Synesthesia as Cosmic Isomorphism: Logostics of q(t) → τ(t) Across Image, Music, and Mind

Douglas C. Youvan

doug@youvan.com

www.youvan.ai

October 29, 2025

This paper advances a thesis: what we call synesthesia is not merely a neurological quirk, but a discovery of cosmic isomorphisms—stable correspondences among structures that already exist across modalities. Using a conservative combinatorial model, we sketched the "design space" of image→music mappings and arrived at a floor on the order of ~10^44 distinct pipelines (and vastly larger if rhythm or synthesis parameters are allowed). Such immensity suggests that brains do not invent cross-modal meaning ex nihilo; rather, they tune to relational invariants—the τ—that prefigure perception. In the Logostics framing, q(t) (observer state) seeks resonant alignment with τ(t) (generative manifold) by minimizing informational divergence under constraints of embodiment and culture. We formalize this by identifying modality-agnostic invariants—periodicities, symmetries, curvature fields, entropy gradients, topology of features—and showing how they induce natural functors between visual and auditory categories. Practical renderers (e.g., spiral traversals with hue→pitch, saturation→velocity, value→duration, gradient→pan, plus hysteresis) serve as empirical probes of these invariants rather than arbitrary aesthetics. The result is a metaphysical and mathematical account: synesthesia becomes evidence of It-from-Bit relations expressed as eigenstructures of transformation, with art and music functioning as laboratories where q(t) → τ(t) converges under resonance rather than rule.

Keywords: synesthesia, Logostics, q(t) -> τ(t), It-from-Bit, information geometry, category theory, invariants, resonance, entropy, curvature, wavelets, Betti numbers, topology, mapping space, hysteresis, spiral traversal, image-to-audio, cosmology of meaning.

Outline

1. Thesis and Scope
   1.1 Synesthesia as discovery, not invention
   1.2 Logostics: q(t) -> τ(t) as alignment rather than control
   1.3 From brain-local to cosmic correspondences

2. A Combinatorial Lower Bound on Mapping Space
   2.1 Traversals, grids, features, assignments, transforms
   2.2 Pitch systems, rhythm grammars, instrumentation, effects
   2.3 A conservative floor: ~10^44; realistic spaces: 10^80–10^120+
   2.4 Implication: abundance argues for pre-given τ-structure

3. Invariants That Survive Modality
   3.1 Periodicity & cyclic variables: hue ↔ pitch class (S^1)
   3.2 Intensity & energy: value/brightness ↔ loudness
   3.3 Spatial frequency ↔ spectral brightness (timbre)
   3.4 Orientation/flow ↔ rhythm/panning
   3.5 Entropy & mutual information as cross-modal "mass"

4. Category-Theoretic Frame
   4.1 Objects: images as elements of a shape/texture topos; sounds as elements of a time-spectral topos
   4.2 Functors that respect invariants (periodicity, order, topology)
   4.3 Natural transformations as "mappings with meaning"
   4.4 Limits, colimits, and compositional renderers

5. Information Geometry and Divergence
   5.1 Joint description length as a criterion (compression before taste)
   5.2 Divergence D(q || τ) in cross-modal mapping selection
   5.3 Geodesics on manifolds of renderers; resonance as shortest-path

6. Topology and Multiscale Structure
   6.1 Wavelet pyramids: low→pads, mid→arpeggios, high→sparkles
   6.2 Persistent homology: Betti curves → phrase lengths & cadences
   6.3 Curvature fields: ornamentation vs legato as geometric response

7. Dynamics and Memory
   7.1 Hysteresis: deadbands and slew limits to prevent jitter
   7.2 Spiral vs raster: holistic traversal as an ordering over τ
   7.3 Section-level thresholds: when structure merits modulation

1. Thesis and Scope

1.1 Synesthesia as discovery, not invention

Synesthesia, in this view, is not a private quirk of cortex but the recognition of pre-existing isomorphisms between modalities. The brain does not mint cross-modal meaning ex nihilo; it selects and stabilizes mappings that already live in the structure of the world. Circular variables align with circular variables (hue class ↔ pitch class as S^1), intensities align with intensities (luminance ↔ loudness as energy scalars), textures align with spectra (spatial frequency ↔ timbral brightness), directed curves align with directed motion (orientation field ↔ pan/rhythm). These are not metaphors but type-level compatibilities: operations preserved across domains.

When we estimated the conservative design space of image→music renderers, we found a lower bound that is astronomically large. Such abundance is not noise; it indicates a rich τ-manifold of correspondences in which many distinct pipelines can still feel "natural." Synesthesia then becomes the perceptual discovery of these correspondences under finite attention and culture-shaped priors. Art is where we test and tune these correspondences; music is where we hear their fixed points.

Practically, this reframes experiments: instead of asking "What trick convinces the brain?", we ask "Which invariants are preserved?" The criterion for a good mapping is not cleverness but structure preservation: does the mapping commute with the relevant operations (e.g., scaling, rotation, tempo change, key modulation) and compress description length across modalities?

1.2 Logostics: q(t) -> τ(t) as alignment rather than control

Logostics treats perception and creation as dynamics of alignment. Let q(t) denote the agent's evolving internal state (models, skills, tastes), and let τ(t) denote the generative manifold of world-level structure (those cross-modal invariants). The task is not to force q(t) to dominate τ(t), nor to dissolve q(t) into τ(t), but to resonate: to reduce divergence while keeping freedom of response.

In this frame, a synesthetic renderer is an alignment operator A that updates q(t) by exposing it to invariants in τ(t). Formally, we privilege moves that reduce a divergence D(q || τ) subject to stability (hysteresis, deadbands, slew limits) and pluralism (multiple valid minima). The key insight: the system behaves better when τ(t) is allowed to move—the world is not static, and alignment must be entrained, not solved once. Hence the emphasis on spirals, wavelet pyramids, and topology: they are not stylistic flourishes but channels of τ(t) that nudge q(t) without coercion.

This ethos has engineering consequences. We choose mappings that: (1) respect type ($S^1 \leftrightarrow S^1$, scalar↔scalar), (2) preserve order and composition (functoriality over feature pipelines), and (3) include memory (hysteresis) so the listener perceives intention rather than jitter. Alignment is thus obedience to structure, not submission to rules.

1.3 From brain-local to cosmic correspondences

If synesthesia is alignment to world-level invariants, its ground is cosmic, not merely neural. The same correspondences that make a hue wheel and a pitch wheel both live on S^1 are rooted in mathematics, not in anatomy. The It-from-Bit claim is modest here: informational structure (groups, metrics, curvatures, entropies, homologies) constrains the space of sensible mappings long before neurons fire. Brains then become receivers with gain control, learning to lock onto a few stable channels within a far larger broadcast.

This shift resolves a paradox. Individual synesthetic pairings vary, yet many feel universally "right." The variance is q-side selection; the rightness is τ-side invariance. Cross-cultural convergences—bright ↔ loud, high ↔ up, smooth ↔ legato—are not arbitrary conventions but low-description-length couplings across modalities. Where cultures diverge (mode, cadence, timbre), we witness different local minima on the same τ-landscape.

For art and music, the consequence is liberating. We can design systems that probe τ directly: spiral traversals (holistic order), multiscale analyses (wavelets ↔ orchestration), topological summaries (Betti curves ↔ phrase form), and information-theoretic controls (entropy ↔ density, mutual information ↔ call-and-response). The studio becomes a laboratory of correspondences, where $q(t)$ learns the shape of $\tau(t)$ by listening to how invariants survive transformation.

2. A Combinatorial Lower Bound on Mapping Space

2.1 Traversals, grids, features, assignments, transforms

We factor the design space into dials that are mostly independent and discretize each into a small, defensible set. Let N be the product of these dial-cardinalities; this yields a lower bound.

- Traversal T (how we order pixels): {raster, serpentine, spiral, Hilbert, Peano, edge-follow, contour, streamline, k-means tour, random-walk} ⇒ $|T| \approx 10$.
- Time grid G (how we quantize time): meters {2/4, 3/4, 4/4, 5/4, 7/8}, subdivisions {quarters, 8ths, 16ths, triplets, swing}, note-bin counts {32, 64, 128, 256, 512} ⇒ $|G| \approx 80$.

- Aggregation A (statistics per time-bin): choose a non-empty subset of {mean, median, variance, entropy, histogram-peak, texture/GLCM} ⇒ |A| = 2^6 − 1 = 63.
- Feature channels C (which image fields are exposed): choose non-empty subset of {H,S,V, L,a,b, grad-magnitude, grad-angle, edges, corners, wavelet-low, wavelet-mid, wavelet-high, segment-ID} (≈14) ⇒ |C| = 2^14 − 1 = 16,383.
  (If we keep the earlier conservative 12, |C| = 4,095; we will use 12 later to stay conservative.)
- Source→control assignment S (which channel drives which control): controls = {pitch, velocity, duration, pan, register, onset-jitter}. If on average 6 channels are in play and each control picks 1 channel, S ≈ 6^6 = 46,656.
- Control transforms F (how values are mapped): per control choose from {linear, log, exp, sigmoid, tanh, rank/quantile, z-score+clip, quantize-to-scale} ⇒ 8 choices per control ⇒ |F| = 8^6 = 262,144.

These alone, even before "music proper," already give a gigantic space.

2.2 Pitch systems, rhythm grammars, instrumentation, effects

We now layer the expressly musical choices, again in coarse counts.

