forum acusticum 2026
8-12 September · Graz, Austria

ORDER-BASED SPATIAL DECOMPOSITION IN HIGHER-ORDER AMBISONICS FOR VEHICLE INTERIOR NOISE ANALYSIS

Zhenxian Li[1]
[1] *LVA, UR677, INSA Lyon, France*

Abstract

In multi-motor battery electric vehicles the loss of combustion masking makes the order-structured tonal noise of each drive unit perceptually salient, and units that share the same machine and single-speed gear ratio, such as the two wheel motors of one axle or the twin units of a symmetric dual-motor vehicle, hold their same-numbered orders within about one percent of each other whenever both are active. This spectral degeneracy prevents tracking-filter separation for any small number of channels, including the binaural pair used in standard NVH practice. We propose a spatial decomposition framework for third-order Ambisonic recordings. Phase-coherent tracked-harmonic extraction, driven by one instantaneous phase per shaft, is applied independently to every spherical-harmonic channel, and full-sphere direction-of-arrival maps attribute each separated component to its source. Where the order families are degenerate, a constrained adaptive spatial prefilter, whose spherical-harmonic steering vectors are independent of frequency, places nulls on the interfering motor and on the background before extraction. The framework is validated on a real anechoic-chamber recording and on two-motor scenes constructed from measured electric-drive order data and measured room impulse responses, and is benchmarked against binaural processing of the same fields. Where per-channel tracking, binaural processing and even exact-phase joint Vold-Kalman estimation remain limited, with weaker-motor correlations of 0.60–0.84, spatial-null preprocessing restores the separation to 0.94, and the recovered components retain their directions of arrival.

Keywords: *higher-order ambisonics, engine order tracking, electric vehicle NVH, spatial filtering*

Corresponding author: *zhenxian.li@insa-lyon.fr.*

1. Introduction

Interior noise remains central to perceived vehicle quality and is conventionally decomposed into powertrain, road/tyre and wind contributions [1]. Electrification removes the broadband combustion noise that masked the tones of rotating machinery: as overall levels fall, the tonal content of the electric drive, comprising electromagnetic force harmonics, gear-mesh whine and inverter switching sidebands, becomes perceptually dominant and a primary determinant of interior sound quality [2], [3]. Such components are termed *orders*, their frequencies being fixed multiples of a shaft speed.

Production battery-electric vehicles increasingly carry two or four drive units, each coupled to its axle through a fixed single-speed reduction, so that all motor speeds are locked to the vehicle speed by constant ratios. The consequences depend on the configuration. Units of different machine type or gear ratio, as in many mixed dual-motor designs, give order families offset by several to tens of percent, which remain spectrally separable. Units sharing machine and ratio, such as the two wheel motors of one axle in a tri- or quad-motor vehicle or the twin units of a symmetric dual-motor vehicle, instead stay within about one percent in a straight line, a budget set by tyre-radius tolerance and differential slip, and within a few percent between the outer and inner wheels of an axle in a corner. Whenever both units drive, as during acceleration, climbing or towing, the beat frequencies are of the order of a few hertz, while the units remain spatially separated from any cabin position. The multi-motor problem is then spectrally degenerate yet spatially well conditioned.

Established tools address the spectral axis only. Close or crossing orders of independently running shafts are the recognized difficult case for tracking filters [4], [5]: two components of the same order k are separable in time-frequency only if their spacing $k\,|\Delta f_{\mathrm{rot}}|$ exceeds the reciprocal of the analysis span, a bound independent of the number of channels. The standard capture instrument is the binaural dummy head [6], often combined with transfer-path analysis [7]; two fixed channels give only lateral cues and no post-measurement re-steering. Microphone arrays have been used in cabins for source-level mapping [8], but rarely to extract per-source order waveforms.

