Evaluation of an ILD-based hearing device algorithm using Virtual Sound Environments

Ruksana GIURDA[1,2], *Laurent S. R. SIMON*[1,3], *Hannes WÜTHRICH*[4], *Norbert DILLIER*[1,3]

[1]ENT department, University Hospital, Zürich, Switzerland
[2]Department of Electrical Engineering, ETH Zürich, Switzerland
[3]University of Zürich, Switzerland
[4]Sonova AG, Stäfa, Switzerland

Abstract

Hearing devices are mostly tested using controlled laboratory environments. These feature simplified auditory representations of the real conditions under which the hearing devices will perform once released, such as one speech signal played from one loudspeaker plus a noise signal playing from another loudspeaker. It can however be insufficiently realistic for some of the processing included in the devices. In order to circumvent this limitation, Virtual Sound Environments (VSE) can be used to simulate complex sound fields using headphones or arrays of loudspeakers, and thus to offer more realistic conditions than most common hearing device evaluation protocols. In this paper, a hearing device algorithm based on Interaural Level Differences (ILD) was evaluated using Binaural Room Impulse Responses (BRIR, used here as a reference), and artificial room effects reproduced via either Higher Order Ambisonics (HOA) or Vector-Based Amplitude Panning (VBAP). Results show that the algorithm does not work when using 5^{th} order HOA, and that the performance of the algorithm is lower when using VBAP than when using BRIR.

Keywords: Hearing devices, 3D audio, Interaural Level Differences, Signal processing

1 INTRODUCTION

Normal Hearing (NH) people show remarkable abilities in listening tasks, even in challenging conditions. Those abilities strongly rely on the use of the binaural cues such as Interaural Level Differences (ILD) and Interaural Time Differences (ITD). By combining the information from the left and the right ears, normal hearing people are able to focus on a single speaker in a mixture of competing/interfering talkers such as cocktail-party situations. They are also able to localize a target sound source in complex auditory environments. However, the cues used for such tasks are degraded in Hearing Impaired (HI) people. Depending on the type and severeness of the impairment, tasks such as sound localization, sound lateralization, discrimination of interaural level and time differences, speaker discrimination are challenging and, in some severe cases, impossible. A brief review of past studies conducted on the effect of hearing impairment on binaural interaction is given in [1].

Initial studies on spatial perception in HI people used limited conditions e.g., no room simulation, sources coming from single loudspeakers. Recently, the development of Virtual Sound Environments (VSE) made it possible to simulate complex environments and use them to understand and improve speech perception in such environments [3]. In this study, VSEs are used to guarantee more realistic test conditions, particularly in terms of variety of sources, source positions, room effects, and dynamic of the auditory scenes. The increased realism might be required for testing the efficiency of e.g., noise reduction or spatial processing algorithms.

The use of Virtual Sound Environments (VSE) for evaluating hearing devices in complex auditory 3D environments was already discussed in [4, 5, 6, 7, 8]. More particularly, the authors discussed the limitations of 3D audio environments for evaluating hearing devices. In [9], the authors compared the reproduction of ILDs with Distance-Based Amplitude Panning (DBAP), Higher Order Ambisonics (HOA), Multiple Direction Amplitude Panning (MDAP), and Vector-Based Amplitude Panning (VBAP). They showed that the ILDs were better reproduced with VBAP by 3dB on average. The ILD errors could reach up to 17dB with DBAP. Given that according to [10], sources located 90° to the side can lead to ILDs up to 25dB at high frequencies, this implies that the errors of ILDs for VSE are not neglectible.

The impact of such errors on hearing device processing is not yet known. Yet it is important to consider these 3D audio reproduction artifacts when evaluating hearing aids: hearing devices offer tools that can help the wearer to focus on a target sound source in complex acoustic environments. Some of these tools are based on algorithms which also rely on the estimation of binaural acoustic features. For that reason, a correct simulation of the binaural cues is fundamental when creating virtual acoustic environments for hearing device evaluation.

The current study therefore illustrates the effects of the ILD reconstruction errors of HOA and VBAP reproduction techniques on a hearing device algorithm for sound localization. This paper describes the experimental setup, shows the results and discusses the limitations of the reproduction systems for the evaluation of hearing devices.

