Do hearing-impaired listeners benefit from spatial and temporal cues in a complex auditory scene?

Virginia Best, Nicole L Marrone, Christine R Mason, Gerald Kidd Jr, and Barbara G Shinn-Cunningham
Hearing Research Center, Boston University, Boston, MA, 02215, USA.

ABSTRACT

In auditory scenes containing many similar sound sources, difficulties with the detection and organization of acoustic information can lead to disruptions in the identification of behaviorally relevant targets. A previous study conducted in young normal-hearing listeners (Best *et al.*, 2007) investigated the benefit of providing simple visual cues for when and/or where a target string of spoken digits would occur in a complex acoustic mixture. Importantly, the visual cues provided no information about the target *content*. A visual cue indicating which loudspeaker (from an array of five) would contain the target improved accuracy, and a cue indicating which time segment (out of a possible five) would contain the target resulted in a smaller improvement. The present study extended this work to young listeners with sensorineural hearing loss. These listeners performed more poorly overall than the normal-hearing group, but did benefit from visual cues indicating where and when to listen for the target. While the magnitude of the temporal cue benefit was comparable between groups, the spatial cue benefit was smaller on average for the hearing impaired-listeners. This result suggests that one component of the difficulties experienced by listeners with hearing loss in complex tasks of this nature is related to directing spatial attention.

INTRODUCTION

In many everyday listening situations, a listener's goal is to hear out one sound of interest from amongst a mixture of other interfering sounds. Normal-hearing (NH) listeners are remarkably adept at this, and make use of many physical properties of the stimulus to accomplish this task. For example, when interfering sounds fluctuate over time, listeners are able to make use of brief "glimpses" of the target (Cooke *et al.*, 2006) and/or of comodulation across frequency in the interferers (Grose and Hall, 1992). When competing sound sources are separated spatially, interaural differences lead to effective increases in the signal-to-noise ratio that aid in target detection (Zurek, 1993) and differences in perceived location that aid in segregating the target from interferers (Freyman *et al.*, 2001; Arbogast *et al.*, 2002).

In highly complex or uncertain settings, top-down selective attention is important for successfully processing a source of interest. For example, providing a-priori information about where to listen in a multiple-talker array enhances target intelligibility, particularly when there are more than two talkers (Ericson *et al.*, 2004; Kidd *et al.*, 2005). Recently, Best *et al.* (2007) examined the benefits of attentional cueing when listeners were confronted with a mixture of five simultaneous streams of speech coming from five spatially separated loudspeakers. The speech streams were unintelligible except for a short intelligible target which occurred from a randomly chosen location at an unpredictable point in time. Listeners received a robust benefit from simple visual cues indicating where (and to a lesser extent, when) to listen in the mixture, even though the cues gave no explicit information about the identity of the target.

Complex, dynamic listening situations of this kind are extremely difficult for hearing-impaired (HI) listeners (Gatehouse and Noble, 2004). In the current study, attention-related aspects of these difficulties were explored using a complex scene with five simultaneous sources (after Best *et al.*, 2007). Based on a substantial body of previous research, it was expected that HI listeners would perform worse overall on the task than NH listeners. For example, HI listeners receive little benefit from amplitude fluctuations present in interferers such as speech (Duquesnoy, 1983; Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Lorenzi *et al.*, 2005) and a greatly reduced benefit from spatial separation of simultaneous sources (Bronkhorst and Plomp, 1989; Marrone *et al.*, these proceedings). However, the effect of hearing impairment on the direction of attention within this kind of scene remains unclear. Listeners with hearing loss normally rely heavily on non-auditory cues (such as those provided by lip-reading) to function in difficult listening situations. For this reason these listeners might benefit *more* than NH listeners from visual cues about timing and location in the listening environment simulated here. On the other hand, reduced spectro-temporal resolution in HI listeners may limit the perceptual segregation of competing sources, which could make it difficult for them to direct spatial attention selectively to the target source (Shinn-Cunningham, these proceedings). If so, HI listeners might benefit *less* than NH listeners from visual cues that guide attention. The overall goal of the current study was thus to determine whether hearing impairment has an impact on improvements in speech intelligibility that are specifically related to attention.

METHODS

Listeners

Seven HI listeners (2 male, 5 female, aged 19 – 42) and eight normal-hearing listeners (3 male, 5 female, aged 19 – 30) participated in the experiment. Listeners were paid for their participation, and the experiment was approved by the Boston University Charles River Campus Institutional Review Board.

The HI listeners had mild to moderately severe, bilateral, symmetric, sloping, sensorineural hearing loss. Six of the seven were regular hearing-aid wearers but participated in the experiment with their aids removed. The NH listeners were screened to ensure that they had pure-tone thresholds in the normal range (no greater than 10 dB HL) for frequencies between 250 Hz and 8 kHz. Mean audiograms for both groups are shown in Fig. 1.

