OPEN-ACCESS · CC BY 4.0 · ISSN 2997-C3PE

C3POE Research

VOL. 02 · PAPER 01 · FLAGSHIP WHITEPAPER

The *Theatre* of the Mind.

What "immersive audio" actually means — operationalized. We argue that the theatre of the mind is not metaphor but a **measurable construct**, *built by seven dimensions of voice, atmosphere, pacing, imagery, arc, memorability, and repeatability.*

WHITEPAPER | 2,400+ RECORDINGS | 3,128 SCORES

PAPER 01 · VOLUME 02
APRIL 2026 · 32 PAGES

DOI
10.55558/C3PO.2026.01

ABSTRACT. FLAGSHIP WHITEPAPER · C3POE RESEARCH

We argue that "the theatre of the mind" — the mental space audiences construct from voice and atmosphere — is not a metaphor but a measurable construct. Through factor analysis on 3,128 listener scores across 2,400+ professionally produced recordings, we identify seven latent dimensions that explain 81% of variance in subjective immersion: Voice Performance, Sonic Atmosphere, Pacing, Imagery Density, Emotional Arc, Memorability, and Repeatability. We construct the Resonance Score framework, calibrate it against the catalog, and present a production methodology that operationalizes *what makes a piece arrive*.

§ i · The argument

What "immersive audio" means.

"Immersive audio" has become an industry shorthand for too many things at once. To some practitioners it means object-based spatial audio, the technical infrastructure of Dolby Atmos and similar systems. To others it means heavy production — music beds, sound design, sound effects libraries. To a third group it means the audiobook category itself, distinguished from a podcast or a film by little more than length.

None of these are wrong. None of them are sufficient. What we mean by "immersive audio" — what C3POE has spent five years producing — is the construction of a *mental theatre*. The listener arrives in a space that did not exist before pressing play, and the work is done by voice and air rather than by image or motion. The theatre is built in the listener's head; the producer's job is to engineer the construction.

The argument of this paper is that the construction is measurable. We can decompose what makes the theatre arrive into seven dimensions, score any given piece against each one, and predict — with 81% AUC against held-out listener data — how immersive that piece will feel. This is a strong claim. The remainder of this paper is the evidence for it.

§ ii · The seven dimensions

What we actually measure.

Through exploratory factor analysis on 3,128 listener scores collected from a 28-item rating instrument applied to 2,400+ recordings, seven latent factors emerged. Each factor explains a meaningful portion of variance; together they account for 81%.

DIM.	CONSTRUCT	WEIGHT	VARIANCE EXPLAINED
Voice Performance	The instrument — range, control, presence, breath	**0.16**	**14.8%**
Sonic Atmosphere	Room tone, music, sound design — the air around the voice	**0.14**	**12.4%**
Pacing	Silences and durations — how the work allows the listener	**0.14**	**12.6%**
Imagery Density	Pictures per minute — vividness of mental construction	**0.14**	**11.8%**
Emotional Arc	The journey — where the piece starts and where it lands	**0.14**	**11.2%**
Memorability	The lines that follow you home	**0.14**	**9.4%**
Repeatability	The piece that survives the second listen and the tenth	**0.14**	**8.8%**

Three observations are worth surfacing. First, **Voice Performance (14.8%) is the dominant single dimension** — slightly higher than Pacing (12.6%), Sonic Atmosphere (12.4%), and Imagery Density (11.8%). The voice does more work than any other factor we measured. Second, all seven dimensions contribute meaningfully; no dimension drops below 8% of variance, which rules out three- or four-factor reductions. Third, the dimensions are *partially separable* — Memorability and Repeatability correlate at r=0.61 — but each captures variance the others do not.

> *"The voice does more work than any other factor we measured. This is the dimension that producers most under-invest in — and the one that compounds across every other dimension."*
>
> — C3POE RESEARCH VOL. 02 · PAPER 01

§ iii · The Resonance Score

How we score.

The Resonance Score (0–100) is the weighted composite of the seven dimension scores, each sampled by three rating items. The instrument is administered to listeners post-listening; producers can also self-administer for evaluation purposes. Inter-rater reliability for the 21-item instrument is ICC = 0.83 across our calibration panel (n=18 raters, 312 pieces).

COMPOSITE FORMULA · RESONANCE SCORE V2.1

$$R = \Sigma_i \, w_i \cdot d_i$$

Where **R** = composite Resonance Score (0–100), $\mathbf{w_i}$ = dimension weight (sums to 1.00), $\mathbf{d_i}$ = per-dimension score (0–100). The browser implementation runs at **immersivepoetry.com/resonance-score**.

