

Reference-dependent C³ responses to birdsong-to-music transformations

Amaç Erdem

Independent researcher

Abstract

Spectral and temporal transformations of natural sounds offer a controlled test of the acoustic distinctions encoded by auditory cognition models. Here, we characterized the responses of C³, the cognition layer of a frozen version of Musical Intelligence (MI), to transformations of a single 2.45-s blackbird phrase. Five paths connected the recording, a synthetic transcription and sample-derived flute and piano endpoints. We analysed 1,146 complete-phrase renderings, including level controls, across 84 mechanisms and 202 present-state dimensions. With whole-phrase audio levels matched, the root-mean-square (RMS) model displacement was larger for piano than flute from the transcription (0.136 versus 0.098), but larger for flute than piano from the bird recording (0.065 versus 0.052), despite identical instrument endpoints. Spectral replacement elicited greater displacement than isolated pitch or onset transformations, and their joint effects were non-additive. Refinement to 97 anchors stabilized mechanism rankings, while interpolation errors remained 4.4–19.6% of path RMS. Inspection of the implemented formulas linked several prominent responses to direct sensory terms. These findings demonstrate reference-dependent model sensitivity and identify contrasts for perceptual experiments. Whether these distinctions predict listener judgments remains to be tested across additional natural sounds.

Introduction

Naturalistic sounds contain frequency trajectories, amplitude fluctuations and event relationships that conventional pitch-and-onset descriptions only partly retain. Perceptual research shows that spectrotemporal representations can support useful models of timbre, but their dimensions require behavioural calibration: learning a distance from listeners' judgments is different from assigning perceptual meaning to an uncalibrated numerical response.¹ Expectation can also change auditory acuity under controlled task conditions.² A model that responds to acoustic change therefore addresses an initial question—what its representation distinguishes—before it can address what a listener distinguishes.

This distinction is consequential. Artificial auditory representations can preserve distinctions that are unimportant to people, or discard distinctions that people readily perceive. Model-metamer experiments provide a direct demonstration that apparently successful systems can nevertheless diverge from human perceptual invariances.³ A meaningful test of a cognition architecture must consequently expose both its sensitivity and the assumptions that turn sensitivity into an interpretation. Large output changes, elaborate mechanism names and dense visualizations are insufficient on their own.

The present experiment grew from sustained compositional study of Olivier Messiaen and the problem of translating birdsong into music. A sonographic examination of Messiaen's Australian birdsong transcriptions documents selective correspondences between recordings and notation, alongside compositional transformations.⁴ This motivates two separable questions: which acoustic details are discarded or reorganized by a musical

abstraction, and whether musical structure can instead be induced from regularities in natural sounds. The rendering operations here implement the first question. They impose selected pitches and event positions; they do not learn musical rules from a birdsong corpus.

Related experimental approaches delimit what this pilot can establish. Manipulating speech modulations has identified information important for intelligibility,⁵ and sound-texture synthesis has tested which auditory statistics preserve recognizable texture.⁶ Task-optimized auditory models further show that the usefulness of precise temporal information depends on the task.⁷ These findings motivate controlled transformations and explicit behavioural targets, rather than a universal spectral or temporal resolution at which sound becomes perceptually different.

We examine C³, MI's cognition-labelled computational layer, while retaining its frozen acoustic and temporal front end. Every anchor is a separately rendered, complete phrase processed from a fresh model state. The primary outcome is the root mean square (RMS) displacement of present-state outputs from the first anchor of each path. We ask whether the leading responses depend on transformation, instrument endpoint, level control and sampling density; whether direct bird-to-instrument paths agree with paths starting from a synthetic transcription; and how the strongest changes relate to the actual implemented formulas.

Across all 84 mechanisms, the central result is a reversal in relative flute–piano sensitivity when the reference changes. It provides a candidate contrast for psychophysics and shows why one ranking cannot summarize the model's response to birdsong becoming music.

