Localization in speech mixtures by listeners with hearing loss

Virginia Best[a] and Simon Carlile
School of Medical Sciences and The Bosch Institute, University of Sydney, Sydney, New South Wales 2006, Australia
virginia.best@sydney.edu.au, simon.carlile@sydney.edu.au

Norbert Kopčo
Department of Cybernetics and Artificial Intelligence, Technical University of Košice, Košice, 04001, Slovakia
kopco@tuke.sk

André van Schaik
Bioelectronics and Neuroscience, University of Western Sydney, Penrith, New South Wales 2751, Australia
a.vanschaik@uws.edu.au

Abstract: The ability of listeners with bilateral sensorineural hearing loss to localize a speech source in a multitalker mixture was measured. Five simultaneous words spoken by different talkers were presented over loudspeakers in a small room, and listeners localized one target word. Errors were significantly larger in this group compared to a control group with normal hearing. Localization of the target presented alone was not different between groups. The results suggest that hearing loss does not impair spatial hearing *per se*, but degrades the spatial representation of multiple simultaneous sounds.

PACS numbers: 43.66.Qp, 43.66.Dc, 43.66.Mk, 43.71.Es [QJF]
Date Received: February 2, 2011 **Date Accepted:** March 7, 2011

1. Introduction

Spatial auditory perception is an important function in human listeners, providing awareness of one's surroundings, enabling localization of useful or dangerous objects in the environment, and enhancing speech communication by enabling attention to be directed selectively to the location of a talker of interest in the presence of other spatially separated talkers or noises (e.g., Broadbent, 1954). While a wealth of studies have described human sound localization abilities in simple environments (e.g., Middlebrooks and Green, 1991; Blauert, 1997), only a handful have examined the sensitivity of listeners to the spatial location of different sources in more realistic mixtures of sounds such as speech (Hawley *et al.*, 1999; Simpson *et al.*, 2006; Kopčo *et al.*, 2010).

There are many suggestions in the literature that spatial hearing is compromised in listeners with sensorineural hearing loss. This comes largely from the much-cited observation that these listeners receive less benefit from the spatial separation of competing sounds when the task is to identify a target sound (e.g., speech identification in noise or in a multitalker mixture, Duquesnoy, 1983; Bronkhorst and Plomp, 1989; Marrone *et al.*, 2008). However, studies that have directly investigated spatial sensitivity in hearing-impaired listeners have produced mixed results. Several studies have reported poorer interaural time difference thresholds in listeners with hearing loss (Hawkins and Wightman, 1980; Smith-Olinde *et al.*, 1998); however, thresholds vary dramatically across individuals and can be as low as in listeners with normal hearing (Durlach *et al.*, 1981; Colburn, 1982; Gabriel *et al.*, 1992). Free-field localization of

[a]Author to whom correspondence should be addressed. Current address: Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA 02215.

isolated noise stimuli in the horizontal plane is impaired only for listeners with unilateral losses or substantial low-frequency losses (Häusler *et al.*, 1983; Rosenhall, 1985; Noble *et al.*, 1994). In the presence of a fixed-location noise masker, Lorenzi and colleagues (1999a,b) found that listeners with hearing loss localized a click-train target more poorly than their normally hearing counterparts. However, they also measured detection of the clicks and found that performance declined in parallel for the localization and detection tasks, suggesting that audibility of the target was the critical issue. To our knowledge, no previous study has examined absolute localization of speech in the presence of competing talkers in hearing-impaired listeners, a situation more akin to that generally used to assess speech intelligibility. This experiment represents an attempt to bridge this gap.

2. Methods

2.1. Participants

Seven subjects with hearing impairment (HI group) participated, one female and six males between the ages of 22 and 69 yr. They were recruited from the local community with the only selection criterion being a bilateral sensorineural loss (or mixed sensorineural/conductive loss in one case). The audiometric profiles of the group were relatively heterogeneous, but tended to be roughly symmetric (no more than 20 dB difference between the ears at any frequency) and sloping with poorer thresholds at higher frequencies (see Fig. 1). Although four of the listeners wore hearing aids regularly, they were tested unaided. Seven subjects with self-reported normal hearing (NH group) group acted as controls, one female and six males between the ages of 18 and 50 yr. The data presented here for these subjects formed part of a larger experiment, the entirety of which has been presented elsewhere (Kopčo *et al.*, 2010).

