VPS VIBRATIONS IN PHYSICAL SYSTEMS

DOI: 10.21008/j.0860-6897.2024.2.03
Vibrations in Physical Systems, 2024, 35(2), 2024203

Azimuth and elevation errors in binaural reproduction of ambisonics sound

Agnieszka Paula PIETRZAK

Warsaw University of Technology, Faculty of Electronics and Information Technology, Institute of Radioelectronics and Multimedia Technology, ul. Nowowiejska 15/19, 00-665 Warsaw

Corresponding author: Agnieszka Paula PIETRZAK, email: agnieszka.pietrzak@pw.edu.pl

Abstract Binaural decoding of an ambisonic sound is reproducing the information about a soundfield over headphones. It is done based on the spherical harmonics representation of the spatial sound and on the use of Head Related Transfer Function (HRTF). Inaccuracies in the decoding process, which can be caused for example by using non-personalized HRTF, may lead to difficulties in localizing the sound source by the listener. Especially in the elevation plane, localization errors can be significant. In this study, listening tests were conducted for naive listeners in order compare azimuth and elevation errors for first and third-order ambisonics recordings of pink noise bursts recorded in anechoic chamber. The tests involved 16 participants who used headphones to listen to ambisonic sound recordings, specifically bursts of pink noise, captured using Sennheiser Ambeo (1st order) and Zylia (3rd order) microphones. These recordings were then converted to B-format and decoded into binaural format for the listening tests. In the azimuth plane, the highest errors occurred at 0°, indicative of front-back confusions. In contrast, the elevation plane exhibited generally larger errors, with third-order ambisonics resulting in notably higher errors compared to first-order.

Keywords: ambisonics, FOA, HOA, localization error, listening test.

1. Introduction

The faithful reproduction of sound source directionality in three-dimensional space is a fundamental pursuit in the domain of spatial audio [1]. The pursuit of immersive auditory experiences has driven innovation in the field of spatial audio, and ambisonics is part of this development. Ambisonics is an audio format based on spherical harmonics [2], and it can record and reproduce three-dimensional soundscapes, providing both azimuthal (horizontal) and elevational (vertical) sound localization cues, which are fundamental for achieving convincing spatial fidelity [3]. Azimuthal localization, which refers to the precise determination of a sound source's left-right position, has been a central focus in the development of ambisonic technologies. However, the immersive auditory experience depends not only on azimuthal accuracy but also on elevational precision. Elevational cues, responsible for conveying the perceived height of sound sources, play an important role in creating a truly three-dimensional auditory environment [4-6].

While ambisonics has exhibited considerable potential in encoding the spatial cues, the process of recording and decoding the audio signal from the multiple capsule microphones to the binaural reproduction has its challenges. Recording ambisonic sound involves capturing the complete three-dimensional soundscape, which requires precision. One challenge lies in the placement and synchronization of multiple microphones, typically arranged in a spherical array. These microphones must be precisely positioned to capture sound from all directions, and their signals must be synchronized to maintain coherence during the recording process. Calibration and synchronization errors can lead to inaccuracies in the captured spatial information. This can complicate the process of faithfully representing the true spatial characteristics of the sound source.

Decoding these recordings to B-format involves complex signal processing to accurately translate raw microphone signals into a 3D sound field representation. This process must ensure compatibility with various playback systems, from headphones to multi-speaker setups, which can be challenging. As the ambisonic order increases, the spatial complexity of the recording also rises. Higher-order ambisonics can represent sound sources with intricate spatial characteristics. However, this complexity introduces challenges in accurately decoding the elevation and azimuth cues. The problems with the recording and decoding process manifest frequently as azimuthal and elevational errors, causing difficulties with listener's

perceptual localization of sound sources [7]. Analyzing and quantifying these errors is important for the ongoing improvement of ambisonic technologies.

The process of reproducing ambisonics sound from B-format to binaural is a critical step [8]. Binaural reproduction of ambisonics relies on the use of spherical harmonics to convey directional sound source information and on the Head-Related Transfer Functions (HRTF) [9] to emulate how listeners perceive sounds from various angles [10, 11]. However, HRTFs are highly individualized, and using generic HRTFs can lead to less accurate spatial perception for the listener [12]. The use of standardized HRTFs within the decoding algorithms, as opposed to the personalized HRTFs of individual listeners can lead to front-back and up-down confusions [13].

