lower than the beam for NH control subjects in that study, consistent with the findings here). However, as in the current study, the localization performance for the beam condition was near chance while the source localization under the triple condition was better than either the KEMAR or beam conditions. That work follows prior studies in which beamforming or other signal processing schemes were proposed to enhance the localization abilities of hearing aid or cochlear implant users while preserving or bolstering speech intelligibility (e.g., Baumgärtel *et al.*, 2015; Best *et al.*, 2015; Moore *et al.*, 2016; Bissmeyer and Goldsworthy, 2017; Dieudonné and Francart, 2018). It is of interest to note that the side beam angles used in the study by Yun *et al.* (2019) with bilateral cochlear implant subjects and NH subjects listening to vocoded speech stimuli were oriented at ±40°, which likely was near the optimal acoustic look directions for maximizing the overall performance given the simple sum of *z*-scores metric used here.

One implication of the comparison of the relative benefits of the various approaches is that it suggests that a device that incorporates multiple algorithms that could be selected or adapted according to the circumstance might be useful. For example, a situation in which the listener wished to attend only to a single source without the need to monitor/localize competing sounds—especially if the competition comprised highly EM sources (e.g., noise)—might best be assisted with a single channel beamformer as in the beam algorithm. This appears to be the case for bilateral CI users in some conditions based on the results from Yun *et al.* (2019). In other circumstances, perhaps when monitoring multiple talkers or during rapidly changing "target" talkers during conversation, a triple beam system might perform better overall. Intermediate cases—for example, mixing the side beam inputs with the primary beam at different relative levels—might provide the best assistance. One observation that may be important for adapting the triple beam algorithm to difference circumstances is that the different acoustic look directions for the two complementary side beams had very different effects on the tasks of source segregation—as inferred by the SOS thresholds—and localization. Depending on the needs of the listener in a particular situation, the ability to enhance performance on one task or the other may be useful and possible through adjusting the side beams. Related, past work on the beamforming algorithm tested in this study has indicated that steering the beam under voluntary control of the listener using visual guidance (e.g., Kidd *et al.*, 2013; Best *et al.*, 2017b; Roverud *et al.*, 2018) can improve performance when the target source changes dynamically. We do not know whether similar benefits of visual guidance would also be obtained for the triple beam algorithm or whether there would be a different effect of side beam orientation.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH)-National Institute on Deafness and Other Communication Disorders (NIDCD) Grant No. R01DC013286 and U.S. Air Force Office of Scientific Research (AFOSR) Award No. FA9550-16-10372. The authors wish to express their appreciation to Christine Mason for her assistance with this project.

APPENDIX: MONAURAL CONTROL

The extent to which the benefit of the triple beam algorithm depends on binaural information, as opposed to simply enhancing the S/N in a single ear, was examined for a subset of conditions. Two young adult listeners with NH (including author A.J.B. and a member of the research staff in the laboratory) participated. Because these data were obtained during the period of time when our laboratory was closed due to the COVID-19 crisis, the findings shown below were acquired using laptop computers and sound cards in the homes of the two subjects and, thus, were attained under less stringent controls than the earlier results.

Other than differences in the equipment used and the listening environment, these monaural control data were measured using exactly the same procedures as in the prior experiments reported in Secs. II and III. The only other difference was the variable of interest here, namely, that stimuli were presented only from a single earphone. Thus, referring back to Fig. 1, the left ear input was identical to that shown in the schematic while the right ear input was disconnected. For this monaural control, only a single side beam angle—±30°—was tested. The masker separations were the same as in the earlier experiments. The KEMAR and triple beam algorithms also were tested. The results are shown in Fig. 10 in the same format as in Fig. 4.

It should be noted that these S/Ns fall well below those that are typical of normal conversation (e.g., Weisser and Buchholz, 2019). The low thresholds are due, in part, to the speech intelligibility measurement procedures used (see methods above) in which the adaptive rule tracked 50% correct responses with a correct response counted as selecting three of the four test words in each target sentence (cf. Swaminathan *et al.*, 2016).