2.2. Environment and stimuli

The experiment took place in a darkened, carpeted office of dimensions 3 m $\times$ 5 m $\times$ 2.5 m. Eleven loudspeakers (V6; Tannoy Ltd., Coatbridge, Scotland) were positioned on stands (height 1.6 m) in a horizontal arc of radius 1.5 m. They were separated by 10° and spanned locations from $-50°$ to $+50°$ azimuth relative to the listener, who stood in the center of the array. The target was presented from one of the 11

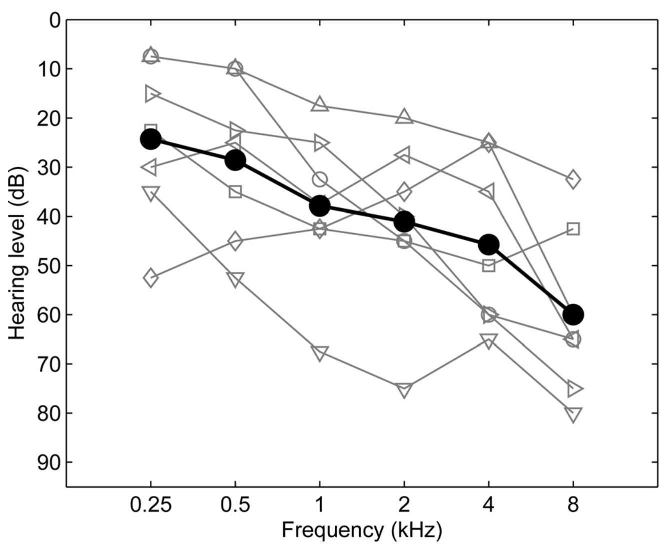

Fig. 1. Audiograms for the seven hearing-impaired listeners. Thin gray lines and open symbols show individual audiograms (mean across left and right ears); thick black lines and solid circles show the across-subject mean.

loudspeakers at random. On mixture trials, four simultaneous maskers were also presented, and they were arranged randomly in one of five configurations such that each masker was presented from a different loudspeaker (specific details of masker locations can be found in Kopčo *et al.*, 2010). Listeners indicated their responses by pointing their head in the perceived direction of the target and pressing a hand-held response button. A headtracker (IC3; InterSense Inc., Billerica, MA) mounted on a plastic headband was used to measure the orientation of the head at the time of response. No feedback regarding accuracy of responses was given.

Speech materials were taken from a corpus of monosyllabic words (Kidd *et al.*, 2008). The target was the word "two" spoken by one of the female voices in the corpus. Maskers were four different non-digit words spoken by four of the eight male talkers in the corpus. The masker words were all longer in duration than the target word. The fixed target word and the gender difference between the target and the maskers minimized the potential ambiguity about which sound to localize. On "catch trials" (see below) the target was replaced by another randomly chosen masker word. Stimuli were created using MATLAB software (MathWorks, Natick, MA) at a sampling rate of 48 kHz and sent via a multichannel soundcard (RME Fireface 400; Audio AG, Haimhausen, Germany), D/A converter (DA-16x; Apogee Electronics Corp., Santa Monica, CA), and amplifier (Powerflex 6250; Ashly Audio Inc., Webster, NY) to the loudspeakers.

2.3. Presentation levels

For the NH listeners, each word (target and maskers) was presented at a level of 57 dB sound pressure level [SPL(A)]. This represents the 0 dB target-to-masker ratio from the previous study (Kopčo *et al.*, 2010). To determine roughly what sensation level (SL) this represented for the NH group, one listener completed a short threshold test. In this test, the target word was progressively attenuated (in 1 dB steps) from 57 dB SPL until the listener could no longer detect the stimulus. This point was reached at 17 dB SPL, indicating that the presented stimulus was approximately 40 dB SL.

For each listener in the HI group, detection thresholds were measured in the same way, and a level of 40 dB SL was chosen as the presentation level. However, in two subjects, this level was judged to be frustratingly soft, and thus levels of 42 and 55 dB SL were chosen following feedback from the subject. In one subject, 40 dB SL could not be reached within the limits of the system and instead 33 dB SL was used. Final presentation levels ranged from 57 dB to 75 dB SPL.

2.4. Procedures

Before a session, the experimenter positioned the subject such that he/she was in the center of the loudspeaker array with his/her head pointing to 0° azimuth, and this location was recorded by the headtracker as the reference position. Before the stimulus was played on each trial, the subject was required to orient his/her head to this position and feedback was given by way of a small LED display.

In mixture runs, it was expected that there would be a number of trials in which the listener would not be able to detect the target. To avoid these trials influencing the localization data, listeners were instructed to give a specific response if they did not detect the female target (miss trials). This response was to point to a location directly above the head—a response that was easily distinguished from regular localization responses that all had an elevation component on or near the audiovisual horizon. To ensure that listeners were following this instruction, a number of catch trials were included in which the target was replaced by another random male masker and false alarms to these trials were monitored.

