Audio Engineering Society

Convention Express Paper 447

Presented at the AES 160th Convention
2026 May 28–30, Copenhagen, Denmark

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Zylia ZM-1 vs. Harpex Spcmic: A Case Study of Higher-Order Ambisonic Recording Performance

Bartłomiej Mróz[1] and Szymon Zaporowski[1]

[1]*Department of Multimedia Systems, Gdańsk University of Technology*

Correspondence should be addressed to Bartłomiej Mróz (bartlomiej.mroz@pg.edu.pl)

ABSTRACT

The Zylia ZM-1 (≈103 mm spherical array, 19 MEMS capsules, 3rd-order) and Harpex Spcmic (170 mm planar array, 84 MEMS capsules, 5th-order capable) represent two distinct geometrical approaches to higher-order Ambisonics capture. Despite widespread adoption, systematic comparison of their performance under real recording conditions is absent from published literature. We present a controlled case study based on simultaneous recordings of solo piano repertoire in a 370-seat concert hall (Aula Politechniki Gdańskiej, RT60 = 1.97 s), with both arrays mounted 17 cm apart on a single stereo bar. After peak-normalization to −0.5 dBTP, four metrics were compared: W-channel spectral response, integrated loudness (ITU-R BS.1770-5), per-order spatial energy distribution, and first-order directional ratios. The headline finding concerns spatial energy: at nominally identical 3rd-order, the ZM-1 attenuates 27.4 dB from 0th to 3rd order, while the Spcmic attenuates only 8.4 dB—a ∼19.0 dB gap that is not driven by aliasing-band content and persists across repertoire. The results indicate that nominal Ambisonics order alone does not predict per-order energy distribution in real recordings; geometry and encoding contribute substantially even when the format label is identical. Recordings are openly available under CC BY-NC-SA 4.0 (10.34808/5xe2-ah94).

1 Introduction

Ambisonics encodes the sound field as a truncated spherical-harmonic expansion around a single reference point [1, 2]. The proliferation of higher-order Ambisonics (HOA) microphone arrays has democratized spatial audio capture, with the Zylia ZM-1 becoming particularly widespread in research and production due to its accessibility and USB-powered operation. Simultaneously, professional-grade arrays like the Harpex Spcmic offer substantially higher channel counts and larger apertures at significantly higher cost.

Related work. Foundational work on validation of higher-order Ambisonic recording was established by Moreau, Daniel and Bertet, who derived the encod-

ing filters and reported objective measurements for a fourth-order spherical array [3]. Two subsequent comparative studies are particularly relevant. Bates and colleagues [4, 5] benchmarked first- and higher-order ambisonic microphones—the SoundField MKV, Core Sound TetraMic, Sennheiser Ambeo, MH Acoustics Eigenmike and Zoom H2n—using MUSHRA listening tests against a reference DPA 4006. Pinardi and Farina [6] compared four spherical arrays (the Zylia ZM-1, the Eigenmike em32, and two B&K arrays of 36 and 50 capsules) using objective spatial-accuracy metrics derived from anechoic measurements. The dataset of McKenzie, McCormack and Hold [7] provides spatial room impulse responses captured simultaneously with the em32 and the ZM-1 in a variable-acoustics room, illustrating that side-by-side multi-array capture is now an accepted methodology. None of these studies, however, includes a planar large-aperture array of the Spcmic class, nor does any compare arrays of substantially different geometry under real concert-hall conditions: that is the gap the present paper addresses.

While both systems nominally produce standardized Ambisonics formats, systematic comparison of their performance characteristics using real musical content in authentic acoustic environments remains absent from published literature. Existing validation studies have focused primarily on anechoic chamber measurements with test signals or frequency-dependent spatial accuracy metrics, leaving questions about energy distribution across Ambisonics orders—a fundamental characteristic affecting reproduction quality—unanswered. This case study addresses this gap through controlled simultaneous recordings of piano recitals, examining spectral response, loudness characteristics, spatial energy distribution, and directional component behavior.

