VPS VIBRATIONS IN PHYSICAL SYSTEMS
DOI: 10.21008/j.0860-6897.2022.2.14
Vibrations in Physical Systems, 2022, 33(2), 2022214

Localization of Sound Sources in Binaural Reproduction of First and Third Order Ambisonics

Amaia SAGASTI[1] , Agnieszka Paula PIETRZAK[2] , Ricardo SAN MARTIN[1] , Rubén EGUINOA[1]

[1] Public University of Navarre, The Higher Technical School of Industrial, Computing and Telecommunication Engineering, Cataluña Avenue, 31006, Pamplona, Navarre, Spain
[2] 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: The use of higher-order ambisonics in spatial sound recordings makes it possible to increase the accuracy of recording information about the direction from which the sound comes to the listener. However, with binaural ambisonic sound reproduction, the listener's ability to locate the sound source accurately may be limited. This paper presents a comparison of the listener's ability to locate a sound source during binaural listening to recordings made with first and third order ambisonic microphones. The analysis was carried out for two types of signal: pink noise and ringing sound. The analysis of localization errors depending on the ambisonics order, azimuth and elevation angles as well as the type of signal is presented. The obtained data indicate that in binaural reproduction of the ambisonic sound the localization errors in the azimuth plane were smaller for the third order ambisonics, compared to the first order. In the elevation plane both for first and third order the errors were significant.

Keywords: ambisonics, localization, spatial audio, listening tests.

1. Introduction

With the increasing popularity of the 360 video format, enabling immersion in virtual reality, it is also required to use an audio format that allows for a reliable reproduction of sound not only in the horizontal plane, but in the full sphere around the listener. Ambisonics is a full-sphere format in which recording and reproducing spatial audio is based on the spherical harmonics [1].

The spatial resolution of an ambisonic recording is dependent on the order, so the number of spherical harmonics. In first-order ambisonics (FOA) the information about the localization of the sound source is described by four channels called W, X, Y, Z. The W channel is the omnidirectional component, and X, Y, Z channels give information about the three dimensional (front-back, left-right, up-down) components of the spatial sound. Higher order ambisonics (HOA) have $K = (N+1)^2$ number of channels, where N is the ambisonic order. Each of the channels represents the coefficient of the series expansion of the sound field around the origin. 1st, 2nd and 3rd order ambisonics are the most common. The higher the ambisonic order, the better the spatial resolution of the recording and it can be supposed that using the higher order microphone will result in better localization accuracy in the reproduced ambisonic sound [2].

To reproduce the sound which was recorded in ambisonic B-format, it has to be decoded, either to a set of speakers, or to a binaural format when using headphones. The binaural reproduction of ambisonics gives the listener the sensation of being present in the spatial audio scene [3], [4], and it is achieved both by the use of the spherical harmonics, which carry the information about the original direction of the sound source, and by using Head Related Transfer Function (HRTF) [5], so the information about how the listeners perceives sounds coming from a given angle [6].

The aim of this study was to measure what is the localization accuracy for the sounds recorded by 1st and 3rd order ambisonic microphone and if the accuracy is better for the 3rd order ambisonics, given that decoding it for binaural reproduction may be causing distortions [7], as the HRTF used in the decoders is not the personal HRTF of the particular listener, but a default HRTF of a dummy head [8].

2. Methods

The study was conducted through listening test, with 19 participants. The participants were listening on headphones to binaural reproduction of ambisonic sound, recorded in anechoic chamber. Recordings were made using Sennheiser Ambeo (1st order) and Zylia (3rd order) microphones. The recorded material was two types of sound: the bursts of pink noise and a sound of a bicycle bell (chosen as a more natural sound), recorded in ambisonic format, generated through speakers placed around the microphone, from different angles: by every 15° from -135° to 135° in horizontal plane (with 0° in front of the microphone, negative values on the right and positive on the left), and by every 15° from -45° to 180° in vertical plane (0° in front of the microphone, negative values meaning below and positive above).

The recorded A-format material was decoded to B-format using Ambeo A-B Format Converter [9] for the Ambeo and Zylia Ambisonic Converter [10] for the Zylia microphone. Then, for both Ambeo and Zylia, B-format was decoded to binaural format using IEM Binaural Decoder [11]. The binaural reproduction was presented to the participants as four tasks, separately for pink noise and bicycle bell and separately for azimuth (horizontal plane) and elevation angles (vertical plane). For each task, there was an audio introduction, which gave the listeners the information about the test and also allowed to test if the right and left channel of the headphones were placed on the proper ear. Audio-technica ATH-M50x headphones were used. For each task there were ten angle samples (5 for Ambeo and 5 for Zylia), and each of them was repeated twice.