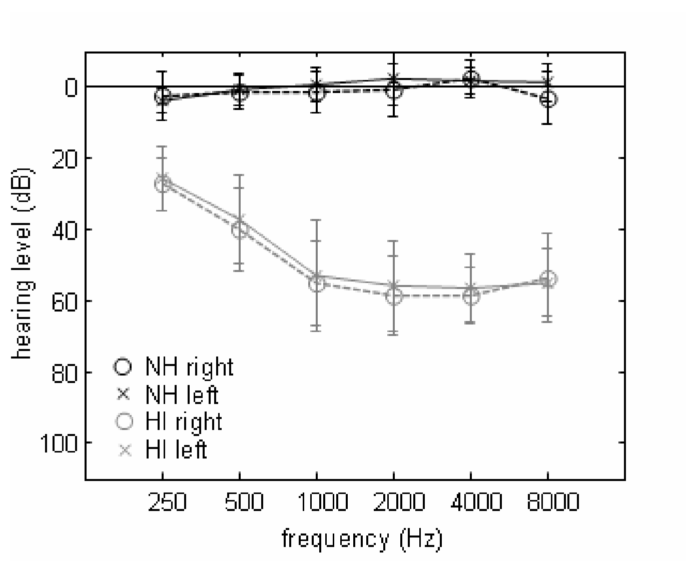

Fig 1: Mean pure-tone thresholds across listeners in the two listener groups (error bars show standard deviations).

Environment

The experiments took place in a single-walled IAC booth with interior dimensions of 12’4” x 13’ x 7’6” (length, width, height), with perforated metal panels on the walls and ceiling and a carpeted floor. The listener was seated on a chair in the center of the room, with a head rest to minimize head movements. No instructions were given to listeners regarding eye fixation during stimulus delivery. Stimuli were presented via five loudspeakers (Acoustic Research 215PS) located on an arc approximately 5 ft from the listener at the level of the ears. The loudspeakers were positioned within the visual field at lateral angles of -30°, -15°, 0°, 15°, and 30°. Listeners indicated their response using a handheld keypad. The booth was kept dark during the experiment, except for a small lamp to illuminate the keypad.

Digital stimuli were generated on a PC located outside the booth and fed through five separate channels of Tucker-Davis Technologies hardware. Signals were converted at 40 kHz by a 16-bit D/A converter (DA8), attenuated (PA4), and passed through power amplifiers (Tascam) before presentation to the loudspeakers. Each loudspeaker had an LED affixed on its top surface, which was controlled from the PC via a custom-built switchboard.

Stimuli

Targets were sequences of spoken digits from the TIDIGIT database. Each sequence comprised five digits from the set 1-9, spoken by one of 20 male talkers. To create a masking stimulus that was spectro-temporally similar but unintelligible, the target sequences were concatenated and reversed in time; individual maskers were then generated by selecting an arbitrary portion of this stimulus and applying a 10-ms cosine-squared ramp to each end.

Each of the five loudspeakers presented an ongoing signal that was divided into five contiguous time segments, giving a 5 x 5 space/time matrix. On any trial, the target occurred in one of these 25 space/time positions; the other 24 contained maskers. In any time segment all maskers were different. All time segments were fixed to the length of the target and were 1600 ms long on average, for a total stimulus duration of approximately 8 s. Signal level was equated across the loudspeakers such that the target was equal in level to each of the maskers.

Procedures

Before testing, each listener’s quiet identification threshold for single digits was measured using an adaptive procedure. In the main experiment, all stimuli were presented at a level 30 dB above this threshold. To verify that listeners could identify the 5-digit target stimuli at the chosen level, a short identification test was conducted in the absence of any maskers. All listeners performed at or near 100% in this test.

Four experimental conditions were tested in the main experiment. While the auditory stimulus was identical across the four conditions, a visual component was varied in order to manipulate the attention of the listener (see Fig. 2):

1. NO CUE: No visual cue was given.
2. WHERE: The LED located on the target loudspeaker lit up synchronously with the onset of the first time segment and remained on for the entire stimulus.
3. WHEN: All of the five LEDs lit up at the start of and were turned off at the end of the time segment containing the target.
4. WHERE+WHEN: The LED located on the target loudspeaker lit up only for the duration of the time segment containing the target.