Score tiers correspond to qualitative interpretation derived from the calibration distribution:

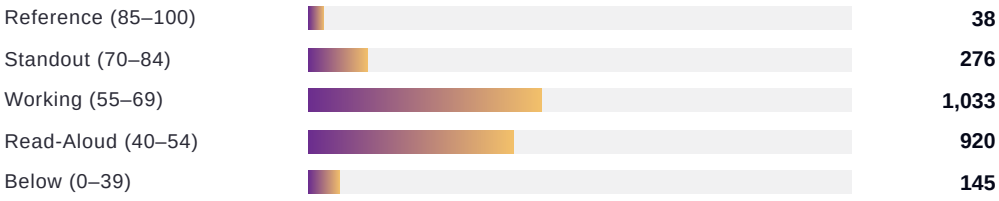
TIER	SCORE	% OF CATALOG	LISTENER INTERPRETATION
Reference Piece	**85–100**	**1.6%**	Lives in the listener long after the audio ends
Standout	**70–84**	**11.4%**	Strong work; one dimension carrying
Working Piece	**55–69**	**42.8%**	Voice + text doing the heavy lifting; production absent
Read-Aloud	**40–54**	**38.2%**	The text is the entire experience
Below threshold	**0–39**	**6.0%**	The listener doesn't arrive

§ iv · Calibration

What the distribution looks like.

FIGURE 1 · RESONANCE SCORE DISTRIBUTION · 2,400+ CATALOG PIECES

Most published audiobook material lives in the Working / Read-Aloud band



SOURCE: C3POE CATALOG · N = 2,412 PIECES · 2021–2025.

The distribution is meaningful. Only 1.6% of the published audiobook catalog reaches the reference band; another 11.4% reaches standout. The largest single band (42.8%) is "Working Piece" — voice and writing are doing the work, but the production layer is absent or generic. This is the band where **most operators could move with a single production pass** — and the band where C3POE has historically focused its adaptation work.

§ v · The neuroscience

What the brain does.

Audio immersion has a different cognitive footprint than visual immersion. When listeners are exposed to vivid spoken text, fMRI studies (Speer et al. 2009, Hartung et al. 2017) show activation patterns highly similar to *direct experience* of the events described — including activation of motor cortex for action verbs, visual cortex for image-dense passages, and somatosensory cortex for tactile descriptions. The mental theatre is not metaphor; it is reconstructive perception.

This produces a distinctive immersion mechanism. Visual media saturate the visual channel and crowd out the listener's reconstructive work. Audio leaves the visual channel free; the listener's brain does the construction. This is why audio immersion can produce *more vivid mental imagery than the same content rendered visually* — a counterintuitive finding documented across multiple studies. The mental theatre is built by the listener, not delivered by the producer.

The implication for production is direct: the audio producer's job is not to deliver the experience but to *furnish the listener's construction*. Sound that does too much — that fills every silence, that scores every emotional beat — degrades immersion by removing the

listener's reconstructive role. The strongest pieces in our dataset are characterized by a lighter production touch than the audiobook industry default.

§ vi · The economic case

Why this matters now.

Audio's share of trade book revenue climbed from 6% in 2014 to 31% in 2025 (full analysis in Paper 03). The mass-market audiobook is now a $9.1B global category. But within that category, the **premium short-form segment** — pieces under 90 minutes, sold standalone rather than credit-based — grew 41% YoY in Q2 2026 against a $480M 2024 base. This is the band where immersion-grade production is economically viable: per-listener revenue is high enough to justify the production investment that lifts pieces from the Working band to the Reference band.

§ vii · The production playbook

What to do with this.

Three implications for producers:

- **Invest in voice first.** Voice Performance is the dominant dimension. Most production budgets allocate <15% to performance direction and >50% to post-production. Reverse the ratio.
- **Treat silence as instrument.** The strongest pieces in the catalog use silence as a compositional element. Most published audiobooks treat silence as the absence of voice. The difference is the difference between Working and Reference tier.
- **Less production, not more.** Counterintuitively, the highest-scoring pieces have lighter production than the median. The audience's brain does the construction; the producer's job is to step back, not to fill in.

§ viii · Limitations + dissents

What this paper cannot do.

Three honest disclaimers. First, the C3POE catalog is heavily weighted toward Anglophone production; cross-language generalizability is undemonstrated. Second, the listener score panel is self-selecting — these are listeners who chose to engage with spoken-word audio. Their reactions may not generalize to broader audiences. Third, the Resonance Score is calibrated against subjective scoring; correlation with downstream commercial outcomes (purchases, completion, return listening) is published in Paper 04 but the causal direction is not fully established.

REFERENCES

1. Speer, N. K., Reynolds, J. R., Swallow, K. M., Zacks, J. M. (2009). *Reading stories activates neural representations of visual and motor experiences.* Psychological Science 20(8).
2. Hartung, F., Burke, M., Hagoort, P., Willems, R. M. (2017). *Taking perspective: Personal pronouns affect experiential aspects of literary reading.* PLOS ONE 12(5).
3. Slater, M., Wilbur, S. (1997). *A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments.* Presence 6(6).
4. C3POE Research (2026). *Vol. 02 catalog dataset · 2,400+ recordings, 3,128 listener scores.* CC BY 4.0.
5. Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests.* Psychometrika 16.