Tested angles:

- Pink noise – azimuth: -105°, -60°, 0°, 90°, 135°
- Pink noise – elevation: -45°, 30°, 90°. 135°, 180°
- Bicycle bell – azimuth: -135°, -90°. 15°, 75°, 120°
- Bicycle bell – elevation: -135°, -15°, 30°, 90°, 165°.

Participants were asked to answer where from the sound is coming and they marked their answers graphically on a diagram with azimuth or elevation plane, and also by writing down the angle in degrees.

3. Results

Answers given by test participants are presented as polar graphs which show how many times a particular angle was pointed as the localization of the sound source. The data was analyzed separately for the azimuth and the elevation plane.

One way to analyze the data is looking at the most common answer. For pink noise bursts results regarding the azimuth plane are shown in Fig. 1. The most accurately recognized localization was the 90° angle (Fig. 1d). In this case, for samples recorded with 3rd order Zylia, 8 participants gave the exactly correct answer. For this particular angle, using 3rd order Zylia resulted in far more accurate answers than the 1st order Ambeo. Next most recognizable angle was -105° (Fig. 1a) for which 3 people gave the answer with only 5° error, and all these three cases were also for the Zylia recording.

For -60° and 135° (Fig. 1b, e) the most repeated answers were also for Zylia and they were off by 30° for the -60° sample, and off by 0-60° for the 135° sample. For 135° there were no right-left misjudgments. Comparing to 3rd order Zylia, the results for 1st order Ambeo are more scattered.

For the 0° angle (Fig. 1c) it can be clearly seen that both for Zylia and Ambeo there was a confusion whether the sound is coming from the front or back. The most common answers indicate that sound source was at the back. That is probably due to the use of a dummy head HRTF in the Binaural Decoder plugin.

Regarding localization of a bicycle bell sound in azimuth plane (Fig. 2) answers were more scattered than for pink noise. For -90° (Fig. 2b) the most common answer was 10° off the correct angle for Zylia and 30° off for Ambeo. For -135° sample (Fig. 2a), the most common answer for Zylia is the closest to the correct angle but off by 15°.

In the case of 75° and 120° samples (Fig. 2d,e) for Zylia the most common answers are off by 30° and more. The error is smaller for Ambeo, but the answers are less consistent. For 15° (Fig. 2c) the front-back effect is also visible for the answers regarding especially Ambeo, but also Zylia.

