Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers

Alan W. Boyd[1,2], **William M. Whitmer**[2], **John J. Soraghan**[1], and **Michael A. Akeroyd**[2]
[1]Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XQ, UK

[2]MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow G31 2ER, UK

Abstract

Hearing-aid wearers have reported sound source locations as being perceptually internalized (i.e., inside their head). The contribution of hearing-aid design to internalization has, however, received little attention. This experiment compared the sensitivity of hearing-impaired (HI) and normal-hearing (NH) listeners to externalization cues when listening with their own ears and simulated BTE hearing-aids in increasingly complex listening situations and reduced pinna cues. Participants rated the degree of externalization using a multiple-stimulus listening test for mixes of internalized and externalized speech stimuli presented over headphones. The results showed that HI listeners had a contracted perception of externalization correlated with high-frequency hearing loss.

Introduction

An externalized sound is perceived by a listener as originating outside the head, while internalized sounds are those that appear inside the head.[1] The perception of sounds as being internalized can occur when listening over headphones and hearing aids.[2, 3, 4] However a listener's perception of externalization and internalization when wearing hearing-aids has received little attention, despite evidence for an increase in the internalization of sounds when wearing one and particularly two hearing-aids.[5] When listening to a static acoustic scene the individual spectral cues provided by the head-related transfer function (HRTF) become important for externalization and resolving front-back confusions.[6] The relative level difference between sound sources and the direct to reverberant ratio (DRR) can also contribute to externalization by providing a sense of depth to the auditory scene perceived by the listener.[7] Wightman & Kistler developed a psychoacoustically validated headphone simulation of open-ear listening[8, 9] that incorporated these cues into a simulation of three-dimensional auditory space. Ohl employed this simulation in a headphone externalization study, showing that hearing-impaired (HI) listeners were sensitive to the spectral cues that externalize sounds, though their performance was more variable and less sensitive than normal-hearing (NH) listeners.[10] This study controlled the prominence of the HRTF cues by mixing stimuli convolved with the listener's individual head-related impulse responses (HRIR) and impulse responses taken from the same spatial location in the absence of the listener. While this method can introduce high-frequency variability due to headphone simulation of HRTFs, this variability can be reduced through repeated headphone equalization. [11] An open-ear headphone simulation was employed in the current experiment

---

alan.boyd@eee.strath.ac.uk, bill@ihr.gla.ac.uk, j.soraghan@eee.strath.ac.uk, maa@ihr.gla.ac.uk.

to test the effect of in-the-ear (ITE) versus behind-the-ear (BTE) microphone position and broadband versus lowpass filtered (tenth-order Butterworth) bandwidth on NH and HI listeners' externalization of sound sources. While previous studies have focused exclusively on the externalization of single sources[12, 13], we also examined the effect of multiple sources on the participant's perception of externalization in a fixed acoustic scene and the effect that wearing hearing-aids might have in these more complex situations. The importance of spectral cues on externalization was investigated on the basis that spectral changes are known to produce a change in the perceived distance of sound sources from a listener.[14]

Method

Binaural room impulse responses (BRIRs) were recorded with microphones at ITE and BTE positions, with the listener present (theoretical maximally externalized condition) and absent, the microphones being placed on a horizontal bar in the same position as the listener's ears (theoretical maximally internalized condition). Participants rated the degree of externalization using a multiple-stimulus listening test for five mixes of the head-absent and head-present BRIRs convolved with speech stimuli presented over headphones. The mix parameter was the fraction of head-present recording power compared to the sum of head-present and head-absent recording power. Eight stimulus conditions were created, comprising all combinations of microphone position (ITE, BTE), frequency response (broadband and lowpass filtered) and number of talkers (1 and 4).

Participants

Seven NH (one female) and 14 HI (six female) listeners participated. NH listeners had a better-ear four-frequency (0.5, 1, 2, 4 kHz) average (4FA) of less than 20 dB HL, and HI listeners had an average 4FA of 34 dB HL and range of 21 to 51 dB HL. Average asymmetry for HI listeners was 6 dB. Seven HI listeners wore one hearing aid and two wore two at time of testing; all wore behind-the-ear (BTE) type aids. The NH listeners had an average age of 32 (22–46) and the HI listeners averaged 60 (48–72) years.

Test room

Listeners were seated along the central short axis of a room measuring $6.5 \times 5 \times 3$ m that was acoustically treated to perform as a HI classroom under BB93 regulations.[15] The reverberation time (RT30), measured at the same position as the listener's head, was 0.35 seconds. Four loudspeakers (JBL Control 1) where placed at a height of 1.2 metres, at ±30° and ±60° at a distance of 3 metres from the listener to limit the effect of small head movements on the interaural cues in the recording.