Figure 11 shows the monaural S/N (target level re level of combined maskers) for all of the microphone conditions and masker angles. The T/M prior to microphone processing

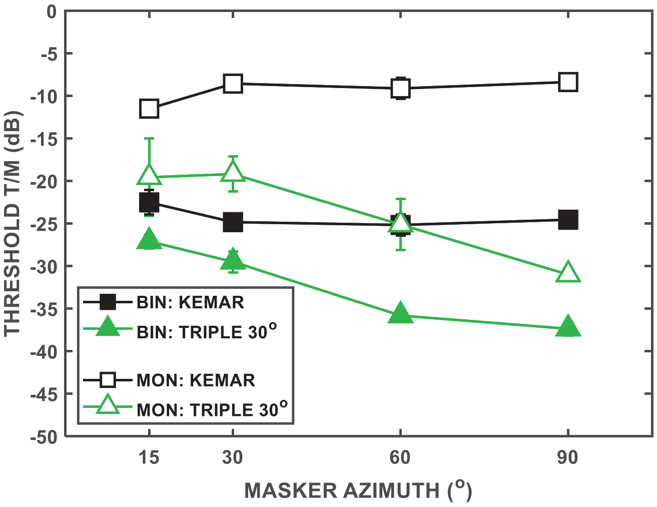

FIG. 10. (Color online) Mean thresholds for two subjects under monaural control conditions (open symbols) plotted along with the corresponding binaural thresholds reported earlier as group means in Fig. 4. For both monaural control and earlier binaural results, both KEMAR and (a single) triple beam microphone conditions are shown.

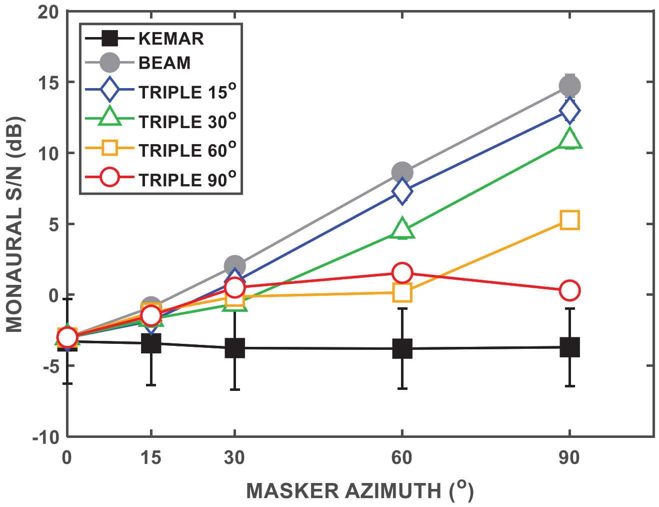

FIG. 11. (Color online) The monaural S/N (either ear) for a T/M of 0 dB plotted as a function of the spatial separation of the maskers for the different microphone conditions. Note that the difference in each point relative to the colocated case (0° separation; not tested in this study) provides an estimate of the magnitude of the improvement in the monaural S/N as the maskers are separated from the target.

(same as specified for the dependent variable in the experiments) was 0 dB. Note that this pattern of S/N values is the same for any of the T/Ms tested but would simply be translated along the ordinate. As expected, there is little change in the monaural S/N for the KEMAR algorithm so the benefits of spatial separation of the maskers are due almost exclusively to binaural processing. As noted above, we did not measure the colocated condition in this study and so cannot directly determine the spatial release from masking. The results in Figs. 10 and 11 demonstrate that monaural listening through the triple beam algorithm can provide a substantial advantage compared to KEMAR listening. Furthermore, the relative contribution of the improved monaural S/N to enhanced binaural cues varies according to the condition with the beam and KEMAR at the two extremes and the different triple beam side beam angles forming intermediate cases. The difference here between our two monaural listeners and the corresponding group mean thresholds was about 10 dB across source locations. Potentially, then, the triple beam algorithm could enhance the speech reception even for a single ear, although presumably the source localization would be poor.