Higher-Order Ambisonics (HOA) provides the complementary spatial dimension. A compact spherical array encodes the field into spherical-harmonic (SH) channels [9], in which the steering vector of a direction is a short real-valued vector independent of frequency; this supports direction-of-arrival (DoA) analysis and spatial filtering after the measurement [10].

This paper examines what the spatial channels contribute to the multi-motor problem. An HOA-domain order/noise decomposition is developed in which tracked-harmonic extraction, driven by one phase per shaft, runs independently in every SH channel, each separated component retaining a direction of arrival; for the degenerate case a constrained adaptive spatial prefilter (LCMV [11], [12]) with frequency-independent SH steering nulls the interfering motor and the background before extraction. The method is validated on a real chamber recording and on two-motor scenes built from measured drive-unit data and room responses, against binaural and joint Vold-Kalman benchmarks, and its validity in a real cabin is analysed in Section 5. Beyond diagnosis, the decomposition has a reproduction-oriented use: since every separated component remains an HOA signal, individual motor orders and the background can be re-rendered binaurally or over loudspeaker arrays, providing order-resolved, spatially faithful source material for target-sound design and interactive replay in NVH driving simulators.

2. Method

2.1. HOA signal model and DoA maps

A third-order spherical microphone array encodes the field into $Q = (N+1)^2 = 16$ SH channels (N=3, ACN ordering, SN3D normalization), collected in $\mathbf{x}(t) \in \mathbb{R}^Q$. For a look direction $\mathbf{\Omega}$ with real-SH steering vector $\mathbf{y}(\mathbf{\Omega}) = [Y_1(\mathbf{\Omega}), \dots, Y_Q(\mathbf{\Omega})]^\top$, the steered-response (Bartlett) energy of a segment with covariance $\mathbf{C} = \frac{1}{L}\mathbf{X}^\top\mathbf{X}$ is

$$E(\mathbf{\Omega}) = \mathbf{y}(\mathbf{\Omega})^\top \, \mathbf{C} \, \mathbf{y}(\mathbf{\Omega}), \quad (1)$$

evaluated on a 2° azimuth-elevation grid and normalized to its peak. The steering vector does not depend on frequency: one set of 16 real coefficients steers the entire band.

2.2. Tracked-harmonic order extraction

The instantaneous rotational frequency $f_{\mathrm{rot}}(t)$ of one shaft defines its phase in revolutions, $\varphi(t) = \int_0^t f_{\mathrm{rot}}(\tau)\,\mathrm{d}\tau$. For each order k, a slowly varying complex amplitude is estimated per frame by demodulating against the tracked phase over a window of length T_w (short-time Fourier-coefficient tracking, STFS [13], [14]),

$$c_k[m] = \frac{\sum_n w[n]\, x(t_m{+}n)\, e^{-j2\pi k\varphi(t_m+n)}}{\sum_n w[n]}, \quad (2)$$

and the order signal is re-synthesized by overlap-add re-modulation, $\hat{o}(t) = \sum_k 2\,\mathrm{Re}\{c_k(t) e^{j2\pi k\varphi(t)}\}$. The residual is $\hat{r}(t) = x(t) - \hat{o}(t)$. In this work $T_w = 80\,\mathrm{ms}$ with 75 % overlap. The estimator generalizes computed and Vold-Kalman order tracking [4], [5] and shares their resolution bound: two sources contribute to the same coefficient c_k, and are indistinguishable, whenever their spacing $k|\Delta f_{\mathrm{rot}}| \lesssim 1/T_w$ (12.5 Hz here) persists. Joint multi-envelope estimators relax this bound, but only until the beat frequency falls within the bandwidth that the envelopes themselves require; this limit is examined in Section 4.3.

2.3. Per-channel decomposition and its limit

Applying Eq. (2) independently to all Q channels with a common phase per shaft yields separated order and residual fields that are themselves HOA signals, so their DoA maps can be compared with the references; this comparison is the spatial validation used throughout. Per-channel extraction exploits no spatial information, however: when two shafts lie within the resolution bound, every channel fails identically, irrespective of Q.