2 Methods

Two MEMS-based microphone arrays with contrasting geometries were compared: the Zylia ZM-1 [8] (19 capsules in a spherical arrangement; outer-shell diameter ≈103 mm with capsules recessed slightly inside the housing—the SPARTA `array2sh` encoder preset models the device as a rigid sphere of radius 49 mm with capsules flush-mounted [9, 10]; 3rd-order maximum) and the Harpex Spcmic [11] (84 capsules arranged 42 per face on a 170 mm × 5.6 mm planar substrate, with on-board temperature compensation and orientation sensing; the manufacturer specifies 3rd-order ambisonic and 5th-order beamforming resolution). Both devices ship with factory per-sensor calibration: the Spcmic stores its calibration table in device-resident non-volatile memory (re-downloadable from the manufacturer), and the ZM-1's A-to-B conversion is performed by the Zylia Ambisonics Converter, which applies the manufacturer's encoding configuration for the device. Independent verification of either calibration—for example by anechoic re-measurement or by comparison against modelled spherical-harmonic encoders such as those in the SPARTA suite [10] or against the modelled-versus-measured framework of Politis and Gamper [12]—was outside the scope of this case study; we therefore treat any residual calibration mismatch as part of the system-level differences observed.

Arrays were mounted on a single stereo bar 17 cm apart, ensuring acoustically identical capture positions while maintaining spatial separation to avoid acoustic interference. Recording sessions occurred in Aula Politechniki Gdańskiej, a 370-seat concert hall with measured reverberation time of 1.97 s, on two dates: 15 August 2024 (César Franck: Prélude, Choral et Fugue; Sergei Prokofiev: Piano Sonata No. 4, Op. 29; total 35.5 minutes) and 30 April 2024 (Alberto Ginastera: Sonata No. 1, Op. 22; 15.4 minutes). Source material comprised solo piano performed on a Steinway D concert grand. Both sessions placed the arrays approximately 1.5 m from the soundboard (Fig. 1). All recordings were processed through manufacturer-provided A-to-B format conversion software (Zylia Ambisonics Converter for the ZM-1; Spacemic VST `1.0.2-alpha` for the Spcmic 5OA path; the Spcmic app `0.9.1-beta` for the 3OA path) and peak-normalized to −0.5 dBTP. All B-format files reported here use the ACN channel ordering with SN3D normalisation (the ambiX convention) [13]. The Spcmic was encoded to native 5th-order (36 channels) for both sessions and additionally to truncated 3rd-order (16 channels) for the August session, enabling direct comparison with the ZM-1 at equivalent order specification; for the April session only the 5th-order render is available[1].

[1]The April session was captured as raw 84-channel WAV in Reaper, treating the Spcmic as a USB audio interface, rather than via the manufacturer's standalone Spcmic app, which writes a native `.spcmic` container with embedded temperature and orientation metadata [11] and renders it to formats up to 3OA. The Spcmic app accepts only its own `.spcmic` files or live microphone input as sources, so it could not retroactively decode the April raw 84-channel



Fig. 1: Recording setup, August session: the Spcmic (centre, planar disc) and one of the two Zylia ZM-1 units used for the comparison (right). A second ZM-1 visible at left belongs to an unrelated parallel comparison and is not analysed in this paper.

Four metrics were analyzed: (1) W-channel spectral response, (2) integrated loudness (LUFS-I per ITU-R BS.1770-5 [14]; multichannel summation as performed by the rendering DAW), (3) spatial energy distribution across Ambisonics orders, and (4) first-order directional component ratios. All numeric values reported below are computed directly from the rendered B-format WAV files by a single analysis script and re-emitted to LaTeX on every run, ensuring full reproducibility.

3 Results

Spectral response examination via W-channel power spectral density (1/6-octave smoothed, normalized to 0 dB at 1 kHz) identified a 5.2–11.0 dB mid-frequency elevation in the 200–600 Hz band in the ZM-1 relative to the Spcmic, with a band-mean of 9.1 dB—consistent with resonance characteristics of smaller MEMS diaphragms in the compact spherical geometry (Fig. 2).

Integrated loudness measurements (LUFS-I per ITU-R BS.1770-5, summing energy across all rendered

files. The Spacemic VST plug-in (version `1.0.2-alpha`), which does accept raw 84-channel input and renders up to 5OA, post-dates the Spcmic app version (`0.9.1-beta`) used for the August session and was used to produce the April 5OA render. We therefore use the August session as the primary 3OA-vs-3OA comparison and the April Ginastera session as a 5OA-only consistency check.