- Pitch system M: 12 chromatic transpositions × 50 common modes/scales (major/minor church modes, harmonic/melodic minors, pentatonic/blues, bebop, Messiaen, heptatonic variants, etc.) ⇒ |M| ≈ 600.
- Harmony/progression grammar H: a small bank of progressors (diatonic triads, modal cadences, quartal/quintal sets, planing, axis-based, chromatic mediants) ⇒ |H| ≈ 20.
- Instrumentation I: choose 4 instruments (with doubling allowed) from a GM-like palette of 128 ⇒ |I| = 128^4 ≈ 2.68×10^8.
  (This ignores articulation; adding 4 articulations per part multiplies by 4^4 = 256.)
- Effects chain E: up to 5 slots, each from {none, LPF, HPF, reverb, delay, chorus, phaser, flanger, compressor, saturator}. If we allow any slot to be none and count coarse presets per effect (~20), a rough count is:
  per-slot ≈ 1 (none) + 9 effects × 20 presets = 181 ⇒ 181^5 ≈ 1.9×10^11.
  To stay conservative we cap E at ~8×10^8 (as in the earlier bound).
- Rhythm style R: density curve (sparse/medium/dense), swing (none, light, heavy), humanization (off/low/med/high), micro-metrical bias (straight, tresillo/clave, hemiola, quintuplet bias) ⇒ |R| ≈ 25.

- Hysteresis/smoothing Y: deadband choices (e.g., 5), slew limits (e.g., 5), pan/amp smoothing (e.g., 5) ⇒ |Y| ≈ 5×5×5 = 125 (earlier we used 10^3; we will keep the conservative 125 here).
- Panning law / summing P: {equal-power, linear, constant-vol} ⇒ |P| = 3.

2.3 A conservative floor: ~10^44; realistic spaces: 10^80–10^120+

To stay tight, use the more conservative channel count (|C| = 4,095) and the smaller effects space (E ≈ 8×10^8) and larger hysteresis (Y = 10^3) from the earlier calculation, yielding:

N = T · G · A · C · S · F · M · H · I · E · R · Y · P

Plugging numbers:

- T = 10
- G = 80
- A = 63
- C = 4,095
- S = 46,656
- F = 262,144
- M = 600
- H = 20
- I = 2.684×10^8
- E = 8×10^8
- R = 25
- Y = 10^3
- P = 3

Compute in stages (rounded for readability):

- T·G·A ≈ 10 × 80 × 63 ≈ 5.04×10^4
- (previous) × C ≈ 5.04×10^4 × 4.10×10^3 ≈ 2.07×10^8
- × S ≈ 2.07×10^8 × 4.67×10^4 ≈ 9.7×10^12
- × F ≈ 9.7×10^12 × 2.62×10^5 ≈ 2.5×10^18

- $\times M \approx 2.5\times10^{18} \times 600 \approx 1.5\times10^{21}$
- $\times H \approx 1.5\times10^{21} \times 20 \approx 3.0\times10^{22}$
- $\times I \approx 3.0\times10^{22} \times 2.68\times10^{8} \approx 8.0\times10^{30}$
- $\times E \approx 8.0\times10^{30} \times 8.0\times10^{8} \approx 6.4\times10^{39}$
- $\times R \approx 6.4\times10^{39} \times 25 \approx 1.6\times10^{41}$
- $\times Y \approx 1.6\times10^{41} \times 10^{3} \approx 1.6\times10^{44}$
- $\times P \approx 1.6\times10^{44} \times 3 \approx 4.8\times10^{44}$

This matches the earlier ~10^{44} floor while still ignoring:

- continuous parameters (cutoffs, ADSR times, detune, modulation depths),
- phrase-structure choices beyond the small H set,
- counterpoint policy, voicing rules, articulation layers,
- learning-based or adaptive controllers.

The moment we add a modest binary rhythm grid over 32 steps × 8 bars (256 toggles), the factor $2^{256} \approx 1.16\times10^{77}$ multiplies N into the 10^{120} vicinity. Similarly, allowing even 10 coarse values for 50 synth/effect parameters yields 10^{50}, again pushing the total beyond 10^{90}.

2.4 Implication: abundance argues for pre-given τ-structure

If "anything goes," most mappings should sound arbitrary. Yet in practice many independent choices lead to renderings that feel coherent—they preserve something we recognize as structure. The sheer abundance of possible pipelines, together with the empirical scarcity of satisfying ones, implies that "good" mappings occupy a thin manifold inside the space: they respect invariants that are not invented by us but discovered.

In Logostics language: there exists a τ-structure of cross-modal invariants (periodicity, intensity, spectral density, curvature, entropy, topology) that defines a low-divergence subspace. Different cultures and individuals (different q) converge on different neighborhoods of this subspace, but the subspace itself precedes the chooser. Abundance does not negate meaning; it reveals the constraints that make meaning portable across carriers. Synesthesia, on this account, is evidence of those constraints—cosmic correspondences that brains learn to tune, not to create.

3. Invariants That Survive Modality

3.1 Periodicity & cyclic variables: hue ↔ pitch class (S^1)

Both hue and pitch class live on circles: S^1 with a chosen origin and orientation. Transpositions in pitch and rotations on the hue wheel are both group actions by S^1. A mapping f: S^1_hue -> S^1_pitch is musically sensible when it is equivariant with respect to rotation/transposition, i.e. f(theta + delta) = f(theta) + delta', where delta' encodes a fixed interval policy (e.g., 1:1 for pure wheel-to-wheel, 12:7 for diatonic compression). This preserves intervals as differences on S^1, so color harmonies (complementary, triadic) correspond to pitch relations (tritone, major/minor triads). In practice: choose a base key, pick a rotation, optionally quantize to a mode. In Logostics, this is a low-divergence move because it commutes with both visual and musical rotations (q-updates do not destroy τ's circular structure).

Probe: rotate the image by 30 degrees in hue; the music should transpose by a consistent interval while preserving internal relationships.

3.2 Intensity & energy: value/brightness ↔ loudness

Brightness (Value) in HSV and loudness in audio are scalar proxies for energy flow. Although human loudness is not linear in amplitude, both domains respect order and contrast: brighter regions should not map to quieter sounds under a monotone transform. Suitable mappings: linear, log, or power-law curves with optional dynamic range compression, preserving rank and relative contrast.

Invariance principle: if a visual gain g rescales brightness by a monotone function, the audio gain should adjust so that local contrast is preserved (e.g., percentile/rank mapping). This supports perceptual coherence when global exposure or mastering levels change. In Logostics terms: the scalar field's ordering is the invariant; q should adapt the transfer curve while τ's order-type remains stable.

Probe: apply global exposure changes to the image; after loudness normalization the relative loudness pattern over time should remain nearly unchanged.

3.3 Spatial frequency ↔ spectral brightness (timbre)

Image spatial frequency (fine vs coarse detail) parallels audio timbre (spectral centroid/brightness). High spatial frequencies map naturally to brighter spectra (more high harmonics, noisier partials); low frequencies map to darker, rounder tones. Wavelet or Laplacian pyramids give multiscale bands: low-band → pads/sub, mid-band → body/partials, high-band → sparkles/noise.

Invariance principle: scale transforms in the image (zoom or blur) should correspond to spectral tilt changes in audio, not to arbitrary pitch jumps. The mapping preserves a frequency-order relation between modalities. In category terms, a multiresolution functor sends image scale to spectral coloration and leaves pitch class untouched.

Probe: Gaussian-blur the image; the result should shift the music toward darker timbres with minimal pitch drift and preserved rhythm.

3.4 Orientation/flow ↔ rhythm/panning

Local orientation fields (edge angles, optic-flow-like cues) carry directedness. In audio, panning and rhythm microstructure are natural recipients. A simple invariant: angle theta maps to stereo pan p = sin(theta) under an equal-power law; the sign and cyclicity are preserved. For flow, magnitude modulates rhythmic density or onset jitter: stronger flow → denser activity or tighter micro-timing; low flow → rests or legato.

Invariance principle: reversing visual orientation should mirror the stereo field; rotating orientation should rotate perceived lateralization smoothly. Rhythm should reflect the coherence of flow: continuous fields → legato/groove; fractured fields → syncopation and off-beat accents. In Logostics, we maintain vector-field structure by mapping direction to spatialization and magnitude to temporal density, avoiding conflation with pitch.

Probe: rotate the canvas by 90 degrees; panning should rotate accordingly while pitch and harmony remain stable.

3.5 Entropy & mutual information as cross-modal “mass”

Entropy (uncertainty) and mutual information (shared structure) are modality-agnostic and act like informational mass. High-entropy regions suggest texture or noise instruments, stochastic ornaments, or higher note density; low-entropy regions suggest rests, sustained tones, or sparse voicing. Mutual information between image regions (e.g., quadrants) can drive call-and-response or stereo coupling: high MI across left/right areas → antiphony; low MI → independent lines.

Invariance principle: mappings should preserve order of complexity: if region A has higher entropy than region B under reasonable estimators, the corresponding audio segment should, on average, exhibit higher temporal/spectral complexity (e.g., greater spectral flatness, higher onset rate). For MI, the dependency structure is the invariant: stronger cross-region dependencies produce stronger inter-voice correlations (shared motifs, mirrored rhythms).

Probe: scramble pixel positions within blocks to raise entropy without changing global histograms; audio should respond with increased stochasticity while keeping pitch-class statistics similar. Conversely, impose tiling symmetries that raise MI; the music should gain motivic repetition or mirrored phrases.