Simulation of open-ear listening

The experiment used a modified headphone simulation of stimuli produced by far-field loudspeakers in a room similar to one used by Ohl[9], after Wightman and Kistler.[7, 8] The listener's head was not fixed, however a fixation point on the wall facing the listener was provided which helped the listeners obey instructions to maintain a fixed head. For the ITE conditions the in-ear microphones (Sound Professional MS-TFB-2) were placed at the entrance to the ear canals. For the BTE conditions, ear-hooks with integrated cable-guides (supplied with the Sound Magic PL30 in-ear headphones) were used to place the microphones above the front portion of the pinna, simulating the microphone position of a BTE hearing-aid. For the head-absent condition the in-ear microphones were placed on a horizontal bar, 18 centimetres apart and at the same height as the loudspeakers (1.2 metres). Eight concatenated swept-sine signals were played from each loudspeaker in succession at 75 dBA and simultaneously recorded by the microphones.

The stimuli were spectrally equalized for headphone playback by presenting the swept-sine signals over headphones and recording them through the microphones in the ITE position, creating headphone-equalized binaural-impulse responses (HEBIR) for headphone playback. In the frequency domain, using the inverse of the extracted impulses for equalization could result in large peaks in the filter and small variations in the position of the headphones in relation to the ITE microphones could vary the filter shape.[11] To reduce these effects, the headphones were removed by the participant after presentation of two swept-sine signals and replaced before recording again (for a total of eight signal presentations). All impulse responses were extracted using the technique given by Smith.[16]

Stimuli

The signals were 3 seconds of concatenated (or truncated) random sentences from the same talker of the IEEE York corpus.[17] The corpus was recorded at 16-bit, 44.1-kHz sampling rate. Spectrograms of the corpus recordings displayed sufficient speech energy above 6.5 kHz up to 15 kHz for high-frequency HRTF cues to be present in the convolved stimuli. For the one talker condition, a male talker was convolved with HEBIRs at −30°. For the four talker condition, female talkers were convolved with HEBIRs at −30° and +60°, and male talkers were convolved with HEBIRs at −60° and +30°. The lowpass condition was created by applying a tenth-order Butterworth filter with a cutoff at 6.5 kHz, mimicking the bandwidth of a standard hearing aid.

The convolved ITE/BTE, broadband/lowpass (BB/LP) sentences were mixed in the time domain with the same sentences convolved with the head-absent responses. The amount of ITE/BTE signal mixed in amplitude with the head-absent signal (the 'mix point') varied from 0% (i.e., hypothetical maximum internalization) to 100% (i.e., hypothetical maximum externalization) in 25% increments. The head-absent responses were time-shifted if necessary before mixing, to produce identical ITDs between the ITE/BTE and the head-absent convolved sentences. The RMS values were standardised before and after each stage of the mixing process for both the one- and four-talker conditions. The playback level was 70 dBA, ensuring greater than 15 dB sensation level for all listeners, using their worse-ear 4FA as a reference.

Experimental procedure

There were eight test conditions (a $2 \times 2 \times 2$ design) consisting of all combinations of the chosen parameters. There were 8 blocks of trials, each block consisting of 5 head-present to head-absent mix points and (if not already included in the mixes) the hidden reference. The parameter combinations were: ITE/BB/1 talker; ITE/BB/4 talkers; ITE/LP/1 talker; ITE/LP/4 talkers; BTE/BB/1 talker; BTE/BB/4 talkers; BTE/LP/1 talker and BTE/LP/4 talkers. In training the listener was played the same sentence in the ITE/BB condition at five mix points in order from 100 to 0%. The participant was asked if the stimuli appeared to move towards them over successive plays, beginning at the loudspeaker. All participants reported this effect, indicating successful creation of an externalized sound for the ITE/BB condition with a mix point of 100% and the viability of the mixing technique. The participant was trained in the use of the modified multiple stimulus with hidden reference and anchor (MuSHRA) test[18] to rate the stimuli. The reference (and hidden reference) for all conditions was the ITE/BB stimulus with mix point of 100%. The response screen consisted of a row of five (in the ITE/BB conditions, as the reference was the same as the 100% mix point) or six "mix" buttons and a slider corresponding to each button; mix points were randomly assigned to each button and slider. Upon pressing a "mix" button, the reference stimulus was played followed after a one second pause by the target stimulus to be rated. The participant was instructed to rate the second target stimulus against the first, using the slider and a 0-100 point scale with five referents: 'At the loudspeaker' (100); 'In the room' (75); 'At the ear'

(50); 'In the head' (25); 'Centre of head' (0). The referents for the scale were modified from Hartmann & Wittenberg's four-point scale.[13] To prevent listeners responding after insufficient listening, each mix had to be played at least twice to enable progression to the next condition. The training was repeated for the four-talker, ITE/BB condition. After training the participants performed the same task unsupervised on all eight conditions, presented in a randomized order.