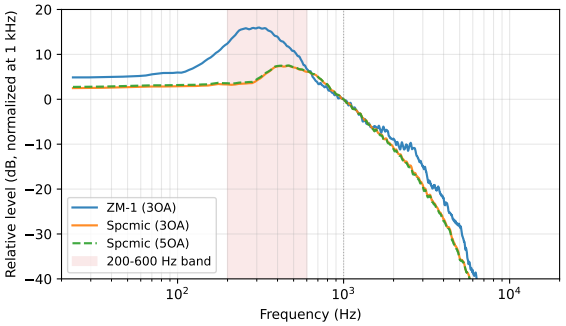


Fig. 2: W-channel power spectral density (1/6-octave smoothed, normalised to 0 dB at 1 kHz) for the ZM-1 and Spcmic on the August session. The ZM-1 shows a 5.2–11.0 dB elevation in the 200–600 Hz band relative to the Spcmic.

Ambisonics channels) demonstrated the Spcmic output yielded 2.3–3.3 dB higher loudness than the ZM-1 despite identical peak normalization (−0.5 dBTP). We note that BS.1770's K-weighting and channel-summation rules were not designed for ambisonic content, and recent work indicates that they under-estimate perceived loudness for scene-based audio [15, 16]; we therefore report multichannel LUFS-I as a comparison metric between two arrays under identical processing rather than as a perceptually calibrated absolute level. Notably, restricting the loudness measurement to the omnidirectional W channel inverts this relationship: on the W channel alone the ZM-1 is 0.8–2.6 dB louder than the Spcmic, confirming that the Spcmic's apparent loudness advantage is carried entirely by its higher-order spherical-harmonic components rather than by the omnidirectional pressure.

Most significantly, spatial energy analysis—computed as the linear-mean per-channel RMS within each Ambisonics order, expressed in dBFS—revealed the ZM-1 exhibits 27.4 dB attenuation from 0th to 3rd order on the Franck recording (26.0 dB on the Prokofiev recording), while the Spcmic 3OA shows only 8.4 dB and 8.5 dB respectively (Fig. 3). This ∼19.0 dB gap (Franck) persists despite both arrays producing nominally identical "3rd-order Ambisonics" formatted output; consistency across both pieces and the additional Ginastera session (ZM-1 $0 \rightarrow 3$ rolloff 22.1 dB) confirms this is systematic array behaviour rather than a repertoire-specific artefact. The same calculation restricted to the sub-aliasing band (below ≈3000 Hz, the

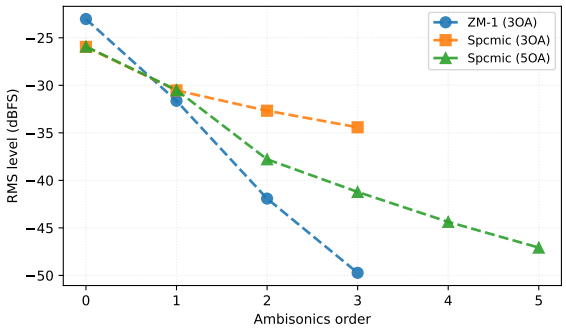


Fig. 3: Per-order spatial energy distribution (linear-mean RMS per Ambisonics order, dBFS) for the ZM-1 and the Spcmic at 3OA and 5OA on the August session. The ZM-1 shows 27.4 dB attenuation from 0th to 3rd order (Franck) versus 8.4 dB for the Spcmic 3OA, despite both producing nominally identical 3rd-order Ambisonics output.

SPARTA-implied ZM-1 spatial-aliasing limit) yields a $0 \to 3$ rolloff gap of 19.3 dB, indicating the difference is not driven primarily by aliasing-band content.

Directional component analysis further revealed the Spcmic exhibits a markedly stronger front-back first-order component (X/W ratio ≈ 0.83) compared to the ZM-1 (≈ 0.55), with similar trends for the elevation (Z/W: 0.57 vs 0.32) and lateral (Y/W: 0.38 vs 0.24) components, all correlating with the gentler overall energy rolloff pattern.

