VPS VIBRATIONS IN PHYSICAL SYSTEMS

DOI: 10.21008/j.0860-6897.2025.2.13
Vibrations in Physical Systems, 2025, 36(2), 2025213

Effects of loudspeaker count and layout on the perception of an ambisonic sound scene

Mateusz ZYCH, Karolina PONDEL-SYCZ, Daniella ŚCIESZKO, Łukasz POKORZYŃSKI, Andrzej BUDNY, Agnieszka Paula PIETRZAK

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

Corresponding author: Mateusz Zych email: mateusz.zych2.stud@pw.edu.pl

Abstract Multichannel loudspeaker arrays form the basis for reproducing spatial sound scenes. When reproducing signals in an ambisonic format, the number and placement of loudspeakers must be matched to the ambisonic order to ensure an adequate representation of the acoustic field. However, cost and space constraints often force a trade-off between the number of loudspeakers and the quality of the listening experience. The aim of this study was to assess the extent to which the number and arrangement of loudspeakers affects the accuracy of virtual sound-source localization and the subjective spatial impression expressed on a Mean Opinion Score (MOS) scale. The experiment was conducted in an anechoic chamber using four configurations containing 4 to 10 loudspeakers distributed around the listener. The participants consisted of non-expert listeners. Test signals consisted of one-second bursts of pink noise, repeated twice and presented from eight azimuth angles. For the quality rating, ambisonics sound scenes were rated. FOA was used for both tasks. Participants indicated the perceived source direction by marking a point on an angular dial and rated the spatial impression on a 1–5 MOS scale. The paper presents the dependence of localization errors on the number and configuration of loudspeakers and indicates the influence of each configuration on overall spatial-impression ratings.

Keywords: ambisonics, spatial audio reproduction, azimuth localization, MOS.

1. Introduction

In recent years, research on the spatial-audio analysis and synthesis has produced a range of methods enhancing the realism and immersive quality of music and speech reproduction. Multichannel loudspeaker configurations remain the foundation for immerse reproduction of sound scenes. Growing expectations for spatial audio draw attention to formats that enable faithful scene reproduction under limited resources. Ambisonics encodes the soundfield into spherical harmonics [1, 2], where the sound scene is described by a set of coefficients (B-format), whose number depends on the order of Ambisonics. At the decoding stage, these coefficients are transformed into signals for any loudspeaker configuration. Unlike traditional multichannel solutions, where each channel requires a dedicated loudspeaker in a specific position, this representation is independent of the layout and quantity of target loudspeakers. In practice however, spatial resolution, the perceived spatial quality and accuracy of sound source localization, depends deeply on the number and distribution of loudspeakers [3]. For a 2-D horizontal rendering, at least 2M+1 loudspeakers are typically required for an ambisonic M-order ambisonics (and $(M+1)^2$ for full 3-D) to provide appropriate spatial resolution with available loudspeaker count [4].

A central physical constraint is spatial aliasing, which emerges when loudspeaker spacing becomes too large relative to wavelength [5]. Aliasing perturbs interaural time and level differences and thereby localization, especially away from the array center and at higher frequencies [3]. Empirically, localization degrades as spacing increases, and different sound-field synthesis methods respond differently to those artifacts across the listening area. Beyond spacing, the placement and layout can also produce measurable differences in localization and listener preference [3].

Methodologically, two types of tests are used to evaluate spatial audio quality: localization tasks and opinion-based ratings. One method for evaluating the quality of sound reproduction is the sound source localization test with broadband stimuli. In such tests listener is seated in an acoustically isolated room (typically an anechoic chamber) at the centre of the loudspeaker array and is tasked with indicating the direction of incoming stimuli. The most used [6] stimuli are twice repeated, several-second sequences of pink noise bursts, thanks to their broadband spectrum and more balanced energy distribution compared with other noises (e.g., white noise). Performance is typically described by the absolute angular error between the true source direction and the listener's response.

