Benefits of Acoustic Beamforming for Solving the Cocktail Party Problem

Gerald Kidd Jr.[1], Christine R. Mason[1], Virginia Best[1], and Jayaganesh Swaminathan[1]

Abstract

The benefit provided to listeners with sensorineural hearing loss (SNHL) by an acoustic beamforming microphone array was determined in a speech-on-speech masking experiment. Normal-hearing controls were tested as well. For the SNHL listeners, prescription-determined gain was applied to the stimuli, and performance using the beamformer was compared with that obtained using bilateral amplification. The listener identified speech from a target talker located straight ahead (0° azimuth) in the presence of four competing talkers that were either colocated with, or spatially separated from, the target. The stimuli were spatialized using measured impulse responses and presented via earphones. In the spatially separated masker conditions, the four maskers were arranged symmetrically around the target at ±15° and ±30° or at ±45° and ±90°. Results revealed that masked speech reception thresholds for spatially separated maskers were higher (poorer) on average for the SNHL than for the normal-hearing listeners. For most SNHL listeners in the wider masker separation condition, lower thresholds were obtained through the microphone array than through bilateral amplification. Large inter-subject differences were found in both listener groups. The best masked speech reception thresholds overall were found for a hybrid condition that combined natural and beamforming listening in order to preserve localization for broadband sources.

Introduction

Among the more difficult communication situations for listeners with hearing loss are those which require them to follow the speech of one particular talker in the presence of multiple-competing talkers. Even when assisted by hearing aids, such complex acoustic environments may be formidable and frustrating, and attempts at spoken communication too often are met with limited success (e.g., Dawes, Munro, Kalluri, & Edwards, 2013; Noble & Gatehouse, 2006). When determined in the laboratory, the primary manifestation of this difficulty is higher than normal speech reception thresholds (SRTs) when competing masking talkers are spatially separated from the target talker. In that case, listeners with sensorineural hearing loss (SNHL) as a group demonstrate higher masked SRTs than normal-hearing (NH) listeners (e.g., Arbogast, Mason, & Kidd, 2005; Best, Marrone, Mason, & Kidd, 2012; Marrone, Mason, & Kidd, 2008b,c; Neher, Laugesen, Jensen, & Kragelund, 2011; Woods, Kalluri, Pentony, & Nooraei, 2013) with some individual SNHL listeners achieving thresholds at positive target-to-masker ratios indicating that they are only able to reliably report the target speech when it is the highest level source.

Typically, listeners with NH are able to rely on binaural information—interaural time and level differences—to locate sound sources and focus attention on a target source while ignoring unwanted masking sources. When both the target and maskers are speech, this ability depends on both peripheral and central processes. Although the magnitude of the benefit of binaural information in overcoming speech-on-speech (SOS) masking varies considerably across studies depending on the stimuli and procedures used, there is compelling evidence suggesting that listeners with SNHL typically do not achieve the same level of performance as their

[1]Boston University, MA, USA

Corresponding author:
Gerald Kidd Jr., Boston University, 635 Commonwealth Ave, Boston, MA 02215, USA.
Email: gkidd@bu.edu

NH counterparts. This difficulty also increases with increasing age (e.g., Gallun, Kampel, Diedesch, & Jakien, 2013; Glyde, Cameron, Dillon, Hickson, & Seeto, 2013).

There are a variety of signal processing strategies implemented by hearing aids that may assist the listener with SNHL. Frequency-specific gain and amplitude compression may improve audibility and loudness perception making the information in the target speech available at comfortable levels. Algorithms that implement noise reduction, which can attenuate unwanted sound sources, and directional amplification, which emphasizes a source originating from a specific azimuth relative to the head, are more directly related to improving signal-to-noise ratio (SNR) and enhancing source selection. Although noise reduction is effective for certain types of unwanted sounds, it inherently cannot assist the listener in choosing among competing talkers because only the listener knows which talker is the target and which are the maskers and that designation changes according to communication intent. On the other hand, directional amplification can provide a significant enhancement to sound source selection as long as the focus of amplification is directed toward the desired source, and the competing sources are sufficiently separated in azimuth (e.g., Gnewikow, Ricketts, Bratt, & Mutchler, 2009; Goldsworthy, 2014). Spatially selective amplification would seem to be particularly useful in multiple-talker situations where, for example, the target source changes frequently during turn-taking in conversation. Because such situations involve the shifting and focusing of selective attention, a system that could easily be steered toward the intended source while achieving a high degree of spatial tuning could be beneficial (e.g., Kidd, Favrot, Desloge, Streeter, & Mason, 2013).

Highly directional amplification, such as that implemented by an acoustic beamforming microphone array (e.g., Greenberg, Desloge, & Zurek, 2003; Greenberg & Zurek, 1992; Stadler & Rabinowitz, 1993), has been available for research purposes for some time. However, the benefits that such amplification could provide in multiple-talker listening situations (e.g., the "cocktail party problem"; see reviews in Bronkhorst, 2000; Carlile, 2014; Mattys, Davis, Bradlow, & Scott, 2012; Yost, 1997) have not been established, despite a great deal of interest recently in studying the ability of SNHL listeners to selectively attend in such situations. Picou, Aspell, and Ricketts (2014) demonstrated a benefit of beamforming under certain conditions for sentence recognition in the presence of unintelligible speech babble. In listeners with cochlear implants, Goldsworthy (2014) found 6.5 to 11 dB of spatial benefit from beamforming for target speech presented against a time-reversed speech masker (*noise*). Also, Kidd et al. (2013) reported large benefits for two listeners with

unilateral deafness using a beamforming microphone array (same as in this study, described later) for either two independent speech or two independent modulated-noise maskers spatially separated (symmetrically, one to either side) from the target location.

In the present study, we compared the benefit obtained from bilateral amplification with that obtained using a highly tuned (in azimuth) beamforming microphone array for the purpose of enhancing target speech reception in multiple-talker sound fields. This comparison is of interest primarily for evaluating possible approaches to the development of auditory prostheses for listeners with hearing loss. However, because past work also has revealed large intersubject differences even among NH listeners in multisource listening environments (e.g., Ruggles & Shinn-Cunningham, 2011; Swaminathan et al., 2015), we examined how beamforming amplification aids these listeners as well. The goal was to determine whether the approach of providing highly spatially selective amplification is more effective at solving the cocktail party problem than conventional bilateral amplification for certain subgroups or individual listeners.

The beamforming microphone array used in the current study is one component of the *visually guided hearing aid* (VGHA) described previously by Kidd et al. (2013). It is a fixed array, meaning that the algorithm that implements the directional pattern of amplification is static and does not change in response to the input. In contrast, beamforming arrays may be adaptive in that the algorithm implementing the directional response adapts dynamically according to the directional properties of the input signals. The relative advantages of these two approaches have been discussed elsewhere (e.g., Desloge, Rabinowitz, & Zurek, 1997; Greenberg et al., 2003; Welker, Greenberg, Desloge, & Zurek, 1997). The VGHA uses visual guidance to steer the orientation of the fixed beamformer; however, in the current study, the acoustic look direction (ALD) of the beam is fixed at $0°$ azimuth (directly in front of the listener, as described later).

