Preprints.org

Article

Perceived Sound Source Localization in Selected Binaural Renderers

Paweł Małecki [*] , Magdalena Piotrowska , Olga Krzyżyńska

Article

Perceived Sound Source Localization in Selected Binaural Renderers

Paweł Małecki [1],*, Magdalena Piotrowska [1] and Olga Krzyżyńska-Hamowska [1]

[1] AGH University of Krakow, al. A. Mickiewicza 30, 30-059 Krakow, Poland
* Correspondence: pawel.malecki@agh.edu.pl

Abstract: The article addresses the issue of localization of a perceived sound source for various binaural renderers, in the context of spatial audio format processing for headphone listening. The study analyzed the accuracy of sound source localization for ambisonic technology and two different versions of the renderer from the Dolby Atmos format. Audio samples were played to participants in a virtual reality (VR) environment, allowing for natural pointing toward sound directions. Results indicate that the Dolby Atmos renderer provides the highest accuracy in reproducing the position of sound sources, especially for samples originating from behind the listener's head. The study also showed that listeners could better recognize sound position in the horizontal plane (left-right) than in the vertical plane (up-down). The use of VR in psychoacoustic research proved not only effective but also engaging for participants, underscoring the potential of this technology for further spatial audio research.

Keywords: binaural rendering; spatial audio; headphone listening; ambisonics; Dolby Atmos; virtual reality; sound localization

---

1. Introduction

Spatial sound plays a crucial role in human auditory perception, allowing individuals to locate sound sources and navigate their environment effectively. Binaural rendering techniques aim to reproduce spatial audio experiences over headphones by simulating the way sound interacts with the human head and ears. These technologies have gained significant attention in various fields, including gaming, virtual reality (VR), and immersive audio experiences [1,2]. However, the accuracy of sound source localization depends on multiple factors, such as the rendering algorithm, head-related transfer functions (HRTFs), and listener adaptation [3,4].

Previous studies have explored the impact of different auditory environments on spatial perception. Mróz and Kostek [5] investigated the differences in perceptual responses to spatial audio presented through headphones versus loudspeakers, highlighting how the reproduction system affects listener perception. Similarly, Moraes et al. [6] examined physiological responses in a VR-based sound localization task, showing how immersion influences auditory accuracy. Research by Tang [7] further emphasized the importance of sound in VR applications, demonstrating that spatial audio significantly enhances realism and presence. Additionally, scene-based audio techniques, such as Higher-Order Ambisonics (HOA), have been widely applied in next-generation audio and 360-degree video experiences [8,9].

Despite advancements in spatial audio rendering [10], challenges remain in accurately localizing sound sources, particularly in headphone-based environments. Previous research has shown that head movements improve localization accuracy, while static listening conditions often lead to perceptual ambiguities [11]. Blauert et al. [12] developed an interactive virtual environment for psychoacoustic research, highlighting the role of controlled experimental setups in evaluating spatial hearing performance.

This study aims to investigate the accuracy of sound source localization across different binaural renderers, including an ambisonic system and two variations of the Dolby Atmos renderer [13]. The

experiment utilizes VR technology to provide an intuitive method for participants to indicate perceived sound directions [14]. By analyzing localization precision in different spatial planes, this research contributes to a deeper understanding of binaural audio perception and its implications for immersive sound applications.

2. The role of virtual reality in psychoacoustic research

Virtual Reality (VR) technology has emerged as a crucial tool in psychoacoustic research, providing immersive and controlled environments for studying spatial hearing and sound localization [14]. VR-based experiments enable researchers to simulate complex auditory scenarios while allowing participants to interact naturally with their surroundings [15]. This approach enhances the accuracy of perceptual studies by reducing the limitations of traditional laboratory settings.

A study by Moraes et al. [6] investigated the role of physiological responses in a VR-based sound localization task, demonstrating that VR environments can effectively capture implicit and explicit metrics related to auditory perception. Their findings indicated that physiological measures, such as heart rate variability, are reliable indicators of cognitive load, reinforcing conclusions drawn from the NASA-TLX questionnaire. Moreover, their research emphasized the importance of user-environment interaction in developing spatial auditory skills, suggesting that future studies should involve larger participant groups and a deeper analysis of learning effects.

Another study [16] evaluated three geometrical acoustic engines to determine their effectiveness in dynamic modeling of acoustical phenomena in interactive virtual environments. The analysis focused on applications in immersive gaming and mobile VR platforms, addressing challenges related to computational constraints in real-time audio processing. The results highlighted the trade-offs between computational efficiency and spatial accuracy in sound rendering, underscoring the need for optimization techniques in VR-based audio research.

Blauert et al. [12] introduced an interactive virtual environment designed for psychoacoustic research, allowing participants to navigate and engage with spatial auditory scenes. The system provided a flexible platform for presenting complex auditory environments, including both direct and reflected sounds from dynamic sources. Such experimental setups facilitate precise localization studies, enabling researchers to assess human spatial hearing with a high degree of ecological validity. The potential for integrating additional sensory modalities further expands the scope of VR applications in auditory perception studies.

One significant obstacle identified is the potential loss of spatial tracking accuracy for Head-Mounted Displays (HMDs) or hand controllers. Such malfunctions can disrupt the virtual environment's spatial coherence, misaligning the calibrated positions within the virtual space [16]. Moreover, the usage of speaker systems in conjunction with VR headgear can introduce acoustic distortions. The physical characteristics of VR goggles—namely their size, shape, and material composition—may obstruct the sound's direct path to the ear, leading to structural reflections or attenuation of sound waves [1,2].

