The Emperor's New Binaural:

Reverse-Engineering Dolby Atmos Binaural Rendering Reveals Minimal HRTF Processing

Andrew Grathwohl
andrew@grathwohl.me

June 2026

Abstract

We reverse-engineer the binaural output of the Dolby Atmos Renderer by building an independent ADM-BWF renderer and measuring Dolby's official binaural re-renders against it on the cues that define binaural audio. Three orthogonal measurements converge. (i) Long-term spectral comparison: full head-related transfer function (HRTF) convolution carves the expected 10–15 dB pinna-notch dips at ∼6.5 kHz and ∼9–10 kHz, the frequencies established as elevation cues by median-plane localization research; these dips are absent from Dolby's output, whose spectrum instead matches pure amplitude panning to within ∼1 dB RMS across five program items. (ii) Interaural time difference (ITD): on a controlled, reconstructed ADM-BWF source our full-HRTF path reproduces the physiologically correct ∼660 μs ITD at 90° azimuth while panning produces exactly zero, and Dolby carries measurable interaural delay on about as few frames as a panner, well below a full-HRTF render of the same material. (iii) Interaural coherence: Dolby reduces it broadband, including below 500 Hz where the pinna imparts negligible interaural difference, the signature of decorrelating reverberation. Dolby applies each cue to a different degree: its timing and timbre sit at the panning floor, while the interaural decorrelation it adds comes from reverberation. Measured against the binaural-rendering and pinna-cue literature, the system behaves as a stereo panner with a reverberation stage.

1 Introduction

Binaural audio has a precise technical meaning. It is the reproduction of sound over headphones using head-related transfer functions (HRTFs) to recreate the acoustic filtering imposed by the listener's pinnae, head, and torso, so that a signal delivered to the two ears carries the interaural time differences (ITD), interaural level differences (ILD), and direction-dependent spectral cues a free-field source would have produced [1, 6]. Among these cues, the high-frequency spectral notches introduced by the pinna are the dominant monaural mechanism for elevation perception and front-back disambiguation [2, 5].

Dolby markets Atmos binaural rendering as "immersive, personalized spatial audio" over headphones and, until July 2025, offered a personalized HRTF capture system based on phone scanning of "up to 50,000 points of the user's head, ears and shoulders" [16]. The marketing invokes individualized HRTF filtering, the mechanism the term "binaural" denotes.

This work began as a debugging exercise. We built an independent ADM-BWF renderer to play Dolby Atmos content for an unrelated project, and with full HRTF convolution engaged its output sounded markedly darker than Dolby's official binaural renders of the same masters. The cause lay in the approach itself: Dolby applies far less HRTF processing than "binaural" implies. We establish this on three signal-domain axes (spectral magnitude, interaural timing, interaural coherence) and anchor the comparison with a controlled, reconstructed ADM-BWF source that fixes the reference scale a true binaural render must reach. We then relate the finding to the published binaural-rendering literature, which has historically assumed full HRTF convolution, and to the acoustics of the pinna notch that lets the spectral measurement be read at all.

2 Background

2.1 The pinna notch and what the 6.5 kHz dip means

The HRTF $H(f, \theta, \phi)$ describes how sound from direction (θ, ϕ) is filtered before reaching the eardrum. Its main perceptual cues are the ITD (up to ∼0.7 ms for lateral sources), the frequency-dependent ILD (up to ∼20 dB above 4 kHz), and a family of direction-dependent spectral peaks and notches in the 5–12 kHz region produced by the pinna [1, 7].

These notches are the cue. Hebrank and Wright [2] first showed that the center frequency of the first pinna notch rises with source elevation, from roughly 6–7 kHz for sources near and below the horizontal to 10 kHz

and above overhead, and that listeners use the shift to judge elevation in the median plane, where interaural cues are ambiguous. Lopez-Poveda and Meddis [3] attributed the notch to destructive interference between direct sound and sound diffracted and reflected within the concha, refining the earlier reflection-only account and reproducing the measured elevation dependence. Raykar et al. [4] developed methods to extract these notch frequencies from measured head-related impulse responses. Iida et al. [5] showed that a parametric HRTF rebuilt from only the first two notches (N1, N2) and the first peak (P1) matches the median-plane localization accuracy of the full measured HRTF. N1 and N2 carry the elevation information; P1, near 4 kHz, stays fixed as elevation changes.