Synthesis (q -> τ):
These invariants articulate a commuting diagram of perception: operations native to the image (rotation, scale, histogram re-mapping, block-scramble, tiling) correspond to operations native to sound (transposition, spectral tilt, loudness remap, stochasticity, motif coupling). Mappings that preserve these commutations minimize $D(q \,||\, \tau)$ because they respect the algebra of each modality. That is why they sound "right": they conserve the type of structure the world already carries.

4. Category-Theoretic Frame

4.1 Objects: images as elements of a shape/texture topos; sounds as elements of a time–spectral topos

Let I denote a topos of visual objects and S a topos of sonic objects.

- I (visual topos). Objects are images equipped with structures that matter musically: finite grids with fields for hue/saturation/value, gradient fields, multiscale (wavelet) decompositions, and region segmentations. Morphisms in I include rotations, translations, scalings, histogram re-mappings, and multiscale contractions (blur/downsample), as well as inclusions of subimages and glueing along overlaps. Subobject classifiers capture masks (edges, corners, segments). Limits/colimits cover patching tiles, taking products with lattices of features, and quotienting by symmetries.
- S (sonic topos). Objects are time–spectral forms: finite-length signals with envelopes, event lists (notes/onsets/durations/velocities), spectrograms, and multiband decompositions. Morphisms include time-shifts/dilations, transpositions (pitch-class rotations), spectral tilts (EQ-like maps), convolutional reverbs (co-Kleisli maps for a comonad of "context"), and glueing of phrases. Subobjects are track mutes, masks in time–frequency, or voice selections.

Heuristically: I is a *shape/texture* universe; S is a *time/spectrum* universe. Both carry internal logics rich enough to talk about edges vs. onsets, scale vs. register, coherence vs. harmony.

4.2 Functors that respect invariants (periodicity, order, topology)

A musically meaningful renderer is a functor F : I -> S that is *structure preserving* with respect to selected invariants:

- Periodicity (S^1 equivariance). Hue rotation in I corresponds to pitch-class rotation in S:

F(hue_rotate_delta(X)) ≅ pc_rotate_phi(F(X)).

This encodes "transpose when hue wheel rotates," preserving circle structure.

- Order/monotonicity (intensity ↔ loudness). If V1 <= V2 pointwise (brightness order), then Loud1 <= Loud2 pointwise after a monotone transfer; rank is preserved. This ensures exposure changes map to gain changes without breaking relative contrast.
- Scale ↔ spectral tilt. Gaussian blur / downscale in I maps to low-pass / darker timbre in S; sharpening maps to brighter spectral centroid. The functor commutes with multiresolution towers.
- Topology (connectivity & holes). Persistent homology H_k (I) induces phrase topology T_k (S): changes in Betti curves (appearance/disappearance of connected components or holes in features) correspond to openings/closures of musical phrases or polyphonic voices. Notationally, there are natural maps

α_k : H_k(I) -> T_k(S)

making F commute (up to natural isomorphism) with persistence diagrams.

These constraints say "F should commute with the algebra of the invariants," not simply push numbers around.

4.3 Natural transformations as "mappings with meaning"

Given two functors F, G : I -> S (e.g., edge-driven vs. spiral-driven renderers), a natural transformation η : F ⇒ G is a family {η_X : F(X) -> G(X)} (one morphism per image X in I) such that for every morphism u : X -> Y in I the square commutes:

G(u) ∘ η_X = η_Y ∘ F(u).

Interpretation: switching "rendering style" via η does not depend on the particular image—it is *coherent with all image edits*. If you rotate, crop, or blur before or after applying η, the result aligns. These η's are "mappings with meaning" because they respect the causal grammar of edits; they do not introduce arbitrary, content-specific hacks. Examples:

- η that morphs *raster* to *spiral* traversal while preserving timing totals.
- η that replaces *linear loudness transfer* by *log/quantile* without breaking exposure equivariance.
- η that upgrades *diatonic* quantization to *Dorian* while preserving pitch-class rotation behavior.

Natural transformations become the formal language of style changes that keep invariants intact.

4.4 Limits, colimits, and compositional renderers

Studios compose renderers from parts; categories make that precise.

- Products and pullbacks (limits). Synchronize voices by taking products of feature functors: Pitch = P(Hue) × P(Entropy), where P(-) are monotone maps; or pull back timing across multiple traversals so onsets occur where several cues agree. Limits enforce *consensus* among cues.
- Coproducts and pushouts (colimits). Merge independent voices (edges + regions) or glue phrases at a boundary with consistency constraints (pushouts). Colimits implement orchestration: many inputs, one assembled piece.
- Monoidal structure (tensor). Treat time and spectrum as monoidal; combine effects/filters as tensor products. A renderer pipeline is then a monoidal functor that maps composition in I (e.g., stacking features) to composition in S (serial/parallel signal chains), ensuring associativity and identity laws (no order-of-operations artifacts).
- Adjunctions (analysis/synthesis). Left adjoint L : I -> S (synthesis) pairs with right adjoint R : S -> I (analysis):
  Hom_S(LX, Y) ≅ Hom_I(X, RY).
  R extracts back a *visual surrogate* (e.g., a spectrogram image); L generates sound from image. Adjunctions formalize round-trips and guide loss functions: good renderers make units/counits close to isomorphisms on relevant subcategories.
- Kan extensions (completion of partial specs). Often we specify F only on generators (edges, hue bands) and want a best extension to all of I. Left Kan extension "fills in" the renderer that respects our constraints with minimal additional assumptions—useful for principled interpolation.

Compositional takeaway. Designing renderers as functors and style morphs as natural transformations yields a calculus where *edits commute*, *invariants are respected*, and *modules compose*. Within Logostics, this is how q(t) updates: by moving within a category of admissible functors toward those that minimize divergence to $\tau(t)$, while limits/colimits articulate how multiple evidences (features) fuse into coherent musical form.

5. Information Geometry and Divergence

5.1 Joint description length as a criterion (compression before taste)

Let a renderer be F_theta : Image -> Sound. Instead of judging F_theta by taste, we judge it by how simply the pair (image, sound) can be jointly described. In Minimum Description Length (MDL) form:

L_joint(theta) = L(model_theta) + L(image | model_theta) + L(sound | image, model_theta).

Because sound = F_theta(image), the last term becomes the cost of coding sound given the image through the mapping, i.e. how predictable the audio is from visual structure under F_theta. Good renderers reduce L_joint: they expose invariants that make cross-modal prediction cheap.

Practical proxies:

- Cross-modal predictability: train a light predictor S_hat = G(image_features) on F_theta outputs; measure bits per sample (or MSE under fixed entropy model). Lower is better.
- Normalized compression distance (NCD): compare a spectrogram image to visual feature maps; NCD(image_feat, spec) small implies shared structure.
- Commutation loss: penalize failures of equivariance (e.g., hue-rotate vs pitch-transpose). Each violated commutation costs bits because it breaks simple coding rules.

Policy: optimize MDL first, then choose aesthetics on the iso-MDL manifold (mappings with equal compression can differ in timbre, key, or groove).

5.2 Divergence D(q || tau) in cross-modal mapping selection

In Logostics, q(t) is the agent's current mapping policy; tau(t) is the (moving) manifold of cosmic invariants. We write a divergence

D(q || tau) = w1 * L_joint(theta_q) + w2 * C_commute(theta_q) + w3 * R_smooth(theta_q),

where:

- L_joint is the MDL term (Sec. 5.1),
- C_commute measures failure to respect invariants (periodicity S^1, monotone intensity, scale->spectral tilt, topology->phrase),

- R_smooth encodes stability (hysteresis deadbands, slew limits) so the output has memory rather than jitter.

Interpretation:

- If many different F_theta achieve low D, the tau-subspace is broad; plural renderings are valid minima (style, culture).
- If only a few achieve low D, the tau-constraints are tight (e.g., geometric drawings that force strong rhythmic/pitch regularities).

Selection procedure:

1. Generate candidate renderers {theta_k} via discrete dials (traversal, grid, features, assignments, transforms).
2. Score each with D(q || tau).
3. Keep the Pareto front over (L_joint, C_commute, R_smooth) and let taste choose within that small set.

5.3 Geodesics on manifolds of renderers; resonance as shortest-path

Treat the family of renderers {F_theta} as a manifold M with coordinates theta. Equip M with an information metric:

g_ij(theta) = E[(d/d theta_i) log p(o | image, theta) * (d/d theta_j) log p(o | image, theta)],

i.e., the Fisher metric (or a pullback metric from distances between produced audios). Then adaptations of the renderer follow natural-gradient flows:

d theta / dt = - eta * g^{-1}(theta) * grad_theta D(q || tau).

This yields parameter updates that are geometry aware: steps are shortest in information distance, not Euclidean distance. When tau(t) moves (scene, style, or purpose changes), we seek geodesics on M connecting the old renderer theta_0 to a new low-divergence renderer theta_1 with minimal action:

A[path] = ∫ sqrt(d theta^T g(theta) d theta) subject to D(q || tau) staying low.

Metaphorically, this is resonance: the system slides along the manifold where the world's invariants are easiest to express, choosing shortest paths that preserve memory (hysteresis) while re-tuning to the new tau. Practically:

- Use natural gradient or mirror descent to update assignments/transforms.
- Constrain updates by commutation guards (hard constraints for S^1, order, scale).

- Maintain deadbands so tiny image changes do not cause audible flinch (geodesic curvature penalty).