A method widely used for evaluating the subjective quality of a spatial sound is Mean Opinion Score (MOS) test, in which listeners rate each stimulus on a five-point scale (1="very poor", 5="excellent") based on the overall spatial quality of the sample [7]. Stimuli are played one by one in random order, at the same level and a short training.

The aim of this paper is to investigate how the number and arrangement of loudspeakers affects the perception of ambisonic scenes. The study focuses on 2-D horizontal rendering and evaluates the sound-source localization using broadband pink-noise bursts, summarized by mean absolute error (MAE), and opinion ratings of perceived spatial quality using a five-point Mean Opinion Score (MOS) scale under matched playback level and brief training.

2. Methodology

The study investigated how loudspeaker configuration influences the accuracy of sound-source localization and perceived spatial impression. Experiments were conducted in an anechoic chamber using planar, horizontal arrays comprising 4, 5, 6, or 10 loudspeakers drawn from a calibrated 10-speaker setup.

2.1. Participants

Seventeen adult volunteers (age range 21–35 years, mixed gender) took part in the study. The group consisted of non-expert listeners with little prior critical-listening experience. All participants self-reported normal hearing with no identifiable hearing deficits. No visual deprivation (e.g., blindfold) was applied. Participants were instructed to fixate forward and minimize head movements during trials.

2.2. Loudspeaker layout and calibration

Playback employed ten Genelec 8320A studio monitors mounted on stands and networked to a workstation via DANTE over Ethernet. The system was calibrated with Genelec GLM 5 [8]. Loudspeaker positions were entered in the GLM user interface, and automated level alignment and frequency-response correction (appropriate to the anechoic environment) were applied. Stimuli were presented from REAPER with FOA ambisonic decoding via the AllRA Decoder plug-in [9]. For each configuration, decoding used the corresponding loudspeaker subset while maintaining identical calibration settings. Overall playback level was kept constant across conditions.

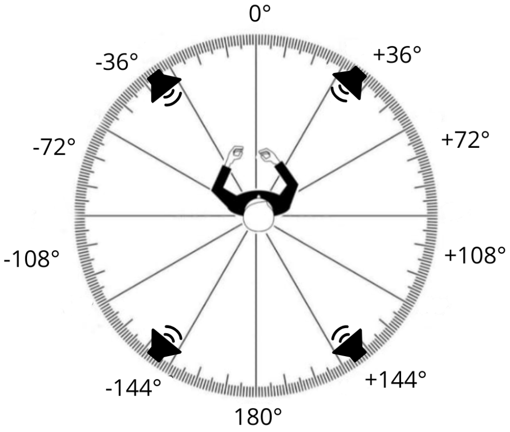
Loudspeakers were arranged on a horizontal circle at 0° elevation with the listener centered at the "sweet spot", with ear height ≈ 1.20 m. The array had a 1.4 m diameter. Azimuth increased clockwise, with positions aligned to a 36° grid (0°, 36°, …, 324°). From the full 10-speaker ring, the 4-, 5-, 6-, and 10-speaker layouts were realized as equiangular subsets around the listener (Fig. 1). The seemingly uneven loudspeaker layouts for the 4- and 6-speaker configurations were derived as equiangular subsets of the full 10-speaker ring. This approach ensured that all conditions shared identical physical loudspeaker positions and calibration parameters, minimizing possible differences in level, distance, and room acoustics

2.3. Test procedure

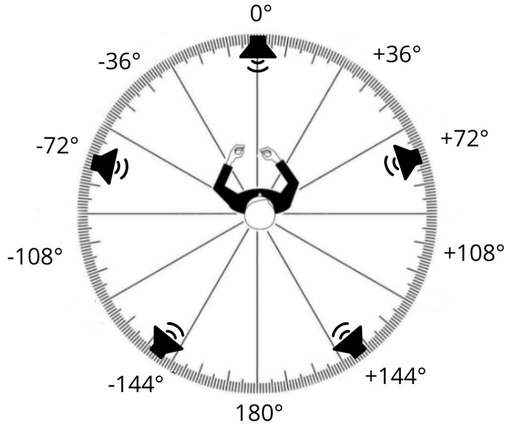
Fifteen participants completed the localization task and seventeen completed the spatial-impression rating task for each loudspeaker configuration. For the localization task, on each trial a 1 s pink-noise bursts were presented twice from one of eight azimuths uniformly distributed around the listener. After each presentation, participants indicated the perceived source direction by marking the localization on a circular chart, with the angles in degrees (0° = straight ahead, increasing clockwise). For each configuration, participants completed 8 azimuth tasks. No feedback was given during the experiment.