There are two potential liabilities to directional amplification that must be considered together with the potential benefit that such an approach provides for enhancing source selection. First, a high degree of spatial selectivity may compromise the ability of the listener to broadly monitor the sound field for new or changing sources. Second, the single-channel output of a typical beamformer eliminates the natural binaural cues that allow listeners to localize and segregate sounds based on location. One solution to these problems was proposed by Desloge et al. (1997) who combined the output of a beamforming microphone array with binaural information obtained by two microphones mounted at the edges of the array confining the two types of input to nonoverlapping frequency

regions. However, this method, which was shown to maintain the ability of listeners to localize sound sources and yielded a modest (approximately 3 dB) improvement in SRT in a quasi-diffuse noise background relative to a diotic control, has not received much subsequent exploration nor has it been applied systematically to SOS masking for listeners with SNHL. In this study, we also tested a hybrid approach similar in concept to that used by Desloge et al. (1997) that combined the acoustic beamforming produced by a microphone array with natural binaural cues that support monitoring of the sound field outside of the beam of amplification. It was unclear, at the outset, whether such a hybrid system would preserve the large spatial release from masking (SRM) that would be expected from either approach separately. This is because the two types of information—a single channel from the beamformer that does not yield spatialization of sound sources and a true (bandlimited) binaural channel where the sound sources are accurately spatialized—might not be integrated perceptually to provide useful source segregation.

These approaches—natural binaural information (as simulated by KEMAR), single-channel (diotic) acoustic beamforming, and a hybrid combining both—were compared under conditions where competing talkers were either colocated with or spatially separated from a target talker in a speech identification task for both NH and SNHL listeners.

Methods

Listeners

A total of 15 listeners participated. Eight of the listeners had stable, bilaterally symmetric SNHL of mild to moderate severity sloping on average from about 25 dB HL at 250 Hz to about 75 dB HL at 8 kHz. The remaining listeners had normal audiometric thresholds. Three of the listeners (one NH and two SNHL) only participated in two of the three microphone conditions (described later) and are not included in the group mean analyses. The SNHL listeners ranged in age from 19 to 39 years (mean = 24.8, $SE = 2.5$) while the NH listeners ranged in age from 21 to 24 years (mean = 22.6, $SE = 0.4$). The group mean audiogram for all eight SNHL listeners is shown in Figure 1.

This research was reviewed and approved by the Boston University Institutional Review Board (protocols 2633E, 2670E, and 3409E). Written informed consent was obtained from all subjects prior to participation.

Stimuli

The stimuli presented on any given trial consisted of five concurrent five-word sentences. The speech materials
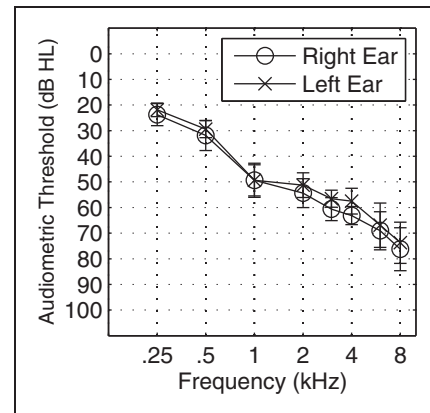

Figure 1. Group mean audiometric thresholds and standard errors of the means for the eight SNHL listeners.

were from a laboratory-developed closed-set corpus (Kidd, Best, & Mason, 2008b) comprising 40 items divided into the categories of *name*, *verb*, *number*, *adjective*, and *object* with eight exemplars within each category. The words were spoken individually with neutral inflection, and in this study, we used the subset of materials spoken by eight young–adult females. The five mutually exclusive target and masker talkers were selected randomly from this set on every trial. Each of the five sentences was presented in correct syntactic order with mutually exclusive selections from the exemplars in each category. The talker or voice and sentence were designated as the target by the first word *Sue*.

Procedures

The stimuli were played through a Tucker-Davis Technology 16-bit digital-to-analog converter at a 25-kHz rate, low-pass filtered at 10 kHz, and presented through Sennheiser 280 Pro headphones to the listener seated in a double-walled IAC booth. The levels of the targets and combined maskers for each ear were controlled separately by Tucker-Davis Technology programmable attenuators (PA4).

The target talker was presented from 0° azimuth (straight ahead), and the four masker talkers were either colocated with or symmetrically separated from the target at ±15° and ±30° or ±45° and ±90°. These spatial conditions were created by convolving the stimuli with the appropriate impulse responses recorded in our mildly reverberant sound field laboratory (standard IAC perforated metal walls and ceiling with carpeted floor; i.e., BARE room condition described in Kidd, Mason, Brughera, & Hartmann, 2005) from loudspeakers located at the source azimuths tested and at a distance of 5 feet.

There were two microphone conditions that were first tested individually and then in combination to produce a hybrid case. One condition, intended to approximate

natural binaural cues, is referred to as KEMAR. In this case, the impulse responses used in creating the stimuli were recorded through the standard microphones located in the ear canals of a KEMAR manikin. The next microphone condition is referred to as BEAM[1], and in this case, the impulse responses were recorded using the acoustic beamforming microphone array mounted on KEMAR's head (see Kidd et al., 2013, for a description of the microphone array used here). The microphone array consists of eight cardioid microphones arranged in four pairs mounted on a headband spanning the top of the head leaving the ears unobstructed. Each microphone in the pair is oriented on the front-to-back axis aimed toward the front (defined as 0° azimuth), and the two microphones in a pair are separated by 3 cm. The four pairs of microphones are spaced 7.1 cm apart so that the total array span is 21.3 cm. It is worth noting that the beamforming array has a specific spatial tuning characteristic (see Favrot, Mason, Streeter, Desloge, & Kidd, 2013; Kidd et al., 2013) that progressively attenuates signals that are further away from the ALD. As a result, the maskers positioned at ±15° and ±30° are attenuated less than the maskers

positioned at ±45° and ±90°. Using measurements from the impulse responses, the gain in broadband SNR provided by the beamformer (i.e., the attenuation of the combined maskers) is approximately 5 dB and 13 dB for the close and far masker configurations, respectively. For the hybrid condition, the output from KEMAR was low-pass filtered at 800 Hz and combined with the output of the BEAM that was high-pass filtered at 800 Hz. This microphone condition is referred to as BEAMAR.

The impulse responses for the KEMAR and BEAM conditions were equated for level at 0° azimuth. For BEAMAR, the levels were left exactly as they would be for the respective regions in KEMAR and BEAM, and the overall level at 0° azimuth was within 1 dB of the values for those conditions. The levels across frequency for each microphone condition and source location obtained from the measured impulse responses are shown in Figure 2. The two left panels are for the KEMAR microphone condition, the two middle panels are for the BEAM condition, and the two right panels are for the BEAMAR condition. The relative spectrum level at one-third octave intervals for the locations of 0°, −15°, −30°, −45°, and −90° are plotted in the upper row
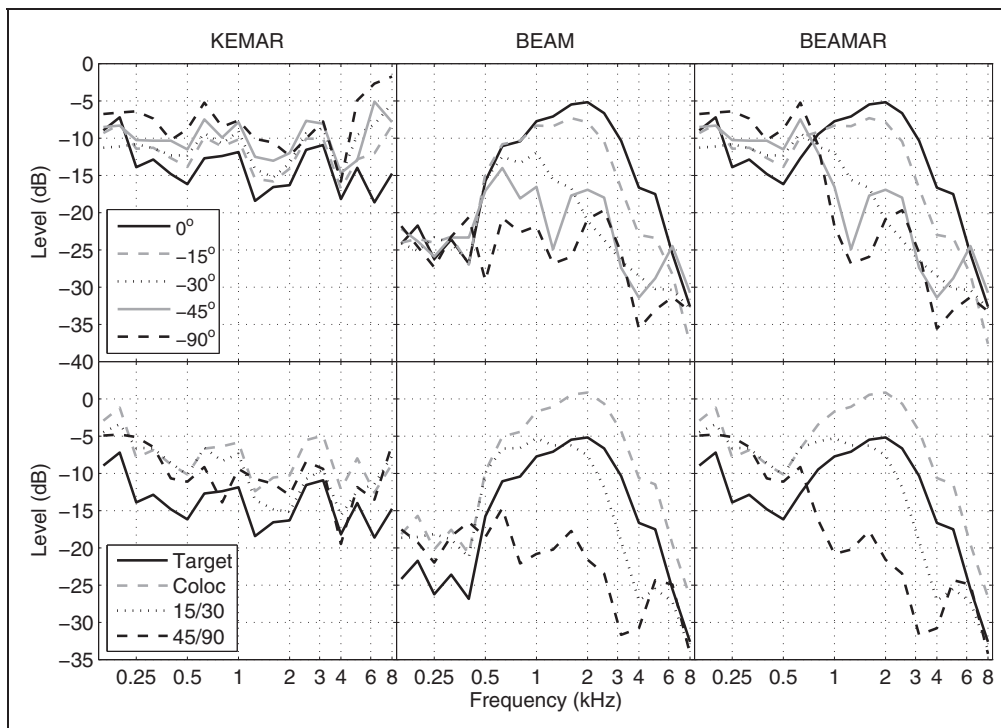

