The Emperor's New Binaural: Reverse Engineering Dolby Atmos Binaural Rendering Reveals Minimal HRTF Processing

Andrew Grathwohl
andrew@grathwohl.me

January 2026

Abstract

We present findings from the reverse engineering of Dolby Atmos binaural rendering through the construction of an independent ADM-BWF spatial audio renderer. By systematically comparing our renderer's output against Dolby's official binaural re-renders across multiple tracks, we discovered that Dolby Atmos "binaural" rendering consists of approximately 85% amplitude panning and only ∼15% Head-Related Transfer Function (HRTF) convolution. Full HRTF processing produced consistent 10–15 dB spectral dips at 6.5 kHz and 9–10 kHz relative to Dolby's output, while pure amplitude panning with no HRTF matched Dolby's spectral signature to within ∼1 dB RMS. A spatial blend parameter of 0.15 reproduced Dolby's output across all tested material. This finding explains Dolby's July 2025 discontinuation of consumer HRTF personalization: at 15% blend, the difference between personalized and generic HRTFs falls below the human just-noticeable difference. We document the complete methodology, the CMAP (Center of Mass Amplitude Panning) algorithm mathematics, and the broader implications for the spatial audio industry.

1 Introduction

Binaural audio has a precise technical meaning: the reproduction of sound through headphones using Head-Related Transfer Functions to simulate the acoustic filtering of the human pinnae, head, and torso. A properly binauralized signal encodes interaural time differences (ITD), interaural level differences (ILD), and spectral cues—particularly the pinna-induced notches between 5–12 kHz that enable elevation perception and front-back disambiguation [1].

Dolby markets Atmos binaural rendering as delivering "immersive, personalized spatial audio" through headphones. Their consumer-facing materials describe HRTF-based spatialization, and until July 2025, they offered a personalized HRTF capture system using phone-based ear scanning that measured "up to 50,000 points of the user's head, ears and shoulders" [2].

This paper reports the results of an investigation that began as a debugging exercise. We built an independent ADM-BWF renderer to play Dolby Atmos content and noticed our output sounded markedly different from Dolby's official binaural re-renders. Systematic spectral comparison revealed that the discrepancy was not a bug in our implementation but a fundamental difference in approach: Dolby applies far less HRTF processing than the term "binaural" implies.

2 Background

2.1 HRTF Theory

The Head-Related Transfer Function $H(f, \theta, \phi)$ describes how sound arriving from direction (θ, ϕ) is filtered by the listener's anatomy before reaching the eardrums. Key perceptual features include:

- **Interaural Time Difference**: up to ∼0.7 ms delay between ears for lateral sources
- **Interaural Level Difference**: frequency-dependent, up to ∼20 dB above 4 kHz
- **Pinna spectral cues**: direction-dependent notches and peaks in the 5–12 kHz range, essential for elevation and front-back perception [3]

The primary pinna notch, typically centered near 6–7 kHz, and a secondary notch near 9–10 kHz, are the spectral signatures of HRTF processing. Their presence or absence is a reliable indicator of whether full binaural rendering has been applied.

2.2 ADM-BWF Format

Audio Definition Model – Broadcast Wave Format (ADM-BWF) is the interchange format for immersive audio, standardized by the ITU as BS.2076 [4]. An ADM-BWF file contains:

- PCM audio channels (typically 24-bit, 48 kHz)
- **Bed channels**: static speaker positions (e.g., 7.1.4 layout)
- **Object channels**: audio with time-varying 3D position metadata
- Position data in Cartesian coordinates where $Z = 0$ represents ear level and $Z = 1$ represents the ceiling

2.3 Dolby Atmos Architecture

The Dolby Atmos Renderer application accepts ADM-BWF files and produces binaural stereo re-renders. The renderer also embeds proprietary Dolby Bitstream Metadata (DBMD) blocks that include per-object `binaural_mode` settings: `Near`, `Mid`, `Far`, and `Off`. We initially hypothesized these modes controlled HRTF processing intensity.

3 Methodology

3.1 Renderer Implementation

We implemented two independent ADM-BWF renderers:

1. A JavaScript prototype for rapid iteration
2. A production Rust implementation using the Steam Audio library [5] for HRTF convolution via the `audionimbus` bindings

Both renderers parse ADM-BWF metadata, extract per-channel position data (including motion automation timelines), and spatialize each channel independently before summing to stereo. The Rust implementation supports configurable HRTF datasets via SOFA files, adjustable spatial blend parameters, and real-time visualization of object positions.