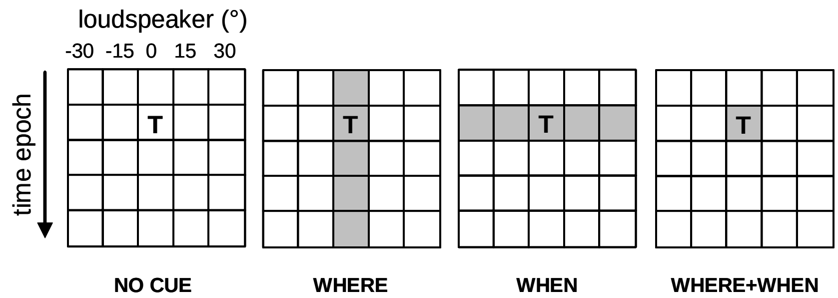

Fig 2: Schematic of the four attention conditions. The target (T) occurred randomly in time and space, and the visual cues (grey regions) could indicate where, when or where and when to listen.

The different conditions were run in tests of 25 trials. Listeners were informed at the beginning of each test as to the kind of visual cue they would receive during that test. A session consisted of one test in each of the four attention conditions. Each listener completed five sessions (approximately an hour each) over the course of 2-3 visits. The order of the four conditions was random and different between sessions and listeners.

RESULTS

Mean percent correct scores are shown in the top panel of Fig. 3. The four bars within a group represent the four attention conditions, as labelled. The first group of bars shows scores for the seven HI listeners, while the second group of bars shows scores for five of the NH listeners. Overall, the pattern of scores is comparable to that described in Best *et al.* (2007), with performance being poorest overall in the NO CUE condition, best in the WHERE+WHEN condition, and intermediate for the other two conditions. HI listeners performed more poorly overall than the NH listeners. In order to have a comparison of cueing effects for the two groups when baseline performance was in a similar range, a group of four NH listeners (including one from the first group) completed an identical experiment in which the target level was reduced by 3 dB (NH -3dB). Mean scores for this group are shown in the right-most group of bars. Performance in the NO CUE condition is similar to the HI group.

To examine directly the benefit of the different visual cues, scores in the NO CUE condition were subtracted from scores in the other conditions for each listener. Mean 'cue benefits' across listeners within each group are shown in the bottom panel of Fig. 3. In the NH group, the WHERE, WHEN, and WHERE+WHEN cues improved percent correct scores by 17%, 9%, and 20%, respectively. In the HI group, the analogous benefits were 10%, 8%, and 18%. Thus, the HI listeners appear to realize less benefit from the WHERE cue than the NH listeners, but receive a similar benefit from the WHEN cue. A repeated measures ANOVA found a significant main effect of condition [$F(2,20)=51.4$, $p<0.001$], no significant main effect of listener group [$F(1,10)=3.6$, $p=0.09$], and a significant interaction [$F(2,20)=6.1$, $p<0.01$]. Separate t-tests with Bonferroni corrections conducted on each cue type found only the WHERE cue benefit to differ significantly between the listener groups ($p<0.05$). Finally, while the WHERE and WHEN cues appear to be additive in the HI group, this is not the case in the NH group. Cue benefits for the NH -3dB group were 19%, 8%, and 26% for the three cued conditions. While the WHERE and WHEN cue benefits are similar to the first NH group, the WHERE+WHEN benefit is larger and closer to the addition of the two component cue benefits, suggesting that the lack of additivity for the first NH group was a ceiling effect.

A repeated measures ANOVA found significant main effects of condition [$F(2,18)=59.6$, $p<0.001$] and listener group [$F(1,9)=6.2$, $p<0.05$] and a significant interaction [$F(2,18)=7.9$, $p<0.005$]. Separate t-tests with Bonferroni corrections conducted on each cue type found both the WHERE and WHERE+WHEN cue benefits to differ significantly between the listener groups ($p<0.05$).

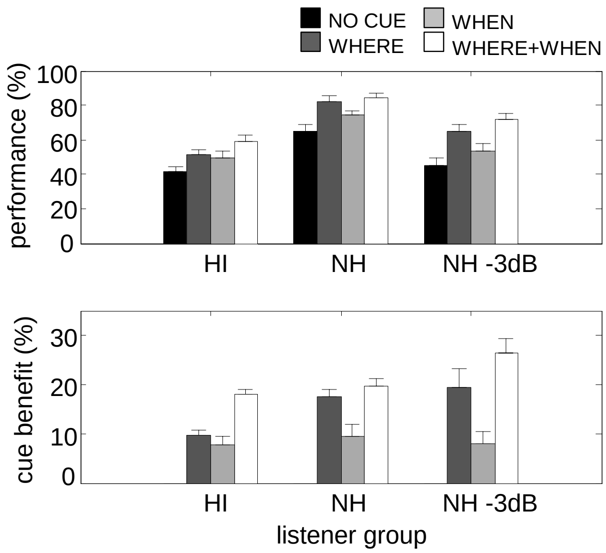

Fig 3: Mean percent correct scores (top panel) and mean cue benefits (bottom panel) for each group of listeners in the four attention conditions (error bars show standard errors).