The aim of this study was to investigate and quantify the azimuthal and elevational errors that occur during the binaural reproduction of ambisonic sound, and check how much the errors differ for first and third-order ambisonic microphones. with a focus on comparing these errors between first and third-order ambisonic microphones. This comparison is essential to understand the effectiveness of ambisonic technologies in spatial audio reproduction, particularly for naive listeners - those without specialized training in auditory localization. The study sought to determine how accurately listeners can localize the sound sources in both the horizontal and vertical planes when experiencing binaurally reproduced ambisonic recordings. This evaluation is crucial in assessing the spatial resolution of ambisonic technologies and their ability to create realistic audio environments. A key question was whether the enhanced detail in higher-order ambisonics effectively translates into better localization accuracy, or if it introduces complexity that might affect the listener's spatial perception, especially when using generic HRTFs.

2. Methodology

The study involved a listening test that included 16 participants, non-expert (naive) listeners, selected to provide an unbiased representation of the general population's ability to localize sound sources in a binaural context. Before the commencement of the listening tests, participants underwent a brief hearing screening to ensure normal hearing levels. This step was crucial to eliminate any biases in the results due to hearing impairments. The participants then received an orientation session, where the purpose of the study and the nature of the listening tests were explained in detail. However, they were not provided with any specific training or feedback regarding sound localization, to maintain their naive status in terms of spatial audio perception.

These participants used headphones to listen to binaural reproductions of ambisonic sound recordings. The sound recordings were captured in an anechoic chamber. This setting was chosen to ensure the elimination of external auditory cues and reverberations that could influence the sound localization proces. Two microphones were used: Sennheiser Ambeo (1st order) and Zylia microphone (3rd order). The recorded material consisted of bursts of pink noise recorded in ambisonic format. Pink noise was generated by playing it through speakers positioned around the microphone from various angles.

In the horizontal plane, the angles covered ranged from -105° to 135°, with 0° representing the front of the microphone. Negative values indicated angles to the right, and positive values indicated angles to the left. In the vertical plane, angles ranged from -45° to 180°, with 0° again indicating the front of the microphone, negative values represented angles below the microphone, while positive values represented angles above it.

The recorded A-format material was then converted to B-format using the Ambeo A-B Format Converter [14] for Ambeo and the Zylia Ambisonic Converter [15] for the Zylia microphone. Subsequently, for both Ambeo and Zylia recordings, the B-format audio was decoded into binaural format using default decoder settings. Audio-Technica ATH-M50x headphones were used for the binaural reproduction. The participants were presented with the binaural audio separately for azimuth (horizontal plane) and elevation angles (median plane), and they were given an audio introduction to explain the test. The tested angles were:

- -105°, -60°, 0°, 90°, 135° (azimuth),
- -45°, 30°, 90°, 135°, 180° (elevation).

For each plane (azimuth and elevation), 10 pink noise samples were presented, with 5 samples from 1st order recordings and 5 samples from 3rd order recordings. Each sample was played twice to ensure consistency and reliability of the responses. Listeners were required to indicate the source's perceived direction and mark their answers graphically on specifically designed diagrams that corresponded to the azimuth or elevation angles. This graphical approach to recording responses was selected for its intuitiveness and ease of understanding for participants.

3. Results

Obtained data was analyzed separately for horizontal and median plane, and for first and third-order. Localization errors were calculated as the absolute error between the presented angle and the answer given by the respondent. Median of the data obtained for every presented angle is presented in Tab. 1.

Table 1. Median localization error for presented angles in azimuth and elevation.

Azimuth:	-105°	-60°	0°	90°	135°
First-order (Ambeo)	15°	50°	180°	20°	25°
Third-order (Zylia)	15°	30°	150°	10°	40°
Elevation:	-45°	30°	90°	135°	180°
First-order (Ambeo)	88°	100°	40°	73°	60°
Third-order (Zylia)	130°	120°	60°	78°	110°

In the horizontal plane, the highest errors occur for 0°. Large localization errors occurring for 0° can be considered front-back confusions, which can arise during the binaural reproduction of ambisonic sound due to distortion of subtle spatial cues and due to variations in individual head-related transfer functions (HRTFs). Front-back errors can be observed particularly in the first-order Ambeo recordings, where the median localization error reached 180°. In contrast, third-order Zylia recordings showed a reduced median error of 150° at this angle, suggesting that higher-order recordings might mitigate front-back confusions to some extent, although they still remain a challenge.

The localization errors were generally smaller for angles where interaural time differences (ITD) and interaural level differences (ILD) were more pronounced, specifically at -105° and 90°. Both first and third-order recordings showed relatively lower median errors at these angles, with the third-order recordings demonstrating more consistent performance across these angles.

