On the Contribution of Target Audibility to Performance in Spatialized Speech Mixtures

Virginia Best, Christine R. Mason, Jayaganesh Swaminathan, Gerald Kidd, Kasey M. Jakien, Sean D. Kampel, Frederick J. Gallun, Jörg M. Buchholz and Helen Glyde

Abstract Hearing loss has been shown to reduce speech understanding in spatialized multitalker listening situations, leading to the common belief that spatial processing is disrupted by hearing loss. This paper describes related studies from three laboratories that explored the contribution of reduced target audibility to this deficit. All studies used a stimulus configuration in which a speech target presented from the front was masked by speech maskers presented symmetrically from the sides. Together these studies highlight the importance of adequate stimulus audibility for optimal performance in spatialized speech mixtures and suggest that reduced access to target speech information might explain a substantial portion of the "spatial" deficit observed in listeners with hearing loss.

Keywords Speech intelligibility · Spatial release from masking · Hearing loss · Amplification · Glimpsing

1 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 acoustic benefits (such as the "head-shadow" advantage) or by effective increases in signal-to-noise

V. Best (✉) · C. R. Mason · J. Swaminathan · G. Kidd
Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, USA
e-mail: ginbest@bu.edu

C. R. Mason
e-mail: cmason@bu.edu

J. Swaminathan
e-mail: jswamy@bu.edu

G. Kidd
e-mail: gkidd@bu.edu

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing,* Advances in Experimental Medicine and Biology 894,
DOI 10.1007/978-3-319-25474-6_10

ratio resulting from neural processing of binaural cues (so called "masking level differences"). In other cases it appears that the perceived separation of sources drives the advantage by enabling attention to be directed selectively.

In many situations, listeners with sensorineural hearing impairment (HI) demonstrate reduced SRM compared to listeners with normal hearing (NH). This observation commonly leads to the conclusion that spatial processing is disrupted by hearing loss. However, convergent evidence from other kinds of spatial tasks is somewhat lacking. For example, studies that have measured fine discrimination of binaural cues have noted that individual variability is high, and some HI listeners perform as well as NH listeners (e.g. Colburn 1982; Spencer 2013). Free-field localization is not strongly affected by hearing loss unless it is highly asymmetric or very severe at low frequencies (e.g. Noble et al. 1994). Other studies have tried to relate SRM in multitalker environments to localization ability (Noble et al. 1997; Hawley et al. 1999) or to binaural sensitivity (Strelcyk and Dau 2009; Spencer 2013) with mixed results. Finally, it has been observed that SRM is often inversely related to the severity of hearing loss (e.g. Marrone et al. 2008). This raises the question of whether in some cases apparent spatial deficits might be related to reduced audibility in spatialized 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 long-term head-shadow benefits to SRM (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. This paper describes related studies from three different laboratories that used the "symmetric masker" configuration to explore the interaction between target audibility and performance under these conditions.

K. M. Jakien · S. D. Kampel · F. J. Gallun
National Center for Rehabilitative Auditory Research, VA Portland Health Care System, Portland, OR, USA
e-mail: kasey.jakien@va.gov

S. D. Kampel
e-mail: sean.kampel@va.gov

F. J. Gallun
e-mail: frederick.gallun@va.gov

J. M. Buchholz · H. Glyde
National Acoustic Laboratories, Macquarie University, Sydney, NSW, Australia
e-mail: Jorg.Buchholz@nal.gov.au

H. Glyde
e-mail: helen.glyde@nal.gov.au

2 Part 1

2.1 Motivation

Gallun et al. (2013) found that the effect of hearing loss on separated thresholds was stronger when one target level was used for all listeners (50 dB SPL) compared to when a sensation level (SL) of 40 dB was used (equivalent to a range of 47–72 dB SPL). They speculated that a broadband increase in gain was not sufficient to combat the non-flat hearing losses of their subjects. Thus in their most recent study (Jakien et al., under revision), they performed two experiments. In the first, they directly examined the effect of SL on SRM, while in a second experiment they carefully compensated for loss of audibility within frequency bands for each listener.

