

Use of a glimpsing model to understand the performance of listeners with and without hearing loss in spatialized speech mixtures

Virginia Best,^{a)} Christine R. Mason, Jayaganesh Swaminathan, Elin Roverud, and Gerald Kidd, Jr.

Department of Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215, USA

(Received 7 March 2016; revised 15 December 2016; accepted 21 December 2016; published online 9 January 2017)

In many situations, listeners with sensorineural hearing loss demonstrate reduced spatial release from masking compared to listeners with normal hearing. This deficit is particularly evident in the “symmetric masker” paradigm in which competing talkers are located to either side of a central target talker. However, there is some evidence that reduced target audibility (rather than a spatial deficit *per se*) under conditions of spatial separation may contribute to the observed deficit. In this study a simple “glimpsing” model (applied separately to each ear) was used to isolate the target information that is potentially available in binaural speech mixtures. Intelligibility of these glimpsed stimuli was then measured directly. Differences between normally hearing and hearing-impaired listeners observed in the natural binaural condition persisted for the glimpsed condition, despite the fact that the task no longer required segregation or spatial processing. This result is consistent with the idea that the performance of listeners with hearing loss in the spatialized mixture was limited by their ability to identify the target speech based on sparse glimpses, possibly as a result of some of those glimpses being inaudible. © 2017 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4973620>]

[DB]

Pages: 81–91

I. INTRODUCTION

In the context of speech communication, spatial release from masking (SRM) refers to an improvement in intelligibility when competing sounds are spatially separated from the talker of interest. This improvement can arise as a result of increases in the acoustic signal-to-noise ratio at one ear caused by the “head-shadow” or by effective increases in signal-to-noise ratio resulting from processing of binaural cues (Bronkhorst and Plomp, 1988). However, in certain multitalker listening situations, where the problem is one of disentangling the speech of competing talkers, it appears that much of the advantage comes from the perceived separation of sources, which facilitates segregation and enables attention to be directed selectively (for review, see Kidd *et al.*, 2008a).

Listeners with hearing impairment (HI) often demonstrate reduced SRM compared to listeners with normal hearing (NH) in multitalker listening situations (e.g., Marrone *et al.*, 2008; Neher *et al.*, 2009; Glyde *et al.*, 2013b), leading to the common notion that spatial processing is disrupted by hearing loss. However, attempts to provide convergent evidence from other kinds of spatial tasks have produced very mixed results. For example, studies that have measured fine discrimination of binaural cues show that individual variability is high, with some HI listeners performing as well as NH listeners (e.g., Colburn, 1982; Koehnke *et al.*, 1995;

Spencer, 2013; Gallun *et al.*, 2014). It has also been reported that free-field localization in the horizontal plane is not strongly affected by hearing loss unless it is highly asymmetric or very severe at low frequencies (e.g., Noble *et al.*, 1994; Otte *et al.*, 2013). However, several studies have shown that older listeners (with typical high-frequency hearing losses) do localize more poorly than young NH listeners (Dobrevá *et al.*, 2011; Neher *et al.*, 2011). Studies that have tried to relate SRM in multitalker environments to localization ability (Noble *et al.*, 1997; Hawley *et al.*, 1999) have had little success. On the other hand, correlational studies looking for a link between multitalker SRM and basic binaural sensitivity have provided an inconsistent picture, with some finding a relationship (Strejczyk and Dau, 2009; Neher *et al.*, 2011) and others not (Spencer, 2013). It seems likely that age-related declines in both basic binaural abilities and speech intelligibility might explain some of the positive relationships (Neher *et al.*, 2012; Moore, 2014). Finally, in studies that have controlled for age, it has been observed that SRM is inversely related to the severity of hearing loss as measured by the audiogram or speech reception in quiet (Marrone *et al.*, 2008; Glyde *et al.*, 2013b; Besser *et al.*, 2015). This raises the question of whether in some cases apparent spatial deficits might actually be related to reduced audibility in speech mixtures.

A popular stimulus paradigm that has been used in recent years consists of a frontally located speech target, and competing speech maskers presented symmetrically from the sides. This configuration was originally implemented to minimize the contribution of head-shadow benefits to SRM

^{a)}Electronic mail: ginbest@bu.edu

(Noble *et al.*, 1997; Marrone *et al.*, 2008) but has since been adopted as a striking case in which the difference between NH and HI listeners is large. In a typical implementation, targets and maskers are drawn from the same set of stimuli and are highly confusable on the basis of their sentence structure or voice characteristics (a situation that can be described as high in “informational masking”; Kidd *et al.*, 2008a). Under these conditions, both groups tend to perform similarly (and poorly) in the co-located condition, but NH listeners are able to achieve much lower thresholds in the spatially separated configuration, and thus demonstrate a much larger SRM (Marrone *et al.*, 2008; Neher *et al.*, 2011; Best *et al.*, 2013; Gallun *et al.*, 2013; Glyde *et al.*, 2013b).

A number of recent studies using the symmetric masker paradigm have provided evidence that stimulus audibility can influence the measured SRM. Gallun *et al.* (2013) found that the effect of hearing loss on performance in the spatially separated configuration was stronger when one target level was used for all listeners [50 dB sound pressure level (SPL)] compared to when a sensation level (SL) of 40 dB was used (equivalent to a range of 47–72 dB SPL). In a recent study (Jakien *et al.*, 2017) this group also tested listeners at two different SLs (20 and 40 dB) and found that SRM was larger at the higher SL. Glyde *et al.* (2013b) measured SRM using the clinical test LiSN-S (Cameron and Dillon, 2007), which includes linear frequency-specific gain applied according to the individual audiogram using the NAL-RP hearing-aid prescription (Byrne *et al.*, 1991). For their subjects, a strong relationship between SRM and hearing status persisted despite this amplification. The authors noted that with the relatively low presentation levels used in their experiment (a fixed masker level of 55 dB SPL, and adaptive target level), the NAL-RP prescription may not have provided sufficient gain especially in the high-frequencies. Thus, in a follow-up experiment (Glyde *et al.*, 2015), they examined the effect of providing systematically more high-frequency gain than that provided by NAL-RP (which is equivalent to increasing the audible bandwidth, see also Ahlstrom *et al.*, 2009; Moore *et al.*, 2010; Levy *et al.*, 2015). They tested HI subjects as well as NH subjects given stimuli filtered to match the audibility of the aided HI group. For both groups, performance in the spatially separated condition (but not the co-located condition) improved systematically with increased high-frequency gain. The fact that increasing the gain only improved thresholds in the separated condition may reflect the fact that in the LiSN-S (and many tests of SRM), the target level is varied adaptively and reaches lower absolute values in the spatially separated condition. In terms of the effect of SRM, while the difference between NH and HI (with NAL-RP) was originally reported at around 8 dB, equating the audibility across the groups reduced this difference to only 2.5 dB.

While these studies provide evidence that audibility is an important factor influencing SRM, it remains a significant challenge to estimate the impact that audibility has for a given individual in a complex multitalker environment. Many experiments include a control condition in which performance is measured for the target alone, in order to confirm that the target itself is audible and highly intelligible

at the presented levels. However, this kind of control does not take into account the fact that when speech is presented in a mixture of other talkers, parts of the signal are energetically masked, and thus redundancy in the speech signal is greatly reduced. It may be that loss of audibility (or bandwidth) is more damaging in this case. Moreover, in spatialized mixtures, different parts of the target may be missing in each of the two ears. In this study, we investigated an approach that isolates the target glimpses that are “potentially available” in a spatialized speech mixture and measures their intelligibility directly in each listener. This approach is based on a glimpsing model that has been applied to non-spatial speech mixtures in several previous studies (e.g., Wang, 2005; Brungart *et al.*, 2006; Cooke, 2006) but applies it separately to the two ears of a binaural stimulus.