Research [1,2,7,16] has demonstrated that VR and AR glasses alter the interaural sound signals and the overall sound spectrum, potentially degrading the listener's ability to accurately localize sounds and perceive spatial audio quality. Nevertheless, it has been observed that visual cues provided by the VR system can mitigate these auditory discrepancies, ensuring that the perceived auditory location remains unaffected in most VR applications. [1] work, offering a database of head-related impulse responses collected with and without HMDs, facilitates the design of compensatory audio filters. These filters aim to restore natural auditory experiences during virtual sound playback, even when VR goggles are worn [2].

Furthermore, the fidelity of sound localization is also contingent upon the acoustic signal's properties. Research [9] evaluated auditory localization using two signal types: pure tones and Gaussian noise, measuring the Minimum Audible Angle (MAA) for each. It was corroborated by [17] that sound localization precision enhances with broader signal bandwidths. Their findings

indicated a significant reduction in MAA values from approximately 10° with tonal stimuli to about 1.5° for wideband signals, varying with the azimuth angle. This underscores the critical role of signal characteristics in optimizing sound localization accuracy within virtual reality settings.

In the VR Tools the pivotal role of auditory delivery is underscored. Sound not only enhances the sensation of presence in virtual environments, as noted by [18] and [19], but also surpasses visual stimuli in fostering immersive experiences [7]. Spatial audio cues in virtual reality environments augment situational awareness, directing user engagement, influencing emotional responses, and disseminating intricate information without burdening the visual sensory channel [7,20]. These auditory landscapes are instrumental in defining the essence and ambience of virtual spaces, contributing to a more comprehensive and immersive user experience [10].

Auditory perceptions during headphone playback significantly differ from those experienced through loudspeaker listening. Notably, there is a sensation of sound originating from within the head, commonly referred to as lateralization. This effect arises from the limitation of sound source localization mechanisms during headphone listening, such as the omission of head-related transfer function (HRTF) considerations. Additionally, factors like bone conduction and environmental reverberation are absent in headphone listening. To mitigate this phenomenon, the effectiveness of employing listener-specific HRTF has been demonstrated. The application of individually tailored HRTF has been observed to substantially reduce the sensation of lateralization among listeners [21]. Research indicates [3] that the accuracy of sound source localization during headphone listening significantly decreases for angles ranging from ±15° to ±30°. A somewhat lesser decline in localization accuracy is also noted for angles extending from ±30° to ±60° .

Auditory signals, due to their omnidirectional and continuous nature, facilitate quicker user responses compared to visual cues, enabling users to perceive environmental alterations without the need to visually scan their surroundings [7,10,20].

The evolution of 360° video formats necessitates audio solutions capable of delivering a full-spherical sound experience. The integration of ambisonic technology, occasionally in conjunction with object-based audio, addresses this requirement by enabling spatial playback around the user [4]. Given the prevalence of headphone usage in VR applications, the ease of converting ambisonic content to a binaural format, combined with the adaptability in sound field manipulation, positions ambisonics as the preferred choice for spatial audio in VR/AR settings [8].

These examples illustrate the growing significance of VR-based interactive systems in psychoacoustic research. By leveraging immersive interfaces and real-time spatial audio rendering, researchers can gain deeper insights into sound perception mechanisms. The flexibility of VR platforms also makes them well-suited for a wide range of applications, from basic auditory science to practical implementations in virtual and augmented reality technologies.

3. Experiments

This project integrated Virtual Reality (VR) and spatial audio technology to study sound localization in an immersive environment. A binaural rendering system processed spatialized audio, and a custom-based platform ensured real-time interaction, data collection, and analysis.

Developed using Unity (2020.2.6f1) with C#, the platform synchronized VR headset movements with real-time audio stimuli. Hand-held controllers allowed precise pointing at sound sources, while Reaper managed auditory stimulus playback via OSC protocol, integrating with Unity through the XR Integration Toolkit (UnityOSC). The system also employed UnityEngine.XR for controller input and Newtonsoft.Json for result serialization.

The experiment utilized the Oculus Rift (CV1) with Touch controllers and sensors, enabling precise spatial audio assessment [15]. Participants engaged with a 3D interactive user interface.

3.1. Audio Processing and Software Integration

Spatial audio was rendered through open-source IEM Ambisonic BinauralDecoder [22], Dolby Atmos Binaural Renderer [23], and Apple Binaural Renderer [23], then imported into Reaper DAW, where pre-rendered samples and markers synchronized with Unity via OSC (Open Sound Control) scripts. The Unity environment managed test administration, user interactions, and data recording, refined through preliminary testing to ensure optimal usability.

Participants (25 individuals) completed a tutorial followed by two experimental sessions:

1.   Session 1: 35 trials (6 minutes)
2.   Break: Up to 5 minutes
3.   Session 2: 40 trials (7 minutes)

Listeners responded by indicating the perceived localization of each sound by pointing with the controllers. Responses were recorded as azimuth and elevation angle pairs, serialized into JSON format for further analysis.