Figure 2. Frequency responses for the three microphone conditions tested: KEMAR, BEAM, and BEAMAR obtained from the measured head-related (KEMAR mannikin and microphone array) impulse responses. The abscissa is frequency while the ordinate is the magnitude of the response in decibels plotted at one-third octave intervals. The parameter is source azimuth in degrees. The three columns show the different microphone conditions (left to right): KEMAR, BEAM, and BEAMAR. The upper row of panels gives the responses for a single source at the different azimuths tested, and the lower row of panels shows the responses for the four-masker cases (at 0 dB T/M) that were colocated at 0° or symmetrically separated at ±15° and ±30°, or at ±45° and ±90°. The black solid lines are for a single (target) source at 0° azimuth and are the same for both panels within each column.

of panels for a single sound source. The lower row of panels shows similar plots for the target (same as 0° in the upper plots) and the four-source symmetrically arranged masker locations used in the experiment for a T/M of 0 dB. The beamformer (BEAM and BEAMAR above 800 Hz) was oriented toward 0° as it was in the experiment, and the inputs to each ear were identical. For KEMAR and BEAMAR (below 800 Hz), the responses in the upper row of panels are plotted for 0° and for the near ear (left ear, closest to the source). The far ear would be attenuated by head shadow (not shown), and of course, the stimulus arrives with interaural time differences in those cases but not for the beamformer. Also, in the lower panels, the difference between target and colocated simply reflects the fourfold increase in intensity for the summation of the four maskers compared with the single-target source. At lower T/Ms, the target spectrum would simply be shifted downward along the ordinate while the maskers remain the same. Of particular note is the increasing benefit of the beamformer with increasing frequency (up to about 6 kHz) for the separated locations apparent in the upper center panel for BEAM (see also Kidd et al., 2013 and Favrot et al., 2013).

Threshold T/Ms were measured by fixing the levels of the maskers and adapting the level of the target using a one-down one-up procedure that estimates the 50% correct point on the psychometric function. For the purposes of adapting the target level, a correct response was counted when three of the four target words (excluding Sue) were identified. The four masker sentences were each set to 55 dB sound pressure level (SPL; for the NH listeners) so that the combined four-talker masker had an expected level of 61 dB SPL. For the listeners with SNHL, individualized gain (NAL-RP formula used; cf. Byrne, Parkinson, & Newall, 1991) was applied to the stimulus input (target plus maskers) at each ear as if a hearing aid were being worn for the purpose of providing simple frequency-dependent linear gain. In this case, the pregain level of the individual masker talkers was 60 dB SPL, except for three listeners who thought those combined maskers were rather loud and so a 55 dB SPL input level was used. As a side note, a procedure using a fixed target level with variable maskers was also considered although in that case in order to perform well, even with a somewhat lower target level, a listener would need to be able to tolerate loud maskers. The limited dynamic range presents challenges at either end of the range of levels.

The target talker was presented at a starting level 10 dB above the level of the individual maskers. The step size was initially 6 dB but was reduced to 3 dB after the third reversal. A minimum of 20 trials and 9 reversals was required for a complete track. The last six reversals were averaged and recorded as the threshold estimate in dB. Four threshold estimates per subject per condition were obtained in all masked conditions. Final thresholds were reported as T/M in dB calculated by subtracting the fixed masker level from the target threshold level. This refers then to the level of the target relative to any single masker talker whereas the target level relative to the combined masker level for four equal-level talkers would be 6 dB poorer (e.g., if the target level at threshold is 55 dB and the masker level is 55 dB/talker, then the T/M at threshold is given as 0 dB whereas the SNR typically would be specified as −6 dB). SRM (the T/M for the colocated case minus the T/M for the spatially separated case in any condition) is also calculated and discussed.

Results

The individual and group mean results for all conditions in the study are contained in Tables 1 (NH) and 2 (SNHL). Group means are tabulated both for all subjects and for the subset of listeners who participated in all three microphone conditions.

Only the results from the six subjects in each group who completed all three microphone conditions will be considered here when discussing the group means. Figure 3 shows group mean T/Ms at threshold for NH (upper left panel) and SNHL (upper right panel) listeners. The thresholds are plotted for KEMAR, BEAM, and BEAMAR microphone conditions for three different masker location configurations: colocated, separated by ±15° and ±30°, and separated by ±45° and ±90°. Two general observations may be made about these results: first, in all cases, the thresholds for the NH listener group were lower than the corresponding thresholds for the SNHL group. Second, in all cases except one, the thresholds for BEAMAR were lower than for either KEMAR or BEAM microphone conditions for both subject groups. The one exception was for the narrowest masker spatial separation for the SNHL listeners where the values were nearly equal for KEMAR (−2.3 dB) and BEAMAR (−2.2 dB).

With respect to the results for the colocated maskers, all of the group mean thresholds for both groups and all three microphone conditions fell within the range of about −0.3 to 2.9 dB. This small range of values is typical of SOS masking experiments for colocated target or maskers using similar procedures (e.g., Best et al., 2012; Dawes et al., 2013; Gallun et al., 2013; Marrone, Mason, & Kidd, 2008a, 2008b). The interpretation of this result is that the listener tends to separate the talkers based primarily on level because of the difficulty with using other segregation cues effectively in this condition. The finding that the mean threshold T/Ms were slightly higher for the SNHL listeners than the NH listeners in the colocated case also has been noted before (Best et al.,
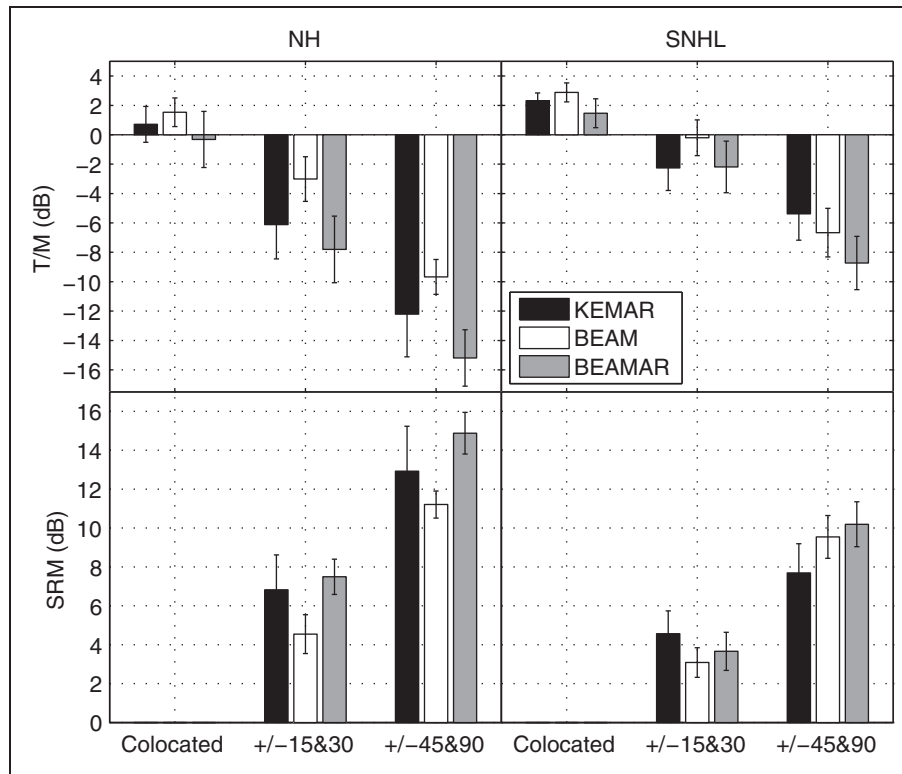
Figure 3. Upper two panels: Group mean *T/M*s in dB at threshold and standard errors for colocated and spatially separated maskers for KEMAR, BEAM, and BEAMAR presentation conditions for NH (left panel) and SNHL (right panel) listeners. The narrower masker spatial separation is ±15° and ±30°, and the wider masker spatial separation is ±45° and ±90°. These results were computed based on the six subjects in each group who were tested in all three microphone conditions. Lower two panels show the SRMs corresponding to the data contained in the upper panels. SRM is computed by subtracting the threshold *T/M* in the spatially separated condition from the threshold in the colocated condition. SRM = spatial release from masking.