2 EXPERIMENTAL SETUP

In this study, a hearing aid localization algorithm based on binaural cues was tested using three different types of sources: a real source, an HOA virtual source, or a VBAP virtual source.

The real source test signals were obtained trough the convolution of Binaural Room Impulse Responses (BRIR) and anechoic speech signals. The BRIRs were recorded with two Behind The Ear (BTE) hearing aid satellites placed on a KEMAR manikin. The database of the BRIRs comprises impulses recorded from 0° to 350°, with a step size of 10° in the azimuth plane. The recordings were conducted in three different rooms with a source-receiver distance of 2m. The acoustical properties of the recording environments are shown in Table 1.

Table 1. Acoustic properties of the rooms.

Room	Volume [m^3]	DRR and RT_{60} at 2m [dB|s]	
Cafeteria	488	5.4	0.61
Listening room	125	5.7	0.17
Classroom	262	4.1	0.59

Both of the HOA and VBAP recordings were made in a listening room (RT30 = 199ms) equipped with an array of 32 8020 DPM Genelec loudspeakers:

- 8 loudspeakers every 45° azimuth at an elevation of -30°
- 12 loudspeakers every 30° azimuth at an elevation of 0°
- 8 loudspeakers every 45° azimuth at an elevation of +30°
- 4 loudspeakers every 90° azimuth at an elevation of +60°

Distance from the loudspeakers to the center of the array was 1.5 m, except for the 4 top loudspeakers, which were 98 cm away from the center.

In order to have a comparison as fair as possible between the methods, only the results obtained with the BRIRs recorded in the Listening room, the one with the most similar RT_{60}, are considered here. In this way, the effect of the reverberation time is reduced and only the geometry of the rooms will affect the results.

Fifth order HOA and VBAP sound reproduction was done using Max[1], a visual programming tool for audio,

[1] https://cycling74.com/

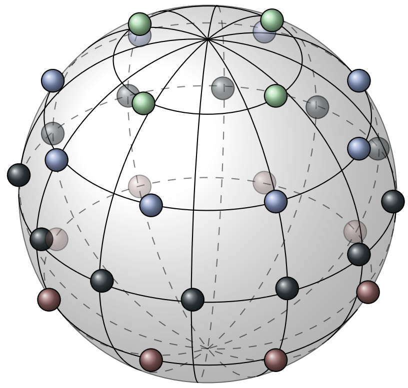

Figure 1. Schematic of the loudspeaker spherical array used to generated HOA and VBAP recordings.

and Spat[2], a spatial audio processing plug-in for Max. No room effect was added in Max nor Spat, meaning the reverberation recorded at the hearing aid satellites was that of the recording room. This means that the RT60 was not the same for all conditions. The smaller RT60 for the VBAP and HOA reproduction room implied that the performance of the tested algorithm should have been better with the simulated sound sources than with the virtual sources.

The stimuli tested in this study included either one or two speech sources. The absolute localization performance of the algorithm was tested using a single talker. The ability to identify and localize multiple sources was tested using two competing talkers placed in specific positions.

The inputs of the tested algorithm are the recordings of the four microphones placed on the hearing devices. The energy level difference between left and right devices is computed for each considered frequency band and only the level differences in the high frequency bands is kept for further analysis. Based on the level difference measured at each time frame and other relevant acoustic parameters, a decision model estimate in which one of the six possible areas the detected sound source is located, as depicted on Fig. 2. The limits between areas 1 and 2 or 1 and 6 are set at $\pm 30°$. The localization performance of the algorithm is azimuth-dependent. In this study, the comparison with different sound reproduction methods was carried out with sound sources placed always in positions where the performance of the algorithm was expected to be good.

3 RESULTS

In this section, the results for the localization algorithm obtained with the three different sound reproduction methods, BRIRs, HOA and VBAP are presented.

Fig. 2 shows the acoustic scene simulated and tested in this study. Two speech sources are located on the left and right side with respect to the hearing aid user. The right sound source was set to be slightly louder than the left one ($\approx 5dB$). The sources were placed at the azimuth positions of $\pm 50°$ with the $0°$ considered at the front of the hearing aid user. This acoustic scene was used to test the three sound reproduction methods here considered. Only for VBAP and BRIRs, a further test with the speakers located at $\pm 60°$ was conducted.