A grid of 25 points was established and distributed across a hemisphere. The selection of their locations was guided by human auditory perception principles (e.g., only a few points were placed within the cone of confusion). The designated points, where the virtual sound source was positioned, are presented in Figure 1, while their exact coordinates are listed in Table 1.
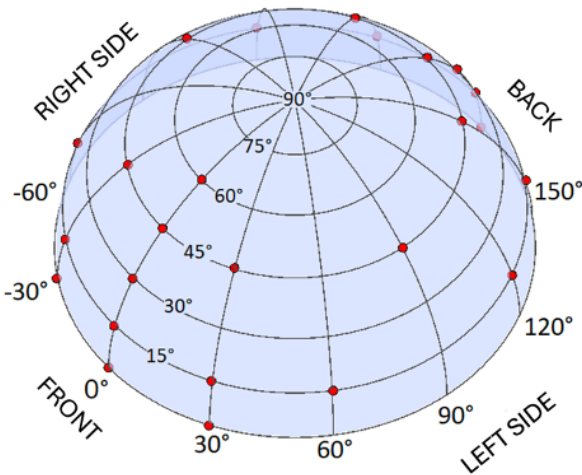

Figure 1. A grid of measurement points where the virtual sound source was positioned.

A spherical coordinate system was used, with the azimuth angle ranging from -180° to 180° and the elevation angle ranging from 0° to 90°.

Table 1. Exact coordinates of the measurement points.

No.	Azimuth [°]	Elevation [°]	No.	Azimuth [°]	Elevation [°]
1	0	0	14	180	0
2	0	15	15	180	15
3	0	30	16	180	30
4	0	45	17	180	45
5	0	60	18	-150	15
6	30	0	19	-150	45
7	30	15	20	-120	15
8	30	45	21	-90	45
9	60	15	22	-60	15
10	90	45	23	-30	0
11	120	15	24	-30	15
12	150	15	25	-30	45
13	150	45			

3.2. Sound stimuli - Preliminary research

Before conducting the main tests comparing the selected renderers, preliminary testing was carried out using one of them — the IEM renderer. The purpose of this phase was to evaluate whether the designed experiment was understandable and comfortable for the participants. During the test, participants were encouraged to report any concerns regarding the procedure. Particular attention was given to the appearance and usability of the VR interface, as well as the overall duration of the test. Additionally, an assessment was made to determine which type of audio signal was easier to localize — a natural source (human voice) or Gaussian noise.

Using the selected renderer (IEM StereoEncoder + BinauralDecoder), 25 samples were rendered for the natural source and another 25 for the Gaussian noise. Each human voice sample lasted approximately 9 seconds, while a single noise sample had a duration of 500 milliseconds, including a 10 ms fade-in and fade-out time. The noise sample was repeated four times, with a 500 ms pause between repetitions. Each signal type was tested in a separate test session. The positions of the sound source were presented in a randomized order.

Two participants took part in the preliminary test, with neither reporting any hearing issues on the day of the session.

The responses collected during the preliminary tests for all positions and both types of audio signals (noise and human voice) are presented in Table 2. Additionally responses for selected positions are presented in Figures 2 and 3. Coordinates of ositions are labeled according to Table 1.

Table 2. Responses provided by participants during the preliminary research.

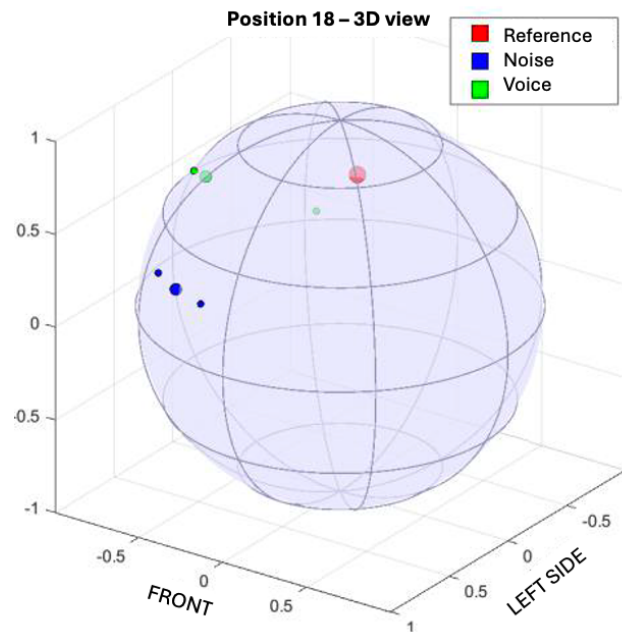| Position	Reference		Noise				Voice			
	Az [°]	El [°]	Subject 1		Subject 2		Subject 1		Subject 2	
			Az [°]	El [°]	Az [°]	El [°]	Az [°]	El [°]	Az [°]	El [°]
1	0	0	-158,4	0,7	169,6	27,7	177,6	47,3	177,5	38,7
2	0	15	-157,6	-6,1	-127,5	11,0	147,4	29,6	135,8	3,1
3	0	30	-151,0	-0,9	-46,2	3,2	84,6	10,2	136,4	0,1
4	0	45	-158,9	57,0	-53,2	68,8	90,6	-0,3	165,8	9,7
5	0	60	171,5	43,6	-155,8	46,7	89,4	-1,0	133,9	10,1
6	30	0	155,6	13,7	44,9	7,1	24,4	23,2	-176,0	43,9
7	30	15	159,1	-12,5	60,5	27,3	43,7	0,5	68,0	1,1
8	30	45	160,6	1,8	113,0	66,7	45,1	44,5	97,8	70,7
9	60	15	162,4	-7,8	116,7	1,1	0,5	14,6	-1,1	22,7
10	90	45	172,4	10,9	84,6	18,7	-0,2	0,0	75,8	12,1
11	120	15	153,3	20,0	146,7	2,9	0,3	6,3	-1,1	43,0
12	150	15	155,2	31,1	81,1	26,4	44,1	32,4	134,8	53,5
13	150	45	18,6	51,9	40,2	53,1	45,5	7,1	44,7	20,2
14	180	0	88,3	-0,4	114,4	11,6	89,1	23,6	172,8	9,8
15	180	15	143,3	-16,8	116,2	-4,6	135,8	7,2	137,5	7,5
16	180	30	97,4	-26,9	89,6	-4,1	155,1	7,7	111,4	10,6
17	180	45	142,2	-10,0	125,2	1,0	91,1	-1,1	126,0	30,7
18	-150	15	-10,9	18,5	-35,1	19,1	-138,3	2,8	-60,1	44,2
19	-150	45	166,9	-22,1	80,5	31,9	148,2	24,5	122,6	45,9
20	-120	15	-43,7	-4,3	-79,8	26,8	-43,6	34,2	-45,5	29,8
21	-90	45	-21,7	5,8	-117,6	19,6	-0,3	15,4	-44,9	55,2
22	-60	15	-53,4	-9,5	-68,4	39,2	-41,6	33,9	-134,4	39,2
23	-30	0	-8,1	-1,1	-44,6	6,8	-18,0	23,2	-45,1	28,3
24	-30	15	-1,6	-0,2	-31,2	3,0	0,5	-0,5	-84,9	52,6
25	-30	45	40,1	20,5	-58,0	1,8	39,6	35,1	-45,4	43,2