Control runs consisted of 55 trials (five trials per target location). Mixture runs consisted of 60 trials as they also included 5 catch trials. All NH listeners completed 10 mixture runs in total. The HI listeners completed as many mixture runs as they had time for (between 6 and 10). This resulted in between 220 and 550 mixture trials in total. A control run was completed at the beginning and end of each visit, and listeners took one or two visits to complete the testing (resulting in 110 or 220 control trials in total).

3. Results

Miss rates in the mixture condition were higher in the HI group than in the NH group (5.8% vs 1.2%), but this was not a significant difference [$t(12) = 1.84$, $p = 0.09$]. False alarm rates in the mixture condition were 18.5% on average in the HI group and 7.1% on average in the NH group. While not significantly different [$t(12) = 1.79$, $p = 0.10$], the sizeable group difference indicates that the HI group were more likely to localize words that were not the target.

Root mean square (rms) errors with respect to the actual target location were calculated across all trials, separately for the control and mixture conditions. Mean errors for the two conditions and the two listener groups are shown in Fig. 2. In the control condition, mean errors were similar for the HI and NH groups (8.0° and 6.9°, respectively). In the mixture condition, however, the HI group produced much larger errors than the NH group (19.2° vs 12.3°). An analysis of variance (ANOVA) with a within-subjects factor of condition and a between-subjects factor of listener group revealed significant main effects of condition [$F(1,12) = 93.5$, $p < 0.001$], listener group [$F(1,12) = 5.5$, $p < 0.05$], and a significant interaction [$F(1,12) = 11.7$, $p < 0.01$]. Post hoc *t*-tests with a Bonferroni correction confirmed that rms errors for the two groups differed significantly in the mixture condition [$t(12) = 2.89$, $p < 0.05$] but not in the control condition [$t(12) = 0.88$, $p = 0.4$].

4. Discussion

Motivated by the observation that target detectability can cause apparent deficits in sound localization in listeners with hearing loss (Lorenzi *et al.*, 1999a), listeners in the current study were instructed to localize the target only on those trials in which they were confident that they could hear it (see also Simpson *et al.*, 2006; Kopčo *et al.*, 2010). The number of miss trials and false alarms were relatively low in both groups. This suggests that the use of a predictable target word, distinguished from the maskers on the basis of talker gender, minimized the effect of the mixture on target detection in both groups. However, the HI group made more false alarms on catch trials than the

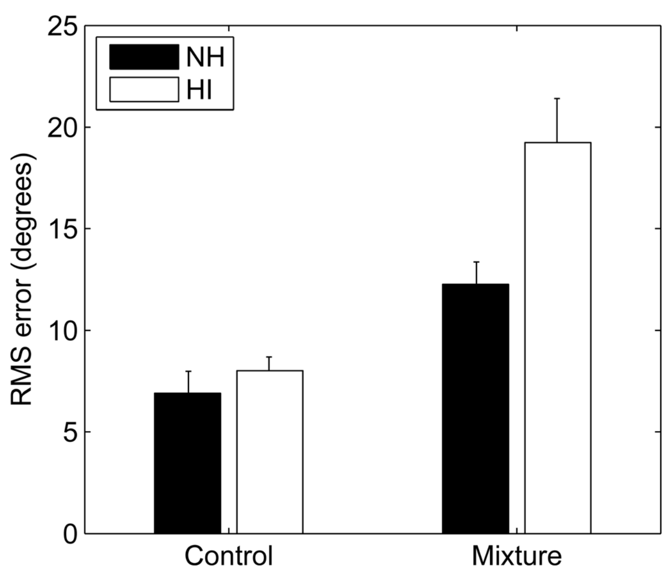

Fig. 2. Mean rms errors in the control and mixture conditions for the two listener groups. Error bars show standard errors of the means.

NH group on average, suggesting that they were less reliable at performing the detection component of the task. The propensity for false alarms varied substantially across listeners, however, and was not significantly higher in the HI group. Moreover, false alarm rates were not correlated with mean rms errors ($r = 0.31$, $p = 0.29$). Nevertheless, it is difficult to rule out the possibility that increased uncertainty about the presence of the target influenced the localization data in the HI listeners. This idea requires further investigation with a larger group of subjects; very few studies have measured speech *detection* in NH listeners (Grant and Seitz, 2000; Balakrishnan and Freyman, 2008) and we are not aware of any studies in listeners with hearing loss.