As mentioned previously, the results obtained with the binaural impulse responses are used as the reference. Fig. 3 shows that sound sources are detected in area 1, 2 and 6 even though the sources are only located in area 2 and 6. The phantom source detected in area 1 comes from the interaction of the two sources. Without further processing, whenever two sources interfere in such a way that the measured level difference is close to zero, the algorithm will interpret these low ILDs as a source located at the front (for which the natural ILDs would be close to zero). This aspect will not be further discussed as it is not relevant for this study and the

[2] https://forumnet.ircam.fr/product/spat-en/

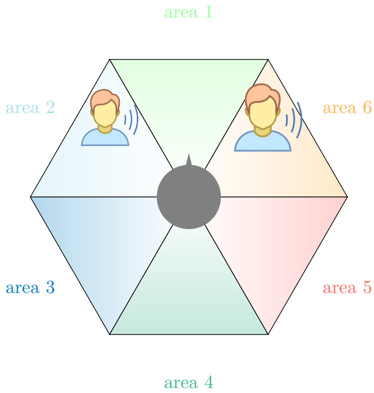

Figure 2. Acoustic scene simulated with the three sound reproduction methods. Sources are placed at the azimuth positions of ±50° with the 0° considered at the front of the hearing aid user.

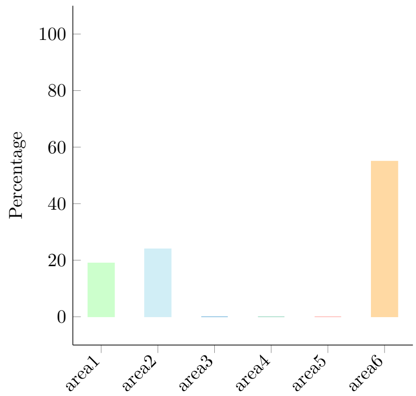

Figure 3. Localization results with BRIRs when the sources are located at ±50°. This output is used as reference in the comparison between the reproduction methods.

result with the phantom source detected in area 1 is still going to be considered as the reference. In addition to the estimated location, the algorithm provides information about which source was louder over time: Fig 3 shows that the source located in area 6 was louder than the source located in area 2. However, no information about the loudness difference is provided.

Fig 4 and 5 show the results obtained with HOA and VBAP, respectively.

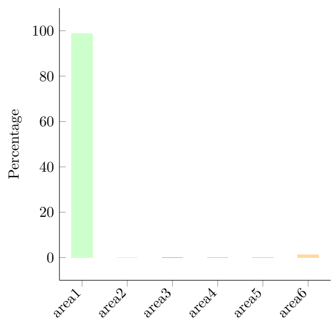

Figure 4. Localization results for HOA when the sources are located at ±50°.

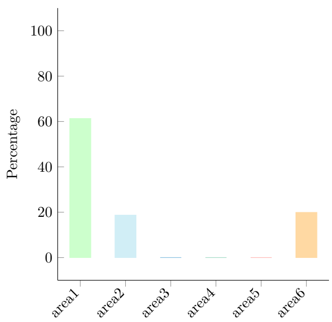

Figure 5. Localization results for VBAP when the sources are located at ±50°.

The results shown on Fig 4 demonstrate that the algorithm failed in localizing the sound sources. With VBAP the algorithm could detect more than one source but the interaction between the sources is quite different from what is obtained with BRIRs. In fact, the percentage of detection in area 1 is more than 60 in case of VBAP but only 19 in case of BRIRs.

To verify whether this difference in the source interaction was mainly due to the reproduction system, a second test was run only for VBAP and BRIRs. Fig. 6 and 7 show the comparison between the localization results obtained with BRIRs and VBAP when the sound sources are located at the true speaker positions of ±60°. In this case, only one speaker per source is active and the difference between the results is smaller. In fact, the percentage of detection in area 1 is 45 for VBAP and 24 for BRIRs. It is worth mentioning here that one may expect to have a smaller detection in area 1 for the BRIRs since the sources are located further to the side.

However, this expectation does not consider the actual geometry of rooms, the actual reflection pattern present in the impulses and the difference , even if small, in the microphone directivity.

