DOI: 10.61782/fa.2023.1103

forumacusticum 2023

LOCALIZATION ERRORS IN BINAURAL REPRODUCTION OF FIRST AND THIRD-ORDER AMBISONIC RECORDINGS

Agnieszka Paula Pietrzak[1*] **Amaia Sagasti**[2] **Ricardo San Martin**[3]
Ruben Eguinoa[3]

[1] Institute of Radioelectronics and Multimedia Technology, Warsaw University of Technology, Poland
[2] Department of Information and Communications Technologies, Pompeu Fabra University, Spain
[3] Science Department, Public University of Navarre, Spain

ABSTRACT

It may be expected that high spatial resolution of a sound image obtained in higher order ambisonics should lead to smaller localization errors compared to first order. However, with binaural reproduction of ambisonic sound, the listener's ability to precisely localize the sound source may be reduced by the inaccuracies and spatial distortions made by the binaural decoders. This study compares the localization errors that occur when listening to pink noise bursts recorded with the use of first and third-order ambisonic microphones and presented binaurally through the earphones. It is discussed how in such a case localization errors vary depending on the ambisonics order and the type of the binaural renderer used. Localization errors were measured separately in azimuth and elevation. The median azimuth error was 25° for first-order and 20° for third-order recordings, with the front-back confusions excluded from the analysis. In elevation the median error for first-order was 65° and for third-order it was 100°, but it should be noted that in this plane front-back and up-down confusions were not omitted. Obtained results indicate no statistically significant differences between the four binaural decoders used.

Keywords: *localization error, ambisonics, FOA, HOA, binaural decoders*

Corresponding author: agnieszka.pietrzak@pw.edu.pl

1. INTRODUCTION

Ambisonics is a full-sphere audio format in which recording and reproducing spatial sound is based on the spherical harmonics [1]. Spatial resolution of sound recorded by ambisonics microphone depends on the number of spherical harmonics, and the higher the order of the microphone, the higher the number of capsules and the more harmonics it records. It can be supposed that using the higher order microphone will result in better localization accuracy in the reproduced spatial sound [2]. However, to reproduce the ambisonics sound, it has to be decoded from B-format, either for a set of speakers, or to a binaural sound when using headphones, and the decoding may cause spatial distortions and inaccuracies [3]. For binaural reproduction, decoding is most often based on the use of Head-Related Transfer Function (HRTF) [4, 5], which is usually not personalized for the particular listener, which can lead to significant localization errors [6].
The aim of this study was to measure the localization error that occurs for the binaural reproduction of sound recorded by 1st and 3rd order ambisonic microphone and to check if the use of different binaural renderers causes significant changes in the error values.

2. METHOD

Listening tests were conducted for 16 participants. The group consisted of telecommunication engineering students, 5 women and 11 men, between the ages of 22 and 26, with no self-reported hearing loss. Participants had no previous experience in spatial sound listening tests. The test material were 1st and 3rd order ambisonics recordings of pink noise bursts, recorded in the anechoic chamber with Sennheiser Ambeo (1st order) and Zylia (3rd order). The recorded angles were: -105°, -60°, 0°, 90°, 135° in azimuth (with 0° in front of the microphone, negative values on the right and positive on the left), and -45°, 30°, 90°. 135°, 180° in elevation (with 0° in front of the microphone, negative values meaning below and positive above).

The B-format was obtained using Ambeo A-B Format Converter and Zylia Ambisonic Converter. Then, B-format was decoded to binaural format using either IEM Binaural Decorder [7], CroPaC Binaural [9], HO-DirAC Binaural [10] or SPARTA AmbiBIN [11] plugins. Each plugin was used with its default settings for 1st or 3rd ambisonics order. Default HRIR's were used. Headphone compensation filters were used if available in the binaural plugin, and it was the case for IEM Binaural Decoder. For the other three there was no headphones compensation. Audio-technica ATH-M50x headphones were used for all tests. Listeners were introduced to four tasks, each for one plugin. The task consisted of 2 parts, regarding horizontal and median plane. For each plane 10 pink noise samples were presented, 5 for 1st and 5 for 3rd order recordings. Every sample was repeated twice. Participants were informed about the specifics of the task, but there was no training session and no feedback, as the aim of this study was to reflect the performance of the listeners unexperienced in localizing sound sources in a listening tests. They were asked to answer where from the sound is coming and they marked their answers graphically on a diagram for azimuth or elevation.

3. RESULTS

Localization errors were calculated as the absolute error between the presented angle and the answer given by the respondent. Data was analyzed for every presented angle, with four renderers, regarding the azimuth and elevation, for first and third-order ambisonic recordings.

Obtained results are presented separately for azimuth (Fig. 1) and elevation (Fig. 2). Median error, 25th and 75th percentile, maximum, minimum and outlier values are shown in the box-and-whiskers graphs. For analyzing the first and third order differences, the rendered type is temporarily disregarded and will be analyzed later.