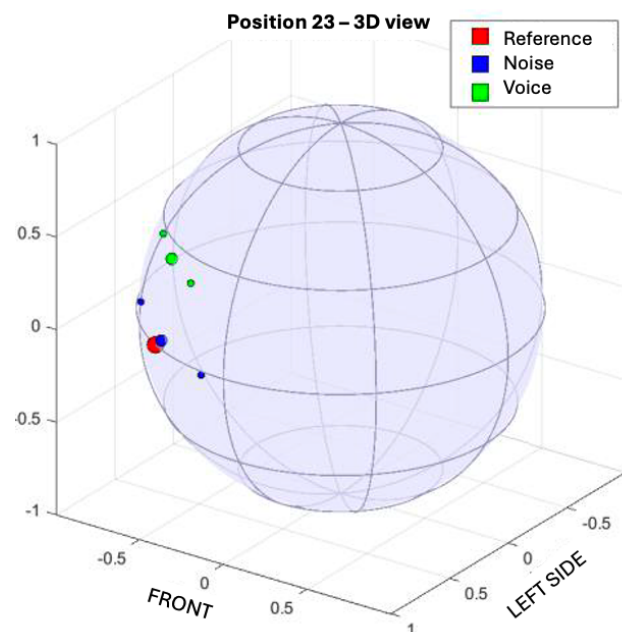**Figure 2.** Participant responses along with the reference point for position 18.

Figure 3. Participant responses along with the reference point for position 23.

The preliminary tests allowed refining the system, the testing setup, and the overall experiment. The participants confirmed that the interface was intuitive and did not cause disorientation during the test. They reported feeling comfortable in the created virtual environment.

During the study, participants observed that, subjectively, it was easier for them to localize the natural source (voice) than the noise. However, the available literature indicates that, when studying localization precision, the stimulus should include frequencies above 7 kHz [24]. In addition, noise is commonly used as a stimulus in psychoacoustic research. For these reasons, Gaussian noise samples were ultimately chosen for the main tests.

Given that only two participants took part, it is difficult to draw conclusions about sound source localization precision. However, the goal of the preliminary tests — to refine the experimental setup — was successfully achieved.

Therefore, gaussian noise samples (500 ms duration, 500 ms pauses, repeated 4 times) were presented from 25 predefined positions on a hemispherical grid. A calibration module ensured accurate tracking, with positional errors observed between 0.5%–4.7% (horizontal) and 0.5%–3.5% (vertical).

4. Results

This chapter summarizes the outcomes of the conducted experiment. In addition to the circular mean, circular standard deviation, and spherical distances between the reference and perceived positions, the accuracy and precision of the listener responses are also discussed. The results are presented graphically on a spherical surface to facilitate the visualization and analysis of listeners' responses, which is essential for studying the characteristics of spatial audio systems.

4.1. Circular mean and Circular standard deviation

Key statistical values were calculated, including the circular mean and the circular standard deviation, using Equations (1) and (2). Calculations were performed within the MATLAB environment using the Circular Statistics Toolbox available [25]. The examples of responses collected for all listeners are presented in Figures 4–6. The mean values have been marked on the plots. The circular standard deviation, which indicates how widely the responses are distributed around the mean, as well as the circular mean itself, are presented in Tables 3 and 4.

$$\bar{\alpha} = \arg\left(\sum_{j=1}^{n} \exp(i \cdot \alpha_j)\right) \tag{1}$$

$$\bar{S}(z) = \sqrt{-2\ln(\bar{R})} \tag{2}$$

where:

$\bar{\alpha}$ – circular mean

$\alpha_1 ... \alpha_n$ – consecutive angles expressed in radians

$\bar{S}(z)$ – circular standard deviation
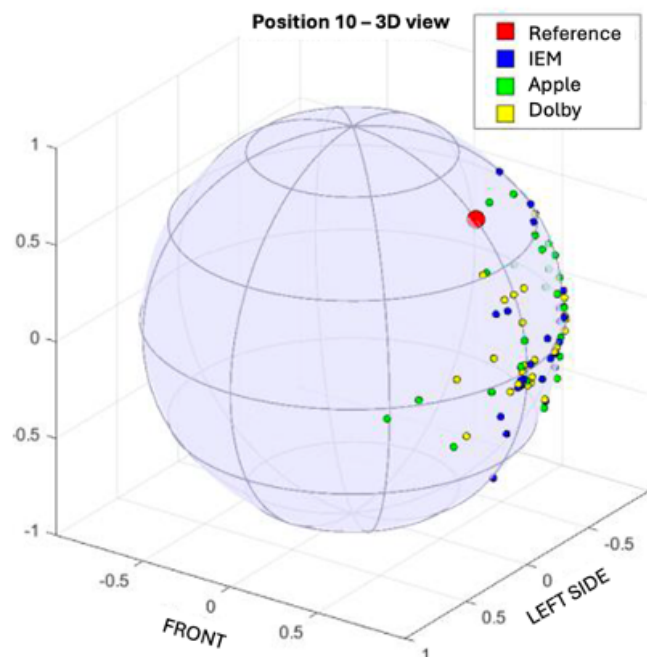
$\bar{R}$ – mean resultant length

Figure 4. Responses collected from all listeners for the reference point no. 10 at an azimuth of 90° and an elevation of 45°.
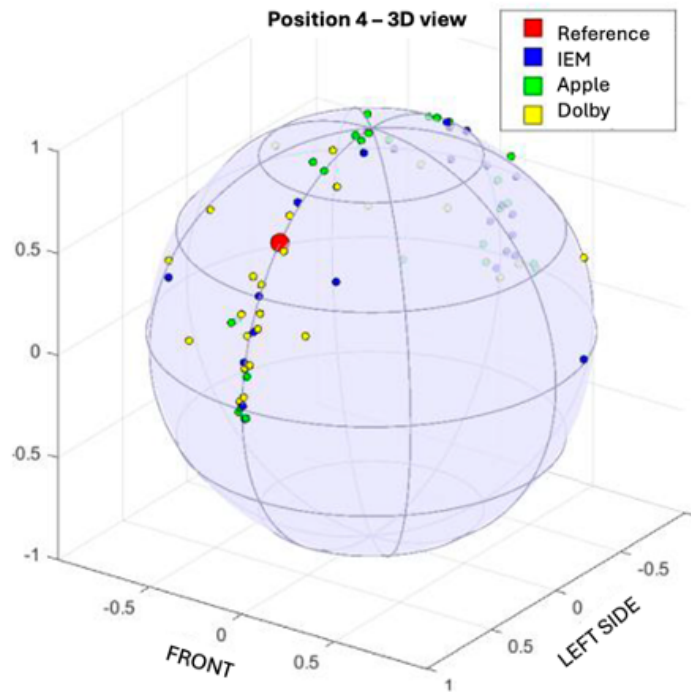
Figure 5. Responses collected from all listeners for the reference point no. 4 at an azimuth of 0° and an elevation of 45°.
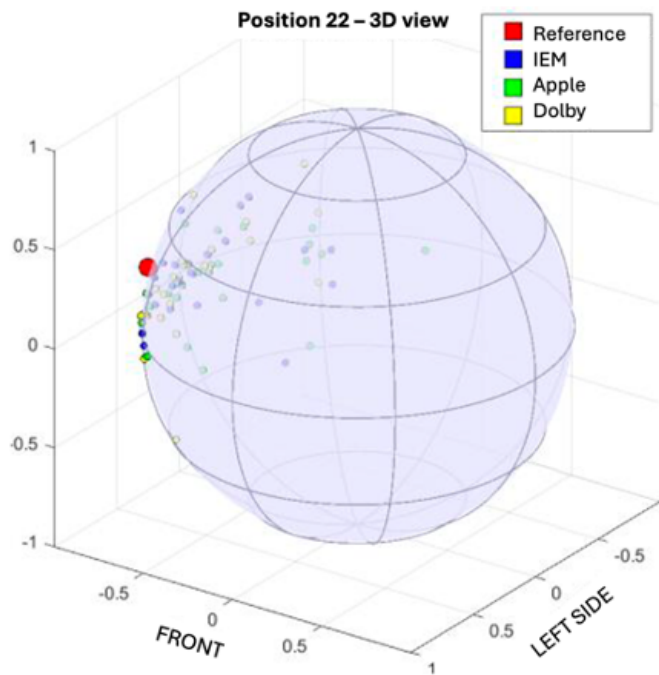

Figure 6. Responses collected from all listeners for the reference point no. 22 at an azimuth of -60° and an elevation of 15°.

Tables 3 and 4 present the mean values and standard deviations for each position, categorized by the tested renderers.

Table 3. The circular mean of the responses provided for each position and each renderer.