2012; Marrone et al., 2008b). With respect to the spatially separated masker conditions, a more complex pattern of results emerged. For the narrowest masker spacing, the group mean threshold *T/M*s under KEMAR were lower than under BEAM in both cases. The thresholds for BEAMAR were lowest of all conditions except for the SNHL in one case as noted above. For the wider masker separation, a somewhat different pattern of results was found. Here, the threshold for the NH group for KEMAR (−12.2 dB) again was lower than for BEAM (−9.7 dB). However, for the SNHL group, this trend was reversed: The threshold for KEMAR (−5.4 dB) was marginally higher than that found with BEAM (−6.7 dB). This meant that under KEMAR, the SNHL listeners only benefitted by about 3 dB when the maskers were moved from the narrower separation (±15° and ±30°) to the wider (±45° and ±90°) separation. In contrast, they improved by 6.5 dB under BEAM for the same increase in source separations. For both NH and SNHL groups, the lowest thresholds in the wider masker spatial separation occurred for BEAMAR. In that case, the SNHL listeners achieved a

threshold of −8.7 dB which was about 3.3 dB lower than was found with KEMAR.

An analysis of variance applied to the *T/M*s at threshold revealed significant main effects of microphone (KEMAR, BEAM, and BEAMAR) [$F(2, 20) = 13.8$; $p < .001$] and spatial condition (colocated, narrow, and wide separations) [$F(2, 20) = 169.9$; $p < .001$] but not group (NH vs. SNHL) [$F(1, 10) = 3.17$; $p = .105$]. Pairwise comparisons for microphone condition indicated that KEMAR and BEAM were not significantly different from one another while BEAMAR was significantly different from both other microphone conditions (at the $p < .05$ level). The two-way interaction of spatial condition by group was significant ($p = .014$) and suggests that the effect of hearing loss depends on spatial condition, which is consistent with the observation that colocated performance is not much different for the SNHL group while larger differences are seen for the spatially separated cases. The interaction between microphone condition and spatial condition was also significant ($p = .015$) suggesting that the effect of microphone condition also depended on spatial configuration

and that too is seen in the smaller difference across microphone conditions for the colocated cases. The remaining two-way interaction of microphone condition by group was not significant at the $p = .05$ level and neither was the three-way interaction.

The lower panels of Figure 3 show the benefit of spatially separating the maskers from the target specified as SRM. In all cases, the SRMs are larger for the NH listeners than for the SNHL listeners. Also, with the exception of the narrower masker spacing for the SNHL group, the SRMs are largest for the BEAMAR microphone condition. Of particular importance is the greater benefit of the BEAM and BEAMAR conditions than the KEMAR condition for the SNHL listeners at the widest masker spatial separation. An analysis of variance applied to the SRMs shown in Figure 3 indicated significant main effects of subject group (NH vs. SNHL) [$F(1, 10) = 148$; $p = .037$] and spatial condition (narrow vs. wide separation) [$F(1, 10) = 260.9$; $p < .001$] with microphone condition (KEMAR, BEAM, and BEAMAR) not being significant. The only significant interaction was between microphone and spatial conditions [$F(2, 20) = 8.8$; $p = .002$].

Two observations about the BEAMAR condition should be made here: First, because the BEAMAR data were obtained after the KEMAR and BEAM conditions were completed, it is possible that the better performance observed for BEAMAR was due to learning or to order of testing. However, all 12 subjects were retested on the KEMAR condition for the wider separation as part of a separate study after the BEAMAR condition was completed. These data were not obtained using the adaptive procedure reported here but rather were measured at fixed T/Ms yielding psychometric functions. We fit these data with logistic functions and then did the same analysis with the trial-by-trial data gathered during the adaptive runs of this study and estimated thresholds from each condition. An analysis of variance conducted on these thresholds showed that they were not significantly different, suggesting that the better performance found for BEAMAR was not likely to reflect a general improvement due to learning effects. Second, informal listening has indicated that the spatialization of off-center sources is preserved under BEAMAR while it is obviously absent in the diotically presented BEAM (cf. Desloge et al., 1997). This raises the possibility that the generally superior performance observed with BEAMAR is due to the preservation of low-frequency spatial information (cf., Wightman & Kistler, 1992) which may have augmented the benefit of beamforming.

Inspection of the individual data contained in Tables 1 and 2 reveals large intersubject differences in thresholds particularly under KEMAR in the spatially separated conditions. Where comparable results exist, this pattern is consistent with previous work (e.g., Best et al., 2012; Marrone et al. 2008a, 2008b; Swaminathan et al., 2015). These differences in the intersubject variability across listening conditions are of interest because they are likely to reflect the viability of specific source segregation cues available to the listener. For the colocated case for KEMAR and for all of the BEAM conditions, the listener hears the stimulus with no interaural difference between sound sources. The generally lower intersubject variability may be related to the difficulty in segregating any one talker until that talker — i.e., the target — is higher in level than the other talkers. Once this occurs, that talker becomes more salient and is most likely to be reported by the listener. The BEAM simply changes the T/M at which this occurs when the sources are spatially separated. Thus, although listeners do differ in the extent to which they can identify the target speech when it sounds colocated with the maskers (and certainly could do a better job if the differences between voices were greater; e.g., different sex talkers), much larger differences are found between listeners when the maskers are spatially separated from the target, and the listener has interaural differences to use for perceptual segregation.

Previous studies have concluded that the ability to utilize the interaural differences created by the spatial separation of sources to obtain a release from SOS masking is adversely affected by hearing loss (e.g., Arbogast et al., 2005; Best et al., 2012; Dawes et al., 2013; Glyde et al., 2013; Gallun et al., 2013; Marrone et al., 2008b). So it seems likely that some of the differences in performance between groups observed here are related to the presence and degree of SNHL. In Figure 4, the individual T/Ms at threshold for all eight SNHL listeners in the two separated cases were plotted as a function of the threshold measured for these stimuli in quiet for KEMAR and BEAM microphone conditions.