2.4. Spatial-null prefiltering

The spatial channels become useful when they are combined. Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_b$ denote the SH steering vectors of the two motors and of the dominant background direction, computed from the known installation geometry (Section 3.2). A linearly constrained minimum-variance (LCMV) weight vector [11], [12] for motor 1 is given by

$$\mathbf{w}_1 = \arg\min_{\mathbf{w}} \; \mathbf{w}^\top \hat{\mathbf{C}}\, \mathbf{w} \quad \text{s.t.} \quad \mathbf{A}^\top \mathbf{w} = [1, 0, 0]^\top, \quad (3)$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_b]$ and $\hat{\mathbf{C}}$ is the diagonally loaded covariance of the mixture [15], with loading $10^{-2}\,\mathrm{tr}\,\hat{\mathbf{C}}/Q$. This filter passes motor 1 with unit gain, places exact nulls on motor 2 and on the background, and uses the remaining $Q - 3 = 13$ degrees of freedom to minimize the residual output power. Since the steering is frequency-independent, $\mathbf{w}_1$ is a single 16-element vector applied by one inner product per sample, and the beam output $y_1(t) = \mathbf{w}_1^\top \mathbf{x}(t)$ is passed to the tracker of Section 2.2. The interfering motor is thus removed spatially, before the temporally degenerate extraction; spatial nulls, unlike main lobes, are not constrained by the order-N beamwidth.

The binaural counterpart makes the difference explicit: with two ear signals the steering vectors are frequency-dependent HRTF pairs, so the analogous filter must be solved in each frequency bin. Two channels and two constraints determine the weights uniquely. No degrees of freedom remain, so a second interferer cannot be nulled and the filter cannot adapt to the actual field; the 2×2 steering matrix also becomes increasingly ill-conditioned toward low frequency, where head shadowing vanishes. These limits follow from the channel count rather than from any implementation choice.

3. Experimental material

All processing is implemented in Python. The tracked-harmonic estimator of Eq. (2) follows the tonal-extraction framework of [14].

3.1. Anechoic-chamber recording

The measurement chain is validated on a real recording: a synthetic ICE-like order signal (orders $\{1, 2, 4, 6, 8\}$, RPM ramp 900–3000) and an independent band-limited noise, reproduced by two loudspeakers in an anechoic chamber and captured by a Zylia ZM-1 third-order array (16 channels, ACN/SN3D, 48 kHz). A synchronization pulse anchors three 10 s segments (noise only, order only, mixture), giving per-channel ground truth for the residual and the separated order.