2.2 Methods

Target and masker stimuli were three male talkers taken from the Coordinate Response Measure corpus. Head-related transfer functions (HRTFs) were used to position the target sentences at 0° azimuth and the maskers either colocated with the target or at ±45° (Gallun et al. 2013; Xie 2013). In the first experiment the target sentences were fixed at either 19.5 dB SL (low SL condition) or 39.5 dB SL (high SL condition) above each participant's speech reception threshold (SRT) in quiet. To estimate masked thresholds (target-to-masker ratio, TMR, giving 50 % correct), the levels of the two masking sentences were adjusted relative to the level of the target sentences using a one-up/one-down adaptive tracking algorithm. In the second experiment the spectrum of the target sentences was adjusted on a frequency band-by-band basis to account for differences in the audiogram across participants. The initial level was set to that of the high SL condition of the first experiment for a listener with 0 dB HL. Target and masking sentences were then filtered into six component waveforms using two-octave-wide bandpass filters with center frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz. The level of each component was adjusted based on the difference between the audiogram of the listener being tested and the audiogram of a comparison listener with 0 dB HL thresholds at each of the six octave frequencies, and then the six waveforms were summed. To estimate thresholds, the levels of the two masking sentences were adjusted relative to the level of the target sentence according to a progressive tracking algorithm which has been shown to be comparable to and more efficient than adaptive tracking (Gallun et al. 2013).

Thirty-six listeners participated in both experiments, and an additional 35 participated in just the second experiment. All 71 participants had four frequency (500, 1000, 2000, 4000 Hz) average hearing losses (4FAHL) below 37 dB HL (mean 12.1 dB ± 8.2 dB) and all had fairly symmetrical hearing at 2000 Hz and below. Ages of the listeners were between 18 and 77 years (mean of 43.1 years) and there was a

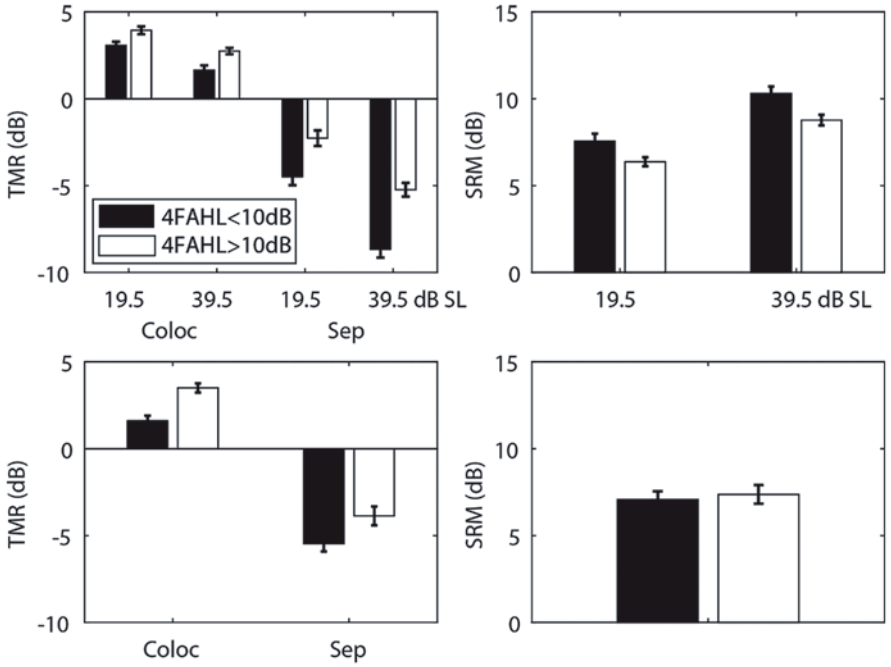

Fig. 1 Group-mean TMRs at threshold in the colocated and separated configurations (*top left panel*) and SRM (*top right panel*) as a function of SL. Group mean TMRs at threshold for the colocated and separated configurations (*bottom left panel*) and SRM (*bottom right panel*) in the equal-audibility condition. Error bars show standard errors

significant correlation ($r = 0.59$, $p < 0.001$) between 4FAHL and age. For simplicity, the listeners have been divided into those with 4FAHLs below 10 dB HL ($n = 22$ in the first experiment; $n = 35$ in the second experiment) and those with 4FAHLs above 10 dB HL ($n = 14$ in the first experiment; $n = 36$ in the second experiment).