For the KEMAR condition, the T/Ms at threshold are significantly positively correlated with the unmasked target (quiet) thresholds. For the narrow masker separation, the correlation was 0.77 ($p = .013$) while for the wider separation the correlation was 0.78 ($p = .011$). For the BEAM condition, the correlations were weaker and were not significant with the value for the narrower separation being 0.49 ($p = .106$) and the value for the wider separation being 0.57 ($p = .072$). Most of the SNHL listeners achieved lower thresholds for the BEAM condition at the wider spacing. This means that the listeners who performed poorly (higher thresholds) in spatially separated conditions with KEMAR received the most benefit from BEAM. Figure 5 emphasizes this point. Here, the benefit of the BEAM condition compared with the KEMAR condition is plotted on the ordinate as a function of the threshold obtained under KEMAR. Because of the wide range of spatially
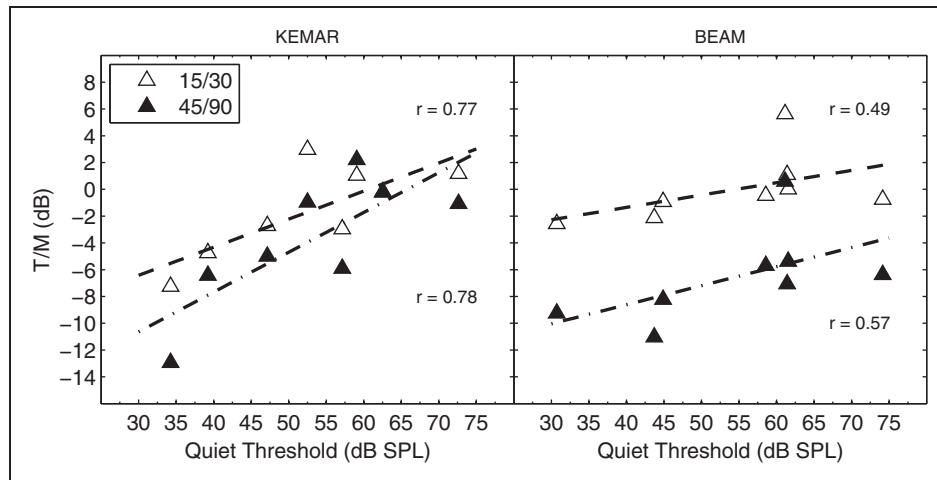
Figure 4. Threshold *T/M*s in dB for individual SNHL listeners plotted as a function of quiet (no masker) speech reception thresholds (in dB SPL) measured for this closed-set corpus. The left panel is for the KEMAR microphone condition while the right panel is for the BEAM microphone condition. The dashed lines (upper) are linear least-squares fits to the thresholds for the narrower masker spacing (open symbols) while the dot-dashed lines are fits to the thresholds for the wider masker spacing (filled symbols). The correlations are given as *r* values for the two least-squares fits in each panel.

separated thresholds found for the NH listeners (see Table 1), their data are included in Figure 5. The slope for the narrower spacing is equivalent to a 4.7 dB increase along the *BEAM* benefit axis for each 10 dB increase in *T/M* at threshold whereas at the wider spacing 6.1 dB more benefit would be predicted for each 10 dB increase in threshold. The data plotted in this fashion emphasize the point that the benefit that may be obtained from the BEAM—and presumably from other types of highly directional amplification—depends on how well the listener is able to overcome the competing talkers in spatially separated conditions using natural binaural information. Values above the horizontal dotted line at 0 dB indicate listeners for whom beam-forming amplification was superior to natural binaural cues (with bilateral amplification for the SNHL listeners) as reflected in these measures.

Discussion

The first conclusion to be drawn from this study is that a large spatial release from SOS masking may be obtained by listeners with SNHL when using acoustic beamforming. This masking release was apparent for the BEAM microphone condition in the form of lower *T/M*s at threshold for spatially separated maskers relative to colocated maskers. For the widest spatial separation, the group mean (*n* = 8) threshold under BEAM was −6.6 dB resulting in a SRM of 9.5 dB. Furthermore, based on group means, these values (threshold and SRM) were about 3 dB better than were found for the same SNHL listeners listening through KEMAR which provided binaural cues following prescription (NAL-RP)
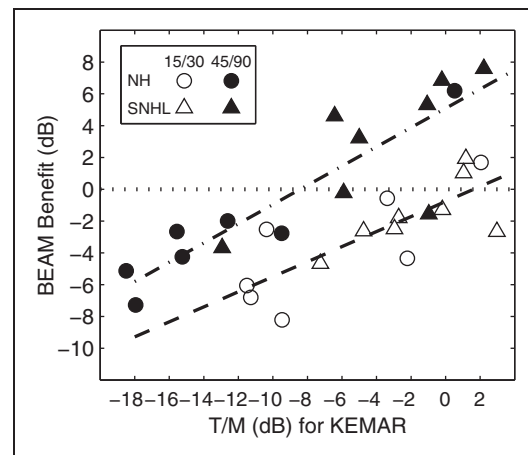

Figure 5. The benefit of listening through the beamforming microphone array compared with natural binaural listening (*T/M* at threshold for the BEAM condition minus *T/M* at threshold for the KEMAR condition) plotted as a function of the threshold *T/M* obtained for the KEMAR condition. Values are plotted for individual NH (circles) and SNHL (triangles) listeners. Masker spatial separations of ±15° and ±30° (dashed line, open symbols) and ±45° and ±90° (dot-dashed line, filled symbols) are shown with linear least-squares fits computed for each set of data. The horizontal dotted line at 0 dB divides cases where the *T/M* was lower for BEAM than for KEMAR resulting in positive values of *BEAM benefit.*

gain applied to the stimulus. This relative benefit (BEAM better than KEMAR for SNHL) was not apparent for the narrower masker separation where the KEMAR condition yielded lower thresholds and larger SRMs than

Table 1. Individual and Group Mean T/Ms and SRMs (bold) in dB for the NH Listeners.

NH Listener	KEMAR						BEAM						BEAMAR					
	Quiet*	Coloc	15/30	45/90	SRM 15/30	SRM 45/90	Quiet*	Coloc	15/30	45/90	SRM 15/30	SRM 45/90	Quiet	Coloc	15/30	45/90	SRM 15/30	SRM 45/90
1	−37.3	2.6	−11.3	−17.9	**13.9**	**20.5**	−40.8	−1.0	−4.5	−10.7	**3.5**	**9.7**		1.8	−9.8	−17.9	**11.6**	**19.8**
2	−44.0	−3.9	−11.5	−18.5	**7.6**	**14.6**	−47.0	1.1	−5.4	−13.4	**6.5**	**14.5**		−4.7	−12.8	−18.5	**8.1**	**13.8**
3	−34.6	3.5	2.0	0.5	**1.5**	**3.0**	−38.9	5.4	0.4	−5.7	**5.1**	**11.1**		6.0	0.7	−6.8	**5.3**	**12.8**
4	−36.2	0.1	−3.4	−12.6	**3.5**	**12.8**	−42.0	0.3	−2.8	−10.6	**3.1**	**11.0**		0.4	−5.9	−14.9	**6.4**	**15.3**
5	−37.7	3.3	−2.2	−9.5	**5.6**	**12.8**	−40.3	3.3	2.1	−6.7	**1.2**	**10.0**		1.4	−4.8	−13.5	**6.3**	**14.9**
6	−39.0	−1.4	−10.4	−15.3	**9.0**	**13.9**	−41.0	0.1	−7.8	−11.0	**7.9**	**11.1**		−6.9	−14.3	−19.5	**7.4**	**12.7**
7	−36.1	2.9	−9.5	−15.6	**12.4**	**18.5**	−39.3	0.1	−1.3	−12.9	**1.4**	**13.0**						
Mean (N=6)	−38.1	0.7	−6.1	−12.2	**6.8**	**12.9**	−41.7	1.5	−3.0	−9.7	**4.5**	**11.2**		−0.3	−7.8	−15.2	**7.5**	**14.9**
SE	1.32	1.22	2.33	2.89	**1.79**	**2.31**	1.14	0.97	1.52	1.18	**1.00**	**0.70**		1.91	2.27	1.92	**0.91**	**1.07**
Mean (N=7)	−37.8	1.0	−6.6	−12.7	**7.6**	**13.7**	−41.3	1.3	−2.8	−10.1	**4.1**	**11.5**						
SE	1.15	1.08	2.03	2.49	**1.71**	**2.11**	1.02	0.85	1.31	1.10	**0.96**	**0.64**						