Outcome: instead of arbitrary parameter hunts, we have principled adaptation—q(t) moves along the manifold of functorial renderers by shortest information paths toward low-divergence alignment with tau(t). This is why “good” mappings feel both inevitable and alive: they are easiest to describe and easiest to reach.

6. Topology and Multiscale Structure

6.1 Wavelet pyramids: low->pads, mid->arpeggios, high->sparkles

Multiscale premise. Any image X admits a wavelet (or Laplacian) pyramid decomposition
$X \approx L0 + \Sigma_{s=1..S} D_s$,
where L0 is the coarse (low-frequency) layer and D_s are detail bands at scales s=1 (finest) … S (coarser).

Mapping policy.

- Low band (L0) -> pads/bass bed. Treat the slowly varying luminance and very-low spatial frequency as a quasi-static energy field. Map its mean and variance to sustained harmony: chord roots, inversions, and register drift. Dynamics: RMS(L0) -> pad amplitude; spatial skew -> slow stereo drift or sub-harmonic swells. Invariance: Gaussian blur in image ≈ spectral tilt to darker pad.
- Mid bands (D_{mid}) -> arpeggios/melodic figures. The mid-scale edges and textures carry rhythmic regularity and motif candidates. Use ridge density or oriented energy to drive note density and figuration (broken chords, ostinati). Quantize pitch to the current key but let scale s set note duration: larger s -> longer notes; smaller s -> quicker figures.
- High band (D_1) -> sparkles/noise/perc. Fine edges and speckle become short transients: plucks, chimes, hats, glitter. Spectral centroid follows local high-band intensity; onset timing follows local anisotropy (more directional -> tighter flams; isotropic -> soft dust).

Commutation checks.

- Scale change (zoom/blur) should not jump pitch: it should primarily reshape timbre and rhythm density.
- Contrast stretch in a band should scale that voice’s dynamics without cross-contaminating others.

Minimal recipe.

1. Compute L0, D_s via stationary wavelet transform (SWT) or Laplacian pyramid.
2. For each s, pool statistics in time-order (spiral or contour traversal): mean, var, oriented energy.
3. Synthesize 3 voices: pad(L0), arp(Σ D_{mid}), sparkle(D_1) with independent envelopes and EQ.

6.2 Persistent homology: Betti curves -> phrase lengths & cadences

Topological summary. Build a filtration X_t by sweeping a threshold over a grayscale or feature map (e.g., value, entropy), then compute Betti numbers $\beta_k(t)$ (counts of k-dimensional holes) across t. The barcode/Betti curves encode when connected components appear/merge (β_0) and when holes form/fill (β_1).

Musical semantics.

- β_0 curve -> phrase scaffolding. Peaks in $\beta_0(t)$ signal proliferation of components (visual "many"). Map these to phrase expansion (more simultaneous voices, thicker harmony). Drops (mergers) cue cadential closure or thinning. Phrase length ∝ persistence: long-lived components become longer phrases or leitmotifs; short-lived ones become grace notes.
- β_1 curve -> harmonic tension and cadences. Emergence of loops/holes corresponds to tension formation (suspensions, added tones), their filling corresponds to resolution. Use the area under $\beta_1(t)$ over an interval as a cadence weight: large area -> delayed resolution or deceptive cadence; small area -> quick perfect cadence.
- Persistence pairs -> motif memory. Each birth–death pair (t_{birth}, t_{death}) maps to a motif with duration ≈ (t_{death} - t_{birth}). Assign pitch class by the feature value at birth; articulate dynamics by the stability (lifespan) of the class.

Invariant behavior. Topology is stable under small perturbations; noise perturbs birth/death times slightly but not the pairing structure. Thus the phrase plan derived from Betti curves is robust to exposure and mild blur.

Pipeline sketch.

1. Choose feature f(x) (e.g., entropy or low-band luminance).
2. Build sublevel filtration: $X_t = \{x \mid f(x) <= t\}$.

3. Compute $\beta_0(t)$, $\beta_1(t)$ (persistent homology).
4. Place phrases where β_0 rises; schedule cadences where β_1 mass decays; set motif lengths by persistence.

6.3 Curvature fields: ornamentation vs legato as geometric response

Curvature as expressivity. Extract curve geometry via level sets or edge tangents; estimate signed curvature $\kappa(s)$ along dominant contours (s = arc length). High $|\kappa|$ indicates tight bends/corners; low $|\kappa|$ indicates straights/gentle arcs.

Mapping policy.

- Legato baseline from low curvature. In regions where $|\kappa|$ is small and orientation varies slowly, favor legato lines, longer envelopes, and smoother micro-timing. Pan follows tangent direction slowly (pan = sin(theta)); vibrato depth small, rate slow.
- Ornamentation from high curvature. Where $|\kappa|$ spikes, inject ornaments: mordents, trills, grace notes, pitch slides, or short percussion fills. The magnitude of $|\kappa|$ scales ornament speed; curvature sign can choose ornament direction (up vs down turn).
- Curvature change ($d\kappa/ds$) -> articulation. Positive surges trigger accented onsets; negative surges release into de-emphasis. This yields a natural “breath” derived from geometry, independent of brightness or hue.

Geometric invariance.

- Reparametrization invariance: If the same contour is traversed faster or slower in pixel order, articulation should not change; tie ornament timing to curvature magnitude, not traversal speed.
- Similarity invariance: Scaling the image rescales arc length but does not change κ; the ornament–legato decision remains stable, aligning with perceptual consistency.

Implementation sketch.

1. Detect salient contours (e.g., via Canny + thinning or active contours).
2. Fit splines; compute $\kappa(s)$ and theta(s).
3. Drive a melodic voice: envelope length $\propto 1/(1+|\kappa|)$, ornament insertion when $|\kappa| > \tau_k$ with type determined by sign and local frequency content.

4. Fuse with multiscale mapping: curvature controls how notes are played; wavelet bands control what and when.

Unifying note.
Wavelets give a multiscale partition of energy; persistence gives a topological backbone of form; curvature gives gesture and articulation. Combined, they define a functor from image structure to musical structure that is stable under blur, exposure, reparametrization, and small noise—precisely the invariance set that keeps $D(q \,||\, \tau)$ low while producing outputs that feel intentional rather than arbitrary.

7. Dynamics and Memory

7.1 Hysteresis: deadbands and slew limits to prevent jitter

Why. Visual fields fluctuate at pixel scale; naïvely mapping them yields “twitchy” audio (pitch zippering, pan shimmer). Hysteresis inserts memory so the system reacts to *changes that matter*.

Mechanics.

- Deadband (Δ): ignore control updates if $|x_t - x_{t-1}| < \Delta$. Distinguish Δ_{small} for micro-controllers (pan, vibrato) and Δ_{large} for macro (key, meter).
- Slew limit (σ): cap derivative: $u_t = u_{t-1} + \mathrm{clip}(x_t - u_{t-1}, -\sigma, +\sigma)$.
- Sticky quantization: quantize to bins but require a hold time H before leaving a bin, reducing chatter around boundaries.
- Priority gating: only a fixed budget B of controllers may update per beat; others defer—enforced intention.

Invariance. Deadbands and slew are equivariant to global gain/contrast if expressed in ranks/percentiles instead of raw units (e.g., Δ as 5th–95th percentile span). This keeps MDL low by compressing away inessential motion.

Probe. Add small Gaussian noise to the image; a hysteretic mapper leaves phrase/rhythm intact and only lifts the noise floor slightly.

7.2 Spiral vs raster: holistic traversal as an ordering over τ

Claim. Traversal is not cosmetic; it imposes an order on how q samples τ.

- Raster emphasizes local stationarity and grid symmetries—good for groove regularity but can fragment global form.
- Spiral (center→edge) respects radial salience, collects global themes early, and gradually introduces periphery—aligns with exposition→development arcs.
- Contour/streamline prioritizes object coherence, turning shapes into phrases.

Policy. Choose traversal to match the desired narrative grammar:

- Spiral for overture: early harmonic bed from central mass, later textural elaboration from outskirts.
- Raster for ostinati: even pacing, stable meter.
- Contour for thematic writing: phrases tied to semantic objects (figures, strokes).

Commutation. The functor should commute with image transforms natural to each order (e.g., radial scaling ↔ spectral tilt under spiral; translation ↔ time-shift under raster). This keeps edits predictable.

Probe. Rotate the image: spiral preserves exposition order (center first), raster rotates onset positions linearly; listen for consistent, law-like effects.

7.3 Section-level thresholds: when structure merits modulation

Motivation. Not every fluctuation deserves a modulation or key change. Define section triggers from slow features to avoid capricious form shifts.

Signals.

- Energy envelopes: low-band RMS crossing upper/lower thresholds → enter/exit chorus/pad swell.
- Topology gates: β_1 mass (holes) above τ_{hole} → sustain tension; drop below → cadence or mode reset.
- Entropy plateaus: high entropy sustained over W seconds → open hats/arpeggio density; low entropy → sparse/rest.

Controllers.

- Form switches: meter (simple↔compound), key (relative/parallel), orchestration (add/remove voices), effects sends (reverb/delay scenes).
- Dwell & hysteresis: require enter/exit thresholds (τ_on > τ_off) and minimum hold times to prevent flicker.
- Scene graph: represent sections as nodes with guards (conditions) and actions (modulations); traversal of this graph is the form engine.