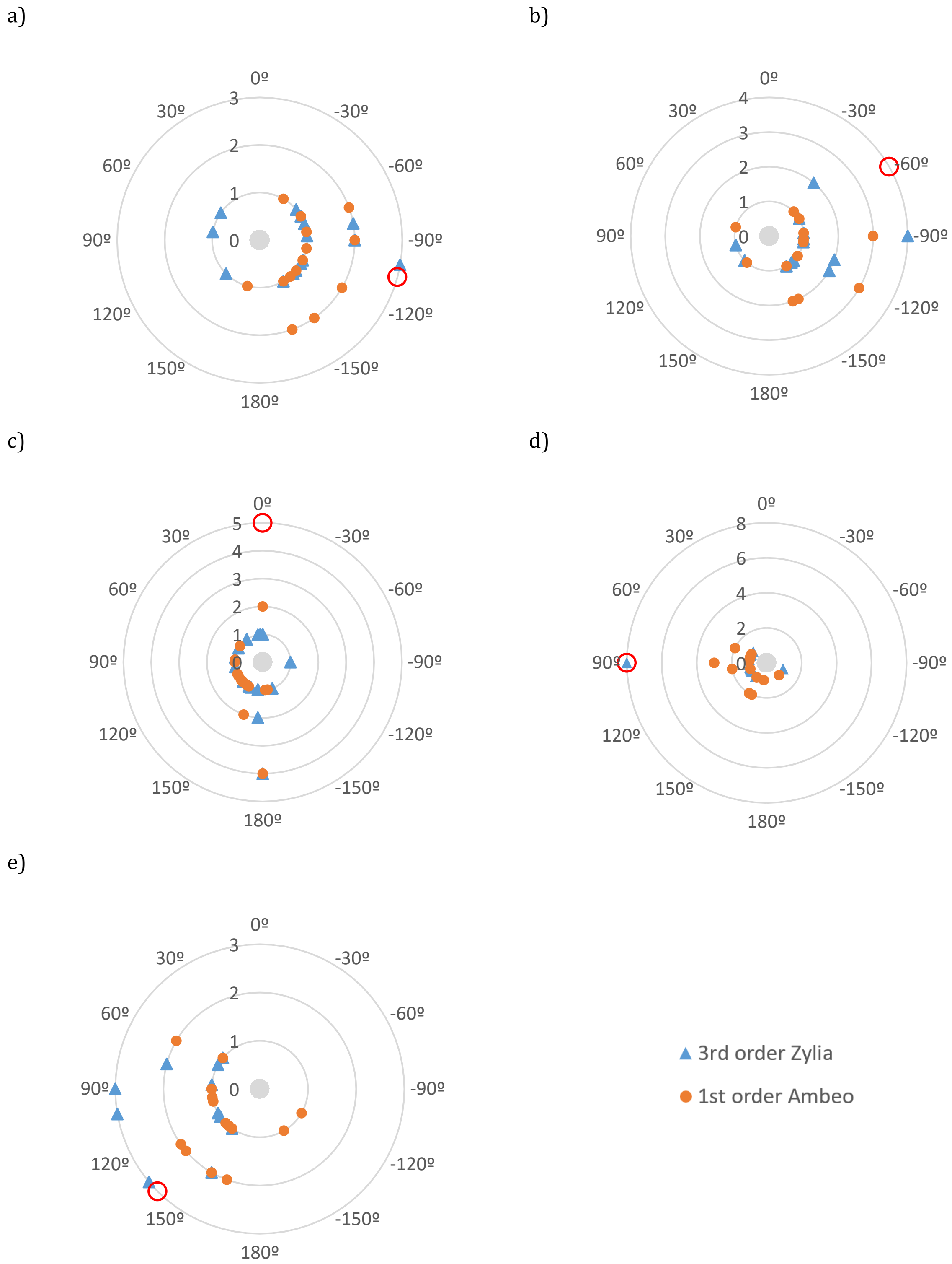

Figure 1. Localization of the sound sources (pink noise) in the azimuth plane. Number of times a given angle was pointed as the sound source localization. Correct answer is marked with a red circle: a) -105°, b) -60°, c) 0°, d) 90°, e) 135°.

The difference in results between the pink noise and bicycle bell sound are due to the nature of each sound. The bell sound was less direct, as it had some reverberation. This affects the perceived size of the sound source, and this broadening directly affects the ability to localize the source correctly.