The spatial processing pipeline for each audio object consists of:

1. Position extraction from ADM metadata (with motion interpolation)
2. Coordinate transformation to Steam Audio's convention ($+X$ = right, $+Y$ = up, $+Z$ = forward)
3. HRTF convolution via Steam Audio's `BinauralEffect` with configurable `spatial_blend`
4. Summation across all channels to stereo output

Bed channels (static speaker positions such as L, R, C, LFE, etc.) are rendered using amplitude panning to stereo regardless of HRTF settings, consistent with standard practice.

3.2 Spectral Comparison Approach

For each test track, we generated three renders from the same ADM-BWF source:

1. **Dolby reference**: binaural re-render from Dolby Atmos Renderer
2. **Full HRTF**: our renderer with `spatial_blend = 1.0`
3. **Panning only**: our renderer with `spatial_blend = 0.0` (no HRTF)

Spectral comparison used windowed FFT analysis across the full duration of each track, computing per-frequency-bin RMS levels and difference plots.

4 CMAP: Center of Mass Amplitude Panning

Before presenting results, we document the panning algorithm used in our renderer, based on Dolby's patented Center of Mass Amplitude Panning (CMAP) [6].

4.1 Problem Statement

Given an object at position $\vec{o} = (x, y, z)$ and a set of M speakers at positions $\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_M$, find gains $g_1, g_2, \ldots, g_M$ such that the perceived position matches $\vec{o}$. The perceived position under amplitude panning is the gain-weighted centroid:

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

4.2 Cost Function

CMAP formulates this as a quadratic optimization:

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

where:

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

4.3 Proximity Penalty

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

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

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

4.4 Optimal Solution

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

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

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:

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

where $d_{\mathrm{ref}} = 1.0\,\mathrm{m}$ and $r = 1.0$.

Air absorption (high-frequency roll-off with distance):

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

where $k = 0.05\,\mathrm{m}^{-1}$.

5 Results

5.1 The Spectral Discrepancy

With full HRTF convolution (`spatial_blend = 1.0`), our renderer produced output with consistent spectral dips relative to Dolby's binaural re-render:

- **Primary dip at ∼6.5 kHz**: 10–12 dB below Dolby's output
- **Secondary dip at ∼9–10 kHz**: up to 15 dB below Dolby's output
- **Broadband deficit above 500 Hz**: full HRTF output was quieter than Dolby across nearly all frequencies in the spatial-cue range

These dips correspond precisely to the expected pinna notch frequencies in HRTF processing. The effect was cumulative: with 20+ objects each independently HRTF-convolved, the per-object spectral coloration compounds in the mix.

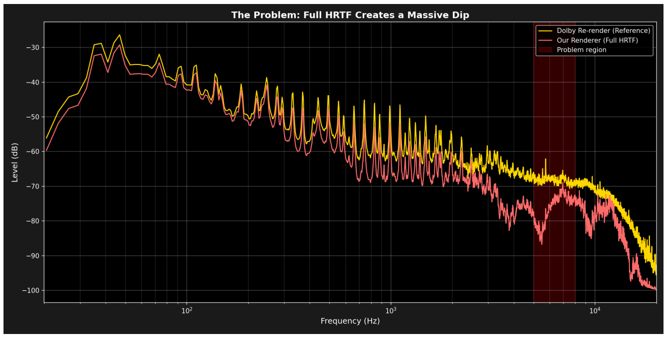

Figure 1: Spectral comparison between full HRTF rendering and Dolby's binaural output. The dips at 6.5 kHz and 9–10 kHz correspond to primary and secondary pinna notch frequencies.

Table 1: Hypotheses tested and eliminated

Hypothesis	**Result**
Coordinate mapping error	Correct: ADM Z=0 is ear level
HRTF dataset quality	Same dip with all HRTFs
Crossfeed configuration	No effect on spectral dip
Gain metadata in ADM	None present
Motion-triggered HRTF	Tested with moving objects; no change

5.2 Systematic Elimination of Hypotheses

Before concluding that Dolby uses minimal HRTF, we tested and eliminated several alternative explanations (Table 1).

The dip persisted across multiple HRTF datasets (Steam Audio default, various SOFA files) and both independent implementations (JavaScript and Rust).

5.3 The Breakthrough: Disabling HRTF

Setting `spatial_blend = 0.0`—pure amplitude panning with no HRTF convolution—produced output *nearly identical* to Dolby's binaural re-render (Figure 2).