In the glimpsing model, which is explained in more detail below, the speech mixture is analyzed to identify regions of the spectrotemporal plane in which the target is more intense than the masker. By retaining only those regions, the model represents what could be achieved by “ideal” segregation. Measuring performance in the glimpsed condition allows us to capture an individual’s ability to identify speech based on clean but relatively sparse glimpses of the target. Because the target has already been segregated from the masker in this condition, any “spatial” factors (such as binaural processing abilities, or the ability to direct spatial attention) should be rendered moot, whereas “non-spatial” factors (such as target audibility) will still affect performance. Thus, if the glimpsing reduces or eliminates the difference between NH and HI groups observed in the spatialized mixture, it would suggest that spatial processing *per se* is disrupted in the HI group. On the other hand, if the group difference persists in the glimpsed condition, it would provide evidence that the difference does not stem from differing abilities to use spatial cues to segregate the sounds and instead could be attributable to non-spatial factors.

II. GENERAL METHODS

A. Participants

Participants were eleven NH listeners (eight female; mean age 23 years, range 19–30 years) and eight listeners with moderate, bilateral sensorineural hearing losses (three female; mean age 25 years, range 20–39 years). Although the groups were not perfectly matched in age, recruitment was deliberately focused on young HI listeners so that effects of hearing loss could be studied without the confounding factor of age that comes with a more typical older HI population. All participants were college students or recent graduates, and all were native speakers of American English. Participants were paid for their participation, and all procedures were approved by the Boston University Institutional Review Board.

The NH listeners had audiometric thresholds <20 dB hearing level (HL) from 0.25 to 8 kHz. The HI listeners had relatively symmetric hearing losses, with differences in threshold between the ears of no more than 10 dB at any frequency from 0.25 to 8 kHz. Individual audiograms (averaged across the ears) for the HI listeners are shown in Fig. 1 (thin

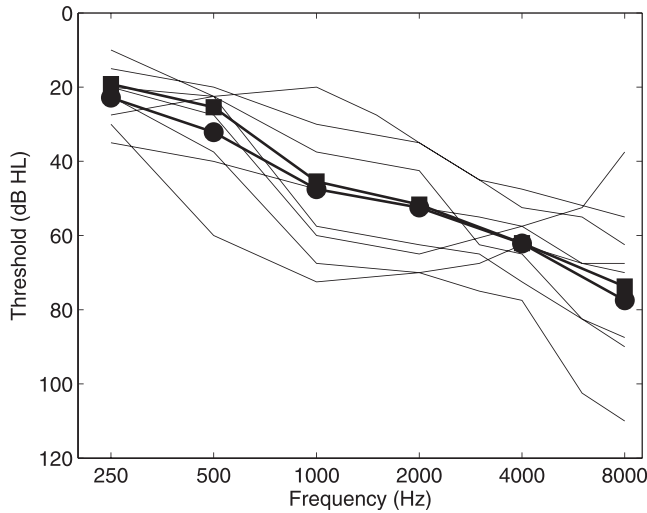


FIG. 1. Individual audiograms for the eight hearing-impaired listeners who participated in the study (thin lines; averaged across left and right ears). Also shown is the across-subject mean for the seven listeners whose data were ultimately included in experiments 1 and 2 (thick lines and circles), and the across-subject mean for the six listeners on which the filtering for experiment 3 was based (thick lines and squares).

lines). Mean four-frequency average (4FAHL) (average thresholds at 0.5, 1, 2, and 4 kHz) in the HI group ranged from 31 to 66 with a mean of 49 dB HL. The etiology of the hearing loss for two listeners was Stickler’s Syndrome (diagnosed at birth) and for the remaining six listeners was unknown (diagnosed in childhood). Six of the eight HI listeners were regular hearing-aid wearers.

B. Stimuli

Speech materials were taken from a corpus of monosyllabic words recorded by Sensimetrics Corporation (Malden, MA). This corpus contains 40 words spoken by eight female and eight male talkers, and has been described previously (Kidd *et al.*, 2008b). From this corpus, five-word sentences are assembled by selecting one word (out of eight options) from each of five categories (e.g., “Sue bought two red toys”). The words in each sentence are concatenated without additional silences, with the result that the overall rhythm and duration of the sentence varies considerably depending on the draw of words. For this study, the target and masker sentences were spoken by different female talkers selected at random from the set of eight. The target was identified by its first word “Sue.”

Stimuli were spatialized using a set of mildly reverberant head-related transfer functions measured on a KEMAR. Specifically, these measurements were done in an IAC booth with perforated walls and carpeted floor, with loudspeakers positioned on an arc at a distance of five feet from the center of KEMAR’s head. The target sentence was presented at 0° azimuth, and two or four different masker sentences were presented at ±90° azimuth, or ±45°/±90° azimuth.

Each masker was fixed in level (see Sec. II C), and the level of the target was varied to adjust the target-to-masker ratio (TMR) to one of five values (in 5 dB steps from -25 to -5 dB, or -20 to 0 dB, depending on the group and

condition). Percent correct scores were calculated at each SNR for each listener in each of the experimental conditions, and logistic functions were fit to these scores. Thresholds, defined as the TMR at 50% correct, were extracted where possible from these fits.

Digital stimuli were generated on a PC using MATLAB software (MathWorks, Inc., Natick, MA) and presented to a pair of Sennheiser HD 280 Pro headphones (Wedemark, Germany). Digital-to-analog conversion was done using TDT equipment (Tucker-Davis Technologies, Alachua, FL) or a 24-bit soundcard (RME HDSP 9632; RME Audio, Haimhausen, Germany). The listener was seated in an audiometric booth fitted with a monitor, keyboard and mouse. Responses were given by selecting five words from a grid of 40 words (five categories × eight options) presented on the monitor.

C. Presentation levels

Three experiments were conducted, that differed primarily in the presentation level of the stimuli. In experiment 1, each masker talker was presented at 55 dB SPL for most listeners. For three of the HI listeners, however, this level was considered too quiet, and thus the masker level was set to 60 dB SPL.¹ Listeners with hearing loss in experiment 1 had linear frequency-dependent gain applied to the stimuli before presentation. This gain was set on an individual basis according to the NAL-RP prescription rule (Byrne *et al.*, 1991; Dillon, 2012), which is a modified half-gain rule that can be used for mild to profound losses. As discussed in the Introduction, amplification using the NAL-RP prescription does not restore full audibility, especially for relatively low presentation levels. Experiment 2 was conducted to determine if the pattern of results obtained in experiment 1 would be the same given a higher overall presentation level, and thus the masker level was increased by 10 dB (to 65 or 70 dB SPL). In experiment 3, the average frequency-dependent audibility of the aided HI group was simulated in a group of NH listeners. For this experiment, a masker level of 57.5 dB SPL was used (half way between the two values used for the HI listeners in experiment 1).

D. Glimpsing model

Simple energy-based analyses have been used to quantify the available target in non-spatial speech mixtures (e.g., the “ideal binary mask” described by Wang, 2005; ideal time-frequency segregation as explored by Brungart *et al.*, 2006; and the glimpsing model of Cooke, 2006). The basic approach is to identify regions in the time-frequency plane where the target energy exceeds the masker energy. To extend this approach to spatialized speech mixtures, in which the signals arriving at the two ears differ, we simply performed this analysis independently for each ear. It is important to emphasize that this approach is based purely on the pattern of energy at each ear, and does not take into account the relationship between the two ears. This is quite different from the “better-ear” glimpsing approach, in which energy is retained in *every* time-frequency region, but is selected from the ear with the better signal-to-noise ratio (Brungart and

Iyer, 2012; Glyde *et al.*, 2013a; Best *et al.*, 2015). The approach should also be distinguished from that described by Roman *et al.* (2003) in which target glimpses are estimated based on binaural cues (in the absence of prior knowledge about the target and masker energies). The important distinction is that our approach makes no assumptions about binaural processes.