Localization errors for median place are higher than for the horizontal plane. Median error varies between 40° and 130°. Generally, localization errors in elevation are often larger than those in azimuth, as, unlike ITD and ILD cues for azimuth, elevation localization relies on more complex mechanisms, such as spectral cues, which are less precise and can be easily influenced by factors like the use of non-personalized HRTF or the frequency response of the headphones used. Surprisingly, elevation errors in third-order ambisonics are notably higher than those in first-order ambisonics.

More specific analysis on the localization errors can be made using histograms of the obtained data, to visualize the distribution of errors, with histograms created separately for azimuth and elevation errors. Focusing first on the horizontal plane (Fig. 1), both first-order and third-order ambisonics exhibit a similar right-skewed distribution. This pattern implies that neither consistently outperforms the other when it comes to azimuth error, suggesting that, in general, for both orders of ambisonics there is some degree of difficulty in accurately localizing sound sources in the horizontal plane. The main difference appears to be that third-order ambisonics have a somewhat narrower distribution with fewer extreme errors compared to first-order ambisonics.

For third-order ambisonics there are less cases with high errors compared to first-order ambisonics. For third-order recordings front-back confusions occurred less frequently. The distribution of the errors allows to easily distinguish the front-back confusions (160-180°), and therefore, taking under consideration the specific nature of those errors, this data can be excluded from further analysis.

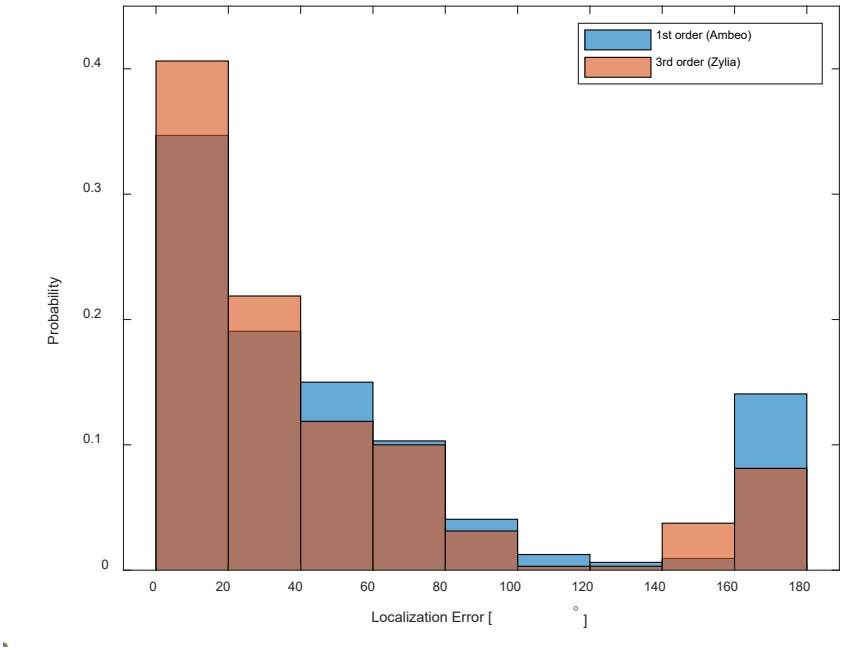

Figure 1. Histogram of localization errors in azimuth for first-order Ambeo (blue) and third-order Zylia (orange) recordings. The brown color indicates the overlap of data.

Analysing the median plane (Fig. 2), a contrast can be seen between the first-order and third-order ambisonics. The histogram representing first-order ambisonics in the elevation domain reveals elevation errors varying across a wide range, indicating that elevational errors vary significantly across different sound source positions. For first-order the distribution is right-skewed, as it was for the horizontal plane, meaning low errors occurring more frequently that high ones. However, for third-order ambisonics in the median plane, the distribution is more left-skewed, meaning most of the elevation errors for third-order ambisonics tend to be higher compared to the fewer instances where they are low.

Data obtained for elevation errors indicate, that for this experiment setup, localizing the sound source was harder for third-order recordings, compared to first-order. This can occur due to several factors. Third-order ambisonics capture a more detailed spatial soundscape compared to first-order. They incorporate higher-order spherical harmonics, which can represent sound sources with greater complexity. However, this complexity can introduce challenges in accurately reproducing elevation cues, as the relationship between the acoustic characteristics and perceived elevation becomes more intricate.