Position	Reference		IEM		Apple		Dolby	
	Az [°]	El [°]	Az [°]	El [°]	Az [°]	El [°]	Az [°]	El [°]
1	0	0	-177	1	176	7	158	12
2	0	15	-92	7	168	21	-180	21
3	0	30	-116	22	173	21	4	23
4	0	45	174	24	-157	31	-13	24
5	0	60	-169	34	28	26	179	21
6	30	0	86	6	110	12	91	6
7	30	15	103	16	108	7	79	8
8	30	45	112	32	78	15	75	23
9	60	15	102	4	99	3	85	-1
10	90	45	107	1	108	7	97	4
11	120	15	99	6	123	2	-159	23
12	150	15	136	-7	121	15	127	-8
13	150	45	118	12	134	20	127	5
14	180	0	-179	1	170	11	177	-6
15	180	15	168	25	178	6	169	1
16	180	30	-150	24	-168	13	-164	5
17	180	45	179	13	-167	18	-167	6
18	-150	15	-131	-12	-148	2	-137	-8
19	-150	45	-119	5	-136	3	-124	2
20	-120	15	-110	-2	-124	1	-132	-6
21	-90	45	-103	2	-111	0	-107	1
22	-60	15	-95	-1	-101	-4	-90	-1
23	-30	0	-91	7	-107	9	-86	2
24	-30	15	-90	13	-123	13	-80	6
25	-30	45	-103	11	-122	13	-94	19

Table 4. The circular standard deviation of the responses provided for each position and each renderer.

Position	Reference		IEM		Apple		Dolby	
	Az [°]	El [°]	Az [°]	El [°]	Az [°]	El [°]	Az [°]	El [°]
1	0	0	116	29	52	36	79	30
2	0	15	120	36	75	30	108	43
3	0	30	97	35	130	28	89	26
4	0	45	95	22	109	34	70	20
5	0	60	107	30	104	31	46	28
6	30	0	36	15	57	21	33	13
7	30	15	36	20	37	21	38	16
8	30	45	38	19	44	18	32	21
9	60	15	27	16	28	17	23	10
10	90	45	20	17	30	16	19	11
11	120	15	17	17	26	20	99	39
12	150	15	32	17	44	22	17	18
13	150	45	49	22	58	27	23	21
14	180	0	36	31	69	36	42	26
15	180	15	66	32	104	28	38	26
16	180	30	94	31	55	25	37	26
17	180	45	97	31	68	21	59	30
18	-150	15	23	15	29	15	29	21
19	-150	45	22	15	27	16	21	20
20	-120	15	21	10	26	15	16	13
21	-90	45	25	13	23	9	18	16
22	-60	15	24	11	26	10	23	11
23	-30	0	34	18	51	23	38	19
24	-30	15	36	19	74	20	33	16
25	-30	45	31	20	51	23	35	18

4.2. Distance between points on a sphere

The distance between two points on a sphere was calculated as the length of the arc defined by those points on the so-called great circle [26]. This is the largest possible circle that can be inscribed in a sphere, and its radius is equal to the radius of the sphere. Figure 7 shows the distance between two points on the sphere determined using a great circle. In the calculations, the radius of the sphere was assumed to be 1.
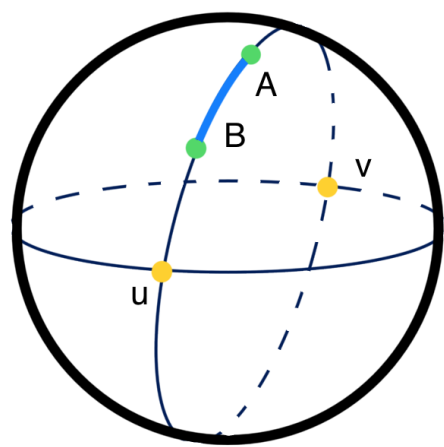

Figure 7. The shortest path between two example points on the surface of a sphere.

Table 5 presents the calculation results along with the shortest distance marked between the reference value and the average value obtained from the responses at each measurement position.

Table 5. The distance between the reference value and the calculated average response for each renderer separately at each measurement position with the assumption of a sphere radius of 1 unit.

Position	IEM	Apple	Dolby
1	3.08	3.00	2.70
2	1.60	2.91	3.04
3	2.03	2.95	0.15
4	2.77	2.67	0.42
5	2.65	0.75	2.46
6	0.98	1.38	1.07
7	1.28	1.36	0.86
8	1.43	0.86	0.80
9	0.73	0.67	0.44
10	0.29	0.32	0.11
11	0.37	0.13	1.42
12	0.38	0.51	0.49
13	0.68	0.44	0.65
14	0.03	0.26	0.11
15	0.28	0.16	0.31
16	0.53	0.36	0.51
17	0.56	0.53	0.71
18	0.49	0.20	0.39
19	0.71	0.62	0.69
20	0.21	0.15	0.30
21	0.23	0.37	0.29
22	0.61	0.70	0.52
23	1.07	1.34	0.98
24	1.05	1.62	0.87
25	1.24	1.54	1.11

4.3. Precision versus Accuracy

To better understand the following conclusions and analyses, it is worth examining the difference between precision and accuracy of measurement. This is illustrated graphically in Figure 8.
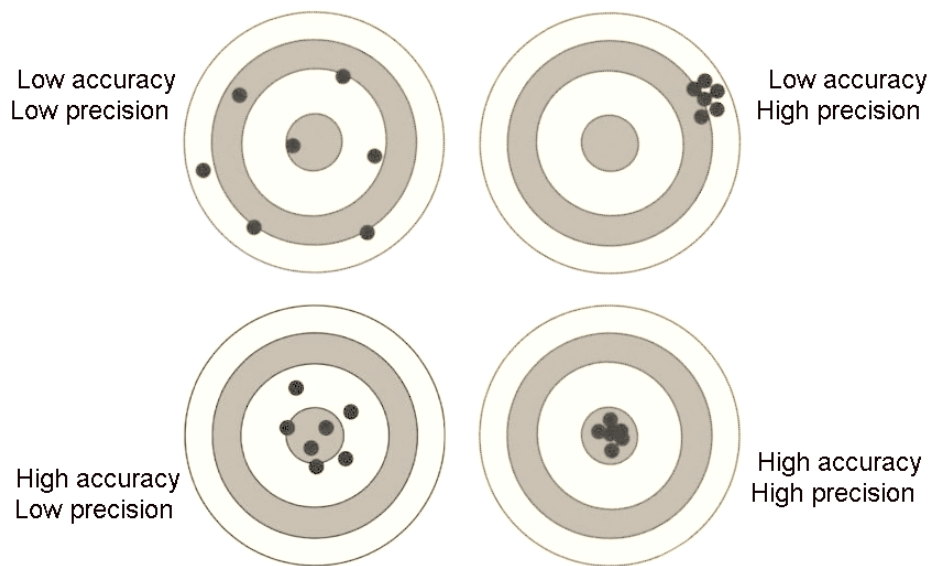