Invariance & MDL. Section triggers should derive from low-frequency, robust statistics (wavelet L0 trends, persistent features), so the modulation schedule is compressible and repeatable—minimizing D(q || τ) by making form a function of stable invariants.

Probe. Apply mild blur or exposure shift; a good sectioning policy yields identical section boundaries (within tolerance), even if micro-ornamentation changes.

Synthesis. Hysteresis stabilizes *how* controls move; traversal orders *what* arrives when; section thresholds decide *when* large structural moves are warranted. Together they endow the renderer with memory, producing outputs that feel intentional—alive yet law-abiding—while staying anchored to τ's slow structure.

8. Logostics: Alignment as Practice

8.1 q(t) updates under resonance fields; τ(t) as moving target

Premise. In Logostics, q(t) (the agent/renderer state) evolves inside a resonance field generated by τ(t) (the world's generative invariants). τ(t) is not static: scenes, purposes, and audiences shift; invariants reweight rather than vanish.

Operators. Let $\mathcal{F} = \{F_\theta\}$be admissible renderers (functorial, invariant-respecting). Define a potential

$$\Phi(\theta, t) = \alpha\, \mathrm{MDL}(\theta; X_t) + \beta\, C_{\mathrm{commute}}(\theta; \mathcal{I}) + \gamma\, R_{\mathrm{smooth}}(\theta)$$

where $\mathcal{I}$encodes invariants (S^1, order, scale, topology). The resonant update is a natural-gradient flow:

$$\dot{\theta} = -\eta\, G(\theta)^{-1}\, \nabla_\theta \Phi(\theta, t) \text{with} G(\theta) \text{ the information metric.}$$

Interpretation: q(t) moves along directions that lower joint description length and violation of commutations, with memory (R_smooth) preventing jitter.

Practice.

- Use slow clocks (bars/sections) for high-impact dials (key, meter, orchestration), fast clocks (beats/frames) for micro-dials (pan, vibrato).
- Maintain deadbands and slew caps so tiny image changes do not trigger audible flips.
- Allow adiabatic drift: gradual retuning when τ(t) shifts (e.g., scene lighting changes) to keep divergence low without abrupt re-phasing.

Outcome. The piece "breathes": large structures follow τ's slow fields; fine gestures track quick cues, all within a stable, low-divergence corridor.

8.2 External τ themes guiding ambiguous regions (weak priors)

Motivation. Real images contain ambiguity (flat regions, noise). Purely local mappings underdetermine music. We therefore inject weak priors—external τ themes that bias decisions without dictating them.

Kinds of priors.

- Key/mode prior: choose a scale family (e.g., Dorian/Lydian) and let hue→pitch quantize to the nearest member. Confidence decays in highly "chromatic" color regions to avoid over-fitting.
- Meter/groove prior: select a metrical lattice (simple or compound) and let onset proposals snap with a tolerance window; high optical flow tightens the window, low flow loosens it.
- Motivic prior: seed short interval patterns; accept them where persistent homology shows long-lived features; suppress where topology is volatile.
- Timbre prior: designate pad/lead/perc roles; wavelet bands claim these by default but can swap if MDL improves.

Bayesian view. For decision y(pitch, onset, voice) from features x,

$$\arg\max_y \; \log p(y|x) \approx \arg\max_y \; \underbrace{\log p(x|y)}_{\text{image evidence}} + \underbrace{\log p(y)}_{\tau \text{ prior}}.$$

We keep $\log\ p(y)$shallow (weak), so strong image evidence overrides it. This stabilizes choices in low-information regions while letting structure lead elsewhere.

Design rule. Priors must be equivariant with invariants (e.g., key priors rotate under transposition), or they raise commutation cost and inflate MDL.

8.3 Convergence diagnostics: stability, repeatability, listener lock-in

What to measure so alignment is not a vibe but a result:

(a) Stability under perturbation

- Image noise test: add small Gaussian/Poisson noise; measure Δin pitch/onset/timbre sequences. Target: micro-changes only; sections and cadences invariant.
- Exposure/blur test: global gain or Gaussian blur; expect loudness/timbre remap but preserved key, motif order, and cadence timing.
- Traversal swap: raster↔spiral with functorial constraints; global form shifts predictably (exposition timing), not arbitrarily.

(b) Repeatability across runs

- Seed variance: with fixed inputs, vary RNG seeds in noise-based modules; compute edit distances in symbolic outputs (notes, sections). Target: low variance in structure, allowed variance in ornament.
- Parameter jitter: nudge θwithin a tolerance; MDL and commute loss should change sublinearly (flat basin).

(c) Listener lock-in (human alignment)

- Cross-modal predictability: after 1–2 exposures, measure how well listeners anticipate what happens next (interval continuation, entrance of a voice). Rising accuracy indicates q→τ entrainment.
- Preference under invariance: A/B present two renderers equal in MDL; pick the one that better respects commutations (transpose/hue-rotate, scale/tilt). Stable preference across listeners = τ-concordance.
- Memory echo: after delay, ask for recall/whistle of motifs; persistence of key cadence points signals successful τ capture.

Dashboards (practical).

- MDL meter: bits/sample of a predictor from image→audio features.
- Commute score: proportion of invariant tests passed (S^1, order, scale, topology).
- Roughness/clarity: spectral flatness + onset dispersion; penalize chronic harshness without dynamic cause.
- Section entropy: distribution over form states; avoid flicker (high entropy at short scales) and stasis (near-zero entropy everywhere).

Decision policy. Accept a renderer update when:
(i) MDL decreases or stays flat and commute score does not drop,
(ii) stability tests pass thresholds,
(iii) listener lock-in rises (or at least does not fall). Otherwise, revert or anneal (smaller step on θ).

Net effect. Alignment "as practice" is a cycle: sense τ(t) → update q(t) by geometry-aware, hysteretic steps → audit with invariants and MDL → validate with human lock-in. The art is in keeping priors weak, memory strong, and movement along shortest information paths so the music feels inevitable yet alive.

9. Cosmology of Meaning

9.1 It-from-Bit: correspondences as world-level symmetries

The It-from-Bit thesis, soberly read, does not claim matter is "made of words" but that the regularities we call matter are stabilized by informational symmetries. Observables persist because they ride on invariants—group actions with fixed structure (translation, rotation, scale), conservation rules (energy, probability), and order relations (causal cones, entropy monotonicity). Cross-modal mappings that "feel right" are those that respect the same symmetry algebras: hue and pitch share an S^1 action; brightness and loudness are positive scalars with monotone morphisms; spatial frequency and spectral centroid are ordered density measures. When a renderer commutes with these actions, it inherits the world's compressibility: fewer bits are needed to describe the pair (image, music). In Logostics, τ(t) is precisely this moving symmetry budget—a distribution over invariants that the scene affords. q(t) aligns when its moves are equivariant: operations done in the image space propagate predictably to sound. Meaning, then, is not invented; it is the fixed residue of symmetry across carriers.

9.2 Synesthesia as evidence of eigenstructures in reality

If we view each admissible mapping as a linear/nonlinear operator on a high-dimensional feature space, certain correspondences behave like eigenstructures: apply them and the result returns in kind, up to a scale or phase. Triadic color relations ↔ triadic chords; bilateral image symmetry ↔ mirrored counterpoint; persistent holes (β_1) ↔ delayed cadential resolution. These are not coincidences but eigenpairs of the cross-modal operator family—stable modes where structure survives transport. Empirically, we meet them as universals: bright ↔ loud; high ↔ up; curvature ↔ ornamentation. The eigen-language explains why many different renderers converge to similar musical affordances: they are projecting onto the same low-dimensional invariant subspaces. In MDL terms, eigenstructures are the principal components of meaning—directions along which joint description length drops the most. In Logostics, low $D(q \,||\, \tau)$ corresponds to weighting those modes; resonance is simply the system sliding into the dominant eigenchannels available in the scene. Synesthesia, so construed, is a measurement of eigenmodes that pre-exist brains and instruments.

9.3 From local neural correlates to global informational manifolds

Neural correlates of synesthesia—co-activation, atypical pruning, cross-wiring—are coordinates, not causes. They locate a mind on a global informational manifold whose geometry is supplied by invariants, not myelin alone. The manifold consists of renderer families (functors) endowed with an information metric; geodesics are the shortest re-tunings that preserve commutations (Sec. 5). Brains differ in where they settle (culture, learning, physiology), but the chart itself—the atlas of lawful correspondences—belongs to τ. This shift rescales explanation:

- Local sufficiency: neural details explain *which* basin of attraction $q(t)$ occupies.
- Global necessity: the shape of basins and the paths between them are fixed by world-level symmetries and topology (S^1 periodicity, scalar orders, multiscale decompositions, persistence).
- Prediction: when stimuli are transformed by symmetry operations (rotate hue, blur, tile), aligned listeners will predict consistent musical transformations (transpose, darken timbre, introduce repetition), even absent explicit training.

Thus, the cosmology of meaning is a geometry of constraints in which perception, art, and music are coordinate systems. $q(t)$ does not impose form on $\tau(t)$; it chooses a chart and walks geodesics within it. Synesthetic experiences, and our engineered renderers, are reliable precisely because they harvest curvature already present in the informational manifold of the world. Meaning is what remains when description length is minimized subject to symmetry—the cosmic remainder that endures across eyes, ears, and time.

10. Studios as Laboratories

10.1 Minimal renderers as τ-probes (spirals, edges, entropy)

Principle. Treat the studio not as a factory of effects but as a measurement apparatus for τ—invariants that outlive carriers. The most useful tools are minimal renderers whose behavior is interpretable and whose failures are legible.