2.3 Results

In the first experiment (top row of Fig. 1), those with lower 4FAHLs had better thresholds, and thresholds and SRM in both 4FAHL groups improved with an increase in SL. In the second experiment (bottom row of Fig. 1), despite equating audibility across listeners, there was a group difference in both the colocated and separated thresholds, but SRM was equivalent between groups.

For the 36 listeners who participated in both experiments, correlations between SRM and 4FAHL were examined. In the first experiment, 4FAHL was negatively correlated with SRM in both the low SL ($r = -0.33$, $p = 0.05$) and high SL ($r = -0.39$, $p = 0.02$) conditions. In the second experiment, 4FAHL was not significantly correlated with SRM ($r = -0.10$, $p = 0.57$).

In summary, increasing SL improved performance and increased SRM for all listeners. Furthermore, careful equalization of audibility across listeners reduced the effects of hearing loss on SRM. On the other hand, no manipulation was able to guarantee equal performance across listeners with various degrees of hearing loss. This suggests that while audibility is clearly an important factor, other factors may impact speech-in-speech intelligibility (e.g. aging, auditory filter width, or comorbidities associated with cognition, working memory and attention).

3 Part 2

3.1 Motivation

Glyde et al. (2013) showed that even with frequency-specific gain applied according to the individual audiogram using the NAL-RP hearing aid prescription, a strong relationship between SRM and hearing status persisted. The authors noted that with the relatively low presentation levels used in their experiment (55 dB SPL masker), the NAL-RP prescription may not have provided sufficient gain especially in the high frequency region. 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. They tested HI subjects as well as NH subjects with a simulated hearing loss.

3.2 Methods

The data is compiled from different studies but each group contained at least 12 NH (mean age 28.8–33.6 years) and 16 older HI (mean age 68.8–73.1 years). The HI listeners had a moderate, bilaterally symmetric, sloping sensorineural hearing loss with a 4FAHL of 48 ± 5dB.

Subjects were assessed with a Matlab version of the LiSN-S test (Glyde et al. 2013), in which short, meaningful sentences (e.g., “The brother carried her bag”) were presented in an ongoing two-talker background. Target and distractors were spoken by the same female talker and target sentences were preceded by a brief tone burst. Using HRTFs, target sentences were presented from 0° azimuth and the distractors from either 0° azimuth (colocated condition) or ± 90° azimuth (spatially separated condition). The combined distractor level was fixed at 55 dB SPL and the target level was adapted to determine the TMR at which 50 % of the target words were correctly understood. Subjects were seated in an audiometric booth and repeated the target sentences to a conductor.

Stimuli were presented over equalized headphones and for HI listeners had different levels of (linear) amplification applied to systematically vary audibility: am-

plification according to NAL-RP, NAL-RP plus 25 % of extra gain (i.e. on top of NAL-RP), and NAL-RP plus 50 % of extra gain. An extra gain of 100 % would have restored normal audibility, but was impossible to achieve due to loudness discomfort. Given the sloping hearing loss, an increase in amplification mainly resulted in an increased high-frequency gain and thus in an increase in audible bandwidth. NH subjects were tested at the same audibility levels. This was realized by first applying attenuation filters that mimicked the average audiogram of the HI subjects and then applying the same gains as described above. No other aspects of hearing loss were considered. Details of the processing can be found in Glyde et al. (2015). The NH subjects were also tested with no filtering.

3.3 Results

Thresholds for the colocated conditions (Fig. 2 left panel) were basically independent of amplification level, and for the NH subjects were about 2.5 dB lower than for the HI subjects. Thresholds for the spatially separated condition (middle panel) clearly improved with increasing amplification for both the NH and HI subjects and were maximal for "normal" audibility. However, thresholds for the NH subjects were on average 4.8 dB lower than for the HI subjects. The corresponding SRM, i.e., the difference in threshold between the colocated and separated conditions (right panel), increased with increasing gain similarly to the spatially separated thresholds, but the overall difference between NH and HI subjects was reduced to about 2.5 dB. It appears that under these conditions, a large proportion of the SRM deficit in the HI (and simulated HI) group could be attributed to reduced audibility.