3.2. Two-motor scenes from measured data

Multi-motor conditions were constructed with exact ground truth. The sources are the order signals of two production electric drive units, synthesized from complex order-amplitude profiles against RPM extracted from interior recordings of production vehicles: unit C1 carries orders $\{4, 8, 12, 16\}$ with a dominant fourth order, and unit C3 carries orders $\{4, 6, 8, 12, 14, 23\}$ with a dominant sixth order; peak amplitudes are 0.02–0.05 Pa. The background combines the road-noise autopower extracted from the same recordings, synthesized with random phase, with band-limited pink noise overlapping the order band (120 Hz–5 kHz), at 0.6 of the combined motor RMS. Each source is then convolved with a measured third-order SH room impulse response from the most absorptive configuration of the Arni variable-acoustics room (Zylia array, ACN/SN3D [16]), using the three source positions of a single receiver. Constructing the scenes by superposition is exact rather than approximate: the room is linear and time-invariant and the three sources are mutually independent, so the sum of the individually convolved signals is identical to the field the sources would have radiated simultaneously; this is the standard construction in spatial-audio evaluation. All SRIRs share one receiver position and array orientation, and their original propagation delays and levels are preserved, so that the relative arrival times and levels between the sources are physically consistent.
Three scenarios cover the range from spectrally separable to fully degenerate. In the *distinct* scenario the two units differ in type and profile, C1 ramping from 1200 to 5200 RPM while C3 rises from 2500 to 7000 RPM and returns, so the order families cross; the sources stand 82° apart. The *cornering axle pair* represents the outer and inner wheel units of one driven axle in a corner, a track-width-to-radius effect: two identical C1 units at a mean speed difference of 3.7 %, a radius of roughly 45 m at 40 km/h, placed at the closest available source pair (24°). In the *speed-locked pair*, two identical C1 units accelerate moderately in a straight line (about 30–60 km/h) with speeds locked 0.5 % apart, the combined budget of tyre-radius tolerance and differential slip; the fourth-order beat is then approximately 1 Hz and the pair occupies the widest available directions (85°). All source directions are surrogates constrained by the three loudspeaker positions of the SRIR set and are conservative for cabin geometry, in which structure-borne paths displace the apparent sources in any case (Section 5). Figure 1 illustrates the resulting degeneracy. The steering vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_b$ used in Eq. (3) are computed from the free-field spherical-harmonic model $\mathbf{y}(\boldsymbol{\Omega})$ of Eq. (1), evaluated at the direction of each source as determined by the measured positions of the loudspeakers and the microphone array; no acoustic calibration measurement and no information from the evaluated SRIRs is used to form them. These vectors are therefore the idealized free-field prediction from geometry alone, and are not adapted to the reflected field of the reverberant scenes used for evaluation: the separation figures of Section 4 already include the steering mismatch expected in a real installation, rather than benefiting from an oracle match to the simulated room.

3.3. Processing chains

Six processing chains are compared; chains A–D share the same tracker and differ in the spatial front end. The binaural baseline **A** renders the HOA scenes to two ears through Neumann KU100 HRTFs [17] with an order-3 SH decoder and tracks each ear signal; chain **B** adds a binaural LCMV stage before tracking, a per-frequency two-channel filter with unit gain on the target motor and a null on the other. On the HOA side, chain **C** applies the per-channel decomposition of Section 2.3 to all 16 channels and chain **D** tracks the beam output of Eq. (3). Chain **E** is a modern single-channel reference in which a second-generation Vold-Kalman filter [18] estimates all eight complex envelopes jointly on the W channel from the exact phases of both motors, the envelope bandwidth being swept from 0.3 to 12.5 Hz and the best setting reported; a joint Kalman/RTS smoother, the exact Bayesian posterior mean for this linear-Gaussian model, and a sparsity-regularized variant reached equal or lower scores. Chain **F** applies the same joint tracker to the beams of chain D. Each chain is evaluated in its own observation domain, estimates being scored against the same rendering or filter applied to the ground-truth component by Pearson correlation, energy-weighted across channels where applicable.

4. Results

4.1. Chamber validation

Table 1 summarizes the single-source chamber experiment. At these low order frequencies ($ka \ll 1$) the source energy concentrates in the low-order SH channels, with 99.7 % of the order energy contained in ACN 0–3; metrics are therefore reported where phys-

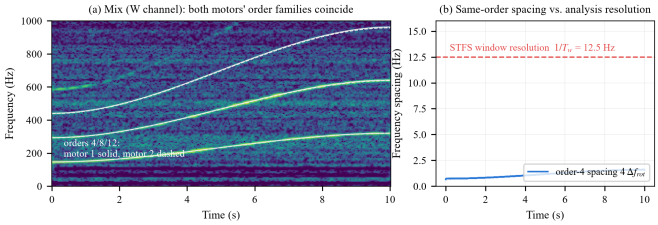

Figure 1: The dual-motor degeneracy (speed-locked scenario). (a) Spectrogram of the mixture (W channel) with both motors' order tracks overlaid; the families coincide. (b) Fourth-order spacing against the $1/T_w = 12.5\,\text{Hz}$ tracker resolution; the sources stay unresolvable in time-frequency throughout.