This sets up the measurement. A deep, persistent notch near 6.5 kHz, with a second near 9–10 kHz, is the acoustic fingerprint of pinna filtering, and its presence in a rendered signal tells us whether HRTF convolution was applied. That is what the spectral comparison of §5.1 reads.

2.2 ADM-BWF format

The Audio Definition Model carried in Broadcast Wave Format (ADM-BWF), standardized as ITU-R BS.2076 [8], is the interchange format for immersive audio. An ADM-BWF file carries PCM audio together with bed channels (static speaker positions) and object channels (audio with time-varying 3D position metadata in Cartesian coordinates, where $Z = 0$ is ear level and $Z = 1$ the ceiling). The format is the input both to Dolby's renderer and to ours, which lets us drive identical content through both pipelines.

2.3 The Dolby Atmos Renderer

The Dolby Atmos Renderer ingests ADM-BWF and produces binaural stereo re-renders. It embeds proprietary Dolby Bitstream Metadata (DBMD) carrying per-object `binaural_mode` settings (`Near`, `Mid`, `Far`, `Off`). We initially hypothesized these set HRTF intensity; the measurements in §6.1 indicate they instead set room reverberation depth.

3 Methodology

3.1 Renderers

We implemented two independent ADM-BWF renderers: a JavaScript prototype for rapid iteration and a production Rust implementation using the Steam Audio library [9] for HRTF convolution. Both parse ADM-BWF metadata, extract per-channel position data including motion timelines, and spatialize each object independently before summing to stereo. Bed channels are rendered by amplitude panning regardless of the HRTF setting, consistent with common practice. A scalar blend mixes the dry panned signal with the HRTF-convolved signal.

3.2 Measurement approach

Binaural rendering depends on several largely independent perceptual cues: the direction-dependent spectral coloration of the pinna notches (5–12 kHz), the ITD, and the interaural level difference and decorrelation between the ears. Each cue needs its own measurement. We characterize every render (the Dolby reference, our full-HRTF render, and our panning-only render, all from one ADM-BWF source) on three axes, anchored by a controlled reference signal. Together they locate where Dolby sits between pure panning and full HRTF.

Spectral coloration. We compare per-frequency RMS levels via windowed FFT over the full duration of each item. This axis targets the pinna-notch signature of §2.1: full HRTF convolution carves 5–12 kHz dips, so the depth of those dips measures how much pinna filtering a render contains.

Interaural timing (ITD). Interaural timing is an independent cue from spectral balance, so we measure it directly. We compute short-time generalized cross-correlation with phase transform (GCC-PHAT) between the channels over 2–12 kHz in 85 ms frames, searching lags within ±1 ms. Because a dense, center-dominant mix yields near-zero *aggregate* ITD for any method, we report the fraction of frames carrying $|\mathrm{ITD}| > 100\,\mu$s, gating frames by interaural coherence (> 0.5) so that diffuse, low-coherence frames (which return spurious lags under PHAT whitening) do not contaminate the statistic.

Interaural coherence. The third axis is decorrelation between the ears, via magnitude-squared coherence per octave band. Both HRTF filtering and decorrelated reverberation reduce coherence, but with different spectral shape: pinna HRTF acts in 5–12 kHz, whereas reverberation decorrelates broadband, including below 500 Hz where the head and pinna impart negligible interaural difference. The shape therefore tells us which mechanism is acting.

Controlled reference scale. To anchor a 0–100% scale we reconstructed an ADM-BWF from an extracted 7.1.2 test bed: we synthesized a BS.2076 `axml` chunk encoding ten objects at their canonical positions, injected a broadband-noise probe at a single known azimuth, and rendered it at full HRTF and at panning. This fixes the ITD a true binaural render should produce for a lateral source, independent of mix content.

4 CMAP: Center of Mass Amplitude Panning

We document the panning algorithm used in our renderer, based on Dolby's patented Center of Mass Amplitude Panning (CMAP) [10], so that the baseline against which Dolby is compared is a faithful Atmos panner.