Figure 8. Represenation of the difference between precision and accuracy of measurement

In the context of this analysis, **precision** refers to how closely grouped the responses are (i.e., how consistent participants were), while **accuracy** refers to how close the average response was to the reference sound source location. In other words, a low **standard deviation** indicates high precision, and a low **distance from the reference point** indicates high accuracy.

Table 4 presents the standard deviation values of the responses. Cases in which a given renderer had the lowest standard deviation for both elevation and azimuth (i.e., the responses were the least scattered) are listed as follows:

- IEM – 3 such instances,
- Apple – 1 such instance,
- Dolby – 9 such instances.

This indicates that, when listening to samples rendered using the Dolby renderer, participants most frequently gave responses that were closely grouped together. Therefore, in the conducted study, the Dolby renderer demonstrated the highest **precision**.

Further, by analyzing the distances between the reference point and the average response for each renderer at every test position (Table 5), the number of times each renderer achieved the **lowest distance** was counted:

- IEM – 5 times,
- Apple – 9 times,
- Dolby – 11 times.

This means that in the case of the Dolby renderer, the participants' responses were most often **closest to the expected (reference) location** of the virtual sound source. Although this time the lead was not as dominant (Apple achieved the best results in only two fewer cases), Dolby still proved to be the most **accurate** renderer.

Additionally, there were 5 specific test positions (3, 4, 8, 9, and 24) for which the Dolby renderer demonstrated both the **highest precision and highest accuracy** among all tested renderers. In comparison, neither IEM nor Apple achieved this for any position.

5. Discussion

The collected data provide valuable insights into the precision of sound source localization in binaural listening, depending on the spatial position of the source. By analyzing the response distributions, circular means, and standard deviations, several key observations emerge:

- **Greater response dispersion for front-facing sources:** Responses for centrally positioned sound sources at the front (positions 1–5) were more scattered, whereas responses for rear sources (positions 11–13) were more concentrated, as indicated by lower standard deviation values. This effect was particularly noticeable in the azimuth plane (Table 4).
- **Correct hemisphere identification with small shifts:** When the sound source was shifted slightly in the horizontal plane (e.g., position 6, offset 30° to the left), participants correctly identified the hemisphere, confirming that even minor shifts in azimuth were perceived accurately.
- **Increased response precision for rear-located sources:** The further behind the listener the virtual sound source was placed, the more concentrated the responses were around it.
- **Tendency to localize sound sources behind the head:** A striking observation is that 63 out of 75 mean response values were positioned behind the listener (azimuth angle between 120° and 180° or -120° and -180°). This suggests that participants had difficulty localizing sound sources in front of them (Table 3).
- **Limited accuracy for front-facing sources**
- **Smaller variations in elevation perception:** Differences between mean values were smaller in the vertical plane (elevation angle) than in the horizontal plane (azimuth angle). This suggests that listeners perceived less variation in the up-down positioning of the sound source (Table 3).
- **Largest errors for front-centered positions:** The greatest differences between mean listener responses and the reference values occurred for front-central positions (1–5) (Table 3).

These findings reinforce existing knowledge about the challenges of front-facing localization in binaural rendering and suggest that sound sources positioned behind the listener tend to be perceived with higher precision.

6. Conclusions

The results of this study demonstrate variations in sound source localization accuracy depending on the spatial position of the source and the binaural rendering method.

Participants frequently reported difficulty in accurately placing the sound source in front of them. In many cases, they perceived the sound as being "inside the head," which aligns with the known phenomenon of lateralization [27]. On the other hand, sounds coming from the sides were easier to localize, and participants had no difficulty determining the primary direction (left or right).

Significant differences were noted among the various binaural renderers. The Dolby Renderer, utilized in Dolby Atmos technology, demonstrated the highest accuracy, with listener responses being the closest to the expected localization points.

Additionally, VR technology and spatial audio play a crucial role in advancing psychoacoustic research. The combination of these technologies enhances experimental control while also significantly improving participant immersion, leading to more reliable and naturalistic data collection.