Our analysis was based on the approach of Wang (2005) and Brungart *et al.* (2006), and for details the reader is referred to Brungart *et al.* (2006). In brief, the signals were analyzed using 128 frequency channels logarithmically spaced between 80 Hz and 8 kHz, and 20-ms time windows with 50% overlap. Windows in which the target energy exceeded the total masker energy (summed across the two or four masker talkers) were assigned a mask value of 1, and the remaining windows were assigned a value of 0. This resulted in a binary description of the favorable spectrotemporal regions (or the ideal binary mask). The masks for each ear were then used to determine whether the energy in a given time-frequency “tile” would be retained (mask value of 1) or set to zero (mask value of 0) before the signal for that ear was resynthesized.

Using this approach, the number of available target “glimpses” in a speech mixture varies with TMR (fewer glimpses at poorer TMRs), and with the number of maskers (fewer glimpses for four than for two maskers). Figure 2 shows the results of a simulation based on 100 randomly generated stimuli like those used in the experiment. In the top left panel, the proportion of tiles (out of the entire time-frequency plane) retained by the glimpsing model is shown for the left ear only as a function of TMR and number of maskers. The data in the top right panel show how these proportions translate into proportion of target energy retained. To obtain these values, the ideal binary mask was applied to the target stimulus alone, and the resynthesized signal was compared to the original unprocessed target signal. The

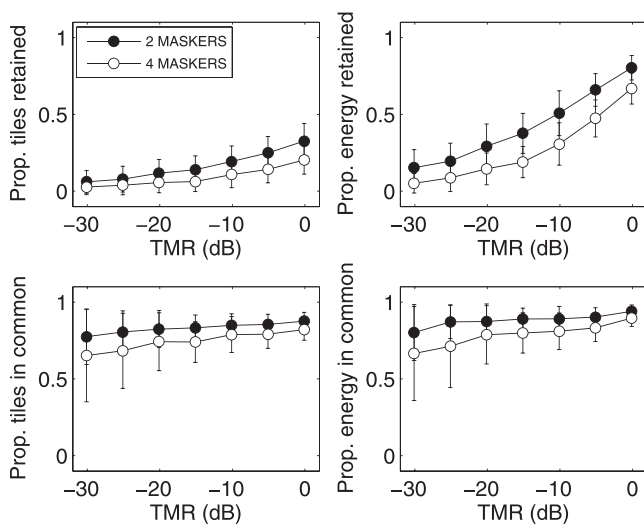


FIG. 2. Top row: Proportion of time-frequency tiles (left) and energy (right) retained in one ear after application of the glimpsing model, as a function of TMR and number of maskers. Bottom row: Proportion of tiles (left) and energy (right) retained that were also retained in the other ear. Error bars indicate standard deviations across 100 simulated trials.

proportional value represents the ratio of the sum of squares for these two signals. For the symmetric-masker configurations examined here, the glimpses could occur in either ear, and often occurred in both ears for a particular time-frequency tile. As shown in the bottom panels of Fig. 2 about 75% of the tiles retained in one ear were also retained in the other ear (or about 80% of the target energy). Figure 3 shows the left and right ear masks for an example target sentence in the presence of two or four masker sentences at a TMR of -5 dB. The gray region indicates tiles that are in common across the ears, and the black regions indicate tiles that were retained in only one ear. Of course, depending on the level of the stimulus and an individual’s hearing thresholds, all of the glimpses may or may not have been audible. Our assumption was that performance in the “glimpsed” condition would provide an estimate of the upper limit of performance in the natural mixture, given the number and distribution of audible glimpses, the quality of the speech in those glimpses, as well as perhaps the individual’s ability to assemble the sparse glimpses and reconstruct the original speech signal.

III. EXPERIMENT 1

A. Procedures

Three conditions were examined in experiment 1. In the natural condition, no glimpsing analysis was done, but the binaural stimuli were processed using “all ones” binary masks so that any potential artifacts introduced by the processing would be present as in the other conditions. In the glimpsed condition, the binary masks were determined from the binaural mixture and applied to each ear as described above. In the glimpsed target condition, the same binary masks were applied to the target stimulus alone. This condition was included as a pure estimate of “target only” intelligibility in case the residual masker energy in the glimpsed

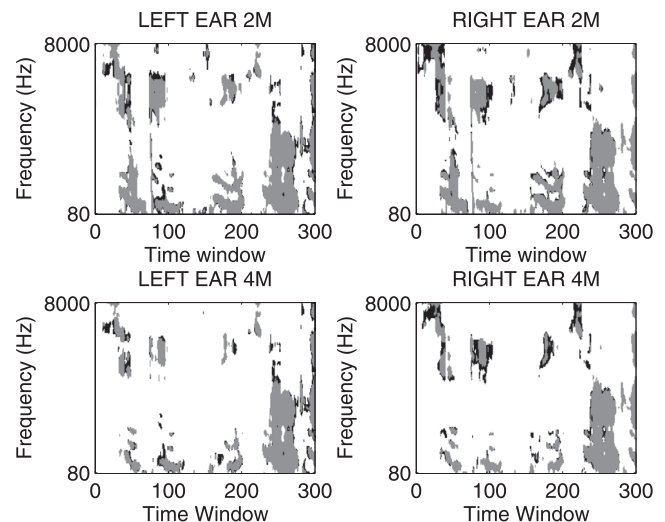


FIG. 3. Left and right ear masks for an example target (the sentence “Sue gave three small gloves”) in the presence of two or four maskers at a TMR of -5 dB. The gray region indicates tiles that are in common across the ears, and the black regions indicate tiles that were retained in only one ear.

stimuli affected different listeners (or groups of listeners) differently.

Listeners completed 20 trials at each of the five TMRs, for both two and four maskers in the three stimulus conditions (natural, glimpsed, glimpsed target), for a total of 600 trials. These trials were presented in blocks of 25 trials (containing five trials at each TMR with masker number and condition fixed within a block), and the blocks were presented in a pseudorandom order such that one block of each unique condition was completed before repeating any condition. In total, testing took 2–3 h and was completed over one or two visits. All of the listeners (eleven NH and eight HI) completed experiment 1. For one NH and one HI listener, performance was particularly poor across the entire tested TMR range, and in most conditions there was not enough data near 50% for sensible thresholds to be extracted. Thus, these two listeners were excluded and data are only shown for ten NH and seven HI. The average audiogram for this group of seven HI listeners is plotted in Fig. 1 (circles).