4 Discussion

The results show that two arrays producing nominally identical "3rd-order Ambisonics" output can differ substantially in per-order energy distribution, with the difference traceable to geometry and encoding rather than to the format label. For users comparing recordings across arrays—whether for engineering, research, or archival purposes—per-order energy distribution may be a more informative parameter than the nominal Ambisonics order alone.

Implications of planar versus spherical geometry. A planar microphone aperture samples the sound field on a two-dimensional manifold and therefore cannot, in principle, distinguish sources mirrored above and below that plane on the basis of pressure alone; the recovery of vertical spherical harmonics depends on diffraction over the array's finite thickness and on the spatial-aliasing properties of the chosen sampling scheme [17, 18, 19]. The Spcmic mitigates this in two ways: a high capsule count (84) pushes the spatial-aliasing frequency upward, and a 170 mm aperture yields a low modal-amplification cut-on. Crucially, however, the present comparison is dominated by horizontal-plane content—a piano located in front of the microphones in a hall whose strong reflections arrive primarily from horizontal surfaces—so the planar limitation is partially masked by the program material. Spherical harmonics that primarily encode elevation will, in general, be reconstructed differently in planar versus spherical geometries even when the nominal Ambisonic order is identical, and this is one factor contributing to the $\sim$19.0 dB order-3 spatial-energy gap between the two arrays despite both being labelled "3rd-order Ambisonics" [3, 2].

The objective measures reported here describe how the two arrays differ but cannot establish whether those differences are audible. A controlled perceptual evaluation—e.g. a MUSHRA protocol [20] on binaural renderings of matched corpus excerpts, along the lines of Bates et al. [5]—is the natural next step.

5 Summary

This case study demonstrates substantial performance differences between the Zylia ZM-1 and Harpex Spcmic in real-world recording conditions, with a $\sim$19.0 dB gap in spatial energy rolloff despite both producing nominally equivalent 3rd-order Ambisonics output. The ZM-1 exhibits 27.4 dB attenuation from 0th to 3rd order on the Franck recording compared to 8.4 dB for the Spcmic 3OA, alongside a 5.2–11.0 dB W-channel spectral elevation at 200–600 Hz and 2.3–3.3 dB lower multichannel integrated loudness under identical peak normalization—an effect that reverses on the W channel alone, where the ZM-1 is 0.8–2.6 dB louder. These findings indicate that the nominal Ambisonics order does not, on its own, predict per-order energy distribution in practical recordings, and that per-order energy may be a more informative parameter when comparing arrays for recording, research, or archival purposes. This case study complements existing anechoic validation by demonstrating that performance differences manifest in authentic recording conditions with real musical content. The

recordings analysed here, together with the broader corpus from which they are drawn, are openly available under CC BY-NC-SA 4.0 from the Bridge of Data (Most Danych) repository at Gdańsk University of Technology [21] (`10.34808/5xe2-ah94`), together with the rendered B-format files. The analysis pipeline that produces every numeric value in this paper is mirrored at `git.pg.edu.pl/p829296/hoa-mic-comparison-aes2026/`.

Acknowledgements

This work was partially supported by the Polish National Centre for Research and Development (NCBR) under grant INFOSTRATEG4/0003/2022, "ADMED-VOICE — Adaptive Intelligent Speech Processing System of Medical Personnel with the Structuring of Test Results and Support of Therapeutic Process."

References

[1] Gerzon, M. A., "Periphony: With-Height Sound Reproduction," *Journal of the Audio Engineering Society*, 21(1), pp. 2–10, 1973.

[2] Zotter, F. and Frank, M., *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, volume 19 of *Springer Topics in Signal Processing*, Springer, Cham, Switzerland, 1 edition, 2019, ISBN 978-3-030-17206-0, doi:10.1007/978-3-030-17207-7, open Access (CC BY).

[3] Moreau, S., Daniel, J., and Bertet, S., "3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of Spherical Microphone," in *Audio Engineering Society Convention 120*, Paris, France, 2006, convention Paper 6857.