Table 1: Chamber experiment: correlation r and NRMSE of the separated order (with respect to the recorded order) and of the residual (with respect to the recorded noise).

	Order		Noise	
Evaluation	r	NRMSE	r	NRMSE
W channel (ACN 0)	0.979	0.207	0.882	0.489
Low-order (ACN 0–3)	0.975	0.224	0.910	0.422
Energy-weighted (16 ch.)	0.975	0.213	0.882	0.469

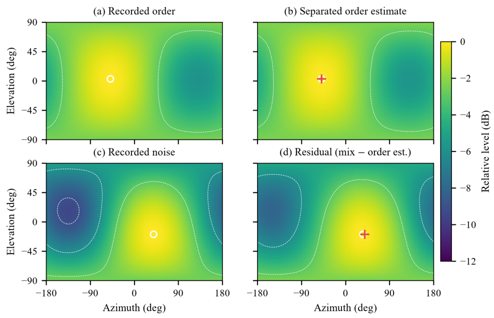

Figure 2: Chamber spatial validation. DoA energy maps of the recorded order (a) and its separated estimate (b), and of the recorded noise (c) and the residual (d). White circles mark map peaks, red crosses the reference peak.

ically meaningful. The separated order reaches an energy-weighted correlation of 0.975 and the residual 0.88. In the spatial domain (Fig. 2) the recorded order and noise peak 90° apart, at azimuths of −49° and +39°, and the separated components reproduce their reference directions to within 0° and 4° on a 2° grid, with map correlations of 1.00 and 0.99. The chain is thus verified on physical hardware.

4.2. Spectrally distinct motors

With two different units on different speed profiles, per-channel extraction alone (chain C) is sufficient: the energy-weighted correlations reach 0.981 for motor 1, 0.882 for motor 2 and 0.866 for the background, against baselines of 0.77, 0.53 and 0.55 for the unprocessed mixture; the NRMSE values are 0.19, 0.46 and 0.51. Figure 3 shows the spatial outcome: every recovered component keeps its direction, with peak errors of at most 2°, map correlations of at least 0.98 and a 82° separation in the recovered field. Cross-capture stays at 0.04, despite the shared orders 4, 8 and 12 and the crossing trajectories.

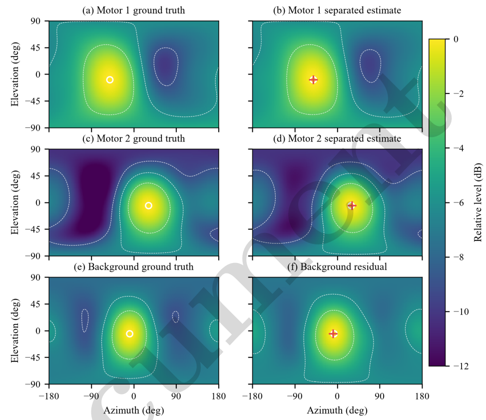

Figure 3: Two-motor scene, distinct scenario: ground truth (left) and separated estimates (right) for motor 1 (a, b), motor 2 (c, d) and the background residual (e, f). Each separated component retains its direction of arrival.

4.3. Spectrally degenerate motors

The situation differs for same-model motors. For per-channel tracking, the correlation of the weaker unit falls from 0.99 at a 13.8 % speed difference to 0.42 at 0.5 %; the transition occurs where the fourth-order beat crosses the $1/T_w$ resolution, and the DoA attribution degrades with it: in the speed-locked scenario the map peak of the motor-1 estimate moves 71° onto the other unit. Joint estimation shifts this boundary without removing it. In the cornering scenario chain E recovers both units almost exactly (1.00/1.00), a 3.7 % spacing being decoupled by joint tracking, as claimed in the Vold-Kalman literature [4], [5], [18]. In the speed-locked scenario, by contrast, every single-channel tracker encounters an identifiability limit (Fig. 4): the order beats (1.2–4.6 Hz) lie within the bandwidth that the true envelope variation itself requires along the RPM ramp, so no setting both follows the envelopes and separates the units. The best joint Vold-Kalman configuration reaches 0.75/0.84, and the Bayesian-exact smoother 0.72/0.83.