[1] For this situation, we will use the more precise designation of target-to-masker ratio, T/M, in much of the discussion rather than S/N with T/M referring to the level of the target relative to each individual masker (i.e., 0 dB T/M corresponds to −3 dB S/N, where *N* is the sum of the two presumably uncorrelated speech masker waveforms). This is because the casual use of the term noise to mean any unwanted sound in the popular literature is distinct from actual Gaussian noise, which has a well-known scientific definition and composition according to specific statistical properties.

Adiloğlu, K., Kayser, H., Baumgärtel, R. M., Rennebeck, S., Dietz, M., and Hohmann, V. (**2015**). "A binaural steering beamformer system for enhancing a moving speech source," Trends Hear. **19**, 1–13.

Baumgärtel, R. M., Hu, H., Krawczyk-Becker, M., Marquardt, D., Herzke, T., Coleman, G., Adiloglu, K., Bomke, K., Plotz, K., Gerkmann, T., Doclo, S., Kollmeier, B., Hohmann, V., and Dietz, M. (**2015**). "Comparing binaural signal pre-processing strategies II: Speech intelligibility of bilateral cochlear implant user," Trends Hear. **19**, 1–18.

Best, V., Baumgartner, R., Lavandier, M., Majdak, P., and Kopčo, N. (**2020**). "Sound externalization: A review of recent research," Trends Hear. **24**, 1–14.

Best, V., Mason, C. R., and Kidd, G., Jr. (**2011**). "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers," J. Acoust. Soc. Am. **129**, 1616–1625.

Best, V., Mejia, J., Freeston, K., van Hoesel, R. J., and Dillon, H. (**2015**). "An evaluation of the performance of two binaural beamformers in complex and dynamic multitalker environments," Int. J. Audiol. **54**, 727–735.

Best, V., Roverud, E., Mason, C. R., and Kidd, G., Jr. (**2017a**). "Examination of a hybrid beamformer that preserves auditory spatial cues," J. Acoust. Soc. Am. **142**, EL369–EL374.

Best, V., Roverud, E., Streeter, T., Mason, C. R., and Kidd, G., Jr. (**2017b**). "The benefit of a visually guided beamformer in a dynamic speech task," Trends Hear. **21**, 1–11.

Best, V., Streeter, T., Roverud, E., Mason, C. R., and Kidd, G., Jr. (**2016**). "A flexible question-answer task for measuring speech understanding," Trends Hear. **20**, 1–8.

Best, V., Thompson, E., Mason, C. R., and Kidd, G., Jr. (**2013**). "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the spectral overlap of competing talkers," J. Acoust. Soc. Am. **133**, 3677–3680.

Bissmeyer, S. R. S., and Goldsworthy, R. L. (**2017**). "Adaptive spatial filtering improves speech reception in noise while preserving binaural cues," J. Acoust. Soc. Am. **142**, 1441–1453.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Brouwer, S., Van Engen, K., Calandruccio, L., Dhar, S., and Bradlow, A. (**2012**). "Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content," J. Acoust. Soc. Am. **131**(2), 1449–1464.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Calandruccio, L., Bradlow, A., and Dhar, S. (**2014**). "Speech-on-speech masking with variable access to the linguistic content of the masker speech for native and nonnative English speakers," J. Am. Acad. Audiol. **25**(4), 355–366.

Calandruccio, L., Dhar, S., and Bradlow, A. R. (**2010**). "Speech-on-speech masking with variable access to the linguistic content of the masker speech," J. Acoust. Soc. Am. **128**, 860–869.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969a**). "Release from multiple maskers: Effects of interaural time disparities," J. Acoust. Soc. Am. **45**, 411–418.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969b**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Chalupper, J., Wu, Y. H., and Weber, J. (**2011**). "New algorithm automatically adjusts directional system for special situations," Hear. J. **64**, 26–33.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and two ears," J. Acoust. Soc. Am. **25**, 975–979.

Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., and Kidd, G., Jr. (**2016**). "Executive function, visual attention and the cocktail party problem in musicians and non-musicians," PLoS One **11**, e0157638.

Culling, J. F., and Stone, M. A. (**2017**). "Energetic masking and masking release," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Springer Nature, New York), pp. 41–73.