4.1 Problem statement

Given an object at position $\vec{o} = (x, y, z)$ and a set of M speakers at positions $\vec{s}_1, \ldots, \vec{s}_M$, find gains $g_1, \ldots, g_M$ such that the perceived position matches $\vec{o}$. The perceived position under amplitude panning is the gain-weighted centroid:

$$\vec{o}_{\text{perceived}} = \frac{\sum_{i=1}^{M} g_i \, \vec{s}_i}{\sum_{i=1}^{M} g_i}. \tag{1}$$

4.2 Cost function

CMAP formulates this as a quadratic optimization:

$$C(\vec{g}) = \vec{g}^{\mathsf{T}} A \vec{g} - 2\,\vec{b}^{\mathsf{T}} \vec{g} + \vec{g}^{\mathsf{T}} D \vec{g}, \tag{2}$$

where $A_{ij} = \vec{s}_i \cdot \vec{s}_j$ is the **speaker geometry matrix** encoding spatial correlation between speakers; $b_i = \vec{o}\vec{s}_i$ is the **object–speaker alignment vector** measuring how well each speaker direction matches the target; and D is a diagonal **proximity penalty matrix**.

4.3 Proximity penalty

The penalty matrix prevents gain explosion when objects approach speaker positions:

$$D_{ii} = \alpha \, d_0^2 \left(\frac{\|\vec{o} - \vec{s}_i\|}{d_0} \right)^{\beta}, \tag{3}$$

with constants $\alpha = 20$, $\beta = 3$, and $d_0 = 2.0\,\text{m}$.

4.4 Optimal solution

Setting $\nabla_{\vec{g}} C = 0$ yields

$$\vec{g}_{\text{opt}} = (A + D)^{-1} \vec{b}. \tag{4}$$

Negative gains are clamped to zero and the result is normalized:

$$g_i \leftarrow \frac{\max(g_i, 0)}{\sum_j \max(g_j, 0)}. \tag{5}$$

4.5 Distance effects

Two distance-dependent effects are applied post-panning. Inverse distance attenuation:

$$\text{atten}(d) = \frac{d_{\text{ref}}}{d_{\text{ref}} + r\,(d - d_{\text{ref}})}, \tag{6}$$

with $d_{\text{ref}} = 1.0\,\text{m}$ and $r = 1.0$. Air absorption (high-frequency roll-off with distance):

$$\text{absorption}(d) = e^{-k(d - d_{\text{ref}})}, \tag{7}$$

with $k = 0.05\,\text{m}^{-1}$.

5 Results

5.1 The spectral discrepancy

With full HRTF convolution our output showed consistent spectral deficits relative to Dolby: 10–12 dB at ∼6.5 kHz, up to 15 dB at ∼9–10 kHz, and a broadband reduction above 500 Hz (Fig. 1). These frequencies coincide with the first and second pinna notches (N1, N2) identified as elevation cues in §2.1. With 20–30 objects each HRTF-convolved and summed, the per-object notches stack into a broadband, audibly dark coloration. The effect emerges from dense object-based rendering; no single filter is misconfigured.

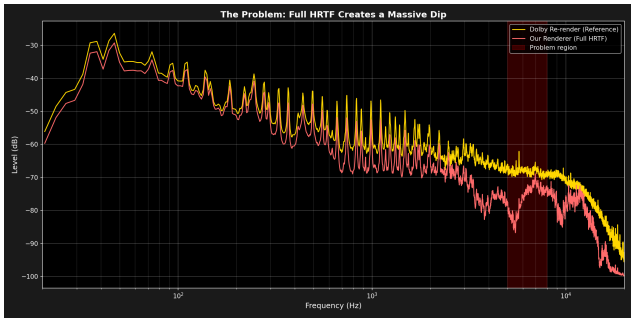

Figure 1: Full-HRTF rendering versus Dolby's binaural output. The deficits at 6.5 kHz and 9–10 kHz coincide with the first and second pinna notches (N1, N2) [2, 5].

5.2 Eliminating alternative explanations

Before attributing the discrepancy to minimal HRTF, we eliminated the obvious alternatives (Table 1). The deficit persisted across multiple HRTF datasets and across both independent implementations, ruling out an implementation-specific artifact.

Table 1: Alternative explanations tested and eliminated.