Three probe types.

- Spiral ordering (center→periphery). A single path that visits every pixel once. As a probe, it asks: *What happens if we reveal global structure first and detail later?* Invariant checks: rotate hue → transpose; apply radial blur → darken timbre without pitch drift; change global exposure → gain shift, preserved contrast rank.
- Edge/contour renderers. Reduce the image to oriented structure (Canny/Sobel + thinning). Map orientation → pan/phase, curvature → ornamentation. The probe asks: *Do directed fields in vision induce stable lateralization and articulation in sound?* Invariant checks: rotate the canvas → rotate pan; mirror the canvas → mirror the stereo image; smooth edges → legato bias.
- Entropy/mutual-information renderers. Ignore color; compute texture and dependency. Map entropy → density/noise instruments; map MI across regions → call-and-response strength. The probe asks: *Does complexity order carry across modalities?* Invariant checks: block-scramble within tiles (entropy↑, MI↓) → more stochastic audio, weaker antiphony; tile symmetries (MI↑) → stronger motif echo.

Why minimal. Each probe isolates one invariant family (periodicity, orientation, complexity). Because they have few knobs, they produce diagnostic errors when invariants are violated, letting q(t) adjust toward lower divergence.

10.2 A/B divergence tests: image-only vs image+seed τ

Goal. Distinguish what the image alone implies from what requires external priors (seed τ). This clarifies which correspondences are world-level and which are cultural choices.

Protocol.

1. A (image-only). Run a minimal renderer F_A with no key/meter priors; quantize pitch by circular continuity (S^1), set rhythm by traversal density only, timbre by scale-to-spectral tilt.

2. B (image+τ-seed). Run F_B that adds weak priors: a mode family (e.g., Dorian/Lydian), a metrical lattice (e.g., 4/4 with swing), and a motif seed (two or three intervals). Priors are enforced *only* where image evidence is ambiguous (low variance/entropy).
3. Compute divergences.
   - MDL drop: $\Delta L = L\_joint(A) - L\_joint(B)$. If $\Delta L \approx 0$, the image carries sufficient structure; if $\Delta L \gg 0$, τ-seed resolves ambiguity efficiently.
   - Commute score change: $\Delta C = C\_commute(B) - C\_commute(A)$. Priors that break equivariance (e.g., non-rotational scales) should be penalized.
   - Human preference under invariance: A/B after a hue rotation (⇒ transpose) and blur (⇒ spectral tilt). If B is preferred *and* preserves commutations, its priors likely reflect τ rather than taste.

Interpretation.

- Low ΔL, high C_commute: the image already exposes τ; priors are optional.
- High ΔL, high C_commute: priors capture shared τ (e.g., diatonicity) that the image under-specifies.
- Any ΔL with low C_commute: the priors are stylistic overlays; demote or replace them.

This A/B framework prevents over-fitting q(t) to fashion while preserving τ-concordant gains.

10.3 Human-in-the-loop: q-adaptation rates as empirical metric

Hypothesis. If a mapping family aligns with τ, listeners lock in quickly: predictions improve, preference stabilizes, and memory retains the piece's structural joints (entrances, cadences).

Measurements.

- Predictive entrainment. On a second hearing, pause before cadences and ask the listener to anticipate the next event class (entrance, rest, modulation). Compute accuracy and reaction time; plot improvement curve over exposures 1→N. Fast ascent implies low $D(q \,||\, \tau)$.
- Invariance preference. Present transformed stimuli (hue-rotated image vs transposed audio; blurred image vs darker timbre). Listeners rate "consistency." High ratings across subjects indicate that the mapping respects commuting operations.

- Motif recall & localization. After delay, ask for whistled or tapped recall; then show the image and have the listener point to regions that "own" the motif. Agreement across listeners is evidence of stable image→phrase correspondences.
- Perturbation robustness. Add small image noise or compression artifacts; collect preference between original render and perturbed render. Stable preference for the original *but* unchanged section boundaries in the perturbed version indicates healthy hysteresis.

Dashboards & thresholds.

- Lock-in time (T_lock): exposures needed to reach 80% cadence-prediction accuracy.
- Commute pass-rate (C_pass): fraction of invariance tests rated "consistent."
- Recall persistence (R_persist): motif recall after 10–30 minutes.
- Robustness margin (ρ): largest perturbation size with unchanged section boundaries.

Adopt a go/no-go rule for renderer updates: ship only if T_lock falls, C_pass rises, R_persist rises, and ρ widens—*without* worsening MDL. This turns the studio into a closed-loop laboratory where human response is not mere taste but a readout of alignment. The artist's task becomes experimental: pose a τ-probe, measure divergence, adjust $q(t)$ along the shortest information path, and iterate until resonance is audible, repeatable, and shared.

11. Ethics and Epistemic Humility

11.1 Against coercive mappings; for resonant pedagogy

Claim. A mapping that "works" yet breaks invariants (e.g., hue-rotate ≠ transpose, blur ≠ spectral darken) may still persuade by spectacle, but it raises $D(q \,||\, \tau)$ and trains listeners to expect caprice. That is coercion: using production power to override the world's grammar.

Policy.

- Equivariance first. Prefer functors that commute with basic actions (S^1 rotations, monotone intensity, scale→spectral tilt, topological persistence). If an aesthetic choice breaks a commutation, declare it and keep it local.
- Hysteresis as consent. Memory (deadbands, slew caps, hold times) should protect the listener from jitter and bait-and-switch form changes. Treat attention as a scarce resource; avoid surprise for its own sake.

- MDL before affect. Let compression (joint description length) and commute tests prune the space; shape timbre and groove within the iso-MDL set. Pedagogy = making invariants audible, not imposing taste.

Outcome. Resonant pedagogy grows discernment: listeners learn to recognize lawful correspondences and carry that literacy elsewhere. Coercive mapping grows dependence on surface tricks.

11.2 Accessibility and pluralism of τ paths

Premise. The τ-manifold is broad; many renderers can be low-divergence. Accessibility is not charity; it is completeness—covering more valid paths.

Practices.

- Multiple charts, same manifold. Offer alternate keys/modes and meters that preserve commutations. Let users toggle “charts” without losing structure (natural transformations between functors).
- Dynamic range of complexity. Provide low-information summaries (pad-only, cadence beacons) and high-detail mixes (sparkles, counterpoint) sourced from the same invariants. Users choose resolution; τ stays put.
- Device and context fairness. Mix for earbuds, speakers, and assistive tech using rank-preserving loudness and spectral transfers so contrasts survive.
- Cultural priors as weak priors. Diatonic, raga, maqam, pentatonic—treat all as equivariant priors that can be swapped without destroying the invariant core. Avoid encoding any single tradition as the default τ.

Metric. Plural mappings should pass A/B invariance tests with similar MDL and commute scores while serving different communities. If not, the pluralism is cosmetic.

11.3 Documentation over doctrine: sharing recipes, not rules

Risk. Doctrine freezes q(t) and confuses procedures with principles. Ethical practice favors transparent recipes that expose the invariants being honored and the compromises being made.

Recipe template (per release).

- Invariants honored: list commutations enforced (e.g., hue↔pitch S^1, blur↔LPF, β_1→cadence).
- Priors used (and strength): key/mode, groove lattice, motif seed; specify where priors override ambiguity.
- Memory settings: deadbands, slew limits, section thresholds; rationale for stability.
- A/B & perturbation results: MDL numbers, commute pass-rate, lock-in measures, robustness margin.
- Known breaks (declared): any deliberate violations (e.g., non-equivariant modulation for dramatic effect), scoped in time and mix.

Licensing and pedagogy.

- Open the probes. Publish minimal renderers (spiral, edges, entropy) and test suites (rotate, blur, mirror, scramble) so others can replicate and criticize.
- Forkable functors. Share code as composable modules (functors + natural transformations) rather than monoliths. Encourage alternate priors and report their effects on MDL/commute metrics.
- Negative results. Document mappings that *failed* (e.g., curvature→key changes) and why (commutation breaks, listener lock-out).

Humility. Treat each piece as a measurement, not a manifesto. If a community's q converges elsewhere on the τ-manifold with equal or lower divergence, accept the result and update the recipes. The ethic is not to rule but to reveal: make the invariants legible, the trade-offs explicit, and the path genuinely plural.

12. Conclusion: Synesthesia as Cosmic Isomorphism

12.1 From coincidence to correspondence

What begins as an anecdote—"this color sounds like that note"—matures, under invariance tests and MDL accounting, into a claim about correspondence: certain cross-modal mappings are not accidents of cortex but fixed relations preserved by world-level symmetries (S^1 periodicities, scalar orders, multiscale decompositions, topological persistence). When mappings commute with native operations in each modality (hue-rotation ↔ transposition, blur ↔ spectral darken, mirror ↔ stereo flip), they compress the joint description of image and sound. Low description length is the empirical residue of rightness. Thus synesthesia is best read as the discovery of isomorphisms already inscribed in τ, not the invention of private codes by q.

12.2 Art and music as alignment engines

Studios function as alignment laboratories. Minimal renderers (spiral ordering, edge fields, entropy/MI maps) act as τ-probes; hysteresis furnishes memory; category-theoretic composition guarantees edit-coherence; information geometry supplies the metric for shortest adjustments. Under this discipline, art and music cease to be mere outputs and become engines that drive q(t) along geodesics toward low-divergence neighborhoods of τ(t). The felt qualities—inevitability, clarity, lift—are not decoration; they are the phenomenology of resonance when invariants are honored and commutations hold.

