The auditory system enables listeners to estimate the localization of auditory events resorting notably to individual binaural cues and room reflections reaching the eardrum. For hearing-impaired listeners using hearing aids, those cues are possibly distorted or not available in various situations such as remote microphone systems. This can make auditory distance evaluation a challenging task and often results in inside-the-head sound perception. The present study aims at evaluating the effect of several binaural rendering strategies on auditory externalization in hearing-impaired and normal-hearing listeners. Within a multiple-stimulus listening paradigm, normal-hearing and moderate-to-highly-profound hearing-impaired listeners were asked to rate the perceived distance of various stimuli presented over headphones, while visual cues were available. Results show that a degree of externalization can be achieved for most of the aided hearing-impaired listeners with a simple generic spatialization method and superimposed early reflections.

Keywords: spatial hearing, hearing aids, externalization

1. Introduction

In certain applications, remote wireless microphone systems are used with hearing aids (HAs) to provide a clean and intelligible speech signal to hearing-impaired (HI) listeners [1]. The voice of the speaker is picked up close to the mouth by a body-worn microphone. It is then transmitted wirelessly to the HAs and rendered diotically, i.e. the same signal is played to both ears. This method allows large improvements in terms of speech intelligibility and reduces listening effort in noisy or reverberant environments [2]. Nevertheless, this technology most commonly discards any spatial cue for localizing the source. A real-time binaural localization and tracking algorithm for HAs and remote microphone (RM) systems was introduced in [3]. The position of the speaker relative to the listener is estimated and the
clean speech signal is spatialized accordingly. Hearing aids process sound within short-time frames (typically less than 10 ms) so as to ensure low-delay rendering of the output signals. Therefore, short impulse responses containing mostly the direct sound with its binaural cues are used. Speaker localization experiments using the aforementioned feature showed that most HA users had similar performance comparing natural and artificial spatial hearing, suggesting that HI listeners keep their localization abilities in the frontal plane when simple generic head-related transfer functions (HRTFs) are used [4]. Nevertheless, in-head localization (lateralization) was reported to occur, and mentioned by HI listeners as a limitation of this new feature.

Including the spatial cues of the filtering caused by the head, pinna and torso, as contained in the HRTFs, as well as reflections occurring in the environment, allows to render a convincing externalized acoustic scene over headphones [5, 6]. The resort to non-individualized HRTFs is known to yield a reduction of the degree of externalization [7]. Reflections have a crucial role in the perception of externalization [5], and more generally in the perceived distance of sound sources [8]. The direct-to-reverberant energy ratio (DRR) is used by the auditory system to estimate auditory distance [9].

Results of listening test conducted in [10] suggests difficulties to resort to DRR to evaluate distance in presbycusic subjects. Wide dynamic range compression (WDRC) is one of the main processing performed by HAs, and is usually non linear. Thus, it could be expected that WDRC might alter cues necessary in a localization task. In [11], WDRC did not have an effect on the performance of experienced HA users in a distance discrimination task, which can be explained by the acclimatization of subjects to their custom settings. In [12], it was suggested that HI listeners experience a contracted perception of externalization, since stimuli were never rated as externalized or internalized as for NH listeners. By varying the amount of head-related binaural information available to listeners during headphone reproduction, it was shown that externalization perception abilities were less homogeneous in HI listeners compared to NH listeners, and that in general, HI listeners are less sensitive to changes in externalization [13]. The influence of a RM mixed with the HA microphone signal on spatial perception was investigated in [14]. It was concluded that the gain of the RM should be reduced as much as possible to optimize localization.

As it is not practically imaginable to measure individual HRIR for every RM system user, the challenge consists in trying to improve externalization with a generic method. Recently, a novel algorithm aiming at improving externalization in HAs was introduced in [15] and relies on the "structural binaural model" introduced in [16]. This model consists in combining separate individual filters representing all the components of the HRIR (head, pinna, ear canal). The study evaluated the effect of various procedures for speech externalization presented through earphones to NH and moderate HI listeners. The addition of room reverberation brought a significant improvement in externalization compared to the anechoic HRIRs. The structural binaural model with simulated reverberation resulted in similar performance as the use of the listener’s own HRIRs.

The present study addressed a similar problematic by evaluating the degree of externalization obtained from several reproduction strategies in NH and HI listeners. In the context of RM systems it is of particular interest to investigate if the addition of early reflections to auralize speech using generic HRIR can provide a substantial improvement in externalization perception. Contrary to the aforementioned studies, HI listeners with a high degree of HL were included in the evaluation.
2. Methods

2.1 Participants

The subject panel consisted of 10 NH (med. age = 22 y.o., 5 female) and 20 HI (med. age = 31 y.o., 10 female) listeners. The mean of the pure-tone average (PTA) was 93 dB at the best hear (range 55-114 dB) for HI listeners. 2 had a moderate HL, 3 had a severe HL, 15 had a profound HL (among which 11 had a PTA at the best ear higher than 100 dB HL). All HI listeners had congenital or pre-lingual HL.

