

# Better-ear glimpsing in hearing-impaired listeners

**Virginia Best, Christine R. Mason, and Gerald Kidd, Jr.**

*Department of Speech, Language and Hearing Sciences and Hearing Research Center,  
Boston University, Boston, Massachusetts 02215  
ginbest@bu.edu, cmason@bu.edu, gkidd@bu.edu*

**Nandini Iyer**

*Air Force Research Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio 45433  
nandini.iyer.2@us.af.mil*

**Douglas S. Brungart**

*Walter Reed National Military Medical Center, Bethesda, Maryland 20889  
Douglas.S.Brungart.civ@health.mil*

**Abstract:** When competing speech sounds are spatially separated, listeners can make use of the ear with the better target-to-masker ratio. Recent studies showed that listeners with normal hearing are able to efficiently make use of this “better-ear,” even when it alternates between left and right ears at different times in different frequency bands, which may contribute to the ability to listen in spatialized speech mixtures. In the present study, better-ear glimpsing in listeners with bilateral sensorineural hearing impairment, who perform poorly in spatialized speech mixtures, was investigated. The results suggest that this deficit is not related to better-ear glimpsing.

© 2015 Acoustical Society of America

[QJF]

**Date Received:** October 27, 2014    **Date Accepted:** January 8, 2015

## 1. Introduction

When competing sounds are spatially separated, listeners can make use of the ear with the better signal-to-noise ratio (the acoustically “better ear,” e.g., [Bronkhorst and Plomp, 1988](#)). For listening situations in which maskers are located on both sides of the head, however, the better ear can alternate between left and right at unpredictable times in different frequency bands according to the natural level variations in the competing sounds.

Two recent studies attempted to understand how these “better-ear glimpses” contribute to binaural performance by taking the better-ear glimpses available in a binaural signal and assembling them into one monaural signal. [Brungart and Iyer \(2012\)](#) demonstrated that listeners with normal hearing (NH) perform as well in this condition as in the natural binaural condition, suggesting that performance in the binaural condition can be explained on the basis of an optimal glimpsing strategy. [Glyde \*et al.\* \(2013a\)](#), however, showed that glimpsing cannot fully explain binaural performance for stimuli high in informational masking.

Listeners with sensorineural hearing impairment (HI) have been shown under many conditions to perform more poorly than NH listeners when competing sounds are spatially separated (e.g., [Marrone \*et al.\*, 2008](#); [Best \*et al.\*, 2012](#); [Glyde \*et al.\*, 2013b](#)). This may be due to various factors such as loss of audibility and/or degraded binaural cues, although attempts to explain the deficit on the basis of these two factors have met with mixed success. Another speculation is that listeners with HI may be less able to make use of better-ear glimpses in complex speech mixtures, perhaps as a result of reduced spectrotemporal resolution (e.g., [Marrone \*et al.\*, 2008](#)). If so, HI listeners may benefit greatly from a processing strategy that accomplishes better-ear glimpsing for them and may achieve better performance than is possible with natural binaural

cues. To test this hypothesis we examined better-ear glimpsing in NH and HI listeners for speech mixtures that were either low or high in informational masking.

## 2. Methods

### 2.1 Participants

Participants were paid listeners with audiometrically normal hearing, or with symmetric bilateral sensorineural hearing impairment. In experiment 1, eight NH (ages 18–36, mean 24 years) and nine HI (ages 18–43, mean 26 years) listeners participated. In experiment 2, eight NH (ages 20–23, mean 21 years) and eight HI (ages 18–39, mean 25 years) listeners participated. Three of the HI listeners were common to both experiments. Reduced audibility in HI listeners was compensated for by applying individualized NAL-RP gain before stimulus presentation. Audiograms for the HI listeners are shown in Fig. 1.

### 2.2 Stimuli and procedures

Listeners were seated in a double-walled audiometric booth (Industrial Acoustics Corporation). Stimuli were delivered via headphones (Sennheiser HD 280Pro). Digital stimuli were generated on a PC outside the booth and then fed through separate channels of Tucker-Davis Technologies System II hardware. This included 16-bit digital-to-analog conversion at 50 kHz (DA8), low-pass filtering at 20 kHz (FT6), and programmable attenuation (PA4).