Table 2 and Fig. 5 present the principal result. In the speed-locked scenario the spatial chains restore the weaker motor to 0.86–0.90 and the background to 0.97, where the best single-channel tracker remains at 0.75; in the cornering scenario they match the joint tracker at 0.94–1.00; and cross-capture between the motors falls to 0.1–0.3. The spatial front end combined with the joint tracker (chain F) performs best overall, so

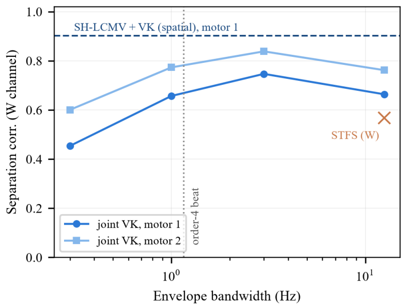

Figure 4: Speed-locked pair: envelope-bandwidth trade-off of the joint Vold-Kalman tracker on the W channel. No bandwidth reaches the spatial chain (dashed); the cross is the per-channel baseline.

the prefilter composes with any downstream tracker. The improvement does not come from steering a main lobe toward the target, since at 24° separation the sources lie well inside the order-3 beamwidth, but from the nulls and from the 13 adaptive degrees of freedom, and replacing the adaptive weights by fixed minimum-norm weights reduces the cornering recovery of chain D from 0.94 to 0.66. The binaural LCMV improves on plain binaural tracking where one null at wide separation suffices, reaching 0.91 on motor 1 in the speed-locked scenario against 0.86 for chain D, which incurs a small self-cancellation penalty under steering mismatch. With no remaining degrees of freedom, however, it can neither null a second interferer nor form a background estimate, and its conditioning deteriorates by two orders of magnitude toward low frequency (Fig. 5c). The SH filter is also a single frequency-independent 16-element vector, whereas the binaural filter needs a per-bin solution from measured HRTFs.

5. Applicability to a real cabin

The validation above uses an anechoic chamber and a strongly damped room; a car cabin adds reflections, distributed structure-borne radiation, near-field geometry and diffuse road noise. This section examines which elements of the method remain valid there.

Per-channel extraction remains valid unconditionally. Every propagation path, whether direct, reflected or panel-radiated, is linear and time-invariant: an order at instantaneous frequency $kf_{\mathrm{rot}}(t)$ emerges as the same order with a complex gain $H(kf_{\mathrm{rot}}(t))$ that varies slowly along the RPM trajectory and is absorbed into the tracked envelope $c_k(t)$ of Eq. (2). Reverberation therefore merely redefines the extracted amplitude as the source signal observed through the cabin transfer function, the quantity of interest in interior-noise analysis.

The DoA maps degrade gracefully. The Bartlett peak follows the strongest coherent path; with cabin reverberation times near 0.05–0.1 s the direct and early-reflected field dominates close to the trim, so the maps still localize the apparent source region. Since the separated components remain HOA signals, sharper parametric SH estimators [10] may replace Eq. (1) without changing the decomposition.

A null on the interferer's steering vector cancels its direct contribution exactly; what leaks through is its reflected field, which the adaptive weights must absorb, so the achievable rejection is bounded by the direct-to-reverberant ratio together with the adaptive suppression. Two properties are favourable. Early cabin reflections are few and coherent, adding a low-rank component to the interference subspace that lies well within the 13 spare dimensions (the mechanism behind most of the recovery in Section 4.3), while the late diffuse residue is broadband and is rejected by the narrowband tracker by construction. In a cabin the rejection will therefore fall below the 0.99 of a damped room, by an amount governed by the interferer's direct-to-reverberant ratio.