Desloge, J. G., Rabinowitz, W. M., and Zurek P. M. (**1997**). "Microphone-array hearing aids with binaural output. I. Fixed-processing systems," IEEE Trans. Speech Audio Process **5**, 529–542.

Dieudonné, B., and Francart, T. (**2018**). "Head shadow enhancement with low-frequency beamforming improves sound localization and speech perception for simulated bimodal listeners," Hear. Res. **363**, 78–84.

Doclo, S., Gannot, S., Moonen, M., and Spriet, A. (**2008**). "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing*

https://doi.org/10.1121/10.0002779

and Sensor Networks, edited by S. Haykin and K. R. Liu (Wiley, New York), Chap. 9.

Ellinger, R. L., Jakien, K. M., and Gallun, F. J. (**2017**). "The role of interaural differences on speech intelligibility in complex multi-talker environments," J. Acoust. Soc. Am. **141**, EL170–EL176.

Favre-Felix, A., Graversen, C., Hietkamp, R. K., Dau, T., and Lunner, T. (**2018**). "Improving speech intelligibility by hearing aid eye-gaze steering: Conditions with head fixated in a multitalker environment," Trends Hear. **22**, 1–11.

Favrot, S., Mason, C. R., Streeter, T., Desloge, J., and Kidd, G., Jr. (**2013**). "Performance of a highly directional microphone array in a reverberant environment," in *Proceedings of the International. Conf. on Acoustics/Acoustical Society of America*, Montreal, Canada, Vol. 18, 8 pp.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**(5), 2112–2122.

Gardner, B., and Martin, K. (**1994**). "HRTF measurements of a KEMAR dummy-head microphone," MIT Media Labs, available at https://sound.media.mit.edu/resources/KEMAR.html (Last viewed 9/1/2019).

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Audiol. **11**, 79–87.

Hauth, C. F., Gößling, N., and Brand, T. (**2018**). "Performance prediction of the binaural MVDR beamformer with partial noise estimation using a binaural speech intelligibility model," Speech Commun. **10**, 301–305.

Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear Hear. **29**, 87–98.

Hladek, L., Porr, B., and Owen Brimijoin, W. (**2018**). "Real-time estimation of horizontal gaze angle by saccade integration using in-ear electrooculography," PLoS One **13**, e0190420.

Hoover, E. C., Souza, P. E., and Gallun, F. J. (**2017**). "Auditory and cognitive factors associated with speech-in-noise complaints following mild traumatic brain injury," J. Am. Acad. Audiol. **28**, 325–339.

Jennings, T., and Kidd, G., Jr. (**2018**). "A visually guided beamformer to aid listening in complex acoustic environments," Proc. Mtgs. Acoust. **33**, 1–8.

Kidd, G., Jr. (**2017**). "Enhancing auditory selective attention using a visually guided hearing aid," J. Speech, Lang. Hear. Res. **60**, 3027–3038.

Kidd, G., Jr., Best, V., and Mason, C. R. (**2008b**). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.

Kidd, G., Jr. and Colburn, H. S. (**2017**). "Informational masking in speech recognition," in *The Auditory System at the Cocktail Party*, edited by J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Springer Nature, New York), pp. 75–109.

Kidd, G., Jr., Favrot, S., Desloge, J., Streeter, T., and Mason, C. R. (**2013**). "Design and preliminary testing of a visually-guided hearing aid," J. Acoust. Soc. Am. **133**, EL202–EL207.

Kidd, G., Jr., Mason, C. R., and Best, V. (**2014**). "The role of syntax in maintaining the integrity of streams of speech," J. Acoust. Soc. Am. **135**, 766–777.

Kidd, G., Jr., Mason, C. R., Best, V., Roverud, E., Swaminathan, J., Jennings, T., Clayton, K. K., and Colburn, H. S. (**2019**). "Determining the energetic and informational components of speech-on-speech masking in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **145**, 440–457.

Kidd, G., Jr., Mason, C. R., Best, V., and Swaminathan, J. (**2015**). "Benefits of acoustic beamforming for solving the cocktail party problem," Trends Hear. **19**, 1–15.