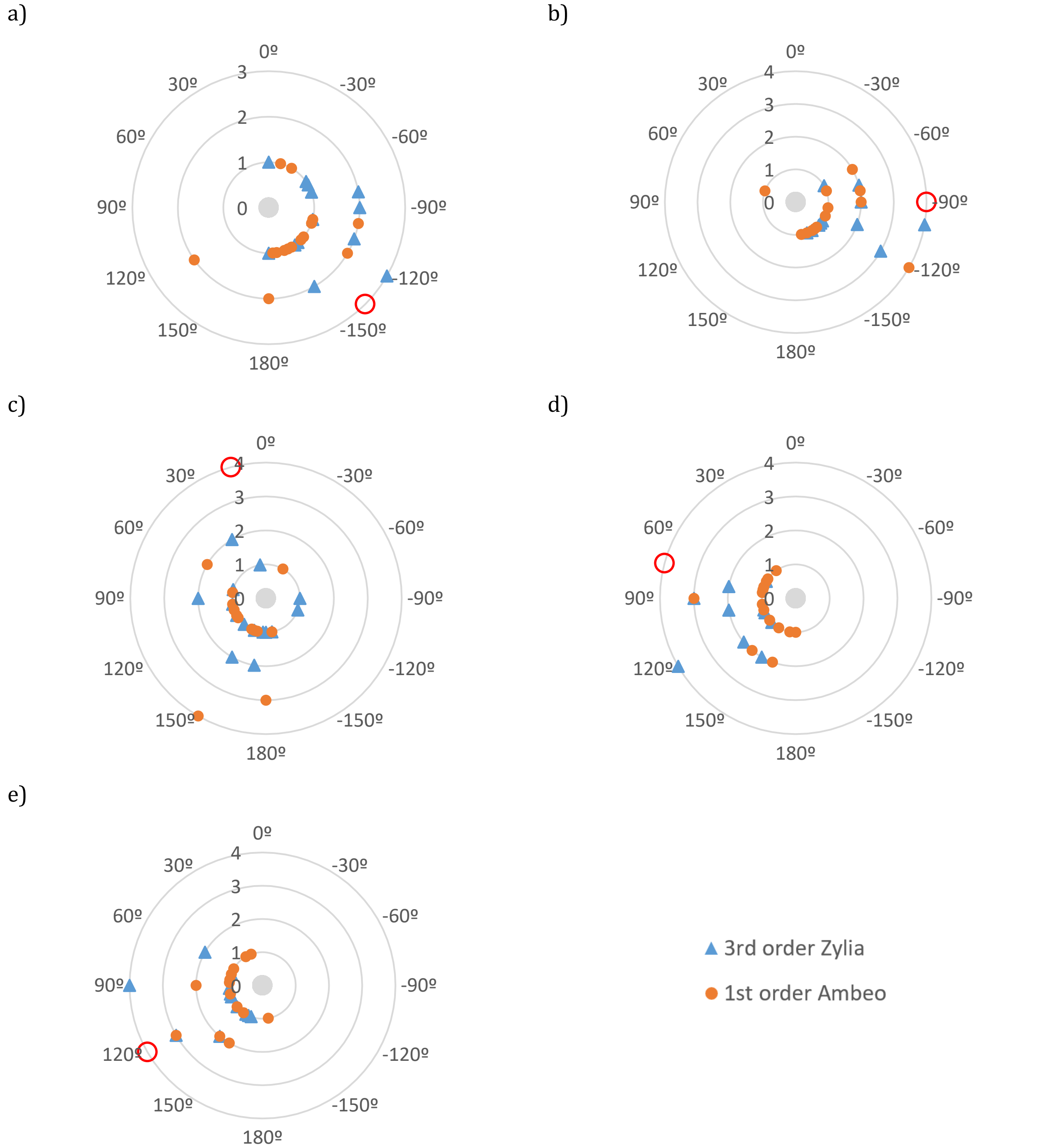

Figure 2. Localization of the sound sources (bicycle bell) in the azimuth plane. Number of times a given angle was pointed as the sound source localization. Correct answer is marked with a red circle: a) -135°, b) -90°, c) 15°, d) 75°, e) 120°.

Analysis of the results was made also for the elevation plane, although the data are difficult to interpret and none of the samples was localized correctly. Regarding the pink noise bursts (Fig. 3) the closest to correct recognition was the 90° sample (Fig. 3c). In this case both for Ambeo and Zylia, the most common answer was off by 30°. For the rest of the samples, the data is very scattered and there is plenty of front-back and up-down misjudgments. In the case -45° sample (Fig. 3a) for example, 5 participants (the most common answer) pointed the 180° as the localization, which means they perceived the sound as coming from the back, when it was meant to be recognized as coming from below and from the front. For 30° sample (Fig. 3b), most participants indicated sound source at the back and below, when it was supposed to be from the front and above. For 135° and 180° (Fig. 3d,e) the misjudgments are smaller, but still significant, and the answers for Ambeo are more consistent and have smaller error then for Zylia.