Azimuth errors (Fig. 1) are the highest for 0°. Median error is 180° for first-order and 150° for third-order recording. This is due to front-back confusions, and it can be seen, that statistically for third-order the confusions occur less frequently. Due to the distinct character of this front-back errors, results for 0° azimuth are omitted in the further analysis. Localization errors in the horizontal plane are the smallest for angles -105 and 90, as expected, because of the highest interaural time differences (ITD) and interaural level differences (ILD). For -105° azimuth, median error is 15° for both 1st and 3rd order, and for 90° azimuth median error is 20° for 1st and 10° for 3rd order. In the case of 60° azimuth median error is by 20° smaller for 3rd order microphone (median error 30°) compared to 1st order (50°). For 135° azimuth, median error is higher for the 3rd order (40°) than for 1st order (25°).

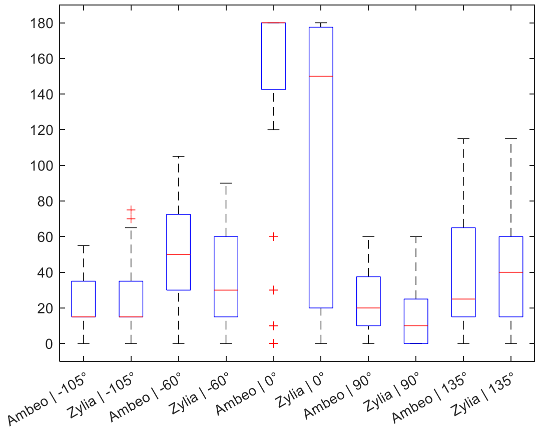

Figure 1. Comparison of localization errors (angle in degrees) in horizontal plane (azimuth) for recordings made with 1st order (Ambeo) and 3rd order (Zylia) ambisonic microphones, regardless of the binaural decoder used.

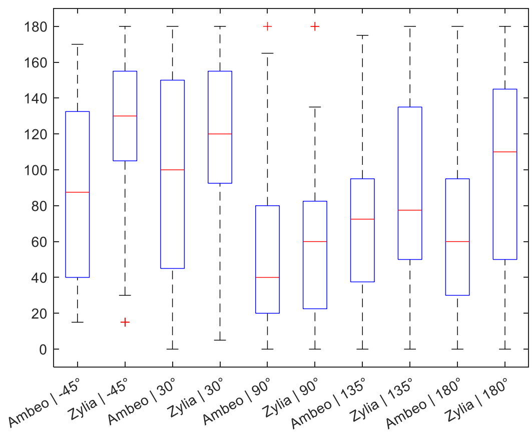

Figure 2. Comparison of localization errors (angle in degrees) in median plane (elevation) for recordings made with 1st order (Ambeo) and 3rd order (Zylia) ambisonics microphones, regardless of the binaural decoder used.

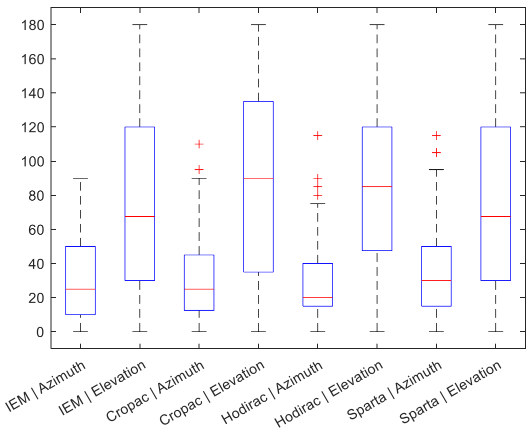

Figure 3. Localization errors for four binaural decoders: IEM Binaral Decoder, CroPaC Binaural, HO-DirAC Binaural and SPARTA AmbiBIN, in azimuth and elevation, regardless of the ambisonics order.

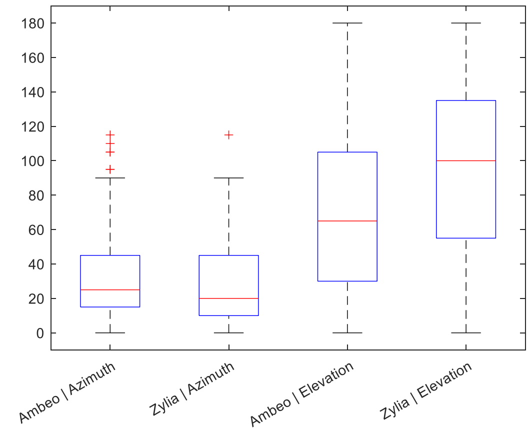

Figure 4. Comparison of the localization errors for 1st (Ambeo) and 3rd order (Zylia) ambisonics recordings in azimuth and elevation, regardless of the binaural decoder used.