Results | A controlled space of complete sound transformations

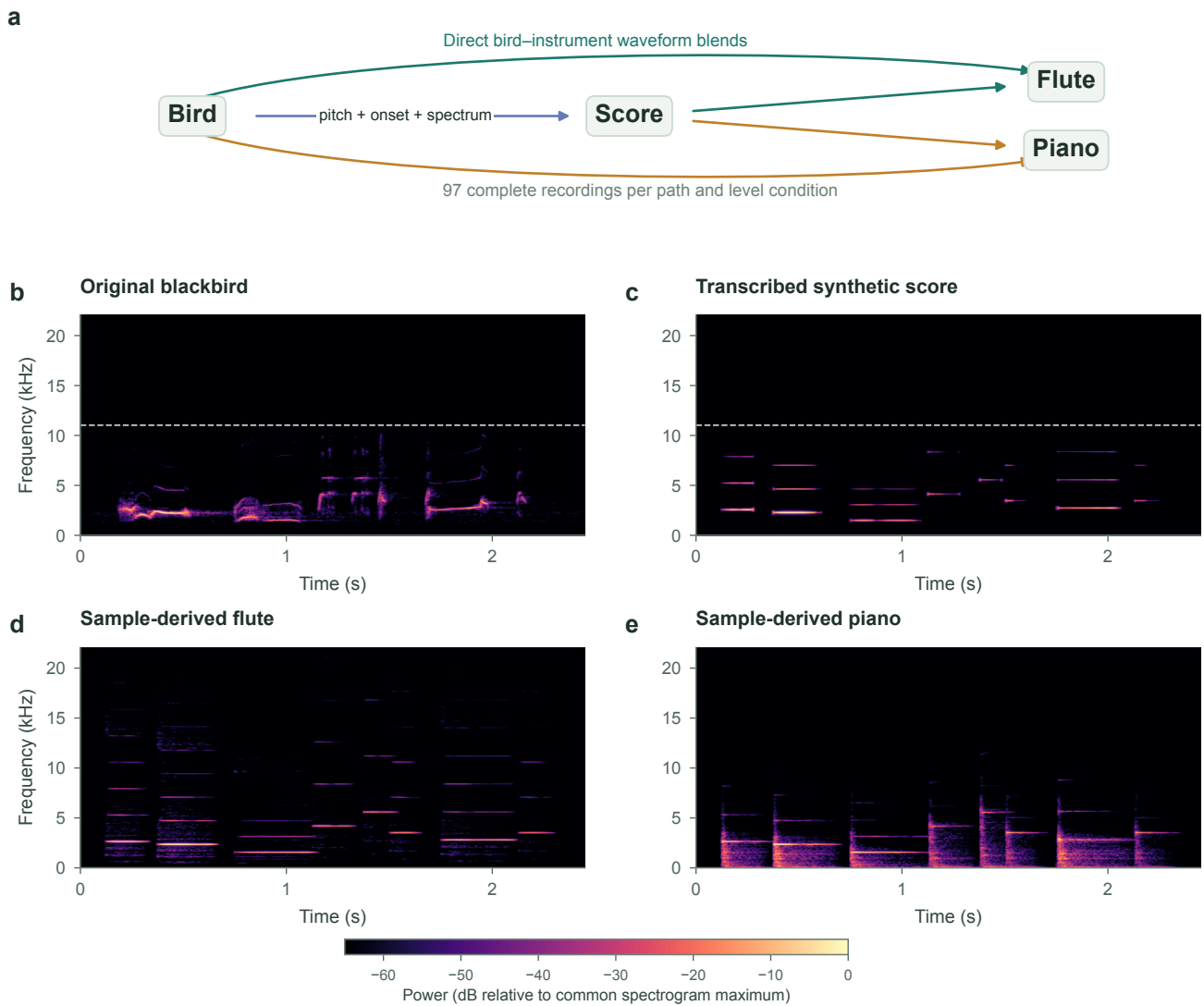


Fig. 1 | Five paths connect a natural phrase, a transcription and two instrument endpoints. **a**, B→S jointly changes pitch, onset placement and spectral content; S→F and S→P mix the synthetic transcription with sample-derived flute and piano. B→F and B→P directly blend the original bird waveform with the fixed instrument endpoints. Every primary path has 97 anchors in each level condition, with identical waveforms shared across paths. **b–e**, Whole-phrase, common-RMS spectrograms; all panels use the same power reference and colour scale. Dashed lines in **b** and **c** mark the original source’s 11.025-kHz Nyquist limit. Instrument resynthesis introduces energy above that source band. These are high-register sample-derived sounds, not unmodified acoustic performances. B, bird; S, score; F, flute; P, piano.