B. Results

Figure 4 shows means (and standard deviations) of the thresholds in each group for each condition. A three-way mixed analysis of variance (ANOVA) conducted on the thresholds (with factors of number of maskers, processing condition, and group) revealed a significant main effect of number of maskers [$F(1,15) = 274.0$, $p < 0.001$], processing condition [$F(2,30) = 107.1$, $p < 0.001$], and group [$F(1,15)$

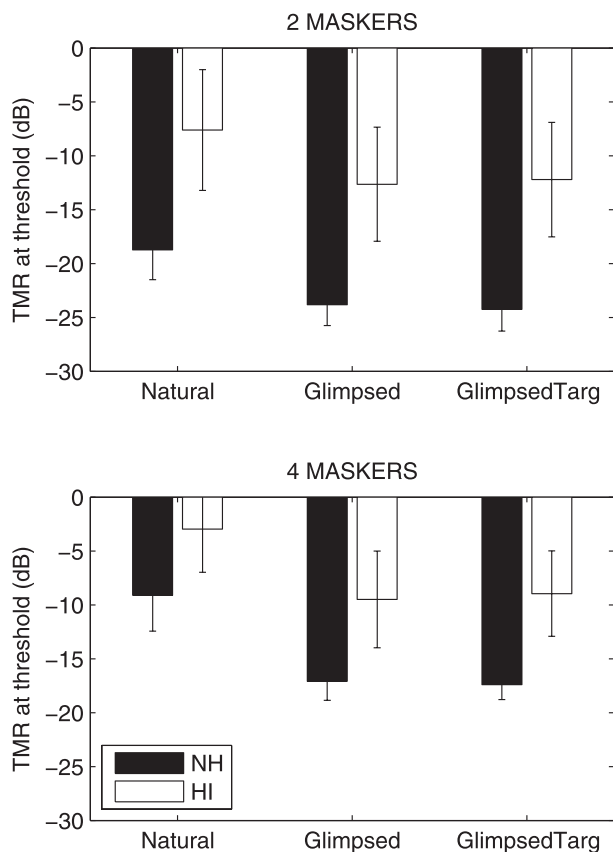


FIG. 4. Mean thresholds (and across-subject standard deviations) for the different conditions in experiment 1.

$= 36.4$, $p < 0.001$]. The two-way interaction between number of maskers and group was significant [$F(1,15) = 34.5$, $p < 0.001$], as was the interaction between number of maskers and processing condition [$F(2,30) = 6.3$, $p = 0.005$]. No other interactions were significant [$p > 0.05$].

As expected, performance was better overall with two maskers as compared to four maskers, and for NH than HI listeners. In all cases, performance was poorest in the natural conditions and improved in the glimpsed conditions, indicating that removing the masker-dominated tiles was generally beneficial. However, performance for the glimpsed and glimpsed target conditions was similar. The interaction between processing condition and number of maskers reflects the fact that the benefit of glimpsing was slightly larger in the four-masker condition (7 vs 5 dB on average).

Of primary interest in this experiment was the difference between the NH and HI groups. The interaction between number of maskers and group shows up here as an overall larger group difference for two maskers than for four maskers (11 vs 7 dB, averaged across conditions). However, the lack of an interaction between processing condition and group suggests that for a given number of maskers the group difference was constant across processing conditions. In other words, HI listeners showed a consistent deficit relative to NH listeners for a given stimulus configuration, even in the absence of any masker energy, when specific regions of the target were simply missing in each ear (as they would be in the mixture due to energetic masking). This result suggests that the deficit HI listeners display in a spatialized speech mixture might be accounted for by a deficit in understanding a target sentence based on only sparse glimpses. This mitigates the need for an explanation based on difficulties with spatial hearing, or with segregating the target from the masker, since these processes are not required in the glimpsed conditions. To investigate this idea further, individual performance in the natural condition was compared with performance for the glimpsed condition (Fig. 5). Thresholds were strongly correlated across these two conditions for two maskers ($r = 0.95$, $p < 0.001$) and four maskers

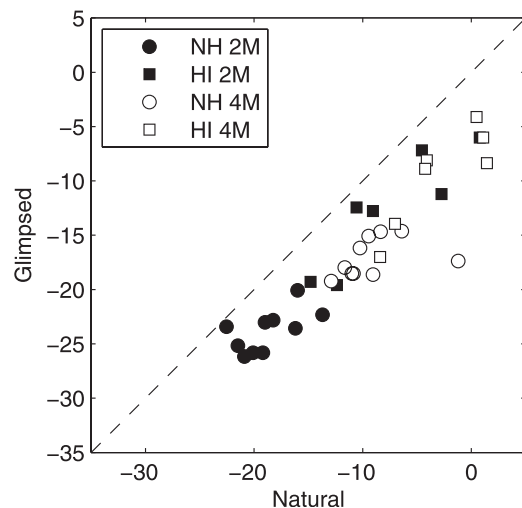


FIG. 5. Scatterplot showing individual thresholds for the glimpsed condition against individual thresholds for the natural condition in experiment 1.

($r = 0.83$, $p = 0.001$). While not conclusive, this is consistent with the idea that performance in the natural condition is limited by the same factors as performance in the glimpsed condition. Within the HI group, thresholds in both the natural and glimpsed conditions for both two and four maskers were significantly correlated with 4FAHL ($r > 0.80$, $p < 0.05$ for all four cases).

C. Discussion

Application of the glimpsing model improved performance for all listeners, leading to thresholds in the glimpsed condition that were better than in the natural condition by around 5–7 dB. As discussed in detail by others (Brungart *et al.*, 2006; Kidd *et al.*, 2016), the magnitude of this improvement can be thought of as a measure of informational masking in the original speech mixture, since the glimpsing aims to largely preserve the effects of energetic masking (by removing tiles that were obscured by the masker). Brungart *et al.* reported improvements of around 3–5 dB for a noise masker, where the informational masking can be assumed to be very low, and improvements of 22–25 dB for highly confusable speech maskers, where the informational masking is high. Wang *et al.* (2009) also observed relatively larger benefits of glimpsing in a multi-talker cafeteria background than in speech-shaped noise (10.5 vs 7.4 dB in NH listeners). The relatively small improvement in our experiment likely reflects the fact that in our baseline natural condition the informational masking was quite low due to the spatial separation between target and maskers.

An interesting finding from experiment 1 was that performance in the glimpsed and glimpsed target conditions was identical. This suggests that the residual masker energy present in target-dominated tiles had very little impact on performance. A similar observation was made by Drullman (1995) for speech in a background of steady-state noise; our study extends this finding to speech maskers and listeners with hearing loss.

Importantly, the benefit of glimpsing was similar for NH and HI listeners. In other words, the glimpsing did not help HI listeners reach normal performance levels, and the group difference remained. This result seems to conflict with the results of Wang *et al.* (2009), who found larger improvements in HI listeners that led to equivalent NH/HI performance in their glimpsed conditions. This difference is puzzling, especially given that their HI group had a very similar average audiometric profile to our HI group, and were also given individualized spectral shaping via NAL-RP. However, there are several methodological details that may explain this difference. First, while Wang and colleagues use a variable target level to calculate thresholds in the natural condition, they switched to a variable masker level in the glimpsed condition, specifically to maintain the overall energy in the increasingly sparse target. Moreover, they applied additional broadband amplification for their HI listeners to ensure that the stimuli were “audible and yet not uncomfortably loud” in all conditions. Both of these adjustments would have minimized the effects of audibility in the

glimpsed condition, especially for the HI group. In our experiment, within-glimpse target levels were deliberately kept constant across the natural and glimpsed conditions so that a direct comparison of the intelligibility of the glimpses could be made, and no additional gain was provided to the HI group beyond NAL-RP shaping.

Arguably then, the simplest explanation for the group difference observed in experiment 1 (for both natural and glimpsed speech mixtures) is that the audibility of the target glimpses was inadequate in the HI group, due to a combination of the presentation level and the relatively severe hearing losses of this group (which are not fully compensated for by NAL-RP amplification). This is supported by the persistent relationship between performance and 4FAHL. In experiment 2, the contribution of the presentation level was examined further by repeating the experiment with the overall level increased by 10 dB for all listeners.