Stimuli in experiment 1 were identical to those used by [Brungart and Iyer \(2012\)](#). Targets were sentences from the Modified Rhyme Test (MRT; [House \*et al.\*, 1965](#)). These sentences have the form “You will mark [keyword] now,” where the keyword is randomly selected from 300 monosyllabic words. Listeners were presented with a set of six rhyming options (e.g., PEAK, PEAS, PEAT, PEAL, PEACE, PEACH) and selected one by clicking on the relevant button on the screen. On each trial the target voice was selected from a set of six (three male and three female). Maskers were sections of continuous discourse spoken by two talkers of the same gender as the target. As discussed by [Brungart and Iyer \(2012\)](#), the targets and maskers were chosen specifically not to be confusable with each other, and therefore by design contain very little informational masking.

In experiment 2 the target and two maskers were all sentences from the Coordinate Response Measure corpus (CRM; [Bolia \*et al.\*, 2000](#); [Brungart, 2001](#)). These sentences have the form “Ready [callsign] go to [color] [number] now,” where

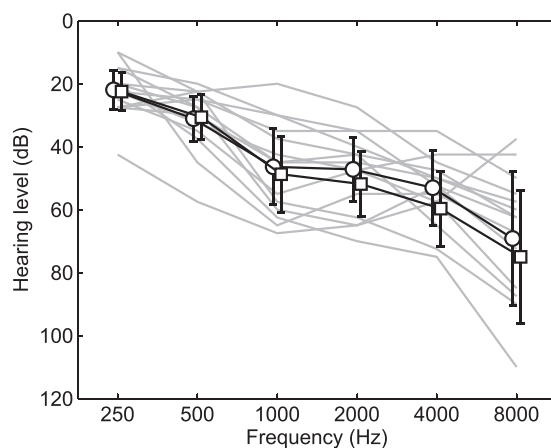


Fig. 1. Individual hearing thresholds (averaged across ears) for the HI listeners and mean thresholds for the groups who completed experiment 1 (circles) and experiment 2 (squares). The NH listeners had thresholds better than 20 dB hearing level at all frequencies.

the call signs, colors, and numbers are selected from sets of 8, 4, and 8, respectively. It is well established that these stimuli may produce large amounts of informational masking, due to the high degree of confusability between the target and masker sentences. Listeners were instructed to follow the sentence with the call sign “Baron” and select the corresponding color/number by clicking a button on the screen. On each trial the target voice was selected from a set of six (three male and three female) and the maskers were spoken by the other two talkers of the same gender.

To generate psychometric functions, a range of target-to-masker ratios (TMRs) was tested by varying the level of the target relative to a fixed-level masker (65 dB sound pressure level, pre-gain in the HI listeners). In experiment 1, the NH group was tested at six TMRs 3 dB apart spanning the range  $-15$  to  $0$  dB. For the HI listeners the range of TMRs was extended to better cover the dynamic range of performance ( $-15$  to  $10$  dB in 5 dB steps). Each TMR was tested a total of 40 times in each of the six masker conditions (see below) for a total of 1440 trials. Before testing with maskers, intelligibility of the target sentences in quiet was tested for each of the levels used in the masked conditions (20 trials per level, 120 trials in total). Total testing time per listener for experiment 1 was approximately two hours.

In experiment 2, the NH listeners were tested at 6 TMRs ( $-15$  to  $10$  dB) and the HI listeners were tested at 5 TMRs ( $-10$  to  $10$  dB). Each TMR was tested a total of 20 times in each of the four masker conditions (see below) for a total of 480/400 trials (NH/HI). For the HI group the intelligibility of the target sentences in quiet was tested for each of the levels used in the masked conditions (10 trials per level, 50 trials in total) before the masker conditions commenced. The NH group was not tested in quiet, but were given 1 quiet trial per target level for familiarization purposes (six trials in total). Total testing time per listener for experiment 2 was approximately one hour.

### 2.3 Better-ear processing

Better-ear processing was described in detail in [Brungart and Iyer \(2012\)](#). Briefly, spatialized stimuli were generated using KEMAR head-related transfer functions (HRTFs). The target was always simulated to be at  $0^\circ$  azimuth, while the maskers were simulated to be colocated with the target (at  $0^\circ$  azimuth) or separated symmetrically on either side (at  $\pm 60^\circ$  azimuth). Better-ear processing was then applied to the spatially separated signals. Left and right ear masker signals were processed by a 128-channel gammatone filterbank (80–5000 Hz). The time-domain outputs from each channel were then divided into 20-ms segments (with 50% overlap) and multiplied by a 20-ms raised-cosine window. Because the HRTFs were symmetric across the median sagittal plane, the target signal was always identical in the two ears. Thus it was possible to identify the “better ear,” independently for each time-frequency segment, as the ear with the lower total root-mean-square (rms) energy. Conversely, the worse ear for that segment was defined as the ear with the higher rms.