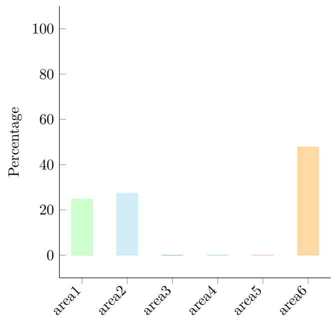

Figure 6. Localization results obtained with BRIRs when the sources are located at the true speaker positions of ±60°.

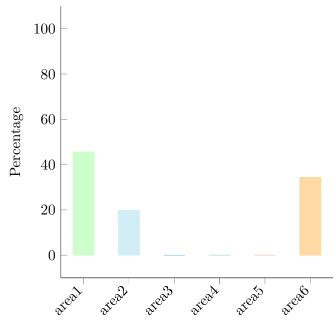

Figure 7. Localization results obtained with VBAP when the sources are located at the true speaker positions of ±60°.

4 DISCUSSION

In [9], an objective comparison of various VSE was done in the same room and similar conditions to those described in this paper. In the study by Simon et al., ILDs produced by virtual sources located every 10° on the horizontal plane were compared to the ILDs of real sources for the same source positions (either one of the loudspeakers of the reproduction system, or an additional loudspeaker used only for the measurement of real source impulse responses). When a virtual source produced with VBAP is located at the same position as one of the loudspeakers of the array, only that loudspeaker is active. Therefore at those positions, the ILDs produced by VBAP virtual sources are the same as the ILDs of the real sources. This is not the case for HOA, for which all loudspeakers are always active, although with different weights.

In the following, the expression 'estimation error' will refer to the difference between the estimation of the algorithm with HOA and VBAP and the estimation with real sources, i.e. BRIRs.

The results shown in Fig. 8 and 9 could explain why the ILD-based algorithm failed to work correctly in the case of acoustic environments reproduced with HOA and VBAP. The plots show the difference in ILDs between the real sources and the two other reproduction methods as a function of frequency and azimuth. The colorbar range has been manually limited between 1 and 7 for an easier comparison. However, in the case of HOA, the difference can reach 10 dB in the high frequency range. The algorithm considered here strongly relies on the actual ILD value measured at any time frame. Whenever the ILD values are distorted, the algorithm will misinterpret the acoustic scene. In particular, as already mentioned in section 2, the algorithm considers only the ILD values in the high frequency range where the sound is more directive and the ILD is the dominant cue. The error in ILDs produced by HOA and VBAP differs in magnitude and nature. In the case of HOA, the sound field in the high frequency range is not faithfully reproduced due to spatial aliasing and to the limited number of real sources producing the sound. Therefore, the directional cues are missing and the algorithm cannot estimate the direction of arrival of the sound or, in case of multiple sources, which one is louder. In the case of VBAP, there is a visible effect of the azimuth position of the source playing. The estimation error of the algorithm was larger in the case of virtual sound sources, i.e. sources that are placed between two speakers. In this case, more than one speaker at a time is active and the correlation between the sound fields generated by each active speaker creates the illusion of a source placed where no physical speaker is present. The effect of the correlation is limited and results in a sound field which is not a faithful reproduction of the desired acoustic scene. Comparatively, the estimation error was relatively small when the source was played at a true speaker position. In this latter case, only one speaker is active and the differences between the estimations with

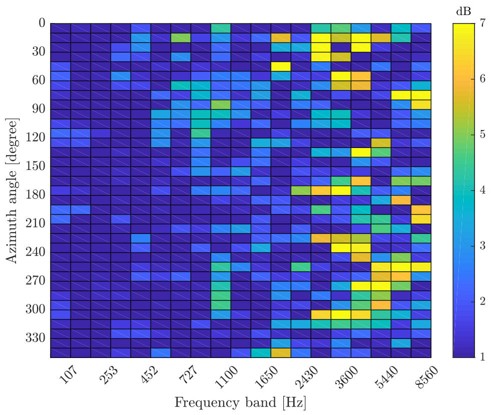

Figure 8. Difference of ILD in dB between the real sources and the HOA5 reproduction system in the same listening room, at the same positions. Measurements were done every 10°.