IV. EXPERIMENT 2

A. Procedures

Experiment 2 was identical to experiment 1 except for the masker presentation level (which was 10 dB higher, as described in Sec. II B). Moreover, in the interests of time, the glimpsed target condition was not included. Experiment 2 took around 1.5 h and was completed in a single visit. Eight of the eleven NH listeners and all eight of the HI listeners completed experiment 2. Again, the data for one NH and one HI subject were excluded because thresholds could not be obtained in several conditions (the same two that were excluded from experiment 1) and thus data are presented for seven NH and seven HI listeners.²

B. Results

Figure 6 shows means (and standard deviations) of the thresholds in each group for each condition. A three-way mixed ANOVA conducted on the thresholds (with factors of number of maskers, processing condition, and group) revealed a significant main effect of number of maskers [$F(1,12) = 495.3$, $p < 0.001$], processing condition [$F(1,12) = 146.2$, $p < 0.001$], and group [$F(1,12) = 50.5$, $p < 0.001$]. The two-way interaction between number of maskers and group was significant [$F(1,12) = 64.1$, $p < 0.001$], as was the interaction between number of maskers and processing condition [$F(1,12) = 6.3$, $p = 0.028$]. No other interactions were significant [$p > 0.05$]. This pattern of results is very similar to those observed in experiment 1. Again, thresholds were strongly correlated across natural and glimpsed conditions for two maskers ($r = 0.90$, $p < 0.001$) and four maskers ($r = 0.95$, $p = 0.001$), as illustrated in Fig. 7. Within the HI group, thresholds in both the natural and glimpsed conditions for both two and four maskers were again significantly correlated with 4FAHL ($r > 0.76$, $p < 0.05$ for all four cases).

A direct comparison of thresholds in experiments 1 and 2 for the listeners who were common to both experiments found a global improvement in performance for the higher presentation level. This was confirmed by a four-way mixed ANOVA conducted on the thresholds (with factors of

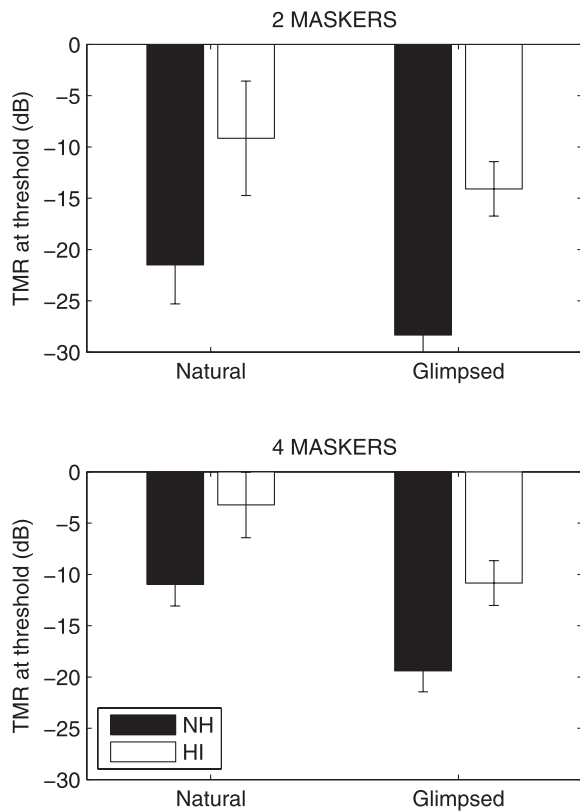


FIG. 6. Mean thresholds (and across-subject standard deviations) for the different conditions in experiment 2.

experiment, number of maskers, processing condition, and group) which revealed a significant main effect of experiment [$F(1,12) = 15.9, p = 0.002$]. The only significant interaction involving experiment was the two-way interaction with number of maskers [$F(1,12) = 13.5, p = 0.003$]. On average, the improvement in thresholds was 2.4 dB for the two-masker condition and 1.2 dB for the four-masker condition. The fact that there was no significant interaction between experiment and group [$F(1,12) = 1.9, p = 0.2$]

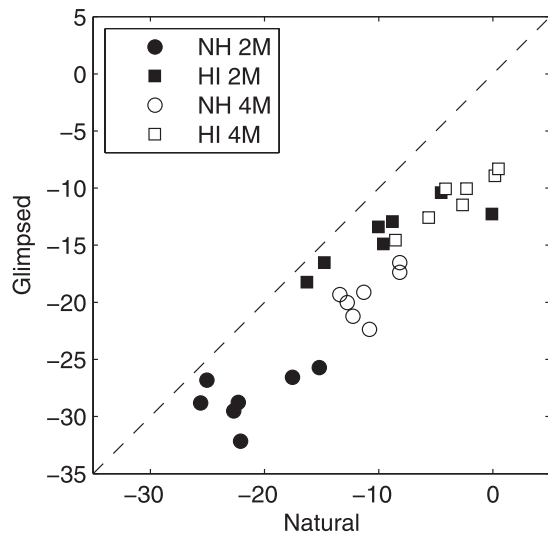


FIG. 7. Scatterplot showing individual thresholds for the glimpsed condition against individual thresholds for the natural condition in experiment 2.

indicates that a similar improvement was experienced by both groups.

A final comparison was made between the NH group tested at the lower level (experiment 1) and the HI group tested at the higher level (experiment 2). A three-way mixed ANOVA conducted on the thresholds (with factors of number of maskers, processing condition, and group) revealed a significant main effect of number of maskers [$F(1,15) = 284.5, p < 0.001$], processing condition [$F(1,15) = 142.4, p < 0.001$], and group [$F(1,15) = 37.6, p < 0.001$]. The two-way interaction between number of maskers and group was significant [$F(1,15) = 22.6, p < 0.001$], as was the interaction between number of maskers and processing condition [$F(1,15) = 12.7, p = 0.003$]. No other interactions were significant [$p > 0.05$]. This pattern of results is similar to that found in each of experiments 1 and 2 and, in particular, the significant main effect of group persisted. In other words, the broadband increase in level in experiment 2, even when given only to the HI group, did not close the gap between the NH and HI groups.

C. Discussion

The results of experiment 1 suggested that non-spatial factors might limit intelligibility for both NH and HI listeners in spatialized speech mixtures. We hypothesized that this limit might be related to inadequate audibility of target glimpses at low target levels, which would especially affect HI listeners, and thus experiment 2 was conducted using a higher overall presentation level. While both groups experienced a small improvement in performance, the level increase did not put the groups on an equal footing.

To further explore the relationship between audibility and performance, and to understand why the increase in overall level did not offer a more robust improvement for HI listeners, it is useful to look closely at the level and spectrum of the presented stimuli in relation to the audiometric profile of our listeners. Figure 8 shows estimated critical band levels of the unprocessed target speech at the eardrum, for

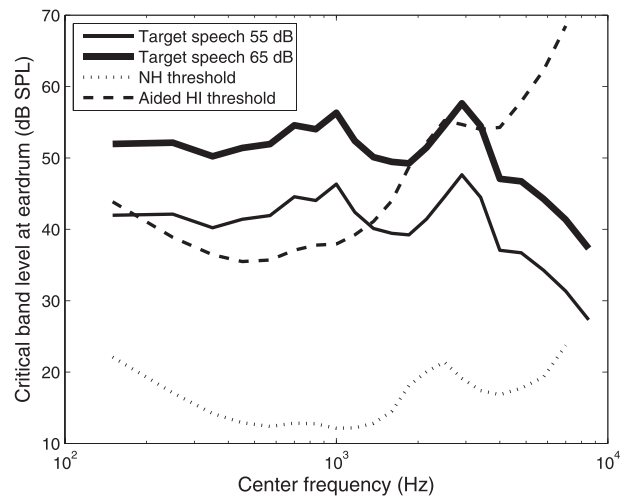


FIG. 8. Critical band levels of unprocessed target speech at 55 and 65 dB SPL (thin/thick solid lines) estimated at the eardrum, shown in comparison to average NH hearing thresholds (dotted line) and average aided HI thresholds (dashed lines).

presentation levels of 55 and 65 dB. Also shown on this figure are average NH thresholds (defined as 0 dB HL at all frequencies) and the average aided HI thresholds. The latter were calculated by subtracting the average applied NAL-RP gain from the average unaided thresholds of the seven HI listeners whose data were included in experiments 1 and 2 (Fig. 1, circles). If one considers the fact that within a target glimpse, the level of the target is as it would be without any processing, then this figure highlights the potential impact of reduced audibility on the availability of the target glimpses.