The results of Section 4 rely on geometric steering vectors alone (Section 2.4); the following remedies address the additional mismatch expected in a real cabin. Point-source steering is the weakest assumption, and it can be replaced. A drive unit reaches the cabin partly through airborne paths and partly through mounts and panels: the effective source is a distributed region, possibly of rank greater than one, and near-field curvature introduces further mismatch relative to any plane-wave $\mathbf{a}(\boldsymbol{\Omega})$. Diagonal loading provides some tolerance [15]; a systematic remedy is to measure the steering vectors directly. The order-domain covariance gated on the phase of one motor yields the spatial signature of that motor in its installed state, and Eq. (3) accepts measured vectors or subspace constraints in place of geometric ones; no free-field assumption then remains.

Finally, the aperture determines the usable band. At $ka \ll 1$ the field information concentrates in the low-order channels (99.7 % of the component energy in ACN 0–3 for the chamber, 56–85 % for the scenes); the array then behaves as a first-order device, whose four dimensions still satisfy the two-null constraint set, with a reduced adaptive margin. For an array of Zylia-class radius ($a \approx 5$ cm), full third-order resolution spans about 0.8–3 kHz, covering the motor fundamentals and the gear-mesh region; inverter sidebands near 10 kHz exceed the spatial-aliasing limit and need larger or higher-order arrays. Per-shaft phase references are production signals (resolver or CAN bus); the diffuse road noise, which no directional constraint can null, is left to the minimum-variance objective and the narrowband tracker.

6. Conclusion

Multi-motor electric vehicles pose an order-separation problem that is spectrally degenerate by construction, since same-model, speed-locked drive units hold same-numbered orders within about one percent of each other, yet the problem remains spatially well conditioned. The proposed framework exploits this structure through phase-coherent tracked-harmonic extraction in each SH channel for attribution, and through a frequency-independent constrained adap-

Table 2: Separation correlation (energy-weighted or in-domain) for the six chains in the two degenerate scenarios. Chain E uses the single W channel with both exact phases, chain F the same tracker on the spatial beams; dashes mark quantities a chain cannot form.

		Cornering axle pair Δrpm 3.7%, 24°			Speed-locked pair Δrpm 0.5%, 85°		
Chain	Ch.	M1	M2	Bg	M1	M2	Bg
A binaural, tracking	2	0.97	0.63	0.35	0.61	0.80	0.29
B binaural LCMV + tr.	2	0.98	0.78	–	0.91	0.95	–
C HOA per-channel	16	0.97	0.67	0.58	0.60	0.82	0.51
D HOA LCMV + tracking	16	0.98	0.94	0.99	0.86	0.94	0.97
E joint Vold-Kalman (W)	1	1.00	1.00	0.97	0.75	0.84	0.91
F HOA LCMV + joint VK	16	1.00	1.00	–	0.90	0.94	–

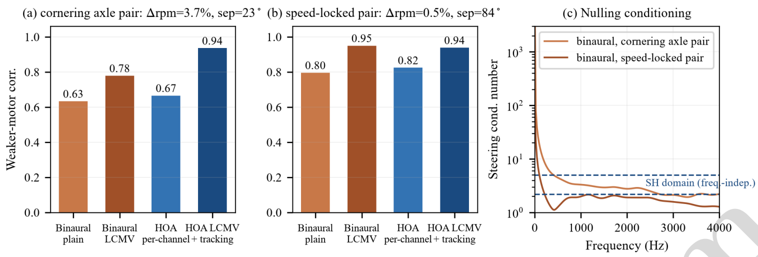

Figure 5: Weaker-motor separation for chains A–D in the cornering-axle-pair (a) and speed-locked (b) scenarios, together with the conditioning of the nulling problem (c). The binaural steering matrix degrades toward low frequency, whereas the SH-domain constraint matrix (dashed) is independent of frequency.