DISCUSSION

As expected based on previous work, HI listeners were poorer overall at identifying a speech target embedded in a mixture of equal-level speech-like maskers. However, HI listeners did benefit from visual cues indicating where and when to listen for the target. While the magnitude of the temporal cue benefit was comparable between groups, the spatial cue benefit was smaller on average in the HI group, even when the groups were matched in terms of their baseline performance.

In the previous study (Best *et al.*, 2007), it was suggested that the spatial and temporal cues invoke different (and independent) modes of attention. This idea is supported by the current study, in that the WHERE and WHEN benefits were affected differentially by hearing loss, and were roughly additive. Temporal cues may have an “alerting” effect, which increases vigilance or arousal during the time epoch containing the target. This effect seems to have a relatively constant impact on performance, regardless of hearing status or overall performance. On the other hand, spatial cues are thought to play an important role in mediating competition between sources. The current study indicates that spatial knowledge is less helpful for listeners with hearing loss than for normal-hearing listeners. Understanding why this is the case is important for understanding why HI listeners have difficulty in multi-source settings, which, in turn, is important for developing strategies to help HI listeners in everyday settings.

Shinn-Cunningham (these proceedings) suggests that reductions in spectral and temporal acuity in HI listeners impair the formation of auditory “objects,” which reduces the effectiveness of selective attention in choosing amongst competing objects. Extending this idea, it may be that reductions in spectral and temporal acuity also lead to degraded (or “blurred”) spatial representations and hence reduce the success with which spatially-directed attention can enhance one source selectively. We plan to test this idea directly in future experiments measuring spatial localization acuity in mixtures for NH and HI listeners.

In conclusion, the current results suggest that HI listeners do benefit from visual cues indicating where and when to listen when hearing out a target from a mixture. However, the benefit they receive from *spatial* information provided by visual cues is significantly worse than in NH listeners. An implication of this finding is that performance deficits shown by HI listeners (relative to NH listeners) on complex tasks of this kind may be larger for tasks in which there is spatial knowledge available.

ACKNOWLEDGMENTS

This work was supported by grants from AFOSR (FN2005-004032), NIH/NIDCD (DC00100, DC004545 and DC004663), and ONR (N00014-07-1-0061).

REFERENCES

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). “The effect of spatial separation on informational and energetic masking of speech.” J. Acoust. Soc. Am., **112**, 2086–2098.

Best, V., Ozmeral, E. J., and Shinn-Cunningham, B. G. (**2007**). “Visually-guided attention enhances target identification in a complex auditory scene.” J. Assoc. Res. Otolaryng., **8**, 294-304.

Bronkhorst, A. W. and Plomp, R. (**1992**). “Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing.” J. Acoust. Soc. Am., **92**, 3132-3139.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). “Informational and energetic masking effects in the perception of multiple simultaneous talkers.” J. Acoust. Soc. Am., **110**, 2527–2538.

Cooke, M. (**2006**). “A glimpsing model of speech perception in noise.” J. Acoust. Soc. Am., **119**, 1562-1573.

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

Ericson, M. A., Brungart, D. S., and Simpson, B. D. (**2004**). “Factors that influence intelligibility in multitalker speech displays.” Int. J. Aviation Psych., **14**, 311-332.

Festen, J. M. and Plomp, R. (**1990**). “Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing.” J. Acoust. Soc. Am., **88**, 1725-1736.

Freyman, R. L., Balakrishnan, U., Helfer, K. S. (**2001**). ”Spatial release from informational masking in speech recognition.” J. Acoust. Soc. Am., **109**, 2112-2122.

Gatehouse S. and Noble, W. (**2004**). “The speech, spatial and qualities of hearing scale (SSQ).” Int. J. Audiol., **43**, 85-99.

Grose, J. H., Hall, J. W. III. (**1992**). “Comodulation masking release for speech stimuli.” J. Acoust. Soc. Am., **91**, 1042-1050.

Kidd, G. J., Arbogast, T. L, Mason, C. R., and Gallun, F. J. (**2005**). “The advantage of knowing where to listen.” J. Acoust. Soc. Am., **118**, 3804-3815.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2005**). “Speech perception problems of the hearing impaired reflect inability to use temporal fine structure." PNAS, **103**, 18866-18869.

Marrone, N. L., Mason, C. R., and Kidd, G., Jr. (**this proceedings**). "Listening in a multisource environment with and without hearing aids."

Shinn-Cunningham, B. G. (**this proceedings**). “Why hearing impairment may degrade selective attention.”

Zurek, P. M. (**1993**). “Binaural advantages and directional effects in speech intelligibility.” In: *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg. Allyn and Bacon, Boston.