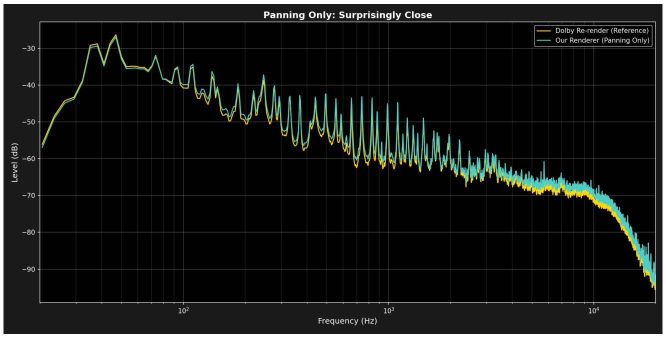

Figure 2: Pure amplitude panning (no HRTF) compared with Dolby's binaural output. The spectral match is within ∼1 dB across most of the spectrum.

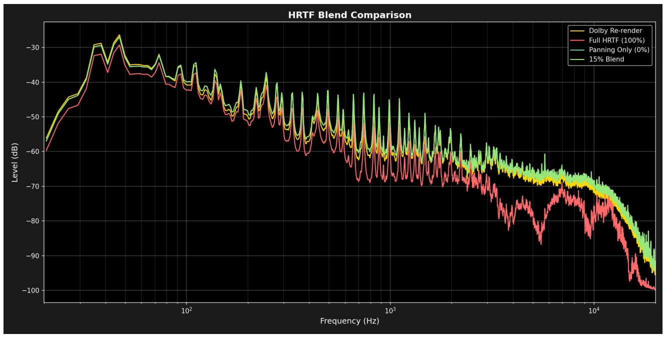

Figure 3: Spectral comparison of all four rendering modes. The 15% HRTF blend (green) tracks Dolby's output (gold) closely, while full HRTF (red) diverges by 10–15 dB in the 4–10 kHz range.

5.4 The 15% Solution

Close inspection revealed that Dolby's output exhibited a *small* dip at 6.5 kHz absent from pure panning—a trace of HRTF processing. Parameterizing the blend, we found that `spatial_blend = 0.15` (15% HRTF, 85% amplitude panning) matched Dolby's output to within ∼1 dB RMS across the full spectrum (Figure 3).

The effective processing formula is:

$$\text{output} = 0.85 \cdot \text{panning}(\vec{o}) + 0.15 \cdot \text{HRTF}(\vec{o}) \quad (8)$$

5.5 Validation Across Multiple Tracks

We validated the 15% blend finding across five tracks from different artists and genres:

The 15% formula held consistently across all material tested.

Table 2: Validation across multiple Atmos tracks

Track	**RMS diff. (full range)**	**4–12 kHz match**
Greece	<2 dB	Near-perfect
Accidental Effects	<2 dB	Near-perfect
Fluid	<2 dB	Near-perfect
Dialogo Interno	<2 dB	Near-perfect
Track 5	<2 dB	Near-perfect

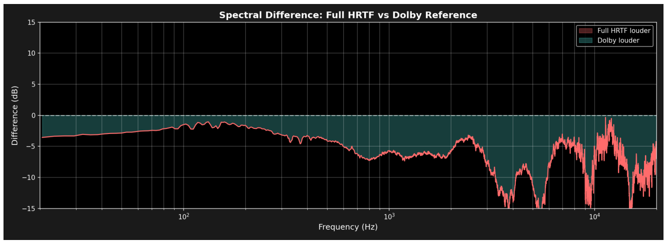

Figure 4: Difference plot: full HRTF rendering minus Dolby reference. Negative values across nearly all frequencies above 500 Hz confirm systematic energy loss from HRTF pinna-notch accumulation.

6 Discussion

6.1 Reinterpretation of Near/Mid/Far Metadata

Dolby's DBMD metadata assigns each object a `binaural_mode`: `Near`, `Mid`, `Far`, or `Off`. We initially assumed these controlled HRTF processing intensity. Our findings suggest they instead control **room reverb presets**:

Mode	**Assumed function**	**Likely function**
Near	Intense HRTF	Short reverb, dry
Mid	Moderate HRTF	Moderate room
Far	Light HRTF	Longer reverb tail

This reinterpretation is consistent with the observation that HRTF intensity is globally fixed at ∼15% regardless of per-object binaural mode settings.