A 2.44998-s excerpt of a common blackbird (*Turdus merula*) recording was annotated as eight events. Pitch trajectories were progressively mapped towards eight equal-tempered target notes; event starts moved towards selected onset positions; and spectral replacement exchanged event waveforms for harmonic tones while attenuating the background. The resulting endpoint is termed the *synthetic score*. It is an operational transcription of this excerpt, not a reconstruction of a Messiaen score (Fig. 1).

Flute and piano versions used the same target sequence and register, MIDI 91–113 (G6–F8), without octave folding. Their attacks, release envelopes and spectra differ. The highest notes exceed ordinary instrumental ranges and require pitch-shifted samples. Consequently, “instrument” denotes the source of the resynthesis material; it does not isolate instrument identity from register, envelope or spectral bandwidth.

Five paths were sampled at 97 anchors. Additional 25-anchor pitch, timing and spectrum paths and a $3 \times 3 \times 3$ factorial grid separated operator effects. Deduplication yielded 569 primary waveforms, 569 common-RMS controls and eight ± 6 -dB endpoint controls: 1,146 recordings in total. Twenty-four diagnostic endpoint perturbations were analysed separately. Direct bird–instrument paths contributed 380 new recordings beyond the original 766-recording assay. Anchors are dependent transformations of one source phrase, not independent biological samples.

The frozen adapter produced 423 frames per recording, 84 distinct mechanisms and 889 declared dimensions, including 202 present-state dimensions. We report deterministic magnitudes and numerical resolution without population-level inference.