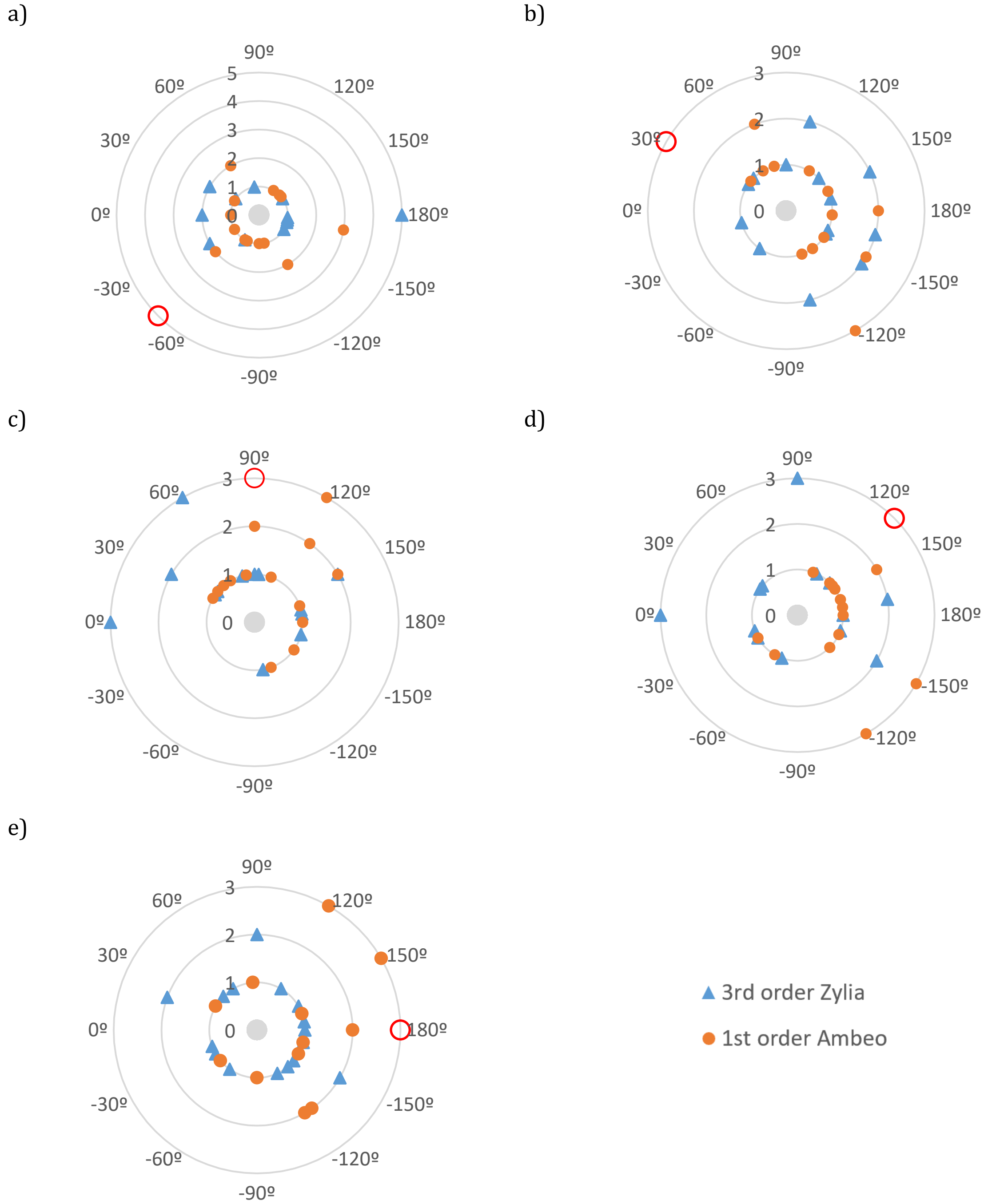

Figure 3. Localization of the sound sources (pink noise) in the elevation plane. Number of times a given angle was pointed as the sound source localization. Correct answer is marked with a red circle: a) -45°, b) 30°, c) 90°, d) 135°, e) 180°.

The bicycle bell sound was also hard to localize in the elevation plane. The results show that there was a great confusion on the direction of the sound source (Fig. 4). The most correctly localized sample was the -135° (Fig. 4a) but the most common answer was off by 30° from the correct angle. For the -15° and 30° samples (Fig. 4b, c), the results are scattered across the whole elevation plane, with more answers indicating the sound coming from the back then from the front. This again shows that for the angles close to 0°, the front-back confusion is really common. For 165° sample (Fig. 4e) the most common answer is for Ambeo and it's off by 45° from the correct angle. Unlike for the pink noise bursts, for bicycle bell sound the 90° sample was poorly localized (Fig. 4d). The most common answers are off by 90° and more.