For the median plane (Fig. 2) the errors are higher than for the horizontal plane. Median elevation error is between 40° and 130°. The high values of this errors are most likely connected with front-back and up-down confusions. As it was done for the horizontal plane, such confusions should be excluded from the analysis. However, as the errors occurred for all the presented angles, and were often not exactly 180°, it was decided to analyze all the median plane results, including all this confusions, to show the overall ability of the listener to correctly localize the source. Surprisingly, for all tested angles, localization error was higher for 3rd order than for 1st. The error was the smallest for 90° elevation, it was 40° for 1st and 60° for 3rd order. The highest errors were observed for angles near the front, it was 130° for -45° elevation and 120° for 30° elevation, and it is once again probably due to the front-back and up-down confusions.

Results obtained for four used binaural decoders (IEM Binaral Decoder, CroPaC Binaural, HO-DirAC Binaural and SPARTA AmbiBIN) are presented in Fig. 3. The one-way analysis of variance (ANOVA) was used to determine whether there are any statistically significant differences between the four used binaural decoders. Results indicated that globally there was no statistically significant difference, neither for azimuth ($F(3,508) = 1.41$, $p = 0.239$) nor elevation ($F(3,636) = 2.28$, $p = 0.078$).

Summary results comparing localization errors for first and third-order recordings are presented in Fig. 4. In azimuth the median error was 25° for first-order and 20° for third-order. In elevation the median error for first-order was 65° and for third-order it was 100°.

4. CONCLUSIONS

Localization of sound sources in binaural reproduction of ambisonic sound can be difficult for the listeners due to spatial distortions occurring on every step of recording and decoding process. In this study localization errors were measured for first and third-order ambisonic recordings, decoded with four binaural plugins.

Obtained results indicate no statistically significant differences between the four binaural decoders used. In the horizontal plane, the median azimuth error was 25° for first-order and 20° for third-order recordings, with the front-back confusions excluded from the analysis. This is consistent with the results found in the literature, e.g. [12, 13]. In [12] these given errors are 26° for first-order and 22° for third-order, and in [13] it is 30° and 23°, respectively.

In the median plane, the analysis was harder to interpret, as there were many front-back and up-down confusions that were not obvious to exclude from the analyzed data. It was decided to calculate the errors for the entire dataset. Taking this assumptions into account, the median elevation error for first-order was 65° and for third-order it was 100°. Surprisingly, even given the confusions, it was expected that localization error would be smaller for third-order recordings. However, the obtained data indicate that the error for third-order recording was by 35° greater than for first-order. The explanation of this results may lay in the recording process and the inaccuracies occurring when using the 3rd order Zylia microphone array. As no reference data was found for this microphone, this issue will be further investigated.

5. REFERENCES

[1] F. Zotter, M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, vol. 19. Springer International Publishing, 2019.

[2] S. Bertet, J. Daniel, E. Parizet, i O. Warusfel, „Investigation on Localisation Accuracy for First and Higher Order Ambisonics Reproduced Sound Sources", *Acta Acustica united with Acustica*, vol. 99, no. 4, pp. 642–657, 2013.

[3] A. Sontacchi, P. Majdak, M. Noisternig, i R. Höldrich, „Subjective Validation of Perception Properties in Binaural Sound Reproduction Systems", *Audio Engineering Society Conference: 21st International Conferencespart,* 2002.

[4] B. Xie, *Head-related transfer function and virtual auditory display*., J. Ross Publishing, 2013.

[5] Z. Ben-Hur, D. Alon, R. Mehra, i B. Rafaely, „Binaural Reproduction Based on Bilateral Ambisonics and Ear-Aligned HRTFs", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021.

[6] E. M. Wenzel, „Localization using nonindividual head-related transfer functions", *J. Acoust. soc. Am.*, vol. 94, no. 1, pp. 858–867, 1993.

[7] Zaunschirm, Markus, Christian Schörkhuber, and Robert Höldrich. "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint." *J. Acoust. soc. Am.*, vol. 143, no. 6, pp. 3616-3627, 2018.

[8] Schörkhuber, Christian, Markus Zaunschirm, and Robert Höldrich. "Binaural rendering of ambisonic signals via magnitude least squares." *Proceedings of the DAGA*. vol. 44, pp. 339-342, 2018.

[9] L. McCormack, and S. Delikaris-Manias, 2019. Parametric First-order Ambisonic Decoding for Headphones Utilising the Cross-Pattern Coherence Algorithm, *Proceedings of the 1st EAA Spatial Audio Signal Processing Symposium*, (Paris, France), 2019.

[10] A. Politis, L. McCormack, V. Pulkki, (2017, October). Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* pp. 379-383, 2017.

[11] L. McCormack, A. Politis, SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods, *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, 2019.

[12] B. Mróz, M. Kabaciński, T. Ciotucha, A. Rumiński and T. Żernicki, "Production of six-degrees-of-freedom (6DoF) navigable audio using 30 Ambisonic microphones," *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Bologna, Italy, pp. 1-5, 2021.

[13] Thresh, L., Armstrong, C., & Kearney, G. "A direct comparison of localization performance when using first, third, and fifth ambisonics order for real loudspeaker and virtual loudspeaker rendering", *143. Audio Engineering Society*, 2017.