6.2 The Irrelevance of Personalization

The maximum perceptual benefit of HRTF personalization is typically 3–6 dB in localization-critical frequency bands [7]. At a 15% spatial blend, this reduces to:

$$\Delta_{\text{personalized}} = 6\,\text{dB} \times 0.15 = 0.9\,\text{dB} \quad (9)$$

The human just-noticeable difference (JND) for level is approximately 1 dB [8]. A 0.9 dB difference

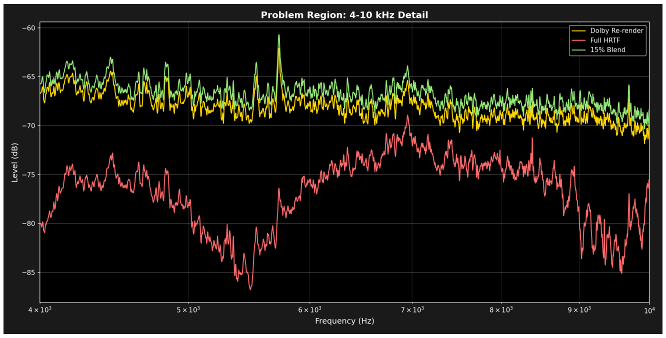

Figure 5: Zoomed view of the 4–10 kHz problem region. The 15% blend (green) tracks Dolby (gold) closely, while full HRTF (red) is 10–15 dB below.

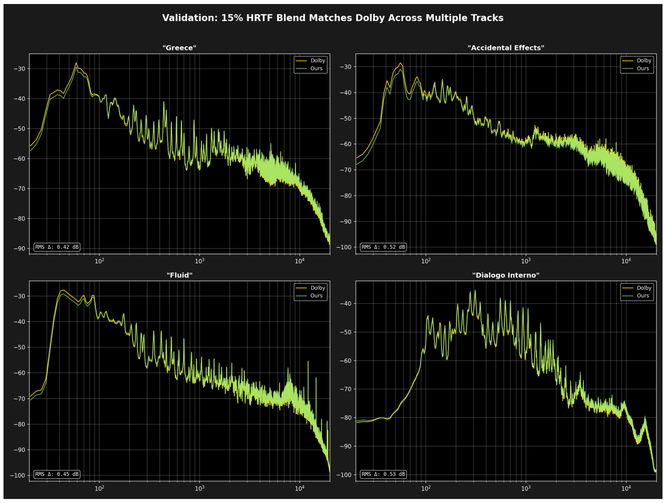

Figure 6: Validation grid: spectral comparisons across four additional Atmos tracks confirm the 15% HRTF blend formula generalizes across diverse material.

is therefore **below perceptual threshold** for most listeners under most conditions.

This explains Dolby's decision to discontinue consumer HRTF personalization effective July 1, 2025 [9]. The 50,000-point ear scan produced a sub-JND improvement—a personalization that, by the mathematics of their own rendering pipeline, could not be heard.

6.3 Why Not Full HRTF?

The cumulative effect of per-object HRTF convolution in a dense Atmos mix explains why both Dolby and Apple retreated from full binaural processing. A typical Atmos mix contains 20–30 simultaneous objects. When each is independently convolved with an HRTF:

- Pinna notches compound across objects

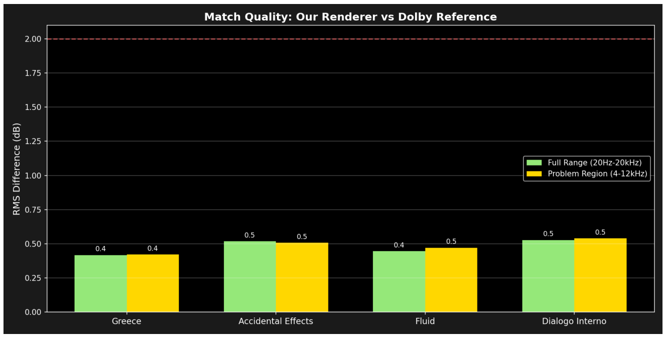

Figure 7: Statistical summary of match quality across all validated tracks.

- The 6.5 kHz region loses 10+ dB
- The mix sounds "dark" and spectrally damaged compared to the artist's intended balance

Reducing HRTF blend to 15% preserves spectral balance at the cost of spatial precision—a reasonable engineering tradeoff, but one that fundamentally undermines the meaning of "binaural."