Another contributing factor is that elevation cues are particularly sensitive to HRTF quality and individual anatomical variations. Inadequate modeling of HRTFs, especially when applied to third-order ambisonics, can amplify elevation errors significantly. Consequently, the influence of the complexity of higher-order ambisonics and the precision required for elevation localization can lead to elevated errors, as observed in the results.

For third-order recordings, localization errors of 120-140 and 160-180 occurred with notable frequency in the median plane. In the horizontal plane (Fig. 1), front-back confusions are relatively straightforward to identify, as the errors predominantly fall within certain angular ranges (160-180°) indicating confusion between the front and back. However, in the median plane, the errors are more evenly distributed, making it challenging to distinguish typical front-back or up-down confusions. This intricacy in error patterns emphasizes the complexity of spatial perception in higher-order ambisonic sound reproduction.

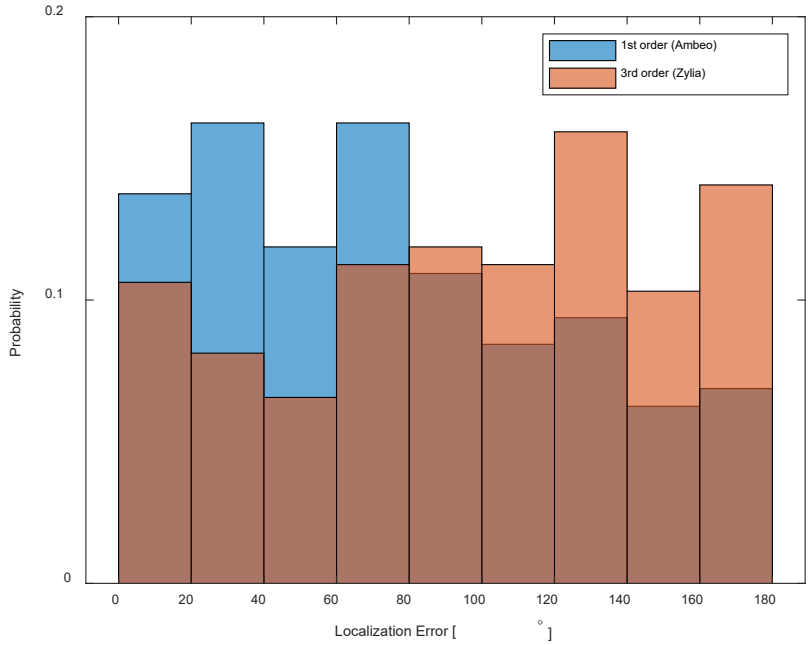

Figure 2. Histogram of localization errors in elevation for first-order Ambeo (blue) and third-order Zylia (orange) recordings. The brown color indicates the overlap of data.

Obtained results can be analyzed statistically based on the box-and-whiskers graph (Fig. 3) [16]. Median error, 25th and 75th percentile, maximum, minimum and outlier values are presented for localization errors in azimuth and elevation for both first-order (Ambeo) and third-order (Zylia) recordings. For azimuth, the front-back confusions are excluded from the data.

In the case of azimuth errors, third-order ambisonics generally exhibits lower errors compared to first-order ambisonics. Remembering this is considering front-back confusions excluded from the data, median localization error for first-order ambisonics is 25°, and for third-order it is 20°. This difference suggests that higher-order ambisonic recordings can potentially offer more precise azimuth localization, likely due to their enhanced spatial resolution capabilities. This precision becomes particularly significant when considering the subtle nuances of sound localization that are vital for immersive audio experiences.

Median elevation for first-order is 65° and for third-order it is 100°. This really high errors are probably related to many front-back and up-down confusions, which were hard to distinguish and exclude. The intricacy of these errors speaks to the potential complexities introduced by higher-order ambisonics in the accurate reproduction of elevation cues.

Examining the outliers in azimuth localization, as depicted in Fig. 3, we notice that they are particularly prevalent in first-order ambisonics. These outliers are significantly higher than the median values and point towards instances where the localization errors deviate from the typical range. Such deviations necessitate a closer examination to uncover the underlying reasons, which could range from variations in recording conditions to the characteristics of the audio equipment used.

These results indicate that while higher-order ambisonics capture more spatial details, this does not straightforwardly translate into improved localization accuracy in binaural reproduction, especially in the elevation plane. The additional spatial detail captured by higher-order ambisonics does not directly equate to improved localization accuracy in the elevation plane during binaural reproduction.