Kidd, G., Jr., Mason, C. R., Best, V., Swaminathan, J., Roverud, E., and Clayton, K. (**2016**). "Determining the energetic and informational components of speech-on-speech masking," J. Acoust. Soc. Am. **140**, 132–144.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2008a**). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Science and Business Media, New York), pp. 143–190.

Lotter, T., and Vary, P. (**2006**). "Dual-channel speech enhancement by superdirective beamforming," EURASIP J. Appl. Signal Process. **2006**, 1–14.

Marrone, N. L., Mason, C. R., and Kidd, G., Jr. (**2008**). "Effect of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," J. Acoust. Soc. Am. **124**, 3064–3075.

Middlebrooks, J. C., Simon, J. Z., Popper, A. N., and Fay, R. R., eds. (**2017**). *The Auditory System at the Cocktail Party* (Springer Nature, New York).

Mills, A. W. (**1972**). "Auditory localization," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 2, pp. 301–348.

Moore, B. C., Kolarik, A., Stone, M. A., and Lee, Y.-W. (**2016**). "Evaluation of a method for enhancing interaural level differences at low frequencies," J. Acoust. Soc. Am. **140**, 2817–2828.

Pollack, I., and Pickett, J. M. (**1958**). "Stereophonic listening and speech intelligibility against voice babble," J. Acoust. Soc. Am. **30**, 131–133.

Rennies, J., Best, V., Roverud, E., and Kidd, G., Jr. (**2019**). "Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and listening effort," Trends Hear. **23**, 1–21.

Rohdenburg, T., Hohmann, V., and Kollmeier, B. (**2007**). "Robustness analysis of binaural hearing aid beamformer algorithms by means of objective perceptual quality measures," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, pp. 315–318.

Roup, C. M., Post, E., and Lewis, J. (**2018**). "Mild-gain hearing aids as a treatment for adults with self reported hearing difficulties," J. Am. Acad. Audiol. **29**, 477–494.

Roverud, E., Best, V., Mason, C. R., Streeter, T., and Kidd, G., Jr. (**2018**). "Evaluating the performance of a visually guided hearing aid using a dynamic audio-visual word congruence task," Ear Hear. **39**, 756–769.

Schubert, E. D., and Schultz, M. C. (**1962**). "Some aspects of binaural signal selection," J. Acoust. Soc. Am. **34**, 844–849.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.

Stadler, R. W., and Rabinowitz, W. M. (**1993**). "On the potential of fixed arrays for hearing aids," J. Acoust. Soc. Am. **94**, 1332–1342.

Swaminathan, J., Mason, C. R., Streeter, T., Best, V., Roverud, E., and Kidd, G., Jr. (**2016**). "Role of binaural temporal fine structure and envelope cues in cocktail-party listening," J. Neurosci. **36**, 8250–8257.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V. A., Kidd, G., Jr., and Patel, A. D. (**2015**). "Musical training, individual differences and the cocktail party problem," Sci. Rep. **5**, 11628.

Villard, S., and Kidd, G., Jr. (**2019**). "The effects of acquired aphasia on the recognition of speech under energetic and informational masking conditions," Trends Hear. **23**, 1–22.

Wang, L., Best, V., and Shinn-Cunningham, B. G. (**2020**). "Benefits of beamforming with local spatial-cue preservation for speech localization and segregation," Trends Hear. **24**, 1–11.

Weisser, A., and Buchholz, J. M. (**2019**). "Conversational speech levels and signal-to noise ratios in realistic acoustic conditions," J. Acoust. Soc. Am. **145**, 349–360.

Williges, B., Jürgens, T., Hu, H., and Dietz, M. (**2018**). "Coherent coding of enhanced interaural cues improves sound localization in noise with bilateral cochlear implants," Trends Hear. **22**, 1–22.

Yun, D., Jennings, T. R., Mason, C., Kidd, G., Jr., and Goupell, M. J. (**2019**). "Benefits from different types of acoustic beamforming in bilateral cochlear-implant listeners," J. Acoust. Soc. Am. **145**(3), 1876–1877.