2.2 Experimental setup and measurement phase

The experimental setup was mounted in a classroom at EPFL (V = 177 m$^3$, RT$_{60}$ = 530 ms). Three loudspeakers (Genelec 1029A), numbered from 1 to 3, were placed at 67 cm, 113 cm and 200 cm respectively from the listener’s position. They were located at an azimuth of 30$^\circ$ on the right side of the listener, as shown on Fig. 1. This angle ensured that all loudspeakers were easily visible at the listener’s position.

The experiment started with a measurement phase. The subjects sat in the center of the room and had their head immobilized by mean of a chin rest. In-the-ear binaural microphones (The Sound Professionals MS-TFB-2) were used to measure their individual BRIRs corresponding to the three loudspeakers, as well as their headphone-to-ear impulse responses (HPIRs). For HI listeners, the real ear aided response (REAR) was measured for each ear using a probe microphone measurement unit (Aurical FreeFit). Thereafter, during the experiment, HI listeners did not use their HAs, and all the stimuli were rendered through a pair of open headphones (Audeze LCD-2C). The custom WDRC settings from their HAs was extracted from the audiologist fitting software (Phonak Target) and simulated in Matlab (The Mathworks). Additionally, a headphones amplifier (Lake People HPA RS 02) was used in order to render high output level with a low total harmonic distortion (THD).

![Figure 1: Schematic representation of the experimental setup mounted in a classroom](image-url)
2.3 Stimuli

The signals were 10-second speech sequences, derived from the concatenation of 5 short phonetically-balanced random sentences (French HINT database). They were processed with different methods described as follows:

- **Reference (Ref):** the speech sequence was convolved with custom BRIRs of Loudspeaker 3.

- **Diotic (Dio):** the original speech sequence was reproduced diotically.

- **Truncated (ER60):** the speech sequence was convolved with a truncated version (60 ms) of custom BRIRs of Loudspeaker 3. It corresponds to the direct sound and early reflections.

- **Generic (Gen):** the speech sequence was filtered by generic minimum-phase 128-sample (5.8 ms) HRIRs, as measured on a KEMAR manikin (G.R.A.S) in an anechoic chamber with a source located at a distance of 2 m, and an azimuth of 30°. The ITD is simulated by a pure delay. This is comparable to a generic anechoic spatialization of a source at the position of Loudspeaker 3.

- **Generic with ER (Gen/ER):** obtained from a mix of the Gen stimulus and additional ER extracted from the HA microphones using a proprietary algorithm. The HA microphone signals were generated with BRIRs of HAs worn by a KEMAR at the position of the listener measured beforehand.

The stimuli were compensated with the frequency response of the headphones measured for every subject and low-pass filtered (cut-off frequency = 6.5 kHz). For HI listeners, all stimuli were processed with the WDRC settings obtained from the fitting software applied after the spatialization processing. For NH participants, no WDRC (i.e. linear amplification) was performed. For NH listeners, the stimuli were rendered at 65 dB SPL. Based on the individual measurements of the REARs, the reproduction level varied between 67 and 120 dB SPL for HI listeners.

2.4 Task

Using a MUSHRA-type graphical user interface (GUI) displayed on a touch-pad, the participants were asked to rate the auditory externalization perceived for the five stimuli presented simultaneously. The instruction consisted in answering the following question: "How far do you perceive each stimulus from your position?". They used a continuous scale displayed as a slider with the following markers: **Center of the head (0), Boundary of the head (20), At Loudspeaker 1 (40), At Loudspeaker 2 (60), At Loudspeaker 3 (80) and Further than Loudspeaker 3 (80 to 100).** The task was repeated over 4 to 6 runs for each participant, depending on the consistency of the ratings of each subject. The experiment was preceded by a training phase, in which the listeners were given the possibility to listen as much as they want to three versions of a speech sequence spatialized with their custom BRIRs corresponding to Loudspeakers 1, 2 and 3. This helped subjects to get accustomed with the task. Additionally, this served to ensure that their auditory spatial perception matched the visual location of the loudspeakers, as well as give them an *a priori* knowledge of the reproduction level used along the experiment.

3. Results

Externalization ratings are reported on Figure 2 for NH listeners (left panel), and HI listeners (right panel). For every subject, only the three runs with the lowest variances on ratings of the Ref and Dio stimuli are considered. The first run was not taken into account and considered as training. Two HI listeners rated the diotic stimuli to be more externalized than the reference, they were thus excluded from the statistical analysis.

---

1 Collège National d’Audioprothèse 2006
Figure 2: Degree of externalization evaluated by NH (a) and HI (b) listeners on a continuous scale with the following markers: Center of the head (0), Boundary of the head (20), At Loudspeaker 1 (40), At Loudspeaker 2 (60), At Loudspeaker 3 (80) and Further than Loudspeaker 3 (80 to 100).