tive prefilter for the degenerate case. Across a real chamber recording and two-motor scenes constructed from measured data, per-channel extraction matches the single-source baseline when the motors are spectrally distinct, with correlations of 0.98, 0.88 and 0.87 and DoA errors of at most 2°. In the speed-locked regime, where even joint Vold-Kalman estimation with both exact phases is held to 0.75–0.84 by the envelope-bandwidth trade-off, spatial-null preprocessing restores the weaker of two same-model motors to 0.90–0.94 and the background to 0.97–0.99. The binaural benchmark shows the limits of the two-channel format: it can place at most one spatial null, retains no spare degrees of freedom, becomes ill-conditioned at low frequency, and offers no full-sphere attribution. The analysis of Section 5 indicates that the decomposition transfers to real cabins unconditionally, and that the spatial recovery transfers with performance bounded by the direct-to-reverberant ratio; spherical-array capture is thus the instrument that makes order-resolved, source-attributed interior noise analysis of multi-motor vehicles possible.

Acknowledgments

The authors thank J. Antoni (LVA) for valuable discussions on tracked-harmonic estimation. The Arni impulse responses are due to McKenzie et al. [16], the KU100 HRTF set to Bernschütz [17].

References

[1] M. S. Qatu, M. K. Abdelhamid, J. Pang, and G. Sheng, "Overview of automotive noise and vibration," *International Journal of Vehicle Noise and Vibration*, vol. 5, no. 1/2, pp. 1–35, 2009.

[2] M. Münder and C.-C. Carbon, "A literature review [2000–2022] on vehicle acoustics: Investigations on perceptual parameters of interior soundscapes in electrified vehicles," *Frontiers in Mechanical Engineering*, vol. 8, p. 974 464, 2022.

[3] K. Horváth, "Multiphysics modeling and simulation of nvh phenomena in electric vehicle powertrains," *World Electric Vehicle Journal*, vol. 17, no. 4, p. 183, 2026.

[4] H. Vold and J. Leuridan, "High resolution order tracking at extreme slew rates, using kalman tracking filters," SAE International, SAE Technical Paper 931288, 1993.

[5] K. R. Fyfe and E. D. S. Munck, "Analysis of computed order tracking," *Mechanical Systems and Signal Processing*, vol. 11, no. 2, pp. 187–205, 1997.

[6] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3–4, pp. 171–218, 1992.

[7] M. V. van der Seijs, D. de Klerk, and D. J. Rixen, "General framework for transfer path analysis: History, theory and classification of techniques," *Mechanical Systems and Signal Processing*, vol. 68–69, pp. 217–244, 2016.

[8] P. Chiariotti, G. Battista, M. Ettorre, and P. Castellini, "Average acoustic beamforming in car cabins: An automatic system for acoustic mapping over 3d surfaces," *Applied Acoustics*, vol. 129, pp. 47–63, 2018.

[9] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Springer Topics in Signal Processing). Springer International Publishing, 2019, vol. 19.

[10] D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 261–270, 2012.

[11] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.

[12] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[13] J. Antoni, "Cyclostationarity by examples," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 987–1036, 2009.

[14] J. Antoni, Q. Leclère, A. Dinsenmeyer, E. Julliard, et al., "Standalone extraction of tonal components from aeroacoustic signals: An approach based on angle-time cyclostationarity," *AIAA Journal*, vol. 60, no. 2, pp. 844–859, 2022.

[15] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[16] T. McKenzie, L. McCormack, and C. Hold, *Dataset of spatial room impulse responses in a variable acoustics room for six degrees-of-freedom rendering and analysis*, arXiv preprint arXiv:2111.11882, 2021.

[17] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proceedings of the AIA-DAGA 2013 Conference on Acoustics*, Merano, Italy, 2013, pp. 592–595.

[18] H. Vold, H. Herlufsen, M. Mains, and D. Corwin-Renner, "Multi axle order tracking with the Vold-Kalman tracking filter," *Sound and Vibration*, vol. 31, no. 5, pp. 30–34, 1997.