4 Part 3

4.1 Motivation

Accounting for audibility effects in speech mixtures is not straightforward. While it is common to measure performance in quiet for the target stimuli used, this does not incorporate the fact that portions of the target are completely masked, which greatly reduces redundancy in the speech signal. Thus a new measure of "masked target audibility" was introduced.

Simple energy-based analyses have been used to quantify the available target in monaural 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. The number of these glimpses is reduced as the SNR decreases, or as more masker talkers are added to the mixture. To define the available glimpses in symmetric binaural mix-

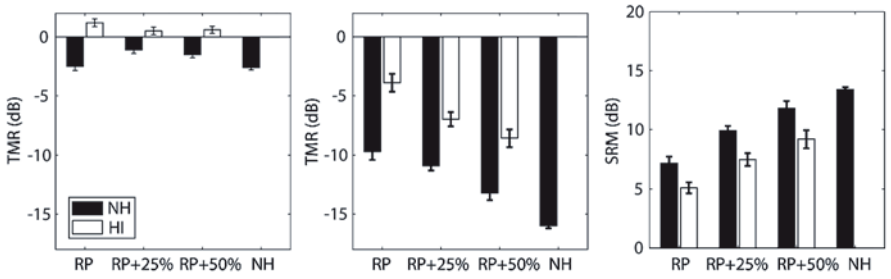

Fig. 2 Group-mean TMRs at threshold for the colocated (*left panel*) and separated (*middle panel*) configurations and the corresponding SRM (*right panel*). Error bars show standard errors

tures, we simply applied a monaural glimpsing model separately to the two ears. For symmetric listening situations, the glimpses can occur in either ear, and often occur in both ears for a particular time-frequency tile. However the glimpses may not all be above threshold, particularly for listeners with hearing loss. Thus we conducted an experiment in which we presented "glimpsed" stimuli to NH and HI listeners to measure their ability to use the available target information. Performance was compared to natural binaural performance to determine to what extent target audibility/availability can explain performance.

4.2 Methods

Six NH (mean age 23 years) and six HI (mean age 26 years) listeners participated. The HI listeners had a moderate, bilaterally symmetric, sloping sensorineural hearing loss with a 4FAHL of 49 ± 14dB.

Speech materials were taken from a corpus of monosyllabic words (Kidd et al. 2008), in which five-word sentences are assembled by selecting one word from each of five categories (e.g., "Sue bought two red toys"). Using HRTFs, one target sentence was presented at 0° azimuth, and two or four different masker sentences were presented at ±90° azimuth, or ±45°/±90° azimuth. All talkers were female, and the target was identified by its first word "Sue". Each masker was fixed in level at 60 dB SPL, and the target was varied in level to set the TMR at one of five values (from −25 to −5 dB in the NH group; from −20 to 0 dB in the HI group). Stimuli for HI listeners had individualized NAL-RP gain applied.

To generate the glimpsed stimuli, an ideal binary mask was applied separately to the two ears of the binaural stimuli using the methods of Wang (2005) and Brungart et al. (2006). In short, the signals were analyzed using 128 frequency channels between 80 and 8000 Hz, and 20-ms time windows with 50% overlap. Tiles with target energy exceeding masker energy were assigned a mask value of one and the remaining tiles were assigned a value of zero. The binary mask was then applied to the appropriate ear of the binaural stimulus before resynthesis. As a control condition, the mask was also applied to the target alone.

Stimuli were presented over headphones to the listener who was seated in an audiometric booth fitted with a monitor, keyboard and mouse. Responses were given by selecting five words from a grid presented on the monitor.

4.3 Results

Figure 3 shows 50% thresholds extracted from logistic fits to the data. Performance was better overall with two maskers as compared to four maskers, and for NH than HI listeners. With two maskers, the difference in thresholds between groups was 11 dB in the natural condition, and 9 dB in the glimpsed mixture condition. For four maskers these deficits were 7 and 9 dB, respectively. In other words, the group differences present in the natural binaural condition were similar when listeners were presented with the good time-frequency glimpses only. In the control condition where the glimpses contained only target energy, the group differences were even larger (12.6 and 10.8 dB). This suggests that the HI deficit is related to the ability to access or use the available target information and not to difficulties with spatial processing or segregation. Individual performance for natural stimuli was strongly correlated with performance for glimpsed stimuli ($r = 0.92$), again suggesting a common limit on performance in the two conditions.