For a 55 dB presentation level, NH listeners have access to the whole bandwidth of the speech materials (up to 8 kHz), and a robust sensation level of around 20 dB across the whole range. The HI group, on the other hand, has a drastically reduced bandwidth (with a cutoff of between 1 and 2 kHz) and a very low sensation level. For the higher presentation level, while the audible bandwidth increases in the HI group, it is still substantially narrower than in the NH group and the sensation level is still inadequate at many frequencies. Thus, it is possible that an increase in overall level of 10 dB is simply not sufficient to combat the sloping losses of this group (see also Humes, 2007). To quantify these observations, we calculated the Speech Intelligibility Index (SII) [ANSI S3.5-1997, ANSI (1997)] using our average NH and aided HI thresholds and speech spectra (as plotted in Fig. 8). The SII provides a value between 0 and 1 that characterizes the audibility of speech information. For the NH group, the SII is estimated at 0.99 and 1.00 for 55 and 65 dB SPL, respectively. For the HI group, these values are 0.41 and 0.69, indicating that while substantially more speech information was made available, audibility was still far from “normal.” It is somewhat surprising that an improvement in behavioral performance was observed in the NH group despite SII values being at ceiling. However, this is probably because the SII values reflect the intelligibility of *intact* speech; for our rather impoverished *glimpsed* speech one would not necessarily expect performance to saturate at high SII values.

We also calculated SII values for each of the seven HI listeners, using their individual hearing thresholds, gain prescriptions, and presentation levels. For experiment 1, these values varied from 0.21 to 0.74, and were significantly correlated with glimpsed thresholds for both two maskers ($r = -0.92$, $p = 0.003$) and four maskers ($r = -0.88$, $p = 0.01$). For experiment 2, these values varied from 0.53 to 0.89, and again were significantly correlated with glimpsed thresholds for both two maskers ($r = -0.83$, $p = 0.02$) and four maskers ($r = -0.88$, $p = 0.009$). Of course with only seven listeners it is hard to draw firm conclusions, but this analysis provides additional support for the idea that audibility of the speech information might explain why the HI listeners show a deficit in the glimpsed condition, which may also extend to the natural condition.

The only way to improve the audibility for HI listeners would be to provide substantially more gain in the high frequencies. However, this raises a general issue with amplification for sloping losses of this nature; it is almost impossible to restore “normal” audibility across the entire speech spectrum because of physical limits on the amount of gain

that can be delivered, limits related to discomfort for loud high-frequency sounds, and a loss of sound quality and intelligibility at high levels (for discussion see Dillon, 2012; Glyde *et al.*, 2015). This makes it difficult to rule out contributions of audibility when comparing the performance of NH and HI listeners. Given this issue, in experiment 3 we took the approach of Glyde *et al.* (2015) and estimated the impact of reduced audibility in NH listeners by simulating the frequency-dependent audibility pattern of our HI group.

V. EXPERIMENT 3

A. Procedures

Spectral shaping was applied to each ear of each stimulus before presentation over the headphones in order to simulate the aided audibility of our HI listeners. Because experiment 3 was commenced before the completion of experiments 1 and 2, the simulation filter was based on the audiograms of only the first 6 HI listeners recruited. Figure 1 shows the average audiogram for this subset of listeners (squares), as well as the average audiogram for the 7 HI listeners whose data were ultimately included in experiments 1 and 2 (circles). Aided thresholds were calculated as per Fig. 8 by subtracting the average applied NAL-RP gain from the unaided thresholds.

Frequency-dependent attenuation was realized using a linear finite impulse response filter of length 256 samples. The stimuli and procedures were otherwise identical to experiment 1 (except that the glimpsed target condition was excluded). This experiment took around 1.5 h and was completed in a single visit. A subset of the NH listeners (seven of eleven) completed experiment 3, and their data from experiment 1 are included in the following analyses for comparison.

B. Results

Figure 9 shows means (and standard deviations) of the thresholds in each condition. A comparison of this figure to Fig. 4 confirms that thresholds for the unfiltered condition for this subset of listeners were similar to those of the full group obtained in experiment 1. A three-way repeated-measures ANOVA conducted on the thresholds revealed a significant main effect of number of maskers [$F(1,6) = 151.8$, $p < 0.001$], processing condition [$F(1,6) = 52.2$, $p < 0.001$], and filtering condition [$F(1,6) = 163.0$, $p < 0.001$]. The interaction between number of maskers and filtering condition was significant [$F(1,6) = 103.4$, $p < 0.001$], suggesting that the filtering had a larger impact on performance with two maskers than with four maskers (similar to the effect of hearing loss in experiments 1 and 2). Unlike in experiments 1 and 2, however, the interaction between processing condition and filtering condition was also significant [$F(1,6) = 16.6$, $p = 0.007$]. As Fig. 9 shows, this interaction reflects the fact that the effect of filtering was slightly larger in the glimpsed condition than in the natural condition. But again, thresholds across these two conditions were related (Fig. 10) with significant correlations for both two maskers ($r = 0.95$, $p < 0.001$) and four maskers ($r = 0.80$, $p < 0.001$). Overall,

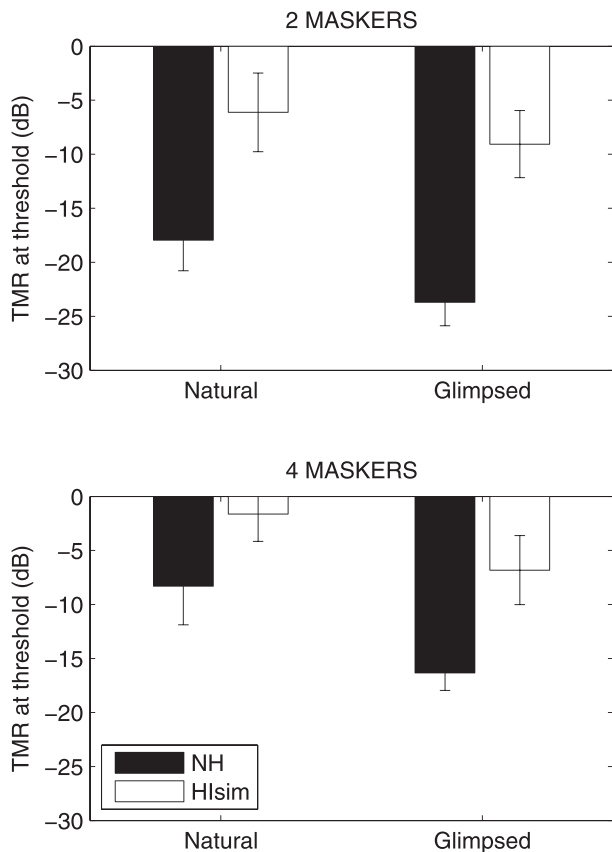


FIG. 9. Mean thresholds (and across-subject standard deviations) for the different conditions in experiment 3.

the detrimental effect of filtering in experiment 3 was similar to the effect of hearing loss observed in experiment 1 (13 vs 11 dB on average for two maskers, 8 vs 7 dB on average for four maskers).