The Mean Opinion Score (MOS) test was designed to assess how the number of loudspeakers influences the perceived spatial impression of the reproduced scene, complementing the objective localization analysis. Each listener rated six short spatial sound samples (music, environmental sounds) in each of the four loudspeaker configurations. For every sample, the four configurations were presented once each in a randomized order. Immediately after each presentation, the listener recorded a MOS on a 5-point discrete scale (1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = excellent). Short practice trials familiarized listeners with the interface.

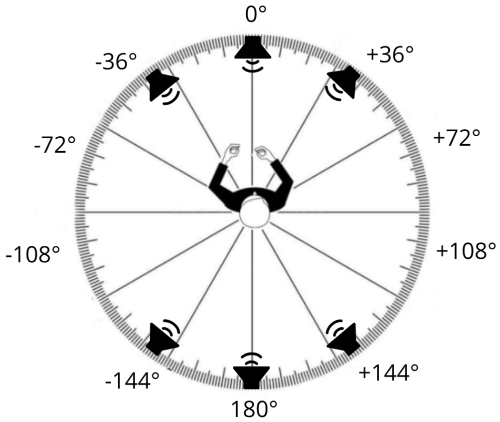
 DOI: 10.21008/j.0860-6897.2025.2.13



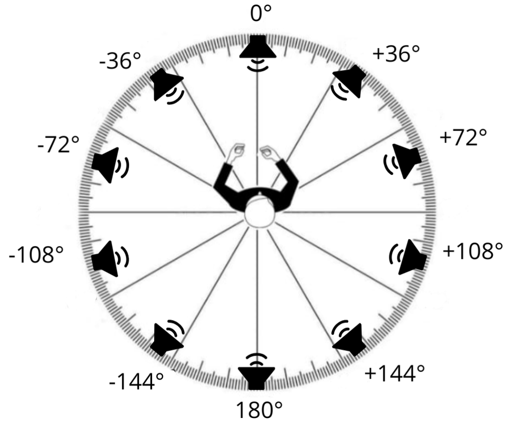
a) 4-speaker configuration



b) 5-speaker configuration



c) 6-speaker configuration



d) 10-speaker configuration

Figure 1. Loudspeaker configurations used in the study:
(a) 4 speakers, (b) 5 speakers, (c) 6 speakers, (d) 10 speakers.

2.4. Measures and statistical analysis

Localization performance was quantified from the angular error between response and target. The signed circular error $\Delta\theta$ was computed as the minimal wrapped difference in (−180°, 180°), and the absolute error was $|\Delta\theta|$. For each participant and configuration mean absolute error (MAE) was derived, as well as median absolute error, bias (mean signed error, positive = rightward/clockwise), and SD of the signed error. The proportion of trials with $|\Delta\theta| \leq 10°$, 15°, and 30° was also reported. Statistical inference operated on within-subject summaries (per-participant metrics). Pairwise configuration differences were assessed with Wilcoxon signed-rank tests. Effect sizes for paired contrasts are reported as Cohen's dz (mean within-subject difference divided by its SD). All angles are expressed in degrees.

The primary MOS outcome for inference was the per-participant mean MOS per configuration, obtained by averaging that participant's six ratings (one per sample) for a given layout. Descriptive statistics (mean, median, SD, IQR) were computed per configuration. Pairwise differences between configurations were evaluated using Wilcoxon signed-rank tests on the per-participant means. For interpretability, paired-samples effect sizes were reported as Cohen's dz.