Results | The reference changes which instrument produces more response

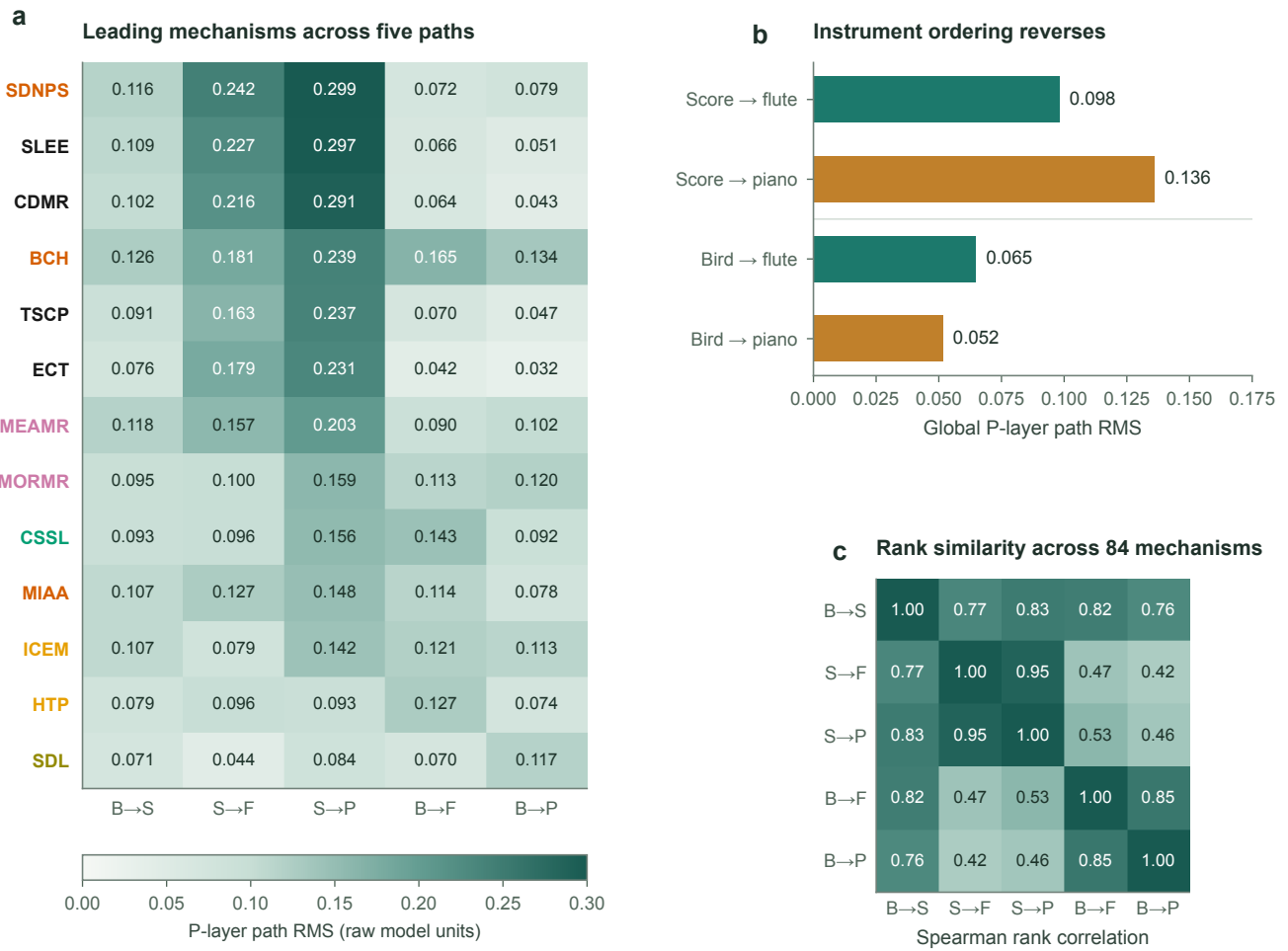


Fig. 2 | Reference-dependent response magnitudes and mechanism rankings. **a**, Path RMS for the union of the five highest-ranked mechanisms on each primary common-RMS path (13 mechanisms); the full 84-mechanism table accompanies the manuscript. Values average squared changes across all anchors, time frames and each mechanism's present-state dimensions before taking the square root. Label colours identify the architecture's F1–F8 function membership, not an additional measured variable. **b**, Global RMS pools all 202 present-state dimensions. Piano exceeds flute from the score, whereas flute exceeds piano from the bird. **c**, Descriptive Spearman correlations of all 84 mechanism rankings. These correlations quantify agreement among this model's response profiles; they are not behavioural prediction scores. All values use common-RMS audio. No error bars or significance tests are assigned to dependent anchors.

Global present-state path RMS was 0.09816 for score→flute and 0.13615 for score→piano. With the bird as origin, the ordering reversed: 0.06488 for bird→flute and 0.05172 for bird→piano (Fig. 2b). Endpoint RMS showed the same ordering within each reference: 0.12373 versus 0.17155 from the score, and 0.13306 versus 0.09191 from the bird. Thus the reversal is not solely a consequence of averaging intermediate anchors.

Each instrument endpoint is the same waveform reused across its two paths. What changes is the origin and the trajectory through stimulus space. The result is reference-dependent geometry and path-specific stimulation, not hysteresis. No previous anchor is retained by the model, and no exposure history is manipulated.

From the score, the largest responses were Stimulus-Dependent Neural Pitch Salience (SDNPS), Synaptic Long-term Encoding

Engine (SLEE) and Context-Dependent Mismatch Response (CDMR). Their flute path RMS values were 0.24177, 0.22657 and 0.21589; corresponding piano values were 0.29945, 0.29683 and 0.29053. Brainstem Consonance Hierarchy (BCH) instead led both direct bird–instrument paths (0.16521 and 0.13450). Names identify frozen code modules, not established physiological processes.

Score-origin flute and piano rankings were highly similar (Spearman $\rho=0.950$), as were their direct-origin counterparts ($\rho=0.855$). Agreement was lower when an instrument was compared across origins: $\rho=0.471$ for flute and 0.455 for piano. A mechanism selected solely from the score-origin ranking would therefore provide an incomplete account of direct birdsong transformation. These comparisons concern the implemented units; cross-mechanism numerical scales have not been perceptually calibrated.