C. Discussion

The results of experiment 3 demonstrate that a simple loss of audibility, approximating that experienced by the HI

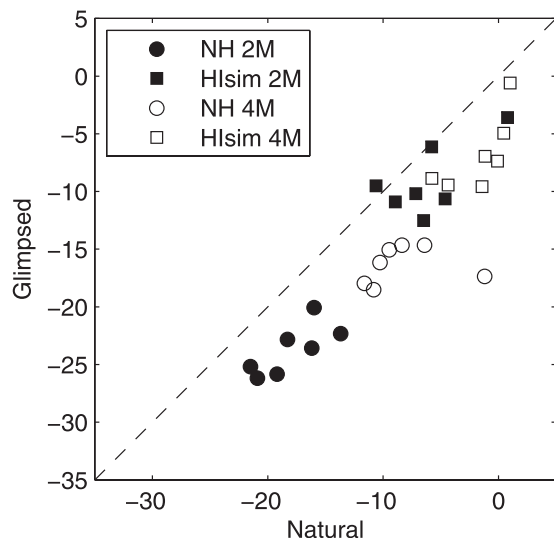


FIG. 10. Scatterplot showing individual thresholds for the glimpsed condition against individual thresholds for the natural condition in experiment 3.

listeners in experiment 1, can drastically reduce performance in a spatialized speech mixture. Thresholds in the natural condition dropped by 7–12 dB as a result of the imposed filtering, which not only reduced the audible bandwidth but also reduced the sensation level at audible frequencies in these young NH listeners. These results echo the results of Glyde *et al.* (2015), who were able to explain a large portion of the deficit in their older HI listeners using a similar simulation. Critically though, the filtering also reduced thresholds in the glimpsed condition by 10–15 dB, suggesting a strong and direct effect on intelligibility of the target glimpses. To the extent that our glimpsing model captures the target information available in the spatialized mixture, this loss of intelligibility of glimpses can more than account for the poor performance observed in the mixture.

VI. GENERAL DISCUSSION

This study was motivated by several recent studies that have demonstrated effects of stimulus level on SRM as measured using a symmetric speech-on-speech masking paradigm (e.g., Gallun *et al.*, 2013; Glyde *et al.*, 2015; Jakien *et al.*, 2017). To investigate the possibility that reduced audibility, particularly in HI listeners, might limit SRM, we wanted a way to estimate target audibility for individual listeners in binaural speech mixtures. Typical control conditions that measure intelligibility for the target speech in quiet may not be sufficient as they often produce ceiling-level performance and do not take into account the fact that only portions of the target speech are available in a speech mixture. Such a reduced representation of speech contains much less redundancy than intact speech, and thus intelligibility might be much more susceptible to audible bandwidth and sensation level. Our solution was to apply a glimpsing model separately to the two ears, in order to isolate the potentially available target information in a binaural speech mixture. By isolating these glimpses, we were able to directly assess their intelligibility in the absence of competing talkers, removing any effects that may come from differences in how well listeners can use spatial cues to segregate competing talkers or are able to direct attention to the target source. Note that this approach, in its original monaural application, could also be useful for estimating target audibility in monaural speech mixtures.

We found that deficits related to hearing loss were similar in natural and glimpsed conditions, and that individual performance across the two conditions was correlated, suggesting that both conditions are subject to the same limits. Given that the glimpsed conditions essentially measure performance for the target alone, these limits are unlikely to reflect a listener’s ability to segregate competing sounds. A similar conclusion was reached by Woods *et al.* (2013) in their recent study using a rather different approach. Woods *et al.* used a speech corpus specifically designed for the application of the SII, but presented it in the context of a spatialized multitalker situation. They were able to predict performance for the majority of their NH and HI listeners based on audibility and a single proficiency factor based on individual performance in quiet. Moreover, because the

glimpsed conditions do not have a strong spatial component, our results also suggest that variations in individual performance are unlikely to be related to deficits in spatial hearing. This adds to mounting evidence from our laboratory that non-spatial factors can influence SRM in listeners with and without hearing loss (Kidd *et al.*, 2010; Best *et al.*, 2012; Best *et al.*, 2013).

By simulating the audibility profile of our HI listeners in a group of NH listeners, we demonstrated that reduced audibility has a dramatic effect on the intelligibility of a processed speech mixture where only the clean target glimpses are presented, just as it does in a spatialized speech mixture. This suggests that audibility may be sufficient to explain the poor performance of HI listeners under these conditions. However, there are several other possible explanations for reduced intelligibility in HI listeners, which we cannot rule out. First, there are many suggestions in the literature that hearing loss is accompanied by poor spectral and/or temporal resolution which contributes to difficulties understanding speech in noise (e.g., Bernstein *et al.*, 2013; Summers *et al.*, 2013). In the context of this study, one might expect that poor resolution would limit one's ability to extract clean glimpses of the target from a speech mixture, thus reducing intelligibility in the natural condition. However, in this case one would also expect that listeners with poor resolution would be less disadvantaged in the glimpsed conditions, where the segregation of targets and maskers is essentially done externally. Our results are not consistent with this, as HI listeners still showed a clear deficit in the glimpsed condition. On the other hand, it is possible that resolution in our HI listeners was good enough for adequate glimpsing, but poor enough to reduce the quality of the glimpses, thus reducing intelligibility of the target glimpses in both natural and glimpsed conditions. Second, it may be that the binding together of sparse target glimpses, or the top-down restoration of missing parts of the target (or "phonemic restoration"), is less efficient in HI listeners (e.g., see Baskent *et al.*, 2010). Related to this, it is also possible that hearing loss somehow disrupts the ability to assemble target glimpses from the two ears into a cohesive whole, although the fact that the majority of the glimpses were present in both ears makes us believe this ability was not a critical factor in our experiment. Whatever the specific difficulty is with making use of target glimpses, our results suggest that it is this difficulty rather than an inability to use spatial cues to segregate sound sources that accounts for the poor performance observed in some HI listeners in spatialized multi-talker mixtures.

In future studies, it will be important to find out whether the results we observed generalize to other kinds of speech materials. Here we used stimuli that were drawn from a closed set, and a task that requires listeners to discriminate between simultaneously presented alternatives from that set. Under these conditions, we found evidence that access to high-frequency information is important. However, it is not known how different frequency regions are weighted in this task, and how that weighting compares to the more natural case of open-set speech. The results also raised the issue of how changes in audibility (e.g., as quantified by the SII)

translate to changes in performance for impoverished speech signals where only sparse spectrotemporal glimpses are available. Finally, it would be of interest to extend this approach to the more typical HI population (who are older), to determine to what extent the intelligibility of glimpsed speech can account for their reduced SRM. Given the growing consensus in the literature that there are specific effects of aging on spatial processing (e.g., Neher *et al.*, 2012; Moore, 2014), our expectation is that performance for glimpsed stimuli in this population will not fully account for the deficit they show in spatialized speech mixtures.

ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD Awards Nos. DC04545, DC013286, DC04663, DC00100 and AFOSR Award No. FA9550-12-1-0171. Portions of the work were presented at the 169th Meeting of the Acoustical Society of America (Pittsburgh, USA, May 2015) and the International Symposium on Hearing (Groningen, Netherlands, June 2015).

¹The three listeners who preferred a higher level did not differ systematically from the other listeners in terms of the severity of their loss, hearing aid usage, etc.

²For another of the HI listeners, a psychometric function could not be fit to the data in one condition only (four maskers, natural). Inspection of her data for the two-masker condition showed that scores were very similar across experiments 1 and 2. Thus, for the missing data point we substituted the four-masker natural threshold from experiment 1.