3. Results

Localization performance and spatial-impression ratings are reported for four loudspeaker layouts (4, 5, 6, 10 speakers).

3.1. Localization accuracy

Considering the localization test, for each configuration, N = 120 localization responses were analysed (15 participants × 8 azimuths). Absolute-error distributions by loudspeaker configuration (per trial) are presented on Figure 2. Box-and-whisker plots summarize the absolute angular error for individual trials in each configuration (4-, 5-, 6-, and 10-speaker). The central line denotes the median, boxes span the interquartile range (IQR), and whiskers extend to 1.5×IQR beyond the quartiles. The distributions shift downward and become tighter as the number of loudspeakers increases, with the 10-speaker layout exhibiting the lowest typical errors and the 4-speaker layout the highest.

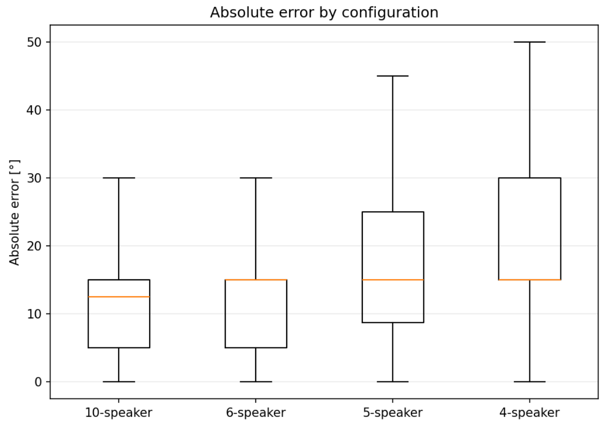


Figure 2. Distribution of localization absolute error across loudspeaker configurations. Central lines mark the median values.

Figure 2 provides an exploratory, per-trial view of variability. In the subsequent analyses, performance is quantified per participant and statistical tests are carried out on these within-subject summaries to ensure correct inference.

Table 1. Localization accuracy metrics by loudspeaker configuration. MAE is the per-participant mean of absolute errors. Bias is the mean signed error (positive = rightward).

Speakers configuration	N	MAE [°]	Median Error [°]	SD [°]	Bias [°]
4 speakers	120	23.6	15.0	32.2	3.9
5 speakers	120	17.3	15.0	21.7	7.8
6 speakers	120	16.0	15.0	24.7	-0.1
10 speakers	120	15.7	12.5	29.2	-2.4

Localization accuracy improved monotonically with the number of loudspeakers (Table 1). Mean absolute error (MAE) decreased from 23.6° with 4 speakers to 15.7° with 10 speakers, with 6 speakers already achieving 16.0° (nearly matching 10 speakers). Median errors were stable at 15° except for the 10-speaker case (12.5°). The signed bias was negligible for 6 speakers (−0.1°) and small for 10 speakers (−2.4°), whereas for 5 speakers a small bias (7.8°) attributable to the sparse discretization of directions was observed. Variability (SD of signed error) was largest for 4 speakers (32.2°) and smaller for 5–6 speakers (21.7–24.7°).

Consistency metrics based on absolute-error thresholds (Table 2) echoed these trends. The proportion of responses within ±15° rose from 63% (4 speakers) to 79–80% (6–10 speakers), while ±30° captures exceeded 90% for 6 and 10 speakers.

Table 2. Proportion of responses within absolute-error thresholds.

Speakers configuration	% of ≤10°	% of ≤15°	% of ≤30°
4 speakers	20%	63%	78%
5 speakers	34%	67%	90%
6 speakers	40%	79%	95%
10 speakers	50%	80%	93%

Pairwise, nonparametric comparisons of per-participant MAE (Table 3) showed that 4 vs 6 speakers yielded a very large benefit (ΔMAE = 7.57°, $p < 0.05$, $d_z = 1.52$). Differences 4 vs 5 (6.28°, $p = 0.0103$, $d_z = 0.71$) and 4 vs 10 (7.85°, $p = 0.0215$, $d_z = 0.69$) were also significant. In contrast, 5 vs 6, 5 vs 10, and 6 vs 10 were not significant ($p \geq 0.21$) with small effect sizes ($d_z \leq 0.16$). These differences indicate that that most of the localization benefit is realized when moving from 4 to 6 loudspeakers, with smaller advantages for further increases to 10.