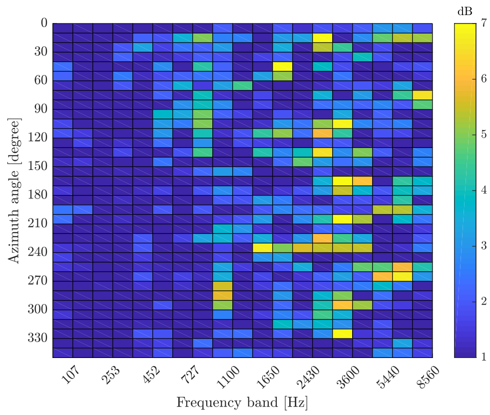

Figure 9. Difference of ILD between the real sources and the VBAP reproduction system in the same listening room, at the same positions. Measurements were done every 10°.

VBAP and BRIRs comes only from the difference in the reflections in the two tests.

Due to the multiple parameters involved in the localization estimation, it is difficult to evaluate the reproduction systems with a single number. In fact, one should take into account the exact geometry of the room, the absorption coefficient of all the elements present in it, their relative location and the exact position of the sources with respect to the receiver. In this study, the comparison was done between two rooms which had a very similar reverberation time but the actual reflection patterns and geometry of the rooms were not considered. However, all these parameters are not likely to produce an ILD difference up to 10 dB. Therefore, the differences in the estimations presented here are still related to the actual reproduction system used.

5 CONCLUSIONS

In this study, the effects of three sound reproduction methods on a hearing-device localization algorithm that is based on ILDs has been conducted. It was shown that the fifth order HOA was not able to faithfully reproduce the sound field in the high frequency range leading to a wrong localization for the whole duration of the sound signals. Better results were obtained with VBAP although it was shown that the mismatch between the reference (BRIRs) and VBAP is smaller when the sources are located at a true speaker position.

ACKNOWLEDGEMENTS

This work was funded by CTI grant Nr. 25276.2 PFLS-LS.

REFERENCES

[1] Durlach, NI; Thompson, CL; Colburn, HS. Binaural interaction in impaired listeners: A review of past research. Audiology journal, vol 20 (3), pp 181-211

[2] Best, V.; Kalluri, S.; McLachlan, S.; Valentine, S.; Edwards, B.; and Carlile, S. A comparison of CIC and BTE hearing aids for three-dimensional localization of speech, International Journal of Audiology, Vol 49 (10), 2010, pp 723-732.

[3] Oreinos, C.; Buchholz, J. M. Evaluation of Loudspeaker-Based Virtual Sound Environments for Testing Directional Hearing Aids. Journal of the American Academy of Audiology, Vol 27 (7), 2016, pp 541-556.

[4] Oreinos, C.; Buchholz, J. M. Objective analysis of higher-order Ambisonics sound-field reproduction for hearing aid applications. Proceedings of Meetings on Acoustics, 2013.

[5] Minnaar, P.; Albeck, S. F.; Simonsen, C. S.; Søndersted, B.; Oakley, S. A. G.;Bennedbæk, J. Reproducing Real-Life Listening Situations in the Laboratory for Testing Hearing Aids, Audio Engineering Society 135th Convention, 2013.

[6] Grimm, G.; Ewert, S.; Hohmann, V. Evaluation of Spatial Audio Reproduction Schemes for Application in Hearing Aid Research, Acta Acustica united with Acustica, Vol 101 (4), 2015, pp 842-854.

[7] Cubick, J.; Dau, T.; Validation of a Virtual Sound Environment System for Testing Hearing Aids, Acustica United with Acta Acustica, Vol 102 (3), 2016, pp 547-557.

[8] Simon, S. R. L.; Wuethrich, H.; Dillier, N. Comparison of Higher-Order Ambisonics, Vector- and Distance-Based Amplitude Panning using a hearing device beamformer, 4th International Conferenceon Spatial Audio, 2017.

[9] Simon, S. R. L.; Wuethrich, H.; Dillier, N. Comparison of 3D audio reproduction methods using a hearing device beamformer, J. of the Acoust. Soc. of Amer., submitted.

[10] Shaw, E. A. G. Ear canal pressure generated by a free sound field, J. of the Acoust. Soc. of Amer., Vol 39 (3), pp 465-470