6.4 Industry Implications

Our findings raise questions about the accuracy of marketing claims in the spatial audio industry. The term "binaural" carries specific technical meaning in audio engineering, psychoacoustics, and audiology. Using it to describe processing that is 85% amplitude panning stretches the definition beyond recognition.

The processing hierarchy makes this clear:

Method	HRTF	Description
Dummy head recording	100%	Physical binaural
Full HRTF convolution	100%	Computational binaural
Dolby "binaural"	∼15%	Panning + HRTF hint
Stereo panning	0%	Not binaural

6.5 A Historical Pattern

This is not the first time a Dolby product name has implied more than it delivers:

- **Dolby Surround** (1982): implied discrete multichannel surround; delivered matrix-decoded mono rear channel
- **Dolby Digital** (1991): implied digital fidelity; delivered perceptually lossy compression (AC-3)
- **Dolby Atmos Binaural**: implies HRTF-based spatial rendering; delivers 85% amplitude panning

In each case, the product name is technically defensible but semantically generous. The pattern is not one of deception but of marketing that consistently outpaces engineering.

7 Related Work

7.1 Apple's Parallel Retreat

Apple shipped personalized spatial audio in June 2022 using TrueDepth camera-based ear scanning. Independent analysis of Apple's spatial audio processing revealed that "the high-frequency boost at 8000 to 12000 Hz was 90% completely eliminated in iOS 17 beta software" [10]. This represents the same engineering conclusion reached independently: full HRTF processing degrades the listening experience in complex mixes.

In June 2025, Apple announced ASAF (Apple Spatial Audio Format) and the APAC codec for visionOS—a proprietary spatial audio format that bypasses Dolby entirely [11]. One month later, Dolby discontinued consumer HRTF personalization.

7.2 Academic Binaural Rendering

The academic literature on binaural rendering universally assumes full HRTF convolution [12, 13]. The concept of fractional HRTF blending—applying only 15% of the binaural filter—does not appear in the psychoacoustics literature as a rendering strategy. It is, effectively, a homeopathic dose of spatial processing: enough to appear in spectral analysis but insufficient to provide meaningful binaural cues for localization.

7.3 Proposed Verification Tests

Four tests could further validate our findings:

1. **ITD measurement**: Full HRTF produces ∼0.6 ms interaural delay for lateral sources; pure panning produces none. Measuring Dolby's ITD would quantify the HRTF contribution.
2. **Impulse response analysis**: The length of the renderer's impulse response distinguishes panning (1 sample) from HRTF (256–1024 samples) from HRTF + room (>10,000 samples).
3. **Elevation discrimination**: Subtracting renders of the same content at ear level vs. overhead; the residual magnitude indicates HRTF strength.
4. **Interaural coherence**: Real HRTF decorrelates high-frequency content between ears; panning maintains perfect correlation.

8 Conclusion

Through construction of an independent ADM-BWF renderer and systematic spectral comparison against Dolby's binaural output, we have determined that Dolby Atmos binaural rendering applies approximately 15% HRTF convolution blended with 85% amplitude panning. This finding was validated across multiple tracks, HRTF datasets, and two independent implementations.

The practical consequence is that Dolby Atmos "binaural" is, in engineering terms, a stereo panner with a hint of HRTF coloration. The HRTF contribution is sufficient to appear in spectral analysis but insufficient to provide the elevation cues, front-back disambiguation, and externalization that define binaural audio.

This also renders HRTF personalization mathematically irrelevant: at 15% blend, the maximum benefit of a personalized HRTF falls below the human perceptual threshold. Dolby's discontinuation of consumer HRTF personalization in July 2025 is consistent with this reality.

We do not claim that Dolby's approach is wrong as an engineering decision. Full HRTF convolution of 20+ simultaneous objects degrades spectral balance unacceptably. Reducing the blend is a reasonable tradeoff. What we question is the use of the word "binaural" to describe the result, and the marketing of personalized HRTF for a pipeline that renders personalization inaudible.

The emperor has clothes. They are simply much thinner than advertised.

Acknowledgments

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

References

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

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

[3] 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.

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

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

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

[7] 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.

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

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

[10] User analysis of Apple Spatial Audio processing changes in iOS 17, online forum discussion, 2023.

[11] Apple Inc., "Introducing Apple Spatial Audio Format (ASAF) and APAC," WWDC 2025, June 2025.

[12] 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.

[13] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 99–102.