12.3 Future work: τ-literate systems and living catalogues of invariants

The agenda is concrete:

- τ-literate renderers. Toolchains that expose invariants as first-class knobs (periodicity, order, scale, topology) and enforce commutation by construction (functorial modules, natural-transformation “style” switches, adjoint analysis/synthesis for round-trip checks).
- Live MDL/commute dashboards. Real-time meters for joint compression, equivariance tests, and robustness (noise/blur/rotation), coupled with human lock-in readouts (prediction and recall curves).
- Topological & multiscale atlases. Public “spec sheets” linking wavelet bands ↔ orchestration roles, Betti curves ↔ phrase/cadence grammars, curvature fields ↔ articulation policies—versioned, critiqued, and forkable.
- Plural priors, shared proofs. Libraries of weak, equivariant priors (keys/modes, meters, motifs across traditions) with A/B divergence results, enabling communities to traverse different τ-charts without sacrificing invariants.
- Learning on the manifold. Natural-gradient adaptation over manifolds of renderers, with geodesic regularization and hysteresis, so systems retune rather than thrash.

A world of τ-literate systems would let listeners and makers co-author the alignment: q(t) becomes agile, τ(t) remains sovereign, and synesthesia—the sense of “this belongs with that”—is revealed as a stable bridge, not a mirage.

References

Foundations: Information, Compression, Geometry

1. Shannon, C. E. (1948). *A mathematical theory of communication*. Bell System Technical Journal.
2. Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*.
3. Chaitin, G. J. (1966). On the length of programs for computing finite binary sequences. *J. ACM*.
4. Rissanen, J. (1978). Modeling by shortest data description. *Automatica*.
5. Grünwald, P. D. (2007). *The Minimum Description Length Principle*. MIT Press.
6. Amari, S.-I. (2016). *Information Geometry and Its Applications*. Springer.
7. Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*
8. Amari, S.-I. (1998). Natural gradient works efficiently in learning. *Neural Computation*.

Category Theory, Topoi, and Functorial Thinking
9. Mac Lane, S. (1998). *Categories for the Working Mathematician* (2nd ed.). Springer.
10. Lawvere, F. W., & Rosebrugh, R. (2003). *Sets for Mathematics*. Cambridge University Press.
11. Johnstone, P. T. (2002). *Sketches of an Elephant: A Topos Theory Compendium*. Oxford University Press.
12. Spivak, D. I. (2014). *Category Theory for the Sciences*. MIT Press.

Topology & Multiscale Structure
13. Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press.
14. Edelsbrunner, H., & Harer, J. (2010). *Computational Topology: An Introduction*. AMS.
15. Ghrist, R. (2014). *Elementary Applied Topology*. Createspace.
16. Zomorodian, A. (2005). *Topology for Computing*. Cambridge University Press.
17. Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete & Computational Geometry*.

Vision Features & Invariants
18. Marr, D. (1982). *Vision*. W. H. Freeman.
19. Canny, J. (1986). A computational approach to edge detection. *IEEE TPAMI*.
20. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *IJCV*.

21. Ruderman, D. L. (1994). The statistics of natural images. *Network: Computation in Neural Systems*.

Audio, Timbre, and Sonification
22. Roads, C. (1996). *The Computer Music Tutorial*. MIT Press.
23. McAdams, S., & Giordano, B. L. (2016). The perception of musical timbre. In *Oxford Handbook of Music Psychology*.
24. Hermann, T., Hunt, A., & Neuhoff, J. G. (Eds.). (2011). *The Sonification Handbook*. Logos.
25. Julius O. Smith III (2010). *Spectral Audio Signal Processing*. W3K Publishing.

Synesthesia, Cross-modal Correspondence, & Aesthetics
26. Cytowic, R. E., & Eagleman, D. (2009). *Wednesday Is Indigo Blue: Discovering the Brain of Synesthesia*. MIT Press.
27. Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*.
28. Ward, J. (2013). *Synesthesia*. Annual Review of Psychology.
29. Marks, L. E. (1974). On associations of light and sound: The basis of synesthesia. *Perception & Psychophysics*.

Music Theory, Scales, and Rhythm
30. Messiaen, O. (1956). *Technique de mon langage musical*. Alphonse Leduc.
31. Tymoczko, D. (2011). *A Geometry of Music*. Oxford University Press.
32. Pressing, J. (1998). Error detection and improvisation. *Music Perception*.
33. London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter* (2nd ed.). Oxford University Press.

Entropy, Mutual Information, and Complexity Measures
34. Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley.
35. Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications*.
36. Hsu, K., & Kuo, C.-C. J. (2017). On information-theoretic measures for image quality. *APSIPA Transactions*.

Hysteresis, Control, and Robust Mapping
37. Altarovici, A., Bertsekas, D. P., & Hadjiconstantinou, E. (2013). Hysteresis and stability in control with quantization and delays. *Automatica*.
38. Julier, S., & Uhlmann, J. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE* (for smoothing/slew analogies).

Cosmology of Meaning: It-from-Bit / It-from-Qubit
39. Wheeler, J. A. (1990). *Information, Physics, Quantum: The Search for Links*. In Zurek (Ed.),

Complexity, Entropy, and the Physics of Information. Addison-Wesley.
40. Tegmark, M. (2014). *Our Mathematical Universe*. Knopf.
41. Preskill, J. (2012). *Quantum information and quantum computation* (lecture notes; “it from qubit” colloquial framing).

Methodological & Cross-cutting
42. Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.
43. Rahimi, A., & Recht, B. (2007). Random features for large-scale kernel machines. *NIPS* (for MDL proxies via simple predictors).
44. Mallat, S. (2016). Understanding deep convolutional networks. *Phil. Trans. Royal Society A* (multiscale invariants perspective).

Practical Tooling & Evaluation
45. Müller, M. (2015). *Fundamentals of Music Processing*. Springer.
46. McFee, B., et al. (2015). Librosa: Audio and music signal analysis in Python. *Proceedings of SciPy*.
47. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *JMLR*.

Notes:

- Items 9–12, 14–17 furnish the categorical/topological backbone for functorial renderers and persistence-driven phrase structure.
- Items 1–8, 34–36, 44 anchor the MDL and information-geometric criteria.
- Items 22–25, 30–33 connect auditory perception, timbre, rhythm, and musical geometry to our mapping policies.
- Items 26–29 address empirical synesthesia and cross-modal correspondences.
- Items 39–41 ground the “It-from-Bit/It-from-Qubit” cosmology that frames τ as symmetry/invariant budget.

Appendices

A. A formal count of mapping families under tighter assumptions

We define a renderer family by the tuple

$$\Theta = (T, G, A, C, S, F, M, H, I, E, R, Y, P)$$

with factors (Traversal, Grid, Aggregation, Channels, Source->Control Assignment, Transforms, Pitch system, Harmony grammar, Instrumentation, Effects, Rhythm style, Hysteresis, Panning law). To avoid double-counting equivalent choices, we quotient by symmetries that do not change the induced mapping on outputs.

Assumptions.

1. Traversal (T). Choose from a set of 10 canonical orderings; identify traversals that differ by reparametrization only (monotone time warp). Count: |T| = 10.
2. Grid (G). Choose meter m in {2/4, 3/4, 4/4, 5/4, 7/8} and subdivision d in {1/4, 1/8, 1/16, triplet, swing}, and note-bin count b in {32,64,128,256,512}. Identify grids that are integer refinements (e.g., 64 at 1/8 vs 128 at 1/16) if the induced onset set is equal up to constant time scaling; keep the coarsest representative. Effective |G| ≈ 80 (as before).
3. Aggregation (A). Non-empty subsets of 6 statistics modulo redundancies: if {mean,z-score} and {mean} produce identical rank order after monotone transfer, identify them. Conservatively keep full 2^6−1 = 63.
4. Channels (C). Choose non-empty subset of 12 features {H,S,V, L,a,b, grad-mag, grad-angle, edges, corners, wavelet-low, wavelet-mid, wavelet-high}. Identify channel sets equivalent under invertible linear transforms within a feature family (e.g., {H,S,V} ~ {L,a,b} if followed by learned affine). To remain conservative, keep 2^12−1 = 4,095.
5. Source->Control assignment (S). Controls = {pitch, velocity, duration, pan, register, onset-jitter}. Each control selects one active channel. Count S = k^6 where k is the number of chosen channels. With expected k=6 (typical), S = 6^6 = 46,656.
6. Transforms (F). Per control choose from {linear, log, exp, sigmoid, tanh, rank/quantile, z+clip, quantize-to-scale}. Identify transforms that differ by an overall affine remap (absorbed by gain staging). Keep 8^6.
7. Pitch system (M). 12 transpositions × 50 scales/modes; identify transpositions modulo absolute key if only intervals matter. We retain |M| ≈ 600 (since absolute key affects orchestration ranges).

8. Harmony grammar (H). 20 progressors (diatonic, modal, quartal, etc.).
9. Instrumentation (I). 4 tracks chosen with repetition from 128 GM classes; identify permutations that only reorder tracks with no mix differences. Since track identity affects mixing/effects, keep 128^4.
10. Effects (E). Up to 5 slots, each from 9 effect types + none, with ~20 coarse presets. Identify commutative reorders of strictly linear time-invariant effects in parallel; still, serial order usually matters. Conservative bound: 8e8.
11. Rhythm style (R). 25 coarse patterns (density, swing, humanization, micrometric bias).
12. Hysteresis (Y). Deadbands x slew x hold times; keep 10^3 distinct combos.
13. Panning law (P). {equal-power, linear, constant-vol}.

