Validation of a Virtual Sound Environment System for Hearing Aid Testing

Jens Cubick¹, Sylvain Favrot², Pauli Minnaar³, Torsten Dau⁴

¹ Centre for Applied Hearing Research, DTU (DK), Denmark, Email: jecu@elektro.dtu.dk
² Dept. of Speech, Language and Hearing Sciences, BU (USA)
³ Oticon A/S (DK)
⁴ Centre for Hearing and Speech Sciences, DTU (DK)

Introduction

Hearing aid (HA) users typically benefit a lot from their HAs in simple acoustic situations like a one-to-one conversation in a quiet room, but they have difficulties following a conversation in challenging listening situations involving multiple talkers, background noise and reverberation [1]. The processing power of HAs has increased dramatically over the last 10 years and advanced signal processing strategies have been applied to help the users, particularly in more complex listening situations. To assess and evaluate the performance of such modern HAs, the test scenarios should be as realistic as possible. Until recently, however, most testing was done either in very basic conditions with simple loudspeaker setups in acoustically dampened rooms or in field studies, where end users wear certain types of HAs for a while and report back via questionnaires after the testing period. The first approach offers much control over the test conditions, but only very limited flexibility regarding the acoustic conditions. In field tests, the participants experience the HAs in their everyday life, but the experimental conditions are difficult to control.

Modern sound field reproduction techniques like wave field synthesis and higher order Ambisonics [3] allow for rendering realistic and reproducible virtual sound environments (VSEs) including room reverberation and multiple sound sources in the laboratory. Unlike in headphone-based systems, listeners are able to wear HAs in a VSE. However, to successfully employ these systems for HA testing, it needs to be known how well experimental results translate to real-life situations.

In this study, speech intelligibility was used as a measure for the authenticity of the simulation. It was assumed that the simulation is sufficiently authentic, if the outcome of the listening experiments in the VSE is close enough to the results obtained in the same experiments carried out in the corresponding real room.

Methods

To assess the sound field reproduction, speech intelligibility experiments were conducted with normal-hearing listeners wearing HAs, both in a real room and in a simulation of that room aurialised via a spherical array of 29 loudspeakers as shown in Figure 1. The benefit with respect to speech intelligibility from a static beamforming (BF) algorithm compared to a HA setting with omnidirectional microphones was tested, an algorithm that has been shown to improve speech intelligibility [7, 8]. The algorithm relies on signals from both HA microphones. The results therefore directly depend on the spatial properties of the sound field and it was expected that inaccuracies in the sound field reproduction could considerably decrease the performance.

Figure 1: Spherical array with 29 loudspeakers that allows for the Auralization of acoustical scenes in virtual rooms. Photo courtesy of Oticon A/S.

The VSE tested in this study was based on a room simulation in the commercial ODEON room acoustic simulation software [2]. The simulation result was processed by the LoRA toolbox [5], that generates a multi-channel room impulse response using either higher order Ambisonics (HOA) or a rendering method, where each reflection is mapped to the nearest loudspeaker available (NLS). The obtained impulse response was then convolved with an anechoic signal to generate the driving signal for each loudspeaker. The room chosen for this study was Room 019, a lecture room at DTU with a volume of about 180 m³ and a reverberation time of about 0.5 s. The ODEON model was carefully matched to the reverberation time and clarity values measured at the listening position in the room (cf. Figure 2).

Eight normal-hearing, native Danish speaking listeners (2 female) with an average age of 27 years participated in
the study. They were supplied with regular production receiver-in-the-ear HAs (Oticon Ino) providing a linear gain of 15 dB across the frequency range of the HA. The HAs were coupled to the ears with mushroom-shaped silicone Oticon power domes to impair the natural hearing ability of the test subjects. Speech intelligibility was determined using the Danish Dantale II speech-in-noise test [6]. The speech target source was placed at 0° at distances of 2 m and 5 m, respectively (cf. Figure 2). Three noise sources (Dynaudio BM6P) were placed at angles of ±112.5° and 180° at a fixed distance of 2 m. The level of the speech-shaped noise was kept constant at 70 dB SPL in all unaided conditions and 62 dB SPL in all HA conditions, while the speech level was adjusted to find the speech reception threshold (SRT) representing the signal-to-noise ratio (SNR) at which 50 % of the words were understood correctly. An overview over the test conditions can be found in Table 1.