Hypothesis	Result
Coordinate mapping error	Correct (Z=0 is ear level)
HRTF dataset quality	Same deficit with all HRTFs
Crossfeed configuration	No effect
Gain metadata in ADM	None present
Motion-triggered HRTF	No change with moving objects

5.3 Magnitude matches amplitude panning

Dolby's long-term spectrum matches our pure panning render to within ∼1 dB RMS (Fig. 2), and is nowhere near the full-HRTF render, whose pinna notches Dolby's output does not contain. On the spectral axis, the timbre a listener actually hears, Dolby is indistinguishable from a panner.

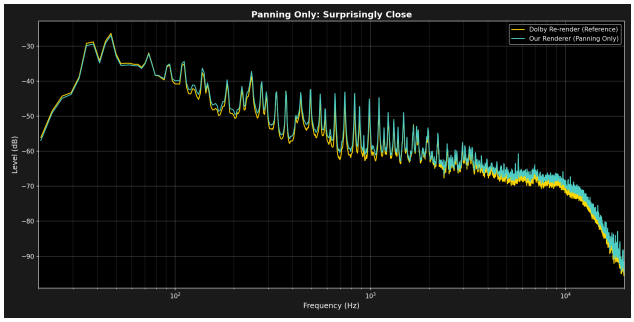

Figure 2: Pure amplitude panning (no HRTF) versus Dolby. The spectra agree within ∼1 dB across most of the band.

5.4 Consistency across program material

The match held across five items of varied genre and density, agreeing with Dolby to within ∼2 dB RMS (Fig. 3). Across that range the perceived tonal character of Dolby's re-render is the same as a pure amplitude pan. The interaural measurements that follow show what Dolby adds beyond timbre.

5.5 The HRTF ceiling is correct (controlled test)

On a reconstructed ADM-BWF carrying a single broadband source at a known azimuth, our full-HRTF path produces physiologically correct binaural cues while panning produces none (Table 2, Fig. 4). The ∼660 μs ITD at 90° is the head-geometry maximum and the accompanying ∼16 dB ILD is likewise in range; 45° yields the correctly smaller ITD. This establishes our renderer as a valid reference and shows that the

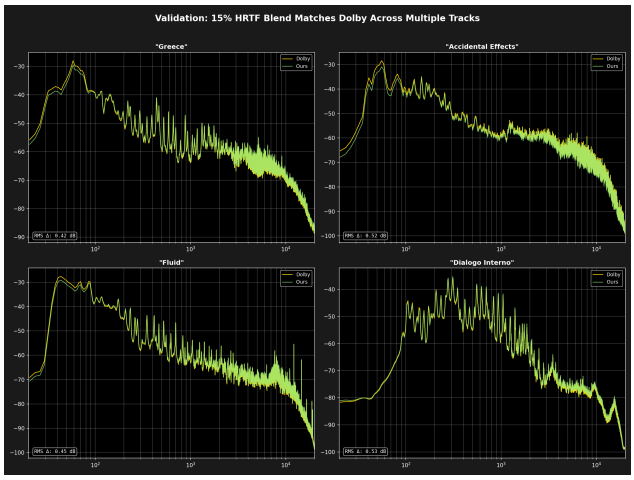

Figure 3: Magnitude comparison across four additional Atmos items: Dolby's spectral balance tracks the panning render consistently, confirming near-identical timbre across varied material.

Table 2: Controlled single-source ITD/ILD ceiling (broadband-noise probe).

Azimuth	Render	ITD	ILD
90° L	Full HRTF	−667 μs	+15.7 dB
90° L	Panning	0 μs	+32.9 dB
90° R	Full HRTF	+604 μs	−17.2 dB
90° R	Panning	0 μs	−32.9 dB
45° L	Full HRTF	−417 μs	+11.7 dB
45° L	Panning	0 μs	+12.6 dB

weak aggregate ITD seen in dense mixes reflects center-dominant content, not a deficient renderer.