Results

Externalization ratings were computed from the average response (with 100 = at loudspeaker and 0 = centre of head) for each mix point in each condition across NH and HI listeners. Each point is based on 14 responses for HI listeners and 7 responses for NH listeners. Table 1 displays the results of a between-subjects analysis of variance on the responses, showing statistically significant main effects and interactions. Three-way interactions were tested, but were not shown in Table 1 as they were not significant.

The top row of figure 1 shows the results for the one-talker conditions for both NH and HI listeners plotted with 95% confidence intervals. In the ITE/BB condition, NH listeners fully externalized ('at the loudspeaker') the 100% ITE mix and perceived the stimuli to move towards the head as the ITE mix decreased, with a minimum externalization between 'at the ear' and 'in the head.' In contrast, the HI listeners did not fully externalize the 100% ITE/ BB stimuli and experienced less internalization at the 0% mix point. NH listeners experience a reduction in maximum externalization across the other conditions. This reduction is similar in these conditions, with responses becoming more variable in the BTE/LP condition. The HI listeners appear relatively unaffected by microphone position and frequency response, placing maximum externalization 'in the room' and minimum externalization 'at the ear.' Externalization also appears to increase in the ITE/LP and BTE/BB conditions when listening to the 100% head-absent condition. They also show greater variability in their placement of the reference condition (In the ITE/BB condition the reference and 100% mix are identical stimuli).

The bottom row of figure 1 shows the average NH and HI listener responses for the four-talker conditions plotted with 95% confidence intervals. NH listeners display similar responses in the four-talker conditions to their one-talker equivalent conditions. Overall, the HI listener results display both a reduced maximum externalization and internalization in comparison to the one talker conditions, with all results lying between 'in the room' and 'at the ear.' Maximum externalization occurs for 50% mix in the ITE/LP condition and 0% mix (100% head absent) in the BTE/LP condition.

The correlation of high-frequency hearing-loss with externalization of the reference condition was statistically significant, (r = −0.61, p < 0.05) but for externalization of the head-absent condition it was not statistically significant (r = −0.13, p > 0.1). The correlation of age with reference externalization rating was also not significant.

Discussion

The NH listener results show that a continuum can be produced between full externalization and internalization using a mixing technique that varies the strength of the HRTF filtering present, holding all other cues constant. Due to the preservation of the ITD cues, the stimuli could still be lateralized in the head, hence the absence of ratings for 'centre of head,' which would have required a diotic stimulus. The results also show the importance of high-frequency pinna cues for the externalization of static acoustic scenes. The removal of these

cues resulted in a reduction in maximum externalization. Therefore, in NH participants, the design and frequency response of a BTE HA produces a reduction in externalization.

The results for the one-talker conditions show that HI listeners are less sensitive to high-frequency HRTF (i.e. pinna) cues with respect to externalization. However, they also show that this insensitivity produces a contracted perception of externalization. A simulated representation of sound in a room at an egocentrically fixed position is not fully externalized (i.e. placed directly at the loudspeaker) for HI listeners, whereas the preservation of the ITD and DRR is enough to partially externalize it in many cases, resulting in a reduced perceptual range. The correlation between high-frequency hearing-loss and reference externalization suggests audibility of pinna cues in addition to sensitivity may be a factor in loss of externalization. The reduction in internalization compared to NH listeners when only the static ITD and DRR cues were present suggests that HI listeners place a greater perceptual emphasis on these cues than pinna cues. The inclusion of dynamic binaural cues (not present in this study) may have resulted in a greater maximum externalization in the HI listeners. Insensitivity to high-frequency HRTF cues results in a smaller variation in maximum externalization across conditions, meaning that the HA microphone position and frequency response have a reduced absolute effect on externalization in HI listeners.

The aforementioned insensitivity and reduced perceptual range are compounded by a more complex acoustic scene of four spatially separated talkers. In these conditions, HI listeners often reported perceiving no change or movement in the stimuli presented and an inability to perceive four distinct talkers. Since the NH listeners did not report this problem in the four talker conditions, the cause may be attributed to the 'cocktail party problem'[19] manifesting as an inability to detect changes in HRTF cues and hence changes in externalization. The results in the low-pass four-talker conditions show that responses become decoupled from the mix played, suggesting that they are more insensitive to the remaining low-frequency cues in these complex listening conditions.