Results | Direct bird–instrument comparisons reveal distinct response profiles

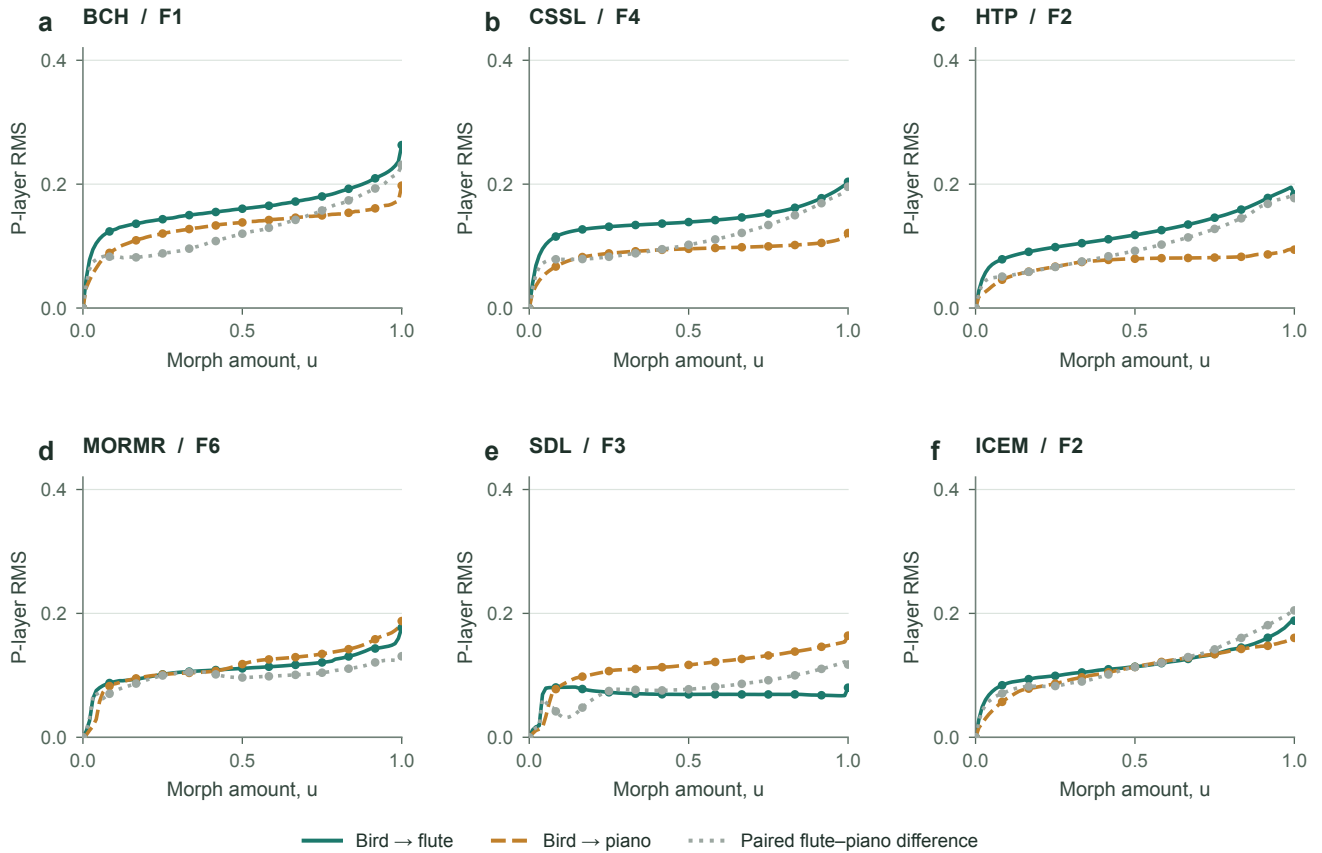


Fig. 3 | Direct flute and piano trajectories cannot be reduced to one scalar sensitivity. **a–f**, Six mechanisms selected by an explicit rule: the union of the three highest-ranked mechanisms on each direct path, plus ICEM, the second-ranked paired flute–piano contrast. BCH, Brainstem Consonance Hierarchy; CSSL, Cross-Species Song Learning; HTP, Hierarchical Temporal Prediction; MORMR, mu-Opioid Receptor Music Reward; SDL, Salience-Dependent Lateralization; ICEM, Information Content Emotion Model. Teal and gold show RMS from the shared bird origin; grey shows the RMS difference between flute and piano tensors at the same morph amount. Grey is not the difference between the other two curves. All 97 computed anchors contribute to each line; every eighth anchor is marked. Lines are visual guides, with a common y scale. F labels indicate architecture membership.