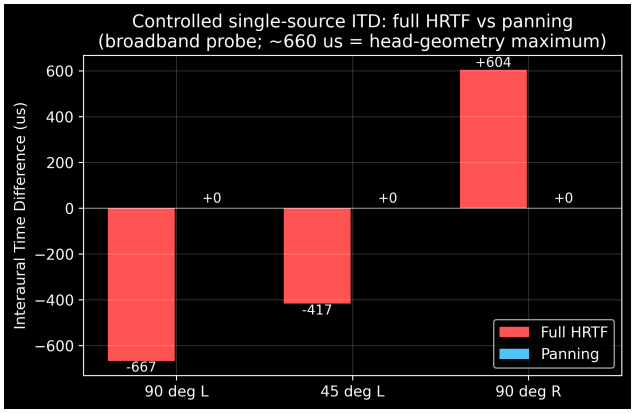

Figure 4: Controlled single-source ITD: full HRTF reproduces the ±660 μs head-geometry maximum; panning yields 0 μs.

Table 3: Frames with |ITD| > 100 μs (2–12 kHz, coherence-gated; mean across validation items).

Render	mean	range
Panning (0%)	0.6%	0.0–2.4%
Dolby	1.6%	0.1–5.2%
Full HRTF (100%)	9.1%	1.3–13.7%

5.6 Dolby suppresses ITD

Against that validated ceiling, Dolby's binaural output carries interaural delay on about as few frames as a panner, and far below a full-HRTF render of the same material (Table 3, Fig. 5). Since a true binaural renderer imposes up to ∼660 μs on any laterally placed source, the near-absence of ITD in Dolby's output indicates that interaural timing is largely absent from the pipeline.

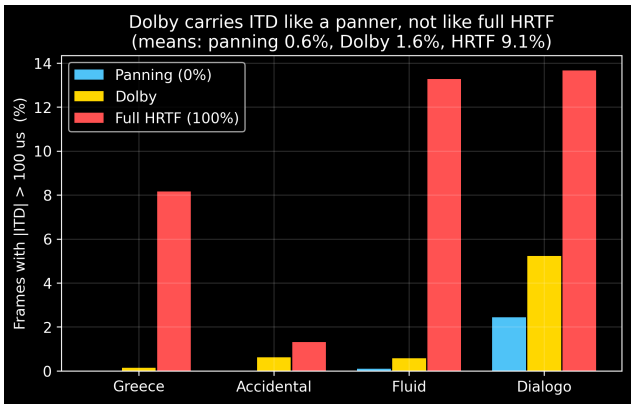

Figure 5: Dolby's ITD-frame count sits at the panning floor, well below full HRTF.

5.7 The residual decorrelation is broadband (reverberation)

Dolby does reduce interaural coherence relative to panning, but the reduction is broadband. Averaged across items, ∼47% of the added decorrelation falls below 2 kHz, including the 200–500 Hz band where the pinna cannot act, and ∼53% in 5–12 kHz (Fig. 6). A pure equalization applied to a panned signal cannot reduce interaural coherence, because the same filter on both channels leaves them scaled copies; the coherence drop therefore comes from processing that changes the interaural relationship. Its broadband shape points to decorrelating room reverberation, since the pinna acts only within 5–12 kHz. This motivates the reinterpretation of the `Near`/`Mid`/`Far` metadata in §6.1.

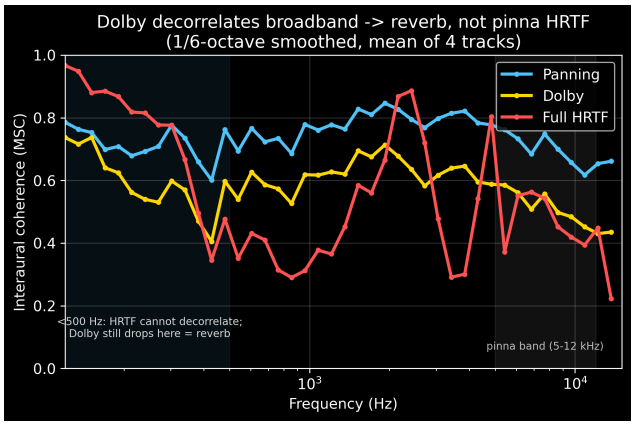

Figure 6: Interaural coherence versus frequency (1/6-octave, mean of four items). Dolby's coherence drops broadband, including below 500 Hz where pinna HRTF cannot act.