The benefit from the BF over the omnidirectional program is shown directly in Figure 4. It can be seen that the results measured in the classroom and in the VSE showed similar tendencies. The BF always yielded a clear advantage in speech intelligibility relative to the omnidirectional microphone setting generally increased the SRT compared to the unaided condition (by up to 4 dB on average in the real room), whereas using HA in the BF program lowered it (by up to 2.7 dB with HOA coding).

Increasing the target source distance from 2 m to 5 m (grey symbols) in Room 019 (left panel), the listeners showed an increase in SNR of about 3 dB at the SRT. This was expected due to the lower direct-to-reverberant sound ratio, which is generally assumed to have an adverse effect on speech intelligibility. Compared to the results for 2 m distance, the SRTs measured for 5 m distance showed a considerably larger spread. At this distance, head movements might have had a rather large effect on the perceived SNR and some listeners might have used them more successfully than others. With NLS coding, the difference in SRT between the two distances turned out to be the same as in Room 019. This strongly suggests that this coding preserves many of the cues that determine speech understanding despite the fairly simple algorithm. The spread in these data was smaller than in the real room, which coincides with the subjective impression of some listeners, who reported that the threshold for speech understanding seemed less gradual than in the classroom. With HOA rendering, the effect of the target source distance was less pronounced. On average, the SRT still increased with increasing distance, but by a smaller amount.

Table 1: Overview over listening test conditions.

<table>
<thead>
<tr>
<th>Room</th>
<th>Distance</th>
<th>HA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R019</td>
<td>2 m</td>
</tr>
<tr>
<td>2</td>
<td>VSE-NLS</td>
<td>5 m</td>
</tr>
<tr>
<td>3</td>
<td>VSE-HOA</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Top view of the room acoustic model with the listening position (1), the three noise sources (P7, P10, P13) and the target speech sources at 2 m (P2) and 5 m (P22).

Figure 3: Average Speech reception thresholds measured in Room 019, the VSE with NLS coding and the VSE with HOA coding. The error bars indicate ± one standard deviation.

Results and Discussion

Figure 3 shows the mean value and standard deviation of the measured SRTs for all combinations of the conditions in Table 1, i.e. the three HA conditions (unaided, Omni and BF) measured in all room conditions (R019, VSE-NLS and VSE-HOA) for the target source distances of 2 m and 5 m. For a target source distance of 2 m (black symbols), the unaided SRTs were found at SNRs of -13.8 dB in the real room (R019, left panel), -11.8 dB in the VSE with NLS coding (middle panel) and -9.4 dB with HOA coding (right panel). The higher SRTs with HOA compared to NLS coding are consistent with findings in an earlier study [4]. Using HAs in the omnidirectional microphone setting generally increased the SRT compared to the unaided condition (by up to 4 dB on average in the real room), whereas using HA in the BF program lowered it (by up to 2.7 dB with HOA coding).
Conclusion

In this study, speech intelligibility in noise was used as a measure to assess the authenticity of a VSE based on a carefully calibrated room acoustic model of an existing classroom. It was shown that the SRTs for normal-hearing listeners with HAs improved in a similar way in the VSE as in the real room, when a BF program was used instead of an omnidirectional one. Furthermore, the dependence of the SRT on the target source distance was found to be very similar in the VSE, when the NLS rendering method was used. This shows that, even though the SRTs differed, all differential results translated well to the real world. Testing new versus traditional HA features typically considers such differential measures. This suggests that the VSE system may prove an invaluable tool for HA testing in the near future, since it can provide valuable information and allows involving end users very early in the HA development process.

References