The findings suggest that factors such as the complexity of the sound field and the quality of HRTF modeling play a significant role in the perceived accuracy of spatial audio. This underscores the importance of considering these factors in the development and implementation of ambisonic technologies, particularly for applications aimed at naive listeners. The study implies that to optimize the spatial audio experience, a balance must be struck between the sophistication of the technology and its practical application in various listening scenarios.

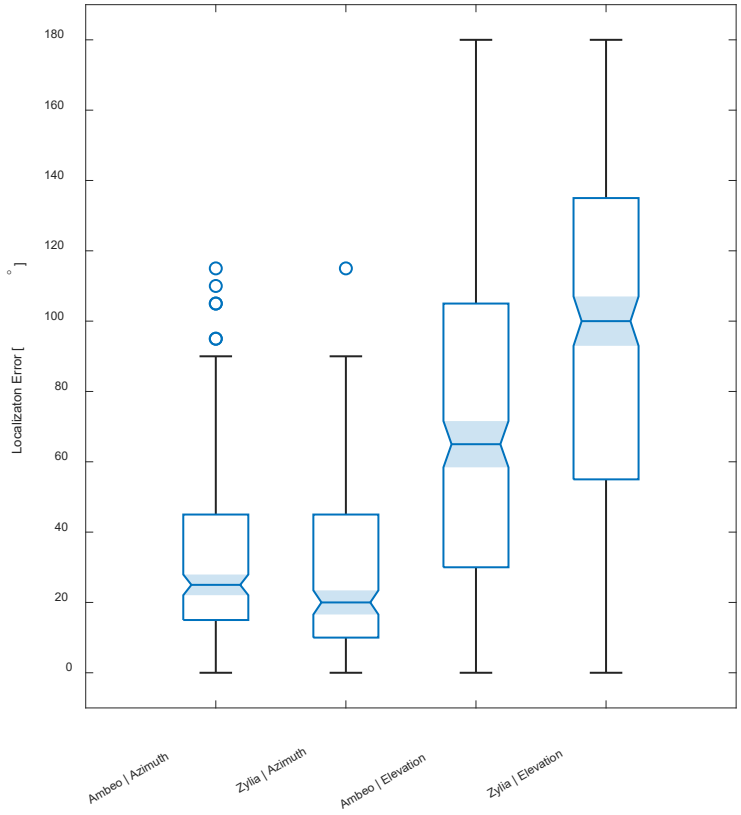

Figure 3. Localization errors in azimuth and elevation for first-order (Ambeo) and third-order (Zylia) recordings.

4. Conclusions

In conclusion, this study investigates the azimuth and elevation errors in binaural reproduction of ambisonic sound, analysing how different recording orders impact sound localization. Azimuthal accuracy, which relates to left-right sound source positioning, has traditionally been a focal point in the development of ambisonics. However, this research illuminated that elevational precision, responsible for conveying sound source height, is equally important in analysing the localisation accuracy. It's this vertical dimensionality that imparts the perception of being enveloped by sound, a quality that is as crucial as the horizontal in creating a fully immersive auditory experience [2, 17].

Localization errors in elevation were found to be notably higher than those in azimuth. The errors were especially high in third-order ambisonics recordings made with Zylia ZM1 microphone. This discrepancy is attributed to several factors, including the complexity of higher-order spherical harmonics used in third-order recordings, which can introduce challenges in accurately reproducing elevation cues [18, 19].

Moreover, the accuracy of elevation localization heavily relies on Head-Related Transfer Function (HRTF) quality and individual anatomical variations. Using a non-individualized HRTF modeling can lead to significant discrepancies in perceived elevation, pointing to the necessity of using personalized HRTF to minimize such errors [12, 20].

Another factor is the listener familiarity and experience with spatial audio. In this study the listeners were unexperienced, so untrained in the nuances of spatial listening. It could have influenced their ability to distinguish sources. Novice listeners may experience more front-back and up-down confusions compared to experienced listeners who have developed better spatial listening skills [3, 21].

To address front-back confusions in binaural reproduction of ambisonic sound, ongoing research focuses on improving HRTF personalization, developing more accurate spatial audio rendering algorithms, and enhancing listener training and education in spatial audio perception. Such advancements are anticipated to not only benefit the naive listener but also to refine the spatial listening capabilities of all users [22, 23].

The study revealed that in azimuth, third-order ambisonics exhibited lower errors compared to first-order, indicating their potential to provide more accurate azimuth localization. Nonetheless, this research emphasizes the importance of further refining HRTF modeling and addressing localization challenges, ultimately enhancing the quality of immersive auditory experiences in ambisonic sound reproduction.