5.8 Anisotropy across cues

The three axes give three different answers (Fig. 7). Placed between the panning floor and the full-HRTF ceiling, Dolby sits near the panner on long-term magnitude and on interaural timing, and well toward full HRTF on broadband decorrelation, which the coherence spectrum attributes to reverberation. No single point on a panning-to-HRTF dial captures this. Timing and timbre track the panner; the decorrelation tracks reverberation.

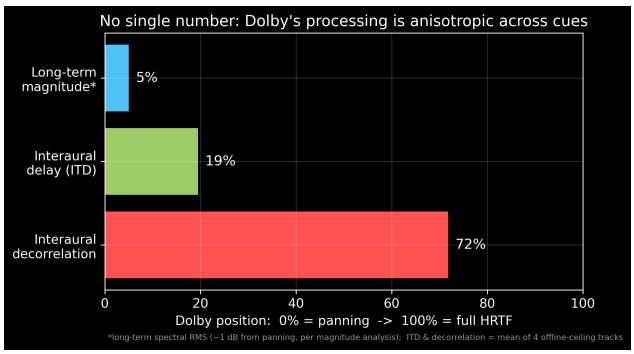

Figure 7: Dolby's position (panning → full HRTF) on three cues. The processing is anisotropic; the decorrelation axis is reverberation, not localization HRTF.

6 Discussion

6.1 Reinterpreting Near/Mid/Far

The broadband, ITD-free decorrelation we measure fits a set of room-reverberation presets (short/dry, moderate, longer-tail) better than it fits HRTF intensity. Reverberation decorrelates the ears without adding interaural delay, which matches the low-ITD, reduced-

Table 4: Processing hierarchy. "Binaural" denotes the top two rows.

Method	HRTF	Description
Dummy-head recording	full	Physical binaural
Full HRTF convolution	full	Computational binaural
Dolby "binaural"	minimal	Panning + reverberation
Stereo panning	none	Not binaural

coherence pattern we see. On this reading the "binaural strength" those modes set is reverberation depth.

6.2 Why a vendor would do this

The engineering rationale is timbre, the same problem the diffuse-field-equalization literature was built to address. Reducing the spatial order or strength of HRTF processing trades localization accuracy for tonal neutrality, and diffuse-field equalization exists because naive convolution colors the signal in ways listeners reject [11, 12]. A dense Atmos mix makes this acute: 20–30 simultaneous objects, each convolved with its own pinna notches, compound into the dark, damaged spectrum of Fig. 1. Substituting reverberation for most of the HRTF holds the tonal balance and gives up spatial precision. This is a defensible product decision. The engineering is sound; calling its output "binaural" is the overreach.

6.3 Implications for personalization

The maximum perceptual benefit of HRTF personalization is typically 3–6 dB in localization-critical bands [13], and it acts on exactly the pinna-notch and ITD cues that this pipeline largely discards. Dolby's July 2025 discontinuation of consumer HRTF personalization [17] is *consistent* with a pipeline in which those cues are minimal; we decline the stronger causal claim, as adoption, support cost, and capture friction are at least as plausible and we have no internal evidence either way.

6.4 Position in the hierarchy

Table 4 places the result on the spectrum from physical binaural capture to plain stereo. "Binaural" carries a specific meaning in audio engineering and psychoacoustics, and a panner with a reverberation send does not meet it. Apple appears to have reached a similar engineering conclusion, reportedly attenuating its high-frequency spatial boost in later iOS releases [18]; we flag that as a community report rather than a measurement and do not rely on it. We measured Dolby, not Apple.

7 Related Work

Prior evaluation of binaural renderers has been predominantly perceptual. Reardon et al. [14] proposed a methodology for comparing binaural renderers on localization, timbre, naturalness, and preference, and applied it to commercial panners including Atmos tools; Dewey, Moore, and Lee [15] surveyed practitioners' perceptions of Dolby and Apple binaural mixes in popular music. These studies establish that listeners can hear differences between renderers but treat the renderer as a black box. We add an instrumental angle: a signal-domain teardown that measures, on the defining cues, how much HRTF a shipping renderer applies.