Note. NH = normal hearing; SRM = spatial release from masking; T/M = target to masker ratio. The three microphone conditions (BEAM, KEMAR, and BEAMAR; see text) are tabulated in the left, middle, and center columns, respectively. All subjects (N = 7) listened under KEMAR and BEAM conditions while a subset of the subjects (N = 6) also listened under BEAMAR. The group mean data are compiled for both cases. *As in the masked conditions, quiet thresholds are given in T/M in dB, as if a masker were present, for use as a lower bound on masked performance.

BEAM. The likely reason for this dependence of the benefit of beamforming on the degree of source separation is that the spatial tuning of the microphone array is wider than the spatial tuning which may be achieved naturally using binaural information for many listeners, and the closer-spaced maskers were on the slope of the BEAM filter. For example, the overall broadband attenuation at $\pm15°$ for this beamforming microphone array was measured to be about 5 dB while the estimated attenuation from spatial tuning achieved from binaural information (based on the SRM from SOS masking data) was about 8 dB on average as reported by Marrone, Mason, & Kidd (2008a). Our results are broadly consistent with these earlier findings.

The second conclusion is that both NH and SNHL listeners vary widely in their ability to make use of binaural information to achieve low thresholds in SOS masking. The variability across subjects also depended on microphone or amplification approaches and likely reflects fundamental differences in the way that the target source may be perceptually segregated from the maskers. Overcoming the masking produced by competing talkers using binaural information involves selective listening and comprehension of the target speech while ignoring or inhibiting the competing speech originating from different spatial locations. The large difference in this ability among listeners within a group, as well as the differences obtained on average between groups, may be attributed, we believe, to intersubject differences in the higher level perceptual and cognitive mechanisms responsible for susceptibility to informational masking (IM) and for the release from IM (for a review of this topic, see Kidd, Mason, Richards, Gallun, & Durlach, 2008a). This ability also varies according to age (e.g., Gallun et al., 2013; Marrone et al., 2008b) but was apparent even for this young–adult SNHL group (cf. Marrone et al., 2008b,c). When the ability to separate and identify target speech in competing speech was measured for colocated target and maskers, that is, when the sources must be separated by means other than binaural processing, the differences between subjects' masked thresholds were reduced. Separating the target voice from the same-sex masker talkers was so challenging that most of the thresholds in the colocated condition were at a positive T/M where the target was the highest level source and thus presumably was the louder and more salient of the five sources. All of the listeners had threshold T/Ms within a few decibels of that level so that the range across listeners was relatively small. Because the BEAM condition yields a similar colocated image of the mixture of talkers, a similarly reduced (relative to the spatially separated masker conditions under KEMAR) range of thresholds across listeners would be expected and, indeed, was observed. For example, the standard error for the NH listeners in the KEMAR condition was about 2.9 dB for the wider masker separation while the standard error for the BEAM condition was only about 1.2 dB. With respect to the thresholds obtained for the BEAM relative to KEMAR on an individual basis, five of the eight SNHL listeners achieved lower thresholds for the wider masker separation under the BEAM condition but only three of eight had lower thresholds under BEAM for the narrower masker spacing.

A third conclusion from this study is that combining beamforming with binaural information can capture some of the benefits of each approach without compromising the usefulness of the BEAM for emphasizing the target location. A priori, one possibility was that the two disparate types of input for the hybrid BEAMAR condition would not yield a unified perceptual image that could be separated from the maskers as well as either approach alone. That is, the BEAM input creates a single image of the mixed sound sources located at the midline (for diotic presentation) while the KEMAR input creates an image of the sound sources distributed at their respective locations. This mismatch could have interfered with the ability of the listener to form an integrated perceptual image of the target or to perceive that image as distinct from the maskers. Based on the current findings, that does not appear to have been the case. The thresholds measured under BEAMAR were equal to, or lower than, the corresponding thresholds under KEMAR or BEAM for every individual listener tested for the wider masker spatial separation and for the majority of listeners for the narrower separation. With respect to preserving the ability to localize sounds under BEAMAR, it is obvious from casual listening that the source locations from which the maskers were presented in the current study could easily be distinguished when the speech stimulus was presented in quiet. These subjective observations are supported by the localization accuracy results presented by Desloge et al. (1997) using a similar hybrid apparatus and also are consistent with the findings of Wightman and Kistler (1992) who emphasized the dominant role of low-frequency interaural temporal cues in preserving sound source localization. That is not the case under BEAM listening in which no interaural differences are present in the stimuli and all sources appear to originate from the same location. However, we do not yet know how accurate localization is under BEAMAR or how that compares to localization based on natural cues under KEMAR, nor do we know the extent to which localization of one speech source among multiple sources is possible using either approach. Furthermore, performance under BEAMAR likely depends on the cutoff frequency separating natural and beamforming components, and this variable has not been examined yet in any systematic way for the SOS masking task.

On a conceptual level, both the acoustic beamforming produced by a multiple-microphone array and selective attention applied to binaural information produce a similar result: tuning in azimuth. The effect of acoustic beamforming is straightforward, enhancing the input to the listener from the direction at which it is aimed (the ALD) and progressively attenuating the inputs from sources separated in azimuth from the ALD until asymptotic attenuation is reached. The spatial filtering characteristics of the beamforming microphone array used in this study have been described in detail elsewhere (Favrot et al., 2013; Kidd et al., 2013) and are shown in Figure 2. In the current results, the spatial filtering caused by the beamformer is apparent in the progressive decline in T/Ms at threshold for the different spatial conditions tested: colocated, narrow, and wide masker separations. On the other hand, the spatial tuning that is achieved by selective attention applied to binaural information is considerably less straightforward although it produces a similar pattern of reduction in T/M with masker separation.

The view of attention acting like a filter applied to auditory stimuli dates at least from Broadbent's seminal "filter theory" (1958; p. 42 and, generally, chapter 3). A later, widely cited example of how the application of selective attention can result in a filter-like pattern of observer responses was provided by Greenberg and Larkin (1968) using the probe-signal method (for related examples, see also Dai & Wright, 1995; Macmillan & Schwarz, 1975; Scharf, Quigley, Aoki, Peachy, & Reeves, 1987). Greenberg and Larkin demonstrated that performance (percent correct in a tone-in-noise detection task) reflected tuning along a simple stimulus dimension—frequency—based on the expectation of the listener about the probability of occurrence of the signal at a point along the frequency dimension. In the spatial domain, an analog of the Greenberg and Larkin result measured along the dimension of sound source azimuth was reported by Arbogast and Kidd (2000). They suggested that a spatial filter tuned in azimuth functioned to reduce the high IM present in complex multiple-source sound fields where the sources were spatially distributed. The importance of the presence of IM in producing spatial tuning was demonstrated in a series of sound field masking experiments by Marrone et al. (2008a). They employed a closed-set SOS masking paradigm based on the *coordinate response measure*—CRM—test (e.g., Brungart, 2001) but comprising two independent masker talkers that were located symmetrically in azimuth around a target source. The reduction in threshold T/Ms that occurred as the maskers were progressively separated in azimuth (referenced to colocated, as here) was taken as an estimate of the attenuation of the attention-based spatial filter (a maximum of about 12 dB). Importantly, a noise-masking control condition only

resulted in a SRM of about 1.5 dB implying that a high degree of IM was necessary to produce large amounts of masker attenuation.[2]