- Ahlstrom, J. B., Horwitz, A. R., and Dubno, J. R. (2009). "Spatial benefit of bilateral hearing AIDS," *Ear. Hear.* **30**, 203–218.
- ANSI (1997). ANSI S3.5-1997, *Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).
- Baskent, D., Eiler, C. L., and Edwards, B. (2010). "Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss," *Hear. Res.* **260**, 54–62.
- Bernstein, J. G., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., and Leek, M. R. (2013). "Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners," *J. Am. Acad. Audiol.* **24**, 293–306.
- Besser, J., Festen, J. M., Goverts, S. T., Kramer, S. E., and Pichora-Fuller, M. K. (2015). "Speech-in-speech listening on the LiSN-S test by older adults with good audiograms depends on cognition and hearing acuity at high frequencies," *Ear Hear.* **36**, 24–41.
- Best, V., Marrone, N., Mason, C. R., and Kidd, G., Jr. (2012). "The influence of non-spatial factors on measures of spatial release from masking," *J. Acoust. Soc. Am.* **131**, 3103–3110.
- Best, V., Mason, C. R., Kidd, G., Jr., Iyer, N., and Brungart, D. S. (2015). "Better-ear glimpsing in hearing-impaired listeners," *J. Acoust. Soc. Am.* **137**, EL213–EL219.
- Best, V., Mason, C. R., Thompson, E. R., and Kidd, G., Jr. (2013). "An energetic limit on spatial release from masking," *J. Assoc. Res. Otolaryng.* **14**, 603–610.
- Bronkhorst, A. W., and Plomp, R. (1988). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.* **83**, 1508–1516.
- Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2006). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," *J. Acoust. Soc. Am.* **120**, 4007–4018.
- Brungart, D. S., and Iyer, N. (2012). "Better-ear glimpsing efficiency with symmetrically-placed interfering talkers," *J. Acoust. Soc. Am.* **132**, 2545–2556.
- Byrne, D. J., Parkinson, A., and Newall, P. (1991). "Modified hearing aid selection procedures for severe-profound hearing losses," in *The Vanderbilt Hearing Aid Report II*, edited by G. A. Studebaker, F. H. Bess, and L. B. Beck (York, Parkton, MD), pp. 295–300.

- Cameron, S., and Dillon, H. (2007). "Development of the Listening in Spatialized Noise–Sentences test (LiSN-S)," *Ear Hear.* **28**, 196–211.
- Colburn, H. S. (1982). "Binaural interaction and localization with various hearing impairments," *Scand. Audiol. Suppl.* **15**, 27–45.
- Cooke, M. (2006). "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**, 1562–1573.
- Dillon, H. (2012). *Hearing Aids* (Boomerang Press, Turramurra), pp. 286–335.
- Dobreva, M. S., O'Neill, W. E., and Paige, G. D. (2011). "Influence of aging on human sound localization," *J. Neurophysiol.* **105**, 2471–2486.
- Drullman, R. (1995). "Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level," *J. Acoust. Soc. Am.* **98**, 1796–1798.
- Gallun, F. J., Diedesch, A. C., Kampel, S. D., and Jakien, K. M. (2013). "Independent impacts of age and hearing loss on spatial release in a complex auditory environment," *Front. Neurosci.* **7**, 252.
- Gallun, F. J., McMillan, G. P., Molis, M. R., Kampel, S. D., Dann, S. M., and Konrad-Martin, D. L. (2014). "Relating age and hearing loss to monaural, bilateral, and binaural temporal sensitivity," *Front. Neurosci.* **8**, 172.
- Glyde, H., Buchholz, J. M., Dillon, H., Best, V., Hickson, L., and Cameron, S. (2013a). "The effect of better-ear glimpsing on spatial release from masking," *J. Acoust. Soc. Am.* **134**, 2937–2945.
- Glyde, H., Buchholz, J. M., Nielsen, L., Best, V., Dillon, H., Cameron, S., and Hickson, L. (2015). "Effect of audibility on spatial release from speech-on-speech masking," *J. Acoust. Soc. Am.* **138**, 3311–3319.
- Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (2013b). "The effects of hearing impairment and aging on spatial processing," *Ear Hear.* **34**, 15–28.
- Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (1999). "Speech intelligibility and localization in a multi-source environment," *J. Acoust. Soc. Am.* **105**, 3436–3448.
- Humes, L. E. (2007). "The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults," *J. Am. Acad. Audiol.* **18**, 590–603.
- Jakien, K. M., Kampel, S. D., Gordon, S. Y., and Gallun, F. J. (2017). "The benefits of increased sensation level and bandwidth for spatial release from masking," *Ear Hear.* **38**, e13–e21.
- 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., Mason, C. R., Best, V., and Marrone, N. (2010). "Stimulus factors influencing spatial release from speech-on-speech masking," *J. Acoust. Soc. Am.* **128**, 1965–1978.
- 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 Handbook of Auditory Research, New York), pp. 143–189.
- Kidd, G., Jr., Mason, C. R., Swaminathan, J., Roverud, E., Clayton, K. K., and Best, V. (2016). "Determining the energetic and informational components of speech-on-speech masking," *J. Acoust. Soc. Am.* **140**, 132–144.
- Koehnke, J., Passaro Culotta, C., Hawley, M. L., and Colburn, H. S. (1995). "Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing," *Ear Hear.* **16**, 331–353.
- Levy, S. C., Freed, D. J., Nilsson, M., Moore, B. C. J., and Puria, S. (2015). "Extended high-frequency bandwidth improves speech reception in the presence of spatially separated masking speech," *Ear Hear.* **36**, e214–e224.
- Marrone, N., Mason, C. R., and Kidd, G., Jr. (2008). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," *J. Acoust. Soc. Am.* **124**, 3064–3075.
- Moore, B. C. J. (2014). *Auditory Processing of Temporal Fine Structure* (World Scientific, Singapore), pp. 103–137.
- Moore, B. C. J., Füllgrabe, C., and Stone, M. A. (2010). "Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task," *J. Acoust. Soc. Am.* **128**, 360–371.
- Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., and van Schaik, A. (2009). "Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition," *Int. J. Audiol.* **48**, 758–774.
- Neher, T., Lougesen, S., Jensen, N. S., and Kragelund, L. (2011). "Can basic auditory and cognitive measures predict hearing-impaired listeners localization and spatial speech recognition abilities?," *J. Acoust. Soc. Am.* **130**, 1542–1558.
- Neher, T., Lunner, T., Hopkins, K., and Moore, B. C. (2012). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners," *J. Acoust. Soc. Am.* **131**, 2561–2564.
- Noble, W., Byrne, D., and Lepage, B. (1994). "Effects on sound localization of configuration and type of hearing impairment," *J. Acoust. Soc. Am.* **95**, 992–1005.
- Noble, W., Byrne, D., and Ter-Host, K. (1997). "Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners," *J. Acoust. Soc. Am.* **102**, 2343–2352.
- Otte, R. J., Agterberg, M. J., Van Wanrooij, M. M., Snik, A. F., and Van Opstal, A. J. (2013). "Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance," *J. Assoc. Res. Otolaryng.* **14**, 261–273.
- Roman, N., Wang, D., and Brown, G. J. (2003). "Speech segregation based on sound localization," *J. Acoust. Soc. Am.* **114**, 2236–2252.
- Spencer, N. J. (2013). "Factors for speech intelligibility in the symmetric and anti-symmetric speech-masker conditions," Ph.D. thesis, Boston University, Boston, MA.
- Strelyck, O., and Dau, T. (2009). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," *J. Acoust. Soc. Am.* **125**, 3328–3345.
- Summers, V., Makashay, M. J., Theodoroff, S. M., and Leek, M. R. (2013). "Suprathreshold auditory processing and speech perception in noise: Hearing-impaired and normal-hearing listeners," *J. Am. Acad. Audiol.* **24**, 274–292.
- Wang, D. (2005). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer, Norwell, MA), pp. 181–197.
- Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (2009). "Speech intelligibility in background noise with ideal binary time-frequency masking," *J. Acoust. Soc. Am.* **125**, 2336–2347.
- Woods, W. S., Kalluri, S., Pentony, S., and Nooraie, N. (2013). "Predicting the effect of hearing loss and audibility on amplified speech reception in a multi-talker listening scenario," *J. Acoust. Soc. Am.* **133**, 4268–4278.