3.1 Externalization in NH listeners

A Friedman test among repeated measures revealed significant differences in the perceived degree of externalization between the five stimuli in NH listeners ($p < 0.001, \chi^2 = 114.96$). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction were conducted. The Ref stimuli were rated as being more externalized than the Dio ($p < 0.001$), Gen ($p < 0.001$), Gen/ER ($p < 0.001$) and ER60 ($p = 0.015$) stimuli. No significant difference was found between the Dio and the Gen stimuli ($p = 0.143$). The Dio stimuli were more internalized compared to the ER60 ($p < 0.001$) and Gen/ER stimuli ($p < 0.001$). ER60 were significantly more externalized than the Gen stimuli ($p < 0.001$), but not significantly different compared to the Gen/ER stimuli ($p = 1.000$). Finally the Gen/ER stimuli were perceived significantly more externalized than the Gen stimuli ($p < 0.001$).

3.2 Externalization in HI listeners

A Friedman test among repeated measures showed that the externalization ratings were significantly different between the five stimuli for the HI listeners ($p < 0.001, \chi^2 = 156.76$). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction were performed. The Ref stimuli were perceived significantly more externalized than the Dio ($p < 0.001$), Gen ($p < 0.001$) and Gen/ER ($p < 0.001$) stimuli, but not from the ER60 ($p = 0.074$) contrary to NH listeners. Similarly as for the NH listeners, the Dio stimuli were more internalized compared to the ER60 ($p < 0.001$) and Gen/ER ($p < 0.001$) stimuli, but not significantly different from the Gen stimuli ($p = 0.051$). The ER60 stimuli were significantly more externalized than the Gen stimuli ($p < 0.001$), but not significantly different from the Gen/ER stimuli ($p = 0.634$). The degree of externalization of the Gen/ER stimuli was significantly higher than in the Gen stimuli ($p = 0.002$).
4. Discussion

In NH listeners, externalization was rated as expected: the reference was perceived at the location of the Loudspeaker 3, and the diotic stimuli was perceived in the center of the head. In HI listeners, a similar tendency was observed in the ratings. Nevertheless, the reference was generally perceived less externalized than in NH listeners and the diotic stimuli were rarely perceived as completely internalized. Thus, HI listeners experienced a contracted perception of externalization, in accordance with the results obtained with mild-to-moderate HI listeners in [12]. In general, the results also revealed that moderate-to-highly-profound HI listeners reported less homogeneous results compared to NH listeners. This is in agreement with conclusions drawn in [13] for mild-to-moderate HI listeners.

In NH listeners, the stimuli including early reflections were perceived more externalized compared to the stimuli using anechoic spatialization. This confirms that early reflections are sufficient to provide a sense of externalization [6, 8]. In HI listeners, the stimuli with early reflections were also perceived more externalized than the diotic and anechoic spatialization. There was no significant difference in externalization between the reference and the truncated BRIRs in HI listeners, while NH listeners rated the reference to be more externalized. This is in accordance with [11], in which the difficulty to resort to DRR to estimate auditory distance for HI listeners was pointed out.

The addition of early reflections picked up from the HA microphones to a simple anechoic and generic spatialization method was sufficient to significantly improve externalization in both HI and NH listeners. This is of special interest in the context of RM systems, for which this method has been already addressed successfully for localization in the azimuthal plan [4]. The method investigated in this paper could therefore solve the problem of missing realism reported by users. Similar conclusions were found in [15] for mild-to-moderate HI listeners and artificial early reflections.

5. Conclusion

This study showed that moderate-to-highly-profound HI listeners have a contracted perception of externalization compared to NH listeners, as was previously observed in mild-to-moderate HI listeners. Previous studies also showed that mild-to-moderate HI listeners perceive externalization less homogeneously, which was found in this study as well. The addition of early reflections to a generic and anechoic spatialization was shown to substantially improve externalization for moderate-to-highly-profound HI aided listeners. This is a promising conclusion in the context of RM systems, for which anechoic spatialization was already proposed [3]. Optimization of speech intelligibility remains the main purpose of RM systems. Hence, further studies should consider the potential effect of additional early reflections on speech intelligibility.

Acknowledgement

This study was partly funded by the Swiss Innovation Promotion Agency (Innosuisse) with grant number 25760.1 PFLS-LS, and partly by Sonova AG, Stäfa, Switzerland. The protocol has been approved by the Cantonal Office of Public Health of the Canton de Vaud (CER-VD), Switzerland, and by the National Authorization and Supervisory Authority for Drugs and Medical Products in Switzerland (Swissmedic). This study is registered under the identifier NCT03512951 on https://clinicaltrials.gov.
REFERENCES