A number of participants found the task very difficult to perform, due to the complexity of the graphical user interface used or an inability to hear any difference between the reference and head-absent conditions during training on the four-talker condition. This produced a large variation across HI listeners, resulting in the large confidence intervals shown in the second row of figure 1. The acoustic scene delivered over the headphones, though acoustically identical to open-ear listening (within the documented limitations[8]), was fixed in space. Therefore, movements of the head resulted in the room moving with it, which could break the externalization illusion. To mitigate against this, participants were asked to listen to the stimuli while looking at the focus point. Combined with the visual localization of the loudspeakers, the externalization remained stable for the duration of the task. Future work could use head tracking to preserve the dynamic interaural cues to study what effect they have on externalization.

Conclusion

The effect of static HRTF cues, hearing-aid microphone placement, bandwidth and number of talkers on participants' ability to externalize speech was examined. Using a headphone simulation of open-ear listening and a modified MuSHRA testing paradigm, the study demonstrated:

1. Externalization can be perceived as a continuum in some conditions for HI listeners (e.g. ITE/BB/1 talker), using a mixing technique varying the strength of HRTF filtering present and keeping all other cues constant.

2. The NH results showed that the microphone placement and frequency response of a BTE hearing-aid adversely affected the perception of externalized sounds, due to the removal of high-frequency pinna cues.

3. The HI results for both the one and four-talker conditions displayed insensitivity to HRTF cues, in both the high and low frequencies.

4. HI listeners experienced a contracted sense of externalization, as stimuli where neither fully externalized nor internalized to the same degree as NH listeners.

These last two findings suggest that binaural cues, such as ITDs and ILDs and by extension dynamic binaural cues, are of greater importance for externalization in HI listeners.

Acknowledgments

References

1. Durlach NI, et al. On the externalization of auditory images. Presence. 1992; 1(2):251–257.

2. Plenge G. Über das Problem der Im-Kopf-Lokalisation. Acustica. 1972; 26(5):241–252.

3. Begault DR, Wenzel EM. Headphone Localization of Speech. Human Factors. 1993; 35:361–376. [PubMed: 8349292]

4. Kim SM, Choi W. On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. J. Acoust. Soc. Am. 2005; 117(6):3567–3665.

5. Noble W, Gatehouse S. Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing Scale (SSQ). Int. J. of Audiology. 2006; 45:172–181.

6. Zhang PX, Hartmann WM. On the ability of human listeners to distinguish between front and back. Hearing Research. 2010; 260:30–46. [PubMed: 19900525]

7. Zahorik P, Brungart DS, Bronkhorst WA. Auditory Distance Perception in Humans: A summary of Past and Present Research. Acta Acustica united with Acustica. 2005; 91:409–420.

8. Wightman FL, Kistler DJ. Headphone simulation of free-field listening. I: Stimulus synthesis. J. Acoust. Soc. Am. 1989a; 85(2):858–867. [PubMed: 2926000]

9. Wightman FL, Kistler DJ. Headphone simulation of free-field listening. II: Psychophysical Validation. J. Acoust. Soc. Am. 1989b; 85(2):868–878. [PubMed: 2926001]

10. Ohl, B.; Laugesen, S.; Dau, T. Externalization versus Internalization of Sound in Normal-hearing and Hearing-impaired Listeners. 2010. p. 136Published as part of: In Fortschritte der Akustik , Deutsche Gesellschaft für Akustik (DEGA)

11. Kulkarni A, Colburn HS. Variability in the characterization of the headphone transfer-function. J. Acoust. Soc. Am. 2000; 107(2):1071–1074. [PubMed: 10687721]

12. Toole FE. In-head localization of acoustic images. J. Acoust. Soc. Am. 1969; 48(4):943–949. 1969. [PubMed: 5480389]

13. Hartmann WM, Wittenberg A. On the externalization of sound images. J. Acoust. Soc. Am. 1996; 99(6):3678–3688. [PubMed: 8655799]

14. Little AD, Mershon DH, Cox PH. Spectral content as a cue to perceived auditory distance. Perception. 1992; 21:405–416. [PubMed: 1437460]

15. School Acoustics Building Bulletin. 2003; 93 http://www.bb93.co.uk/bb93.html Retrieved 19/12/11.

16. Berdahl, E.; Smith, JO. Impulse response measurment toolbox: Real Simple Project. 2008. http://cnx.org/content/m15945/latest/Retrieved 19/12/11