Future research could explore personalized HRTF (Head-Related Transfer Function) adjustments to enhance front-localization precision and investigate the role of head movements in refining spatial perception.

applied. The study and experiments were conducted with complete scientific integrity and neutrality. This research was neither funded nor influenced by the manufacturers of the tested solutions, ensuring the absence of any conflict of interest.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lee, H. A Conceptual Model of Immersive Experience in Extended Reality, 2020. https://doi.org/10.31234/osf.io/sefkh.
2. Sauer, J. Acoustic Virtual Reality: An Introductory Research for Application. Technical report, Delft University of Technology, 2020.
3. Sharma, N.K.; Gaznepoglu, Ü.E.; Robotham, T.; Habets, E.A.P. Two Congruent Cues Are Better Than One: Impact of ITD–ILD Combinations on Reaction Time for Sound Lateralization. *JASA Express Lett.* **2023**, *3*, 054401. https://doi.org/10.1121/10.0017936.
4. Sagasti, A.; Pietrzak, A.; Martin, R.; Eguinoa, R. Localization of Sound Sources in Binaural Reproduction of First and Third Order Ambisonics. *Vib. Phys. Syst.* **2023**, *33*, 2022214.
5. Mróz, B.; Kostek, B. Pursuing Listeners' Perceptual Response in Audio-Visual Interactions—Headphones vs. Loudspeakers: A Case Study. *Arch. Acoust.* **2022**, *47*, 71–79.
6. Moraes, A.N.; et al. The Role of Physiological Responses in a VR-Based Sound Localization Task. *IEEE Access* **2021**, *9*, 122082–122091.
7. Tang, J. Research on the Application of Sound in Virtual Reality. *Highlights Sci. Eng. Technol.* **2023**, *44*, 206–212.
8. Olivieri, F.; Peters, N.; Sen, D. Scene-Based Audio and Higher Order Ambisonics: A Technology Overview and Application to Next-Generation Audio, VR, and 360 Video. Technical report, EBU Tech., 2019.
9. Sochaczewska, K.; Małecki, P.; Piotrowska, M. Evaluation of the Minimum Audible Angle on Horizontal Plane in 3rd Order Ambisonic Spherical Playback System. In Proceedings of the Immersive and 3D Audio: From Architecture to Automotive, 2021.
10. Bosman, I.; Buruk, O.; Jørgensen, K.; Hamari, J. The Effect of Audio on the Experience in Virtual Reality: A Scoping Review. *Behav. Inf. Technol.* **2023**. https://doi.org/10.1080/0144929X.2022.2158371.
11. Gari, S.V.A.; Schissler, C.; Robinson, P. Perceptual Comparison of Efficient Real-Time Geometrical Acoustics Engines in Virtual Reality. In Proceedings of the AES Int. Audio Games Conf., 2024.
12. Blauert, J.; et al. An Interactive Virtual-Environment Generator for Psychoacoustic Research. I: Architecture and Implementation. *Acta Acust. United Acust.* **2000**, *86*, 94–102.
13. Małecki, P.; Stefańska, J.; Szydłowska, M. Assessing Spatial Audio: A Listener-Centric Case Study on Object-Based and Ambisonic Audio Processing. *Arch. Acoust.* **2024**, *49*.
14. Bahu, H.; Carpentier, T.; Noisternig, M.; Warusfel, O. Comparison of Different Egocentric Pointing Methods for 3D Sound Localization Experiments. *Acta Acust. United Acust.* **2016**, *102*, 107–118.
15. Shum, L.C.; Valdés, B.A.; Loos, H.M.V.D. Determining the Accuracy of Oculus Touch Controllers for Motor Rehabilitation Applications Using Quantifiable Upper Limb Kinematics: Validation Study. *JMIR Biomed. Eng.* **2019**, *4*, e12291. https://doi.org/10.2196/12291.
16. Ahrens, A.; Lund, K.D.; Marschall, M.; Dau, T. Sound Source Localization with Varying Amounts of Visual Information in Virtual Reality. *PLoS ONE* **2019**, *14*, e0214603.
17. Chandler, D.; Grantham, W. Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. *J. Acoust. Soc. Am.* **1992**, *91*, 1624–1636.
18. Dalgarno, B.; Lee, M.J.W. What are the learning affordances of 3-D Virtual Environments? *Br. J. Educ. Technol.* **2010**, *40*, 10–32.
19. Kapralos, B.; Jenkin, M.; Milios, E. Virtual Audio Systems. *Presence Teleoperators Virtual Environ.* **2008**, *17*, 527–549.
20. Shilling, R.; Shinn-Cunningham, B.G. Virtual Auditory Displays. In *Handbook of Virtual Environments*; Stanney, K., Ed.; Lawrence Erlbaum Associates Inc, 2000.
21. Hartmann, W.M.; Wittenberg, A. On the externalization of sound images. *J. Acoust. Soc. Am.* **1996**, *99*, 3678–3688. https://doi.org/10.1121/1.414965.
22. Zotter, F.; Frank, M. All-Round Ambisonic Panning and Decoding. *J. Audio Eng. Soc.* **2012**, *60*, 807–820.
23. Pfanzagl-Cardone, E., The Dolby 'Atmo' System. In *The Art and Science of 3D Audio Recording*; Springer Int. Publ.: Cham, 2023; pp. 143–188.

24. Roffler, S.K.; Butler, R.A. Factors that influence the localization of sound in the vertical plane. *Journal of the Acoustical Society of America* **1968**, *43*, 1255–1259.

25. Circular Statistics Toolbox (Directional Statistics). Available at: https://www.mathworks.com/matlabcentral/fileexchange/10676-circular-statistics-toolbox-directional-statistics, accessed: Feb. 21, 2025.

26. Distance between two points on sphere **2025**. Accessed: 2025-03-20.

27. Freyman, R.L.; Balakrishnan, U.; Zurek, P.M. Lateralization of Noise-Burst Trains Based on Onset and Ongoing Interaural Delays. *J. Acoust. Soc. Am.* **2010**, *128*, 320–331. https://doi.org/10.1121/1.3436560.