Table 3. Pairwise comparisons of MAE between configurations.
Wilcoxon signed-rank tests on per-participant MAE. Positive ΔMAE values (A–B > 0) indicate configuration A has a larger (worse) MAE than configuration B.

Comparison	Δ MAE [°]	W (Wilcoxon)	p-value	N-paired	Cohen's dz
4 vs 5 speakers	6.28	15.5	0.01	15	0.71
4 vs 6 speakers	7.57	3.0	<0.05	15	1.52
4 vs 10 speakers	7.85	20.0	0.02	15	0.69
5 vs 6 speakers	1.28	36.5	0.21	15	0.16
5 vs 10 speakers	1.57	44.5	0.42	15	0.13
6 vs 10 speakers	0.28	47.5	0.52	15	0.02

3.2. Mean Opinion Score

Mean Opinion Scores (MOS) increased with the number of loudspeakers, with 10 speakers achieving the highest ratings (Mean = 4.1, Median = 4.0, SD = 1.0). The 5 speakers (3.8, 4.0, 1.0) and 6 speakers (3.7, 4.0, 1.0) configurations achieved slightly lower results. The 4-speaker layout yielded the lowest scores (3.4, 3.5, 1.2). Thus, listeners generally perceived more spacious and convincing scenes as the array became denser, while 5- and 6-speaker layouts delivered similar subjective quality (Table 4).

Table 4. Mean Opinion Scores (MOS) by loudspeaker configuration: number of ratings (N), mean, median, standard deviation (SD) and inter-quartile range (IQR).

Speakers configuration	N	Mean MOS	Median MOS	SD	IQR
4 speakers	102	3.4	3.5	1.2	1.0
5 speakers	102	3.8	4.0	1.0	1.0
6 speakers	102	3.7	4.0	1.0	1.0
10 speakers	102	4.1	4.0	1.0	1.0

Paired comparisons of per-participant MOS indicate that using more than 4 loudspeakers improves perceived spatial quality. Relative to 4 speakers, every other setup achieved higher MOS. The differences were statistically significant for 4 vs 5 speakers (ΔMOS = −0.4, W = 11.0, $p = 0.03$, dz = 0.66) and 4 vs 10 speakers (ΔMOS = −0.6, W = 0.0, $p < 0.05$, dz = 1.17). While 6 speakers also tended to be preferred over 4 (ΔMOS = −0.3, W = 32.5, $p = 0.07$, dz = 0.52), the difference was not significant.

In contrast, pairwise differences among the higher-count layouts were small and not statistically significant in this sample, with 5 vs 6 (ΔMOS = +0.01, $p = 0.71$, dz = 0.09), 5 vs 10 (ΔMOS = −0.3, $p = 0.11$, dz = 0.41), and 6 vs 10 (ΔMOS = −0.4, $p = 0.10$, dz = 0.38) showing effect sizes in the negligible range. Listeners clearly preferred 10 over 4 loudspeakers and showed a moderate preference for 5 or 6 over 4.

Table 5. Pairwise comparisons of Mean Opinion Scores (MOS) between loudspeaker configurations. Mean ΔMOS (A–B) = MOS (A) – MOS (B), with negative values meaning configuration B was favoured.