The academic rendering literature, by contrast, generally assumes full HRTF convolution [19, 20], optionally with diffuse-field equalization to manage the resulting coloration [11, 12]. Reverberation-substituted "binaural" rendering does not appear there as a localization strategy; it is, in effect, a product decision rather than a published technique. Our measurements locate a mass-market product well outside the assumption that underlies that literature.

8 Limitations

Several caveats bound the claims. (1) We have no Dolby binaural re-render of a known hard-lateral source. The controlled test (Table 2) characterizes *our* ceiling, not Dolby on identical content; the decisive experiment (driving the controlled test bed through the Dolby Atmos Renderer and measuring its ITD) requires Dolby's proprietary software and remains open. (2) The 100% ceiling is realized with one HRTF set (Steam Audio); a different set would shift absolute values, and on some items Dolby's decorrelation exceeds that ceiling, which makes "percent of full HRTF" ill-defined above it. (3) The corpus is a small number of items from a single renderer version. (4) We attribute the broadband decorrelation to reverberation from its spectral shape and by elimination, not by recovering an impulse response from the black box.

9 Conclusion

The Dolby Atmos Renderer's binaural output applies minimal localization-grade HRTF. Its long-term spectrum matches amplitude panning, down to the missing pinna notches that carry elevation cues. Its interaural timing falls to near a panner's zero against a validated $\sim$660 μs ceiling. The interaural difference that remains is broadband decorrelation consistent with room reverberation. "Binaural" denotes HRTF-based spatialization with ITD, ILD, and pinna spectral cues, and the pipeline delivers little of it.

This is not poor engineering. Full HRTF on dozens of simultaneous objects sounds bad, and trading most of it for reverberation protects the timbre. The vocabulary is the problem. The emperor is, in fact, dressed; he is just wearing a reverberation stage instead of a head-related transfer function.

Acknowledgments

The author thanks the Steam Audio team at Valve for their open-source HRTF library.

Data and Code Availability

The analysis scripts (ADM-BWF reconstruction, ITD and coherence measurement, and figure generation) and the rendered test assets are available from the author on request.

AI Usage Disclosure

Measurement code, figure generation, and a draft of this manuscript were prepared with AI assistance under the author's direction. All measurements were executed on real audio renders and verified by the author.

References

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, rev. ed. Cambridge, MA: MIT Press, 1997.

[2] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.*, vol. 56, no. 6, pp. 1829–1834, 1974.

[3] E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Am.*, vol. 100, no. 5, pp. 3248–3259, 1996, doi:10.1121/1.417208.

[4] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 364–374, 2005.

[5] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, vol. 68, no. 8, pp. 835–850, 2007.

[6] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 868–878, 1989.

[7] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 6th ed. Leiden: Brill, 2012.

[8] International Telecommunication Union, "Audio Definition Model," Rec. ITU-R BS.2076-2, 2019.

[9] Valve Corporation, "Steam Audio," `https://valvesoftware.github.io/steam-audio/`, 2024.

[10] Dolby Laboratories, "Center of Mass Amplitude Panning," U.S. Patent Application, 2018.

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

[12] T. McKenzie, D. T. Murphy, and G. Kearney, "Diffuse-field equalisation of binaural Ambisonic rendering," *Applied Sciences*, vol. 8, no. 10, art. 1956, 2018.

[13] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, 1993.

[14] G. Reardon, A. Roginska, *et al.*, "Evaluation of binaural renderers: A methodology," in *Proc. Audio Engineering Society Convention*, 2017.

[15] C. Dewey, J. Moore, and H. Lee, "Practitioners' perspectives on spatial audio: Insights into Dolby Atmos and binaural mixes in popular music," *J. Audio Eng. Soc.*, vol. 72, no. 7/8, pp. 504–516, 2024.

[16] Dolby Laboratories, "Dolby Head Tracking and Personalized Spatial Audio," White paper, Mar. 2022.

[17] Dolby Laboratories, "Discontinuation of Dolby Head Tracking Personal Profile Capture," Support notice, July 2025.

[18] User analysis of Apple Spatial Audio processing changes in iOS 17, online forum discussion, 2023. *(Community report; cited as such.)*

[19] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Rendering localized spatial audio in a virtual auditory space," *IEEE Trans. Multimedia*, vol. 6, no. 4, pp. 553–564, 2004.

[20] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE WASPAA*, 2001, pp. 99–102.