5 Conclusions

These studies demonstrate how audibility can affect measures of SRM using the symmetric masker paradigm, and suggest that reduced access to target speech information might in some cases contribute to the "spatial" deficit observed in listeners with hearing loss. This highlights the importance of adequate stimulus audibility for optimal performance in spatialized speech mixtures, although this is not always feasible due to loudness discomfort in HI listeners, and technical limitations of hearing aids.

Acknowledgments This work was supported by NIH-NIDCD awards DC04545, DC013286, DC04663, DC00100 and AFOSR award FA9550-12-1-0171 (VB, CRM, JS, GK), DC011828 and the VA RR&D NCRAR (KMJ, SDK, FJG), the Australian Government through the Department of Health, and the HEARing CRC, established and supported under the Cooperative Research Centres Program—an initiative of the Australian Government (VB, JB, HG).

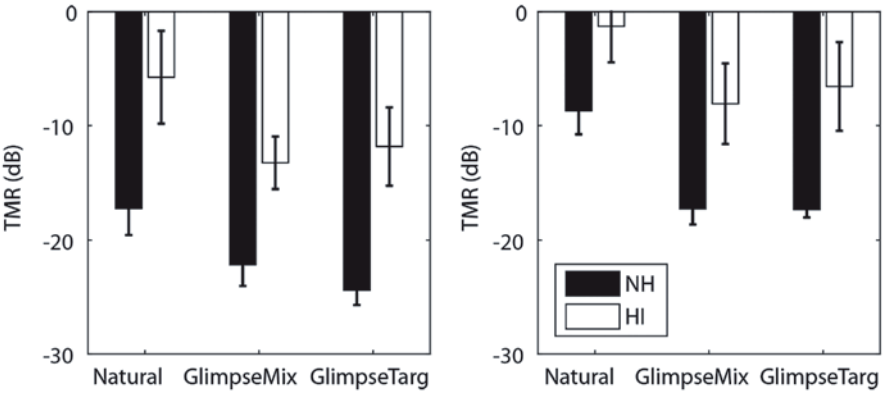

Fig. 3 Group-mean TMRs at threshold for the two-masker (*left panel*) and four-masker (*right panel*) conditions. Error bars show standard errors

References

Brungart DS, Chang PS, Simpson BD, Wang DL (2006) Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. J Acoust Soc Am 120:4007–4018

Colburn HS (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

Gallun FJ, Kampel SD, Diedesch AC, Jakien KM (2013) Independent impacts of age and hearing loss on spatial release in a complex auditory environment. Front Neurosci 252(7):1–11

Glyde H, Cameron S, Dillon H, Hickson L, Seeto M (2013) The effects of hearing impairment and aging on spatial processing. Ear Hear 34(1):15–28

Glyde H, Buchholz JM, Nielsen L, Best V, Dillon H, Cameron S, Hickson L (2015) Effect of audibility on spatial release from speech-on-speech masking. J Acoust Soc Am 138:3311–3319

Hawley ML, Litovsky RY, Colburn HS (1999) Speech intelligibility and localization in a multi-source environment. J Acoust Soc Am 105:3436–3448

Kidd Jr G, Best V, Mason CR (2008) Listening to every other word: examining the strength of linkage variables in forming streams of speech. J Acoust Soc Am 124:3793–3802

Marrone N, Mason CR, Kidd Jr G (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

Noble W, Byrne D, 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, 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

Spencer N (2013). Binaural benefit in speech intelligibility with spatially separated speech maskers. PhD Dissertation, Boston University

Strelcyk O, Dau T (2009) Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. J Acoust Soc Am 125:3328–3345

Wang DL (2005) On ideal binary mask as the computational goal of auditory scene analysis. In: Divenyi P (ed) Speech separation by humans and machines. Kluwer Academic, Norwell, pp 181–197

Xie B (2013). Head-related transfer function and virtual auditory display, 2nd edn. J. Ross Publishing, Plantation