Six masker conditions were created as per [Brungart and Iyer \(2012\)](#).

- (1) Colocated: target and maskers presented from  $0^\circ$  azimuth.
- (2) Binaural: target presented from  $0^\circ$  azimuth and maskers presented from  $\pm 60^\circ$  azimuth.
- (3) Better-ear: for every time/frequency segment, the better ear was chosen from the binaural stimulus and the resynthesized mixture was presented diotically.
- (4) Worse-ear: for every time/frequency segment, the worse ear was chosen from the binaural stimulus and the resynthesized mixture was presented diotically.
- (5) Hybrid: the better-ear segments were presented to one ear (chosen randomly on every trial), and the worse-ear segments were presented to the other ear. This presentation meant that the two ears received the original left and right ear signals within a segment, and thus the original interaural differences were preserved (unlike in the diotic better-ear and worse-ear signals where the interaural differences were always zero).
- (6) Monaural: only one ear (randomly chosen) from the binaural stimulus was presented.

In experiment 1, all six conditions were tested. In experiment 2, only the collocated, binaural, better-ear, and hybrid conditions were tested. Note that better-ear processing effectively resulted in a frequency-dependent increase in the SNR at the ear relative to the collocated and binaural conditions (see Fig. 2).

### 3. Results

#### 3.1 Experiment 1

Logistic functions were fit to the performance data for each listener. No systematic differences in slope were observed across conditions, and thus performance was summarized simply by extracting the TMR corresponding to 60% correct for each condition (referred to from here on as speech reception thresholds or SRTs). Group mean SRTs are shown in Fig. 3(A). The overall pattern of NH results was a reasonable replication of Brungart and Iyer (2012). The best performance occurred for the binaural, better-ear, and hybrid conditions, which were similar. Performance for worse-ear was the poorest. Performance for collocated and monaural conditions was intermediate. The HI group showed a relatively constant deficit relative to the NH group in all conditions (mean 5.8 dB; range 5.1–6.7). A mixed analysis of variance (ANOVA) confirmed that there was a significant effect of condition [ $F(5,75)=95.8$ ,  $p < 0.001$ ], and group [ $F(1,15)=79.8$ ,  $p < 0.001$ ], but no significant interaction [ $F(5,75)=1.7$ ,  $p = 0.2$ ]. Spatial release from masking (collocated SRT–binaural SRT) was 5.1 dB on average in the NH group (cf. 5.5 dB reported by Brungart and Iyer), which was not significantly larger than the 3.5 dB measured in the HI group [ $t(15)=1.7$ ,  $p = 0.1$ ].

To more closely inspect the two key conditions of interest, Fig. 3(B) shows individual better-ear SRTs plotted against binaural SRTs. SRTs in the two conditions were strongly correlated [ $r = 0.91$ ,  $p < 0.001$ ] and fell close to the diagonal (slope of least squares fit = 0.7 dB/dB).

#### 3.2 Experiment 2

Group mean SRTs are shown in Fig. 3(C). For the NH group, the lowest threshold occurred for the binaural condition, intermediate thresholds were found for the better-ear and hybrid conditions, and the highest threshold was found for the collocated condition. The HI group showed no deficit in the collocated condition (mean SRT 0.4 dB less than NH), but small deficits in the better-ear and hybrid conditions (2.0 and 2.2 dB), and the largest deficit in the binaural condition (4.0 dB). A mixed ANOVA revealed a significant effect of condition [ $F(3,42)=29.0$ ,  $p < 0.001$ ], no main effect of

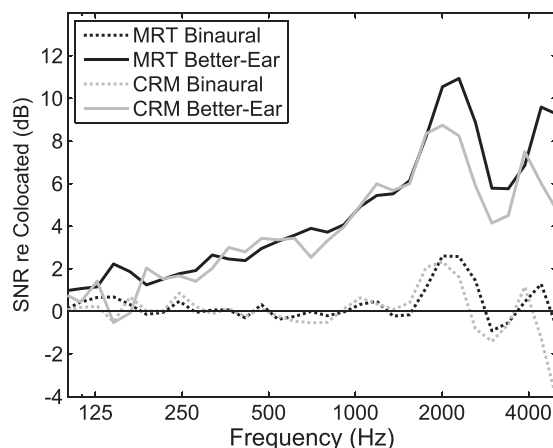


Fig. 2. Average spectra of the MRT stimuli from experiment 1 and the CRM stimuli from experiment 2 in the binaural (dotted line, averaged over left and right ears) and better-ear (solid line) conditions, relative to spectra in the collocated condition.