17. Stacey PC, Summerfield AQ. Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. J. Acoust. Soc. Am. 2007; 121(5):2923–2935. [PubMed: 17550190]

18. Recommendation ITU-R BS.1534-1 Method for subjective assessment of intermediate quality level of coding systems. 2003. http://www.itu.int/rec/R-REC-BS.1534/en . Retrieved 19/12/11

19. Cherry EC. Some experiments on the recognition of speech with and with two ears. J. Acoust. Soc. Am. 1953; 25:975–979.
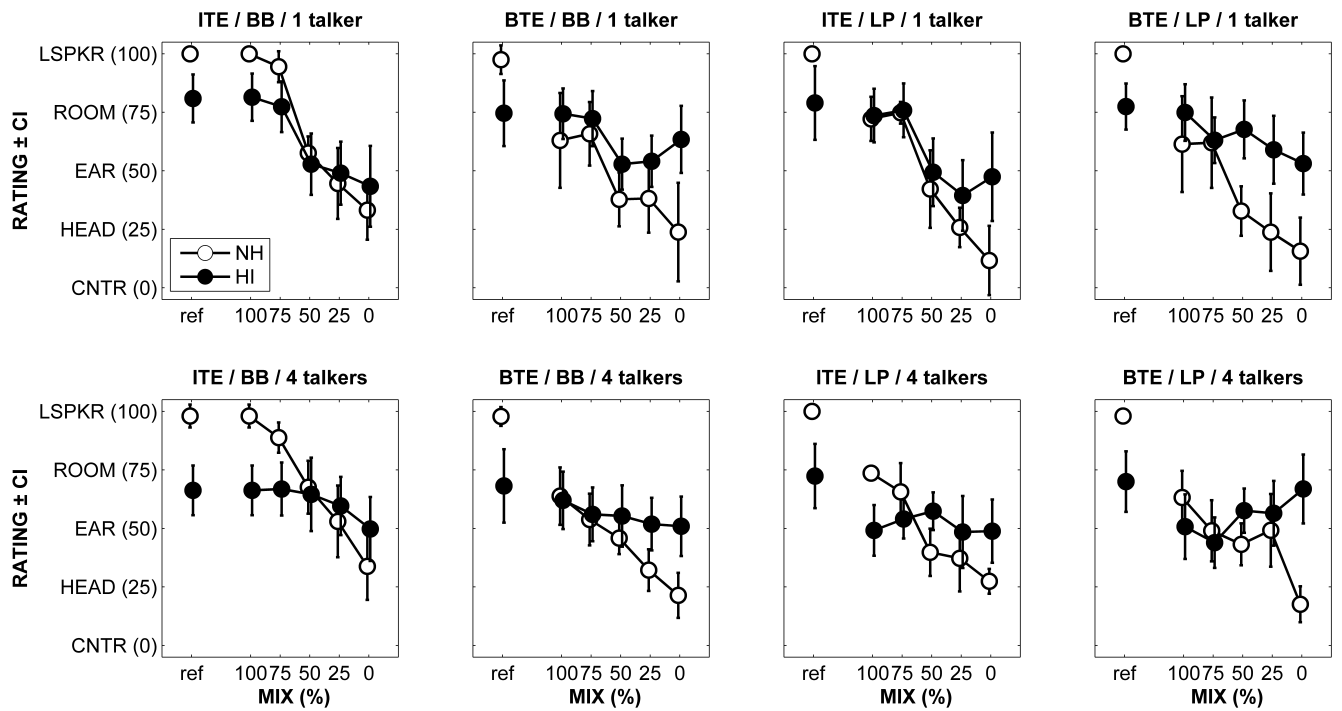
Figure 1.

Table 1

Source	d.f.	F	P	Source	d.f.	F	P
HI/NH	**1**	**26**	**4.44E-07**	**HI/NH*Bandwidth**	**1**	**6**	**0.015**
Mix	**4**	**57**	**0**	**Mix*Talkers**	**4**	**6.1**	**7.80E-05**
Talkers	1	1	0.3	**Mix*MicPos**	**4**	**5**	**0.00058**
MicPos	**1**	**13.9**	**0.00021**	Mix*Bandwidth	4	0.9	0.48
Bandwidth	**1**	**20.2**	**8.0 E-06**	Talkers*MicPos	1	1.2	0.28
HI/NH*Mix	**4**	**18**	**5.40E-14**	Talkers*Bandwidth	1	0	0.86
HI/NH*Talkers	**1**	**5.3**	**0.022**	**MicPos*Bandwidth**	**1**	**13.4**	**0.00028**
HI/NH*MicPos	**1**	**22.1**	**3.20E-06**				
				Error	**631**		