Assuming that impaired detection was not the primary cause of differences in errors across groups, hearing loss appears to degrade speech localization in a multitalker mixture while having no effect on speech localization in quiet. One explanation is that there is a redundancy of localization cues in quiet that conceals the disruption of one or more localization cues in HI listeners. For example, the relative attenuation of high frequencies due to sloping audiograms in the HI listeners might be more critical in a mixture than in quiet where there are robust low frequency cues available. However, Lorenzi *et al.* (1999a,b) had only limited success in simulating the degraded localization of click trains in noise by using appropriately filtered stimuli in NH listeners. A more likely explanation is that the reduced spectrotemporal sensitivity exhibited by listeners with hearing loss (Moore, 1985; Dillon, 2001) leads to more "mutual masking" between simultaneous speech sounds, which in turn leads to a distortion of the spatial cues available for localizing the target.

A lack of spatial acuity in competing speech settings may be a factor in the documented difficulties in which listeners with hearing loss have in multiple-person conversations (Gatehouse and Noble, 2004). Uncertainty about the spatial location of different talkers must certainly interfere with one's ability to orient quickly and appropriately in these dynamic settings. Moreover, poor spatial resolution likely contributes to the reduced benefit of spatially directed attention shown by listeners with hearing loss in multitalker mixtures (Best *et al.*, 2009). A reduced ability to spatially enhance a target of interest further compounds the loss of intelligibility that already results from hearing loss.

Acknowledgments

Work supported by a University of Sydney Postdoctoral Research Fellowship to V.B. and a grant to S.C. from the Australian Research Council (DP0881471).

References and links

Balakrishnan, U., and Freyman, R. L. (**2008**). "Speech detection in spatial and nonspatial speech maskers," J. Acoust. Soc. Am. **123**, 2680–2691.

Best, V., Marrone, N., Mason, C. R., Kidd, G., and Shinn-Cunningham, B. G. (**2009**). "Effects of sensorineural hearing loss on visually guided attention in a multitalker environment," J. Assoc. Res. Otolaryngol. **10**, 142–149.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge), pp. 1–6.

Broadbent, D. E. (**1954**). "The role of auditory localization in attention and memory span," J. Exp. Psychol. **47**, 191–196.

Bronkhorst, A., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1374–1383.

Colburn, H. S. (**1982**). "Binaural interaction and localization with various hearing impairments," Scand. Audiol. Suppl. **15**, 27–45.

Dillon, H. (**2001**). *Hearing Aids* (Thieme, New York).

Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Durlach, N. I., Thompson, C. L., and Colburn, H. S. (**1981**). "Binaural interaction of impaired listeners. A review of past research," Audiology **20**, 181–211.

Gabriel, K. J., Koehnke, J., and Colburn, H. S. (**1992**). "Frequency dependence of binaural performance in listeners with impaired binaural hearing," J. Acoust. Soc. Am. **91**, 336–347.

Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial and qualities of hearing scale (SSQ)," Int. J. Audiol. **43**, 85–99.
Grant, K. W., and Seitz, P. F. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am. **108**, 1197–1208.
Häusler, R., Colburn, H. S., and Marr, E. (**1983**). "Sound localization in subjects with impaired hearing. Spatial-discrimination and interaural-discrimination tests," Acta Oto-Laryngol., Suppl. **400**, 1–62.
Hawkins, D. B., and Wightman, F. L. (**1980**). "Interaural time discrimination ability of listeners with sensorineural hearing loss," Audiology **19**, 495–507.
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.
Kidd, G. Jr., Best, V., and Mason, C. R. (**2008**). "Listening to every other word: examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.
Kopčo, N., Best, V., and Carlile, S. (**2010**). "Speech localization in a multitalker mixture," J. Acoust. Soc. Am. **127**, 1450–1457.
Lorenzi, C., Gatehouse, S., and Lever, C. (**1999a**). "Sound localization in noise in hearing-impaired listeners," J. Acoust. Soc. Am. **105**, 3454–3463.
Lorenzi, C., Gatehouse, S., and Lever, C. (**1999b**). "Sound localization in noise in normal-hearing listeners," J. Acoust. Soc. Am. **105**, 1810–1820.
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.
Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.
Moore, B. C. J. (**1985**). "Frequency selectivity and temporal resolution in normal and hearing-impaired listeners," Br. J. Audiol. **19**, 189–201.
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.
Rosenhall, U. (**1985**). "The influence of hearing loss on directional hearing," Scand. Audiol. **14**, 187–189.
Simpson, B. D., Brungart, D. S., Iyer, N., Gilkey, R. H., and Hamil, J. T. (**2006**). "Detection and localization of speech in the presence of competing speech signals," in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, pp. 129–133.
Smith-Olinde, L., Koehnke, J., and Besing, J. (**1998**). "Effects of sensorineural hearing loss on interaural discrimination and virtual localization," J. Acoust. Soc. Am. **103**, 2084–2099.