Direct bird→flute emphasized BCH, Cross-Species Song Learning (CSSL; 0.14323) and Hierarchical Temporal Prediction (HTP; 0.12673). Direct bird→piano emphasized BCH, mu-Opioid Receptor Music Reward (MORMR; 0.12015) and Salience-Dependent Lateralization (SDL; 0.11733). Information Content Emotion Model (ICEM) ranked fourth on both direct paths. These differences survive a common whole-phrase RMS, but this control does not match instantaneous envelope, spectral balance or perceived loudness.

The trajectory shapes further separate the endpoints (Fig. 3). CSSL and HTP show larger displacement towards flute across much of the path. SDL is more displaced towards piano after an initial transition, whereas MORMR responses remain comparatively close. ICEM’s two distances from the bird remain similar over much of the interior despite a non-zero paired contrast. Equal distances from one origin do not imply equal model representations.

Computing the difference between matched flute and piano output tensors gives a global path RMS of 0.06079 and an endpoint RMS of 0.11463. The strongest paired changes include BCH (0.13314), ICEM (0.12301), Musical Imagery Auditory Activation (0.11852) and CSSL (0.11618). This paired analysis is needed because subtracting two non-negative distances discards direction within the response space.

A rising RMS curve indicates increasing displacement, not increasing activation, reward or salience. Signed means and individual dimensions are retained in the companion tables so that decreases and opposing dimension changes remain inspectable. None of these curves localizes a human category boundary. The sharpest numerical changes near endpoints are candidates for renderer and front-end scrutiny before they are candidates for behavioural threshold testing.