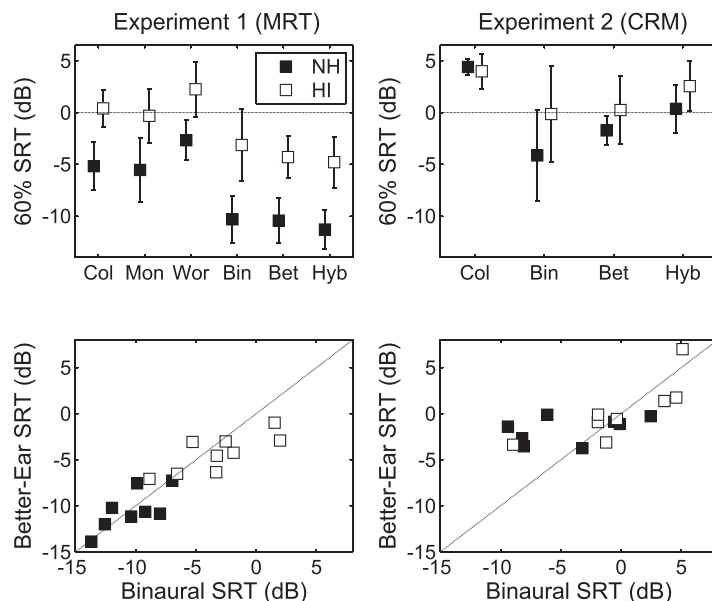


Fig. 3. (A) and (C) Average SRT values for the NH listeners (filled symbols) and HI listeners (open symbols) for the MRT speech materials in experiment 1 and for the CRM speech materials in experiment 2. Error bars indicate across-subject standard deviations. (B) and (D) Scatter plot of SRT values in the better-ear condition versus those in the binaural condition for each listener in experiments 1 and 2.

group [ $F(1,14) = 2.8$ ,  $p = 0.12$ ], and a significant interaction [ $F(3,42) = 3.1$ ,  $p = 0.04$ ]. In contrast to experiment 1, the threshold for the hybrid condition appeared to increase relative to that found in the better-ear condition, whereas, in experiment 1 the two conditions were equivalent. Also different from experiment 1, spatial release from masking was larger on average in the NH group than in the HI group (8.6 dB vs 4.1 dB), and this difference was significant [ $t(14) = 2.3$ ,  $p = 0.04$ ].

To more closely inspect the two key conditions of interest, Fig. 3(D) shows individual better-ear SRTs plotted against binaural SRTs. As in experiment 1, SRTs in the two conditions were correlated [ $r = 0.71$ ,  $p = 0.002$ ] but the least-squares fit was considerably flatter than in experiment 1 (slope = 0.39 dB/dB). When the one HI outlier was excluded from the analysis (rightmost data point), the correlation remained statistically significant [ $r = 0.69$ ,  $p = 0.004$ ] but the slope was even shallower at 0.26 dB/dB. The figure shows that most listeners had similar better-ear thresholds for an extremely wide range of binaural thresholds. Notably, some NH listeners and one HI listener were able to perform much better in the binaural condition than in the better-ear condition.

#### 4. Discussion

In experiment 1, using MRT stimuli, we replicated the previous finding (Brungart and Iyer, 2012) that NH listeners can achieve binaural-equivalent performance with spatially separated maskers if better-ear time/frequency tiles are assembled from the left and right ears and presented diotically. In other words, better-ear glimpses appear to be sufficient to explain spatial release from masking under these conditions. When this experiment was repeated in HI listeners, the HI listeners exhibited poorer performance overall, with SRTs that were on average 5.8 dB higher than in NH listeners. However, the pattern of performance across conditions for the two groups was similar, and in particular, the HI group also showed near-equivalent binaural and better-ear thresholds. Thus, the results of this experiment provided no compelling evidence to suggest that HI listeners benefit more than NH listeners from processing that

implements near-optimal better-ear glimpsing for them, which might have resulted in lower better-ear than binaural thresholds. It seems that the reduced spectral and temporal resolution associated with hearing loss does not have a large impact on the integration of time-frequency glimpses in binaural stimuli. This is consistent with recent modeling work (Glyde *et al.*, 2013a) that suggests broadening the auditory filters has very little impact on the theoretical SNR advantage obtained from better-ear glimpsing.