The current findings suggest that listeners with SNHL are less able to benefit from either type of spatial filter than are listeners with NH. This difference between subject groups is more pronounced for the conditions in which tuning must be accomplished by selective attention applied to binaural input than by beamforming. The decline in thresholds from colocated to widely separated maskers was about 13.7 dB for the NH listeners (Table 1, $n = 7$) under KEMAR and about 6.1 dB for the same conditions for the SNHL listeners (Table 2, $n = 8$). In terms of attenuation that translates roughly into about -3 dB/10° of separation for NH and -1.3 dB/10° for SNHL over the range spanned by the nearest of the maskers in each case (i.e., 0°, ±15°, and ±45° for colocated, narrow, and wide separations, respectively). For BEAM the corresponding values were -2.5 dB/10° for NH and -2.1 dB/10° for SNHL. That the difference between NH and SNHL listeners should be greater for KEMAR than for BEAM is not surprising given that the microphone array implements spatial tuning at the sound input and thus operates external to the listener. Making use of binaural cues to achieve spatial tuning under KEMAR is an operation that is internal to the listener so it is more likely to be affected by hearing loss.

On a practical level, the large individual differences in performance between subjects — specifically, the SRTs obtained when the competing sources were spatially subjects — specifically have important implications for the type of amplification that is best suited to an individual. Although we did not employ commercially available hearing aids in this study, we did compare two fundamentally different approaches to amplification. As discussed previously, the single-channel input to the listener from beamforming limits the source segregation or selection cues principally to those that are available in listening to colocated sounds (with the exception of possible timbre cues due to the frequency-dependent spatial tuning of the array; cf. Kidd et al., 2013 and Figure 2). In contrast, typical binaural amplification requires the listener to perform internal processing to attenuate sources outside of the focus of attention. The potential clinical implication of this distinction is that, for those listeners who are unable to achieve low thresholds when target and masker sources are spatially separated, the beamforming approach that implements spatial selectivity for the listener may be more effective than an approach based on binaural amplification. This point was emphasized in Figure 5 in which those listeners with values falling on the positive side of the dashed horizontal line fared better with the beamformer than with the binaural amplification approach used here. There are, of course, many factors that have to be considered in making a

Table 2. Individual and Group Mean T/Ms and SRMs (bold) in dB for the SNHL Listeners.

SNHL Listener	KEMAR						BEAM						BEAMAR					
	Quiet*	Coloc	15/30	45/90	SRM 15/30	SRM 45/90	Quiet*	Coloc	15/30	45/90	SRM 15/30	SRM 45/90	Quiet	Coloc	15/30	45/90	SRM 15/30	SRM 45/90
1	−23.6	1.4	−2.7	−5.0	**4.2**	**6.4**	−27.4	2.7	−0.9	−8.2	**3.7**	**10.9**		0.4	−3.0	−9.5	**3.4**	**9.9**
2	−12.6	2.5	−0.2	−0.2	**2.7**	**2.7**	−16.3	3.7	1.1	−7.1	**2.6**	**10.8**						
3	−14.2	2.3	1.2	−1.1	**1.1**	**3.4**	−15.3	4.1	−0.8	−6.4	**4.9**	**10.5**		0.3	−0.8	−6.4	**1.1**	**6.7**
4	−28.5	2.4	1.0	2.2	**1.3**	**0.2**	−29.3	2.4	0.0	−5.4	**2.4**	**7.8**						
5	−29.7	4.8	3.0	−1.0	**1.9**	**5.8**	−25.1	5.1	5.6	0.6	**−0.5**	**4.5**		5.9	5.1	−1.6	**0.8**	**7.6**
6	−31.8	2.4	−4.8	−6.4	**7.1**	**8.8**	−30.1	1.1	−2.1	−11.0	**3.2**	**12.1**		0.9	−5.0	−13.2	**5.9**	**14.1**
7	−19.5	1.6	−3.0	−5.9	**4.6**	**7.5**	−21.6	3.1	−0.5	−5.7	**3.6**	**8.8**		2.2	−2.0	−8.3	**4.1**	**10.4**
8	−29.7	1.3	−7.3	−12.9	**8.5**	**14.2**	−36.5	1.2	−2.6	−9.3	**3.7**	**10.4**		−0.9	−7.5	−13.4	**6.6**	**12.5**
Mean (N = 6)	−24.7	2.3	−2.3	−5.4	**4.6**	**7.7**	−26.0	2.9	−0.2	−6.7	**3.1**	**9.5**		1.5	−2.2	−8.7	**3.7**	**10.2**
SE	2.81	0.54	1.54	1.80	**1.18**	**1.50**	2.96	0.65	1.22	1.65	**0.76**	**1.10**		0.98	1.76	1.81	**0.98**	**1.15**
Mean (N = 8)	−23.7	2.3	−1.6	−3.8	**3.9**	**6.1**	−25.2	2.9	−0.0	−6.6	**2.9**	**9.5**						
SE	2.64	0.39	1.21	1.69	**0.97**	**1.52**	2.55	0.49	0.90	1.22	**0.56**	**0.85**						

Note. SNHL = sensorineural hearing loss; SRM = spatial release from masking; T/M = target to masker ratio.
All subjects (N = 8) listened under KEMAR and BEAM conditions while a subset of the subjects (N = 6) also listened under BEAMAR. The group mean data are compiled for both cases. *As in the masked conditions, quiet thresholds are given in T/M in dB, as if a masker were present, for use as a lower bound on masked performance.

determination about the amplification approach that is best suited to an individual including the audibility that is achieved, gain and compression, loudness comfort, and so on. However, our findings suggest that acoustic beamforming, particularly in combination with natural binaural cues, may afford a number of advantages in multiple-talker listening situations. Furthermore, Kidd et al. (2013) found large SRM for both NH and unilateral loss listeners when the maskers were high in energetic masking (speech-shaped speech-envelope modulated noise) as well as for speech. This is important because it suggests that spatially selective amplification may provide advantages for many conditions in which SRM cannot be obtained via natural binaural listening (e.g., for cochlear implant-processed speech; cf. Goldsworthy, 2014; Swaminathan, Mason, Streeter, Best, & Kidd, 2013) or for other conditions degrading the quality of the neural representation of sounds (e.g., hidden hearing loss; cf. Kujawa & Liberman, 2009; also see recent review by Plack, Barker, & Prendergast, 2014). There are other factors that may affect whether to fit highly spatially selective amplification to individual listeners. For example, beamforming works best under low reverberation although the limited testing with the current array has found that the reduction in spatial benefit using the beamformer is similar to that which occurs for natural binaural cues (cf. Favrot et al., 2013). Other factors—such as the ability to monitor the broad sound field while still obtaining the benefit of directional amplification (the case motivating the BEAMAR condition)—may be important as well. Furthermore, for the VGHA that incorporates the beamformer tested here with visual guidance, special considerations could apply such as the perceptual effect of amplifying certain directions preferentially and maintaining the focus of the beam on the desired source. These considerations await further study.