Results | Spectral replacement dominates the tested operators

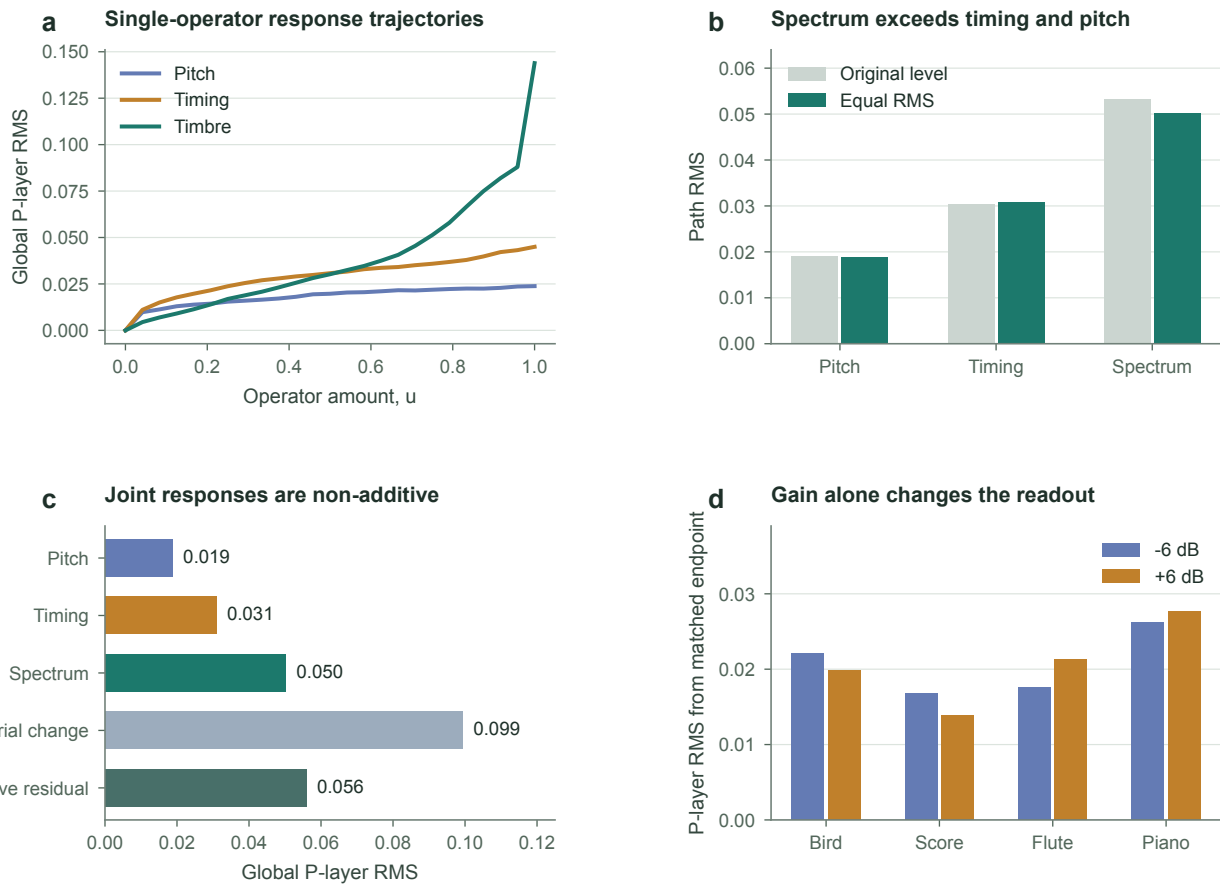


Fig. 4 | Operator and level controls constrain cognitive interpretations. **a**, Twenty-five-anchor, common-RMS paths varying only pitch mapping, onset displacement or spectral replacement. **b**, Path RMS before and after whole-phrase RMS matching. **c**, Single-operator responses compared with the 27-setting factorial change and its residual from the origin-based additive prediction. The residual-to-total RMS ratio is a descriptive non-additivity measure, not a fraction of explained variance. **d**, Present-state RMS changes caused by ± 6 -dB gain around each matched endpoint without clipping. These controls demonstrate residual level sensitivity despite per-clip normalization in the mel front end. All quantities are pooled across the same 202 present-state dimensions. Operator coordinates are implementation parameters and are not equal-sized perceptual or acoustic perturbations.

Common-RMS path RMS was 0.01875 for pitch, 0.03090 for timing and 0.05021 for spectral replacement (Fig. 4a,b). Spectral replacement was therefore 2.68 times the pitch response and 1.63 times the timing response under these particular operator definitions. This ordering does not establish that spectral information is universally more important to cognition: the operations have different physical magnitudes, and spectral replacement also removes background energy and changes within-event structure.

Leading modules differed by operator. MORMR, BCH and ICEM led the pitch path; ICEM, SDL and Melodic Processing Gradient led the timing path; Music-Evoked Autobiographical Memory Reward, SDNPS and SLEE led spectral replacement. These labels describe numerical locations of change. No autobiographical material, reward report or listener memory was supplied to the model.

Combined responses were not reproduced by adding the three isolated effects. Across the 27 factorial settings, global RMS change was 0.09933 and the residual from the additive prediction was 0.05590, a ratio of 0.563 (Fig. 4c). This non-additivity can arise from the renderer, nonlinear acoustic feature extraction, clipping and downstream combinations. The present assay does not uniquely assign it to an interaction among cognitive mechanisms.

Gain changes alone produced global RMS differences of 0.01382–0.02762, overlapping the magnitude of the isolated pitch manipulation. Common-RMS matching nevertheless preserved the broad ranking structure: raw-to-matched rank correlations ranged from 0.915 to 0.985 across the eight analysed paths. For the original joint, score→flute and score→piano paths, C^3 distance covaried with log-mel distance (descriptive correlations 0.785, 0.984 and 0.926). Co-movement along a shared morph coordinate is not evidence that C^3 predicts behaviour beyond acoustic baselines.

Results | Stable rankings coexist with unresolved local sensitivity