Additional information

The author(s) declare: no competing financial interests and that all material taken from other sources (including their own published works) is clearly cited and that appropriate permits are obtained.

References

1. K. Pulkki, V. Pulkki; Virtual Sound Source Positioning Using Vector Base Amplitude Panning; Journal of the Audio Engineering Society, 2023, 45(6), 456-466
2. F. Zotter, M. Frank; Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality; t. 19. Springer International Publishing, 2019; DOI: 10.1007/978-3-030-17207-7
3. J. Blauert; Spatial Hearing: The Psychophysics of Human Sound Localization; Cambridge MA, USA, MIT Press, 1997
4. G. Kearney, T. Doyle; Height perception in Ambisonic based binaural decoding; Audio Engineering Society Convention 139, Audio Engineering Society, New York, October 2015
5. M. Gorzel, G. Kearney, C. Masterson, H. Rice, F. Boland; On the perception of dynamic sound sources in ambisonic binaural renderings; Audio Engineering Society 41st International Conference: Audio for games, London, 2011
6. M.A. Gerzon; Periphony: With-Height Sound Reproduction; Journal of the Audio Engineering Society, 2023, 21(1), 2-10
7. A. Sontacchi, P. Majdak, M. Noisternig, R. Höldrich; Subjective Validation of Perception Properties in Binaural Sound Reproduction Systems; Audio Engineering Society 21st International Conference: Architectural Acoustics and Sound Reinforcement, Audio Engineering Society, St. Petersburg, 2002.
8. C. Gribben, E. M. Wenzel; Advanced Techniques in Binaural Audio Processing and Spatial Hearing; Journal of the Acoustical Society of America, 2022, 130(4), 2334-2346
9. B. Xie; Head-related transfer function and virtual auditory display; J. Ross Publishing, 2013
10. Z. Ben-Hur, D. Alon, R. Mehra, B. Rafaely; Binaural Reproduction Based on Bilateral Ambisonics and Ear-Aligned HRTFs; IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 21, 901-913; DOI: 10.1109/TASLP.2021.3055038
11. K. Stanney ed.; Handbook of Virtual Environments: Design, Implementation, and Applications; CRC Press, 2023
12. P. Majdak, M. Goupell, B. Laback; 3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training; Attention, Perception, & Psychophysics, 2010, 72(2), 454-469
13. S. N. Yao.; L. J. Chen; HRTF adjustments with audio quality assessments; Archives of Acoustics, 2013, 38(1), 55-62
14. AMBEO A-B FORMAT CONVERTER; https://www.sennheiser-sites.com/responsive-manuals/AMBEO_VR_MIC/EN/index.html#page/AMBEO%20VR%20Mic/VR_MIC_04_Software_EN.4.1.html (accessed on 5 September 2023)
15. ZYLIA Ambisonics Converter, ZYLIA - 3D AUDIO RECORDING & POST-PROCESSING; https://www.zylia.co/zylia-ambisonics-converter.html (accessed on 5 September 2023)
16. A.P. Pietrzak, A. Sagasti, R. San Martin, R. Eguinoa; Localization Errors in Binaural Reproduction of First and Third-Order Ambisonic Recordings; Proc. Of 10th Convention of the European Acoustics Association; Turin, Italy, 11–15 September 2023
17. A. Kulkarni, H. Colburn; Role of Spectral Detail in Sound-Source Localization; Nature, 1998, 396, 747-749
18. G. Kearney et al.;Third-Order Ambisonics: A Study of the Advantages and Disadvantages for Spatial Audio Design and Auditory Display; Journal of the Audio Engineering Society, 2016, 64(4), 251-262
19. E.M. Wenzel et al.; Effect of HRTF Personalization on the Perception of Virtual Sounds; Journal of the Acoustical Society of America, 2003, 114(4), 2263-2271

20. S. Spagnol et al.; HRTF Measurement of Spherical Heads for Personalized Binaural Audio; IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(3), 549-560
21. D.R. Begault; 3-D Sound for Virtual Reality and Multimedia; Academic Press, 2000
22. Z. Ben-Hur et al.; Personalized HRTFs for Improved Spatial Perception in Virtual Environments; IEEE Transactions on Audio, Speech, and Language Processing, 2019, 27(5), 1036-1049
23. J.M. Algazi, R.O. Duda, D.M. Thompson, C. Avendano; The CIPIC HRTF Database; Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, 2001, 99-102