Conclusions

- Acoustic beamforming provided a large (about 9–10 dB on average) spatial release from SOS masking for SNHL listeners. Most individual listeners with SNHL performed better using the beamformer than natural cues with frequency-specific gain for the wider masker separation but not for the narrower masker separation.
- SRTs in spatially separated maskers were lower for listeners with NH using natural binaural cues than using beamforming. The thresholds for the NH group were lower in all cases than the corresponding thresholds for the SNHL group.
- The lowest thresholds in spatially separated maskers were obtained using a hybrid natural-beamformer combination which subjectively preserved spatial information.

- A possible clinical implication of these findings is that candidacy for highly spatially tuned amplification may depend on performance with conventional bilateral amplification, with individuals exhibiting poor performance using natural cues more likely to benefit from beamforming or other highly directional amplification strategies.

Notes

1. The term BEAM is used here to denote a specific listening condition and should not be confused with the BEAM® technology from Cochlear Corporation.
2. The slopes of the underlying psychometric functions for noise maskers typically are much steeper than for speech maskers (e.g., Arbogast, Mason, & Kidd, 2002) meaning that a small SRM in decibels may still translate into a large improvement in performance as reflected in percent correct scores.

References

Arbogast, T. L., & Kidd, G. Jr. (2000). Evidence for spatial tuning in informational masking using the probe-signal method. *The Journal of the Acoustical Society of America*, *108*, 1803–1810.

Arbogast, T. L., Mason, C. R., & Kidd, G. Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, *112*, 2086–2098.

Arbogast, T. L., Mason, C. R., & Kidd, G. Jr. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *117*, 2169–2180.

Best, V., Marrone, N., Mason, C. R., & Kidd, G. Jr. (2012). The influence of non-spatial factors on measures of spatial release from masking. *The Journal of the Acoustical Society of America*, *131*, 3103–3110.

Broadbent, D. E. (1958). *Perception and Communication*. Oxford: Pergammon Press.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica United With Acustica*, *86*, 117–128.

Brungart, D. S. (2001). Evaluation of speech intelligibility with the coordinate response measure. *The Journal of the Acoustical Society of America*, *109*, 2276–2279.

Byrne, D. J., Parkinson, A., & Newall, P. (1991). Modified hearing aid selection procedures for severe-profound hearing losses. In G. A. Studebaker, F. H. Bess, & L. B. Beck (Eds.), *The Vanderbilt Hearing Aid Report II* (pp. 295–300). Parkton, MD: York Press.

Carlile, S. (2014). Active listening: Speech intelligibility in noisy environments. *Acoustics Australia*, *42*, 98–104.

Dai, H., & Wright, B. A. (1995). Detecting signals of unexpected or uncertain durations. *The Journal of the Acoustical Society of America*, *98*, 798–806.

Dawes, P., Munro, K. J., Kalluri, S., & Edwards, B. (2013). Unilateral and bilateral hearing aids, spatial release from masking and auditory acclimatization. *The Journal of the Acoustical Society of America*, *134*, 596–606.

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

Favrot, S., Mason, C. R., Streeter, T., Desloge, J., & Kidd, G. Jr. (2013). Performance of a highly directional microphone array in a reverberant environment. *Acoustical Society of America*. POMA, *19*, 050145, 1–8.

Gallun, F. J., Kampel, S. D., Diedesch, A. C., & Jakien, K. M. (2013). Independent impacts of age and hearing loss on spatial release in a complex auditory environment. *Frontiers in Neuroscience*, *252*, 1–11.

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*, 15–28.

Gnewikow, D., Ricketts, T., Bratt, G. W., & Mutchler, R. C. (2009). Real-world benefit from directional microphone hearing aids. *Journal of Rehabilitation Research and Development*, *46*, 603–616.

Goldsworthy, R. (2014). Two-microphone spatial filtering improves speech reception for cochlear-implant users in reverberant conditions with multiple noise sources. *Trends in Hearing*, *18*, 1–13. doi:10.1177/2331216514555489

Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method. *The Journal of the Acoustical Society of America*, *44*, 1513–1523.

Greenberg, J. E., Desloge, J. G., & Zurek, P. M. (2003). Evaluations of array processing algorithms for a headband hearing aid. *The Journal of the Acoustical Society of America*, *113*, 1646–1657.

Greenberg, J. E., & Zurek, P. M. (1992). Evaluation of an adaptive beamforming method for hearing aids. *The Journal of the Acoustical Society of America*, *91*, 1662–1676.

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

Kidd, G. Jr., Favrot, S., Desloge, J., Streeter, T., & Mason, C. R. (2013). Design and preliminary testing of a visually-guided hearing aid. *The Journal of the Acoustical Society of America*, *133*, EL202–EL207.

Kidd, G. Jr., Mason, C. R., Brughera, A., & Hartmann, W. M. (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica United With Acustica*, *91*, 526–536.

Kidd, G., Mason, C. R. Jr., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008a). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143–190). New York, NY: Springer Science+Business Media, LLC.

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after temporary noise-induced hearing loss. *J. Neurosci*, *29*, 14077–14085.

Macmillan, N. A., & Schwartz, M. (1975). A probe-signal investigation of uncertain frequency detection. *The Journal of the Acoustical Society of America*, *58*, 1051–1058.

Marrone, N. L., Mason, C. R., & Kidd, G. Jr. (2008a). Tuning in the spatial dimension: Evidence from a masked speech identification task. *The Journal of the Acoustical Society of America*, *124*, 1146–1158.

Marrone, N. L., Mason, C. R., & Kidd, G. Jr. (2008b). Effect of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *The Journal of the Acoustical Society of America*, *124*, 3064–3075.

Marrone, N. L., Mason, C. R., & Kidd, G. Jr. (2008c). Evaluating the benefit of hearing aids in solving the cocktail party problem. *Trends in Amplification*, *12*, 300–315.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*, 953–978.

Neher, T., Laugesen, S., Jensen, N. S., & Kragelund, L. (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *The Journal of the Acoustical Society of America*, *130*, 1542–1558.

Noble, W., & Gatehouse, S. (2006). Effects of bilateral vs. unilateral hearing aid fitting on abilities measured on the speech, spatial, and qualities of hearing scale. *International Journal of Audiology*, *45*, 172–181.

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*, 339–352.

Plack, C. J., Barker, D., & Prendergast, G. (2014). Perceptual consequences of 'hidden hearing loss'. *Trends in Hearing*, *18*, 1–11.

Ruggles, D., & Shinn-Cunningham, B. G. (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*, 395–405.

Scharf, B., Quigley, S., Aoki, C., Peachy, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Perception & Psychophysics*, *42*, 215–223.

Stadler, R. W., & Rabinowitz, W. M. (1993). On the potential of fixed arrays for hearing aids. *The Journal of the Acoustical Society of America*, *80*, 1332–1342.

Swaminathan, J., Mason, C. R., Streeter, T., Best, V., & Kidd, G. Jr. (2013). Spatial release from masking for noise-vocoded speech. *POMA*, *19*(050071), 1–8.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V. A., Kidd, G. Jr., & Patel, A. D. (2015). *Musical training, individual differences and the cocktail party problem*. Nature Scientific Reports, in press.

Welker, D. P., Greenberg, J. E., Desloge, J. G., & Zurek, P. M. (1997). Microphone-array hearing aids with binaural output — Part II: A two-Microphone adaptive system. *IEEE Transactions on Audio, Speech, and Language Processing*, *5*, 543–551.

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, *91*, 1648–1661.

Woods, W. S., Kalluri, S., Pentony, S., & Nooraei, N. (2013). Predicting the effect of hearing loss and audibility on amplified speech reception in a multi-talker listening scenario. *The Journal of the Acoustical Society of America*, *133*, 4268–4278.

Yost, W. A. (1997). The cocktail party problem: forty years later. In R. Gilkey, & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 329–347). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.