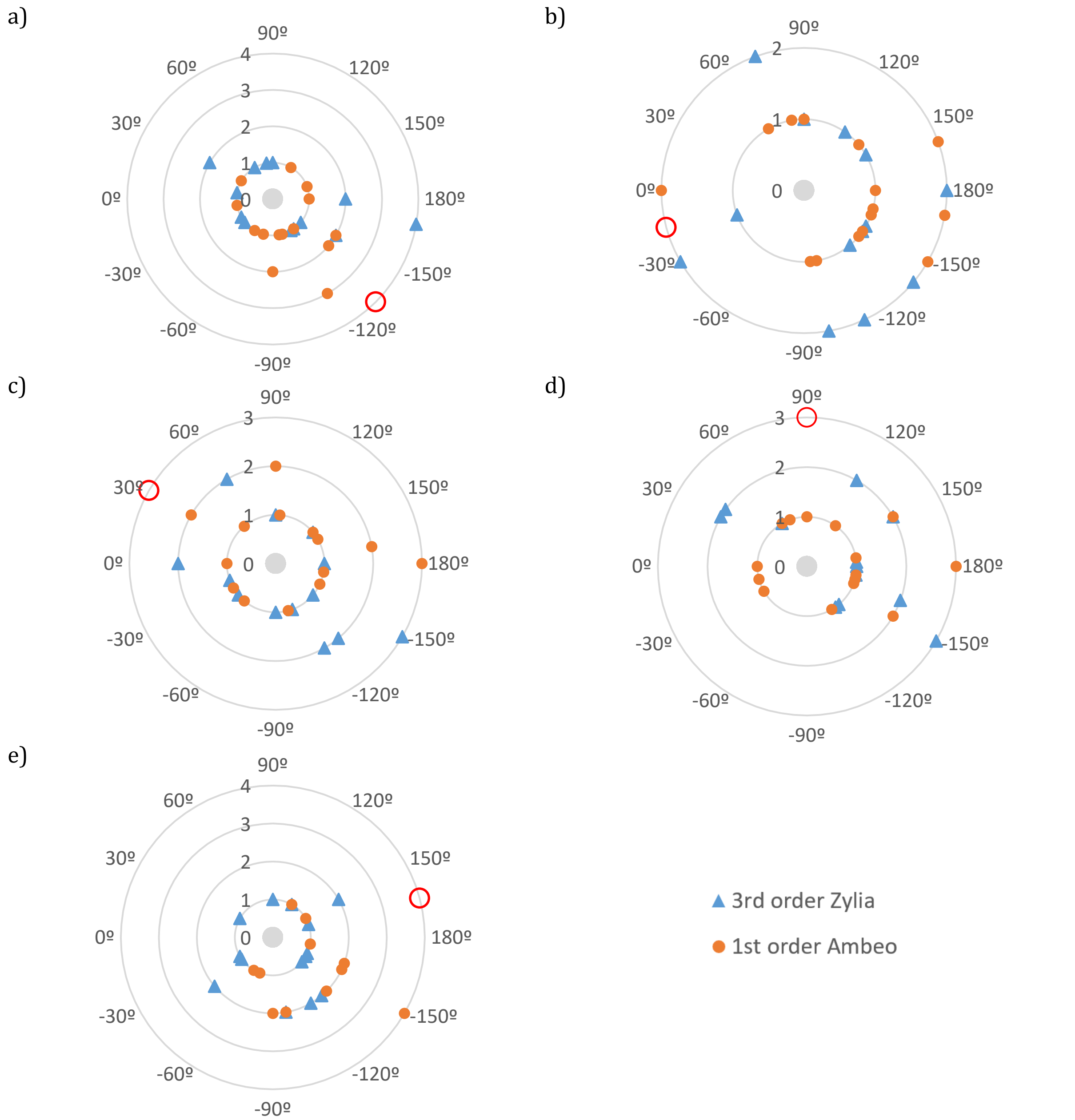

Figure 4. Localization of the sound sources (bicycle bell) in the elevation plane. Number of times a given angle was pointed as the sound source localization. Correct answer is marked with a red circle: a) -135°, b) -15°, c) 30°, d) 90°, e) 165°.

The main conclusion regarding the test elevation plane is that the task was too difficult for the participants. There is no consistency in the answers and the data is really hard to interpret. The reason for this difficulty is that in the elevation plane, when the sound decoded for binaural reproduction, the signal for the right and left ear differs only in the spectrum, having different frequency components for different angles. There are no other cues such as interaural time or intensity difference. The localization cues coming from the dummy head HRTF used in the Binaural Decoder plugin can be not enough to properly localize the sound, especially when they are distorted by headphones frequency characteristics.