Importantly, however, the HI group showed no deficit in terms of spatial unmasking for the MRT task in experiment 1 (see Marrone *et al.*, 2008; Best *et al.*, 2012; Glyde *et al.*, 2013b). This may be a result of the relatively low informational masking for the MRT task performed in a masker that was continuous discourse. It has been reported previously that when informational masking is low, the difference in performance between groups is substantially reduced (e.g., by using time-reversed maskers, Best *et al.*, 2012). In experiment 2, we used CRM stimuli which are known to be high in informational masking, and which have previously been shown to produce a difference in spatial release from masking between NH and HI groups.

Compared to experiment 1, overall performance differed less between groups in experiment 2. This is believed to reflect the fact that performance in the CRM task is driven much more by the confusability of the competing voices/words, where the NH and HI listeners are on a relatively even footing, than by the audibility or detectability of the competing messages, where the NH listeners tend to show an advantage. However, the groups differed significantly in the binaural condition, resulting in reduced spatial release from masking in the HI group as reported previously. It is worth noting that there was an extremely large intersubject variability in thresholds in the binaural condition for both groups, consistent with past work on the spatial release from masking under high-informational masking conditions.

In experiment 2, binaural and better-ear thresholds were *not* equivalent for all listeners; the “better” listeners in the binaural condition did not perform equally well with better-ear processing. Presumably, this finding indicates that the better listeners were able to use differences in perceived location afforded by natural binaural cues to segregate confusable talkers and achieve very low SRTs in the binaural condition. This difference is, of course, lost in the diotic presentation of the better-ear stimulus, even though it offers a more favorable SNR. On the other hand, the “poorer” listeners did not show a superiority of binaural listening over better-ear glimpsing, possibly because they were not able to optimally use binaural cues for segregation under natural conditions. It is rather striking that only one-half of the NH group fell into the better-listener group, while the others fell in line with the HI listeners. Conversely, one HI listener outperformed most of the NH listeners and clearly fell in the better-listener category.

Additional evidence for the importance of binaural spatial cues in the highly informational CRM task is provided by performance in the hybrid condition of this experiment. Here the worse-ear stimulus was presented in the ear opposite the better-ear stimulus in order to reinstate the original interaural cues within each time-frequency segment. Critically, this meant that *differences* in interaural cues between the target and masking signals were available to the listeners, which theoretically could facilitate their segregation. This processing slightly improved performance for the NH listeners on the MRT task in the earlier study by Brungart and Iyer (2012), but appeared to have no effect for that task in the current study. On the other hand, the hybrid condition adversely affected performance for both the NH and HI groups in the CRM task. One explanation for this finding is that when integrated across time and frequency, the interaural cues in the maskers corresponded to opposite sides of the head, resulting in “blurred” apparent locations that made them less distinct from the target talker.

Overall the results are consistent with a model of binaural speech perception in which it is assumed that listeners are able to integrate the glimpses of locally favorable SNR that occur across the ears. This seems to afford a large and robust

advantage for speech intelligibility, which might explain a significant portion of the advantage of natural binaural listening under some conditions. However, the mechanism by which this glimpsing is achieved is not well understood. While some form of rapid switching between the ears is possible, it is hard to imagine that such a mechanism could operate automatically and independently in different frequency channels on the time-scale envisioned here (e.g., [Culling and Mansell, 2013](#)). Another possibility is that the signals at the two ears are processed concurrently, but that additional weighting is given to the ear that contains the time-frequency units that more closely match some internal representation of the expected target speech. Further work is needed to obtain a better understanding of the processes by which listeners make use of glimpses in complex binaural stimuli.

### Acknowledgments

Work supported by NIH/NIDCD Grant No. DC04545 and AFOSR Grant No. FA9550-12-1-0171. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, the Department of the Navy, the Department of the Air Force, the Department of Defense, nor the U.S. Government.

### References and links

- 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.
- Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). "A speech corpus for multitalker communications research," *J. Acoust. Soc. Am.* **107**, 1065–1066.
- 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., and Iyer, N. (2012). "Better-ear glimpsing efficiency with symmetrically-placed interfering talkers," *J. Acoust. Soc. Am.* **132**, 2545–2556.
- Brungart, D. S. (2001). "Evaluation of speech intelligibility with the coordinate response measure," *J. Acoust. Soc. Am.* **109**, 2276–2279.
- Culling, J. F., and Mansell, E. R. (2013). "Speech intelligibility among modulated and spatially distributed noise sources," *J. Acoust. Soc. Am.* **133**, 2254–2261.
- 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., 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.
- House, A., Williams, C., Hecker, M., and Kryter, K. (1965). "Articulation testing methods: Consonantal differentiation with a closed response set," *J. Acoust. Soc. Am.* **37**, 158–166.
- 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.