Product (lower bound).
N = 10 * 80 * 63 * 4095 * 46656 * 262144 * 600 * 20 * 2.684e8 * 8e8 * 25 * 1e3 * 3 ≈ 4.8e44.

Tightening with quotienting (illustrative).
If we quotient out absolute key (divide M by 12) and count only 10 strict effects “skeletons” per chain (rather than presets), we get a trimmed
N' ≈ N / 12 * (10 / 8e8) * (2e3) ≈ still >> 1e38.
Conclusion: even with strong quotienting, the family count remains astronomically large.

B. Example renderer specs (spiral ordering, hue->pitch, hysteresis)

Renderer name: Spiral-HSV-Equivariant v1

Invariants honored (commutations):

- Hue rotation in image -> pitch-class transposition in audio (S^1).
- Monotone value change -> monotone loudness remap (order).
- Gaussian blur (scale up) -> spectral darken (timbre tilt).
- Mirror (left/right) -> stereo mirror; rotate -> pan rotation (orientation).

Traversal: Center-out spiral; 1 visit per pixel; split into 1/16-note bins at 100 BPM.

Channels: H,S,V; grad-angle; wavelet-high.

Assignments:

- pitch <- H (circular variable)

- velocity <- S (scalar)
- duration <- V (scalar)
- pan <- grad-angle (circular)
- sparkle density <- wavelet-high RMS

Transforms:

- pitch: wrap H in [0,1) -> pitch class; quantize to chosen mode; transpose = hue rotation.
- loudness: rank-based mapping of V to [−12 dB, 0 dB] then compressor at 3:1 (to preserve order).
- duration: V -> note length in {1/8, 1/4, dotted 1/8} via 2 thresholds with hysteresis.
- pan: equal-power pan p = sin(theta).

Harmony grammar: Diatonic with modal interchange (borrow bVII, iv) gated by β_1 mass (topology).

Instrumentation: pad, arp, lead, perc (GM approximations).

Hysteresis & memory:

- Deadbands: Δ_{pitch} = 1 semitone; Δ_{vel} = 0.05 (S units); Δ_{pan} = 0.08 (unit circle chord).
- Slew: pitch <= 2 st per beat; pan <= 0.2 per 1/16; loudness <= 6 dB per beat.
- Sticky quantization: hold time H = 2 onsets before key/mode change.
- Section thresholds:
  Enter chorus if low-band RMS > τ_{on} for >= 2 bars; *exit* when < τ_{off} for >= 1 bar ($\tau_{on} > \tau_{off}$).

Effects skeleton: gentle LPF on pad, short plate reverb on arp/lead, tempo-synced delay on lead at 1/8 dotted; compressor bus with -3 dB GR.

Declared breaks: none (v1 aims for strict equivariance).

C. Metrics: compression, roughness, tempo stability, MI across voices

C.1 Joint compression (MDL proxy).
Train a predictor \hat{S} = G(feat(X)) for the produced audio spectrogram S. Use an entropy model H(·) to score bits:

- MDL_theta = L(model) + H(S | feat(X); theta).
  Use a fixed lightweight G (e.g., linear + ReLU) to avoid overfitting; lower MDL is better.

C.2 Commutation score.
Apply test transforms on the image:

- hue-rotate(Δ), blur(σ), mirror(LR), rotate(ϕ).
  Measure distances between audio pairs and the expected audio transforms (transpose, spectral tilt, stereo flip, pan rotate). Score = fraction within tolerance (e.g., pitch alignment <= 15 cents, spectral centroid delta consistent within 5%, pan correlation r >= 0.9).

C.3 Spectral roughness / clarity.

- Roughness: Vassilakis-style or Sethares measure on frames; report median and IQR.
- Clarity: spectral flatness inverse and spectral centroid stability; penalize chronic harshness unless driven by high entropy in image.

C.4 Tempo stability.
Beat-track the audio; compute instantaneous BPM(t); report stddev over the piece. Lower is better unless traversal dictates rubato; compare to intended grid.

C.5 Cross-voice mutual information.
Compute discrete event streams per voice (onset bins, pitch-class bins). Estimate MI(voice_i, voice_j) using k-NN or histogram estimators with bias correction. Desire:

- High MI for intended call-and-response pairs (e.g., left/right regions).
- Low MI for independent textures meant to be decorrelated.

C.6 Robustness margin.
Find the largest perturbations (noise std, blur sigma, exposure shift) such that section boundaries and cadences remain unchanged (within 1 beat). Report as ρ_noise, ρ_blur, ρ_exposure.

C.7 Listener lock-in.

- T_lock: exposures to reach 80% correct cadence prediction.

- R_persist: motif recall rate after 10–30 minutes.
- C_pass: human-rated consistency under invariance tests.

D. Pseudocode for a τ-guided renderer with listener feedback

```
# Inputs:
#   X: image
#   priors: {key_family, meter, motif_seed}  # weak, equivariant
#   params θ: renderer dials (T, G, A, C, S, F, H, I, E, Y, P)
#   invariants ℐ: {S1_equiv, order_mono, scale_tilt, topo_persist}
#   feedback stream ℒ from listener (optional): predictions, ratings

function RENDER_TAU_GUIDED(X, priors, θ, ℐ):
  # 1) Precompute feature stacks
  feats = EXTRACT_FEATURES(X)
    # H,S,V; grad-mag, grad-angle; wavelet pyramid; entropy; topology (β0, β1 curves)

  # 2) Choose traversal and time grid
  order = TRAVERSAL(θ.T, X.shape)        # e.g., spiral indices
  grid  = TIME_GRID(θ.G)                 # meter, subdivision, bins

  # 3) Aggregate features per time bin
  bins = SPLIT(order, grid.num_bins)
  stats = []
  for b in bins:
    stats.append(AGGREGATE(feats, b, θ.A))   # mean/var/entropy/oriented energy
```

```
# 4) Map channels -> controls via assignments and transforms
controls = {}
controls.pitch     = MAP_CIRCULAR(stats.H, θ.F.pitch, priors.key_family)   # hue->pitch class
controls.velocity  = MAP_MONOTONE(stats.S, θ.F.velocity)
controls.duration  = MAP_MONOTONE(stats.V, θ.F.duration)
controls.pan       = MAP_CIRCULAR(stats.grad_angle, θ.F.pan)
controls.sparkle   = MAP_DENSITY(stats.wavelet_high)

# 5) Enforce invariants (commutation guards)
controls = ENFORCE_EQUIVARIANCE(controls, 𝓘)   # S1 rotation, order-preservation, scale->tilt

# 6) Hysteresis / memory smoothing
controls = HYSTERESIS(controls, θ.Y)         # deadbands, slew, hold

# 7) Section engine from slow stats
sections = SECTION_SCHEDULE(stats.low_band_RMS, feats.topology, θ.section_thresholds)
controls = APPLY_SECTIONS(controls, sections, priors, θ.H, θ.I)  # key/mode swaps, orchestration

# 8) Synthesis and effects
audio = SYNTHESIZE(controls, θ.I)            # pad/arp/lead/perc
audio = APPLY_EFFECTS(audio, θ.E, θ.P)

# 9) Evaluation metrics (MDL, commute, robustness)
metrics = EVAL_METRICS(X, audio, 𝓘)
```

```
    return audio, controls, metrics
```

Listener-in-the-loop update (natural-gradient step).

```
function UPDATE_WITH_FEEDBACK(θ, metrics, ℒ):
    # Build divergence objective
    D = α * metrics.MDL + β * (1 - metrics.commute_score) + γ * metrics.instability

    # Incorporate listener lock-in (aim to decrease lock-in time and increase invariance preference)
    if ℒ available:
        D = D - δ1 * ℒ.lock_in_gain - δ2 * ℒ.invariance_preference

    # Natural-gradient parameter update on manifold of renderers
    g = INFORMATION_METRIC(θ)
    grad = GRADIENT(D, θ)
    Δθ = -η * INVERSE(g) * grad
    θ_new = PROJECT_TO_ADMISSIBLE(θ + Δθ)  # keep functorial constraints

    return θ_new
```

Notes.

- ENFORCE_EQUIVARIANCE implements the commuting diagrams: e.g., calibrate hue->pitch so a hue-rotation test exactly transposes pitch classes; calibrate blur->EQ tilt.
- HYSTERESIS applies deadbands/slew/hold in percentile space to preserve order under exposure changes.
- SECTION_SCHEDULE uses low-band RMS and β_1 mass with enter/exit thresholds ($\tau_{on} > \tau_{off}$) and minimum dwell.
- EVAL_METRICS computes MDL proxy via a fixed predictor, commutation tests, roughness, tempo stability, MI across voices, and (optionally) listener metrics.

Deployment loop.

1. Initialize θ from a minimal, strict-equivariance preset.
2. Render, evaluate metrics; collect listener feedback.
3. Update θ via UPDATE_WITH_FEEDBACK.
4. Repeat until MDL stops decreasing and commute score/lock-in plateau.
5. Ship with a recipe card (Sec. 11.3): invariants honored, priors used, memory settings, A/B results, declared breaks.

This turns the renderer into a τ-literate instrument: compositional, testable, corrigible.