Another way to interpret the results in more statistical way is to calculate the average localization errors for every tested angle. The errors were calculated for 1st order Ambeo and 3rd order Zylia, separately for azimuth (Fig. 5a, b) and elevation plane (Fig. 5c, d). In the azimuth plane (Fig. 5a, b) the biggest mean error can be observed for 0° (pink noise) and 15° (bicycle bell). That is due to the front-back confusion, so the cases where the error was close to 180°. The smallest errors can be seen for 90° and -90° samples, so when the sound source was localized on the right or left side of the listener. For the pink noise bursts, mean localization error for Ambeo was 59°, and for Zylia it was 54°, so for 3rd order Zylia it was by 5° smaller than for 1st order Ambeo. In the case of bicycle bell the localization error was 55° for Ambeo and 47° for Zylia, so the difference was 8°, also in favor of 3rd order Zylia. For the whole azimuth test errors are smaller for 3rd order Zylia by 6,8° on average. The data indicate that for the azimuth plane, using 3rd order microphone when recording the ambisonic sound gives better localization accuracy than when using 1st order microphone.

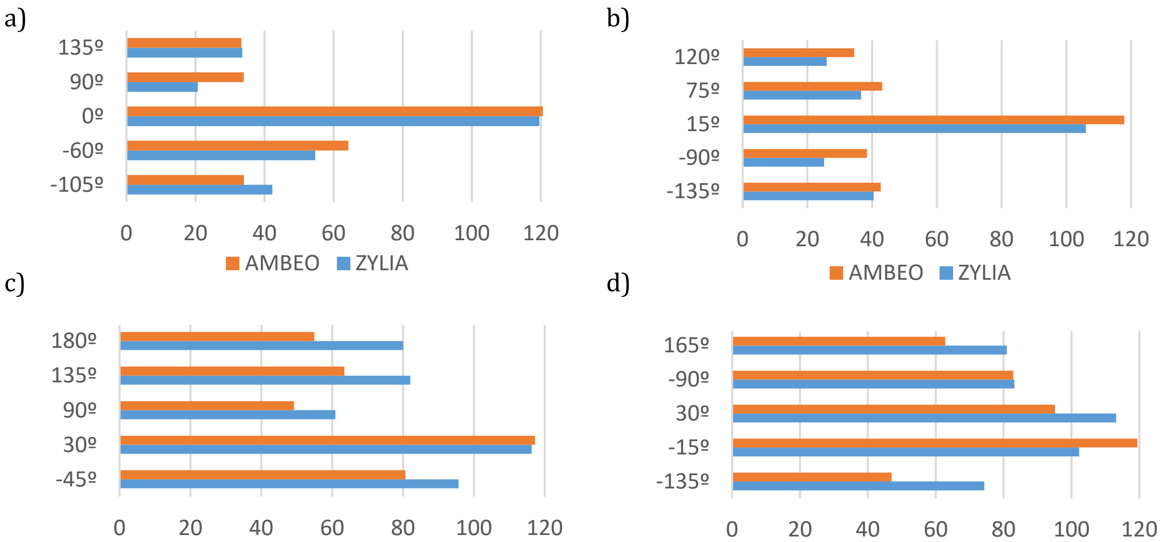

Figure 5. Mean localization errors (angle in degrees) for a) pink noise in azimuth plane, b) bicycle bell in azimuth plane, c) pink noise in elevation plane, d) bicycle bell in elevation plane.

For the elevation plane (Fig. 2c, d) the errors were greater than for the azimuth. The mean error for pink noise was 87° for Zylia and 73° for Ambeo. In the case of bicycle bell sound it was 91° for Zylia and 81° for Ambeo. This means the 1st order Ambeo gave by 14° better localization accuracy for the pink noise and by 10° better for the bicycle bell sound, comparing to 3rd order Zylia. Averaged for both sound samples, it was 11,6° better accuracy. This indicates that in the elevation plane, using 3rd order Zylia microphone doesn't increase the accuracy of the localization of binaurally reproduced sounds.

Table 1. Statistical analysis of the localization errors - median, dominant and standard deviation (angle in degrees).