Comparison	Mean Δ MOS	W (Wilcoxon)	p-value	N-paired	Cohen's dz
4 vs 5 speakers	-0.4	11.0	0.03	17	-0.66
4 vs 6 speakers	-0.3	32.5	0.07	17	-0.52
4 vs 10 speakers	-0.6	0.00	<0.05	17	-1.17
5 vs 6 speakers	0.1	59.5	0.71	17	0.09
5 vs 10 speakers	-0.3	12.0	0.11	17	-0.41
6 vs 10 speakers	-0.4	36.0	0.10	17	-0.38

Taken together, the MOS results indicate that the principal subjective improvement occurs when moving from 4-speaker setup to denser arrays, with 10 speakers clearly preferred over 4. Beyond 5 speakers, diminishing returns are observed: 5-, 6-, and 10-speaker layouts receive comparably high ratings, and their pairwise differences are modest in magnitude and not statistically significant.

4. Conclusions

This study examined how loudspeaker count, and layout can influence the azimuthal localization accuracy and the perceived spatial impression, measured by MOS. Across the four equiangular layouts (4, 5, 6, 10 speakers), both objective and subjective outcomes improved as the array became denser, but the gains were non-linear.

On the localization task, the most consequential step was increasing the array from 4 to 6 speakers, which reduced mean absolute error by roughly 7–8° and produced the most significant pairwise improvement. The 10-speaker layout delivered accuracy close to the 6-speaker setup, indicating less significant improvement under the presented conditions. Variability was highest with 4 speakers, lower for 5–6, and somewhat higher again for 10, suggesting that adding a few well-placed sources stabilizes responses more effectively than simply maximizing count.

On the MOS test, ratings increased with loudspeaker count. The 10-speaker setup was clearly preferred to 4, while 5 and 6 achieved similar scores and did not differ reliably from 10. Together with the localization results, this points to a cost-effective operating point around six equiangular loudspeakers. Such setup captures most of the perceptual and objective benefits for the horizontal-only 1-order ambisonic scenes.

These findings are specific to the anechoic, horizontal-plane, pink-noise paradigm with the AllRA decoding pipeline and a fixed array diameter. Generalization to reverberant rooms, program material, other decoders or elevation components warrants further work. Future studies should also separate 1-order vs 2-order ambisonics outcomes, explore non-uniform layouts optimized for listener bias, include head tracking or visual control (e.g., blindfolds), and test larger, more diverse panels. Within these limits, the present evidence supports the practical recommendation that a well-calibrated 6-speaker array offers an attractive balance between accuracy, listener preference, and system cost, while a 10-speaker ring provides incremental, but smaller, gains primarily in subjective quality over sparse layouts.

Acknowledgments

This research was funded by the Rector's Grant for student research groups at Warsaw University of Technology (WUT) (Decision No. 59/2024 of the Rector of Warsaw University of Technology dated March 25, 2024), awarded to the Electroacoustics Student Research Group.

Additional information

The authors 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. D. Arteaga; Introduction to Ambisonics; Escola Superior Politècnica, Universitat Pompeu Fabra, 2015
2. F. Zotter, M. Frank; Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality; Springer Nature, 2019; DOI: 10.1007/978-3-030-17207-7
3. H. Wierstorf, A. Raake, S. Spors; Assessing localization accuracy in sound field synthesis; J. Acoust. Soc. Am., 2017, 141(2), 1111–1119; DOI: 10.1121/1.4976061
4. S.-N. Yao; Influence of loudspeaker configurations and orientations on sound localization; Arch. Acoust., 2022, 47(1), 57-70; DOI: 10.24425/aoa.2022.14073
5. S. Spors, R. Rabenstein; Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis; In: Proceedings of the 120th Audio Engineering Society Convention, Paris, France, 20–23 May 2006; Audio Engineering Society, 2006
6. O. Santala, V. Pulkki; Directional perception of distributed sound sources; J. Acoust. Soc. Am., 2011, 129(3), 1522–1530; DOI: 10.1121/1.3543949
7. ITU-T Recommendation P.800.1; Mean Opinion Score (MOS) Terminology; ITU-T, 2016
8. Genelec; GLM software; https://www.genelec.com/glm (accessed on 2025.08.30)
9. F. Zotter, M. Frank; All-round ambisonic panning and decoding; J. Audio Eng. Soc., 2012, 60(10), 807–820