[4] Bates, E., Gorzel, M., Ferguson, L., O'Dwyer, H., and Boland, F. M., "Comparing Ambisonic Microphones: Part 1," in *Audio Engineering Society International Conference on Sound Field Control*, Guildford, UK, 2016.

[5] Bates, E., Dooney, S., Gorzel, M., O'Dwyer, H., Ferguson, L., and Boland, F. M., "Comparing Ambisonic Microphones – Part 2," in *Audio Engineering Society Convention 142*, Berlin, Germany, 2017, convention Paper 9730.

[6] Pinardi, D. and Farina, A., "Metrics for Evaluating the Spatial Accuracy of Microphone Arrays," in *2021 Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–9, IEEE, Bologna, Italy, 2021, doi:10.1109/I3DA48870.2021.9610887.

[7] McKenzie, T., McCormack, L., and Hold, C., "Dataset of Spatial Room Impulse Responses in a Variable Acoustics Room for Six Degrees-of-Freedom Rendering and Analysis," 2021, dataset: Zenodo, doi:10.5281/zenodo.5720724.

[8] Zylia Sp. z o.o., *ZYLIA ZM-1 User Manual*, Zylia Sp. z o.o., 2018, version 0.5.

[9] McCormack, L., Delikaris-Manias, S., Farina, A., Pinardi, D., and Pulkki, V., "Real-Time Conversion of Sensor Array Signals into Spherical Harmonic Signals with Applications to Spatially Localized Sub-Band Sound-Field Analysis," in *Audio Engineering Society Convention 144*, Milan, Italy, 2018, convention Paper 9939.

[10] McCormack, L. and Politis, A., "SPARTA & COMPASS: Real-Time Implementations of Linear and Parametric Spatial Audio Reproduction and Processing Methods," in *AES International Conference on Immersive and Interactive Audio*, York, UK, 2019, engineering Brief 111.

[11] Harpex Audio GmbH, *spcmic User Manual*, Harpex Audio GmbH, 2024.

[12] Politis, A. and Gamper, H., "Comparing Modeled and Measurement-Based Spherical Harmonic Encoding Filters for Spherical Microphone Arrays," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 224–228, New Paltz, NY, USA, 2017, doi:10.1109/WASPAA.2017.8170028.

[13] Nachbar, C., Zotter, F., Deleflie, E., and Sontacchi, A., "ambiX – A Suggested Ambisonics Format," in *Proceedings of the 3rd Ambisonics Symposium*, Lexington, KY, USA, 2011.

[14] International Telecommunication Union, "Recommendation ITU-R BS.1770-5: Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," Technical report, ITU Radiocommunication Sector, Geneva, Switzerland, 2023.

[15] Peters, N. and Epain, N., "Loudness Perception of Scene-Based Audio across Loudspeaker Configurations and HOA Orders," in *Audio Engineering Society Convention 154*, Espoo, Finland, 2023, convention Paper 10658.

[16] Norcross, S. G., Nanda, S., and Cohen, Z., "ITU-R BS.1770 Based Loudness for Immersive Audio," in *Audio Engineering Society Convention 140*, Paris, France, 2016, convention Paper 9500.

[17] Rafaely, B., "Analysis and Design of Spherical Microphone Arrays," *IEEE Transactions on Speech and Audio Processing*, 13(1), pp. 135–143, 2005, doi:10.1109/TSA.2004.839244.

[18] Rafaely, B., Weiss, B., and Bachmat, E., "Spatial Aliasing in Spherical Microphone Arrays," *IEEE Transactions on Signal Processing*, 55(3), pp. 1003–1010, 2007, doi:10.1109/TSP.2006.888896.

[19] Poletti, M. A., "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics," *Journal of the Audio Engineering Society*, 53(11), pp. 1004–1025, 2005.

[20] International Telecommunication Union, "Recommendation ITU-R BS.1534-3: Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," Technical report, ITU Radiocommunication Sector, Geneva, Switzerland, 2015.

[21] Mróz, B. and Zaporowski, S., "Higher-Order Ambisonics Recording Corpus," 2026, doi:10.34808/5xe2-ah94, dataset deposit, CC BY-NC-SA 4.0.