Pink Noise – Azimuth		-105°		-60°		0°		90°		135°	
		Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo
	Median	25	35	40	60	153	148	8	23	35	20
	Dominant	15	15	30	60	180	180	0	0	35	10
	Std. Dev.	50	23	44	41	66	57	36	35	25	32
Bicycle bell – Azimuth		-135°		-90°		15°		75°		120°	
		Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo
	Median	30	35	20	30	125	135	40	35	30	30
	Dominant	15	35	30	30	15	135	15	15	30	30
	Std. Dev.	32	36	19	37	53	44	22	34	18	29
Pink Noise – Elevation		-45°		30°		90°		135°		180°	
		Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo
	Median	122	75	120	150	60	35	75	70	65	56
	Dominant	135	125	120	150	60	30	45	105	30	60
	Std. Dev.	49	58	52	62	40	41	51	51	56	41
Bicycle bell – Elevation		-135°		-15°		30°		90°		165°	
		Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo	Zylia	Ambeo
	Median	45	35	105	135	140	110	85	90	75	45
	Dominant	165	15	85	155	30	60	60	90	75	45
	Std. Dev.	61	44	48	48	59	57	38	39	51	35

Statistical analysis (Tab. 1) shows the median, dominant and standard deviation of the localization errors. For the azimuth plane, the median of the errors is on average 53° and it is lower for Zylia (51°) then for Ambeo (55°). For the elevation plane the median of the error is on average 85°, so much higher than for the azimuth plane. It is lower for Ambeo (80°) than for Zylia (90°). The dominant of the localization error shows that the biggest errors in the azimuth plane are made for the angles close to 0° due to front-back confusion, and in the elevation plane there are many up-down errors as the dominant values for the negative angles are often higher than 120°. The standard devation of the localization error is smaller for the azimuth plane (37°) and same for Ambeo and Zylia, and in the elevation plane it is greater (50°) but also really similar (48° for Ambeo and 51° for Zylia.)

4. Conclusions

The differences between the localization accuracy of the sound sources between audio samples recorded with 1st order Ambeo and the 3rd order Zylia microphone in a binaural reproduction of ambisonic sound depend on the characteristics of the sound and have a different character for the azimuth and the elevation plane. For the azimuth plane, using 3rd order Zylia microphone for the recordings gave by 6,8° better localization accuracy in the listening test compared to 1st order Ambeo. In this plane, the mean localization error was 54°. In the elevation plane, the data was scattered and hard to interpret but on the average, using 1st order Ambeo gave by 11,6° better localization accuracy than the 3rd order Zylia microphone. The mean localization error for elevation plane was 83°.

The prepared listening test turned out to be difficult for the participants although they were informed about the course of the test, they got clear instructions and every sample was repeated twice. Especially in the elevation plane, the localization cues coming only from the spectral differences between sound from different angles was to little information for not trained listeners to give a reliable answer about the elevation angle. To achieve more consistency in the results, the number of participants should be increased in order to gain more statistical data and also the group should include experts. The test should also examine different ways of decoding the ambisonic sound to binaural reproduction as well as the influence of participants' HRTF to the localization accuracy.

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. 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. DOI:10.1007/978-3-030-17207-7
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, 2013, 99, 4, 642–657. DOI: 10.3813/AAA.918643
3. G. Kearney, T. Doyle;Height perception in Ambisonic based binaural decoding; Audio Engineering Society Convention 139; Audio Engineering Society, 2015.
4. M. Gorzel, G. Kearney, H. Rice, F. Boland; On the perception of dynamic sound sources in ambisonic binaural renderings; Audio Engineering Society Conference: 41st International Conference: Audio for Games; Audio Engineering Society, 2011.
5. B. Xie; Head-related transfer function and virtual auditory display; J. Ross Publishing, 2013.
6. 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. DOI: 10.1109/TASLP.2021.3055038
7. A. Sontacchi, P. Majdak, M. Noisternig, R. Höldrich; Subjective Validation of Perception Properties in Binaural Sound Reproduction Systems; Audio Engineering Society Conference: 21st International Conference: Architectural Acoustics and Sound Reinforcement; Audio Engineering Society, 2002.
8. E.M. Wenzel, M. Arruda, D. J. Kistler, F.L. Wightman, Localization using nonindividual head-related transfer functions; J. Acoust. Soc. Am., 1993, 94, 1, 111-123. DOI: 10.1121/1.407089
9. „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 (access: 22 July 2022).
10. „ZYLIA Ambisonics Converter", ZYLIA - 3D AUDIO RECORDING & POST-PROCESSING. https://www.zylia.co/zylia-ambisonics-converter.html (access 22 July 2022).
11. D. Rudrich, „Plug-in Descriptions", 7 czerwiec 2018. https://plugins.iem.at/docs/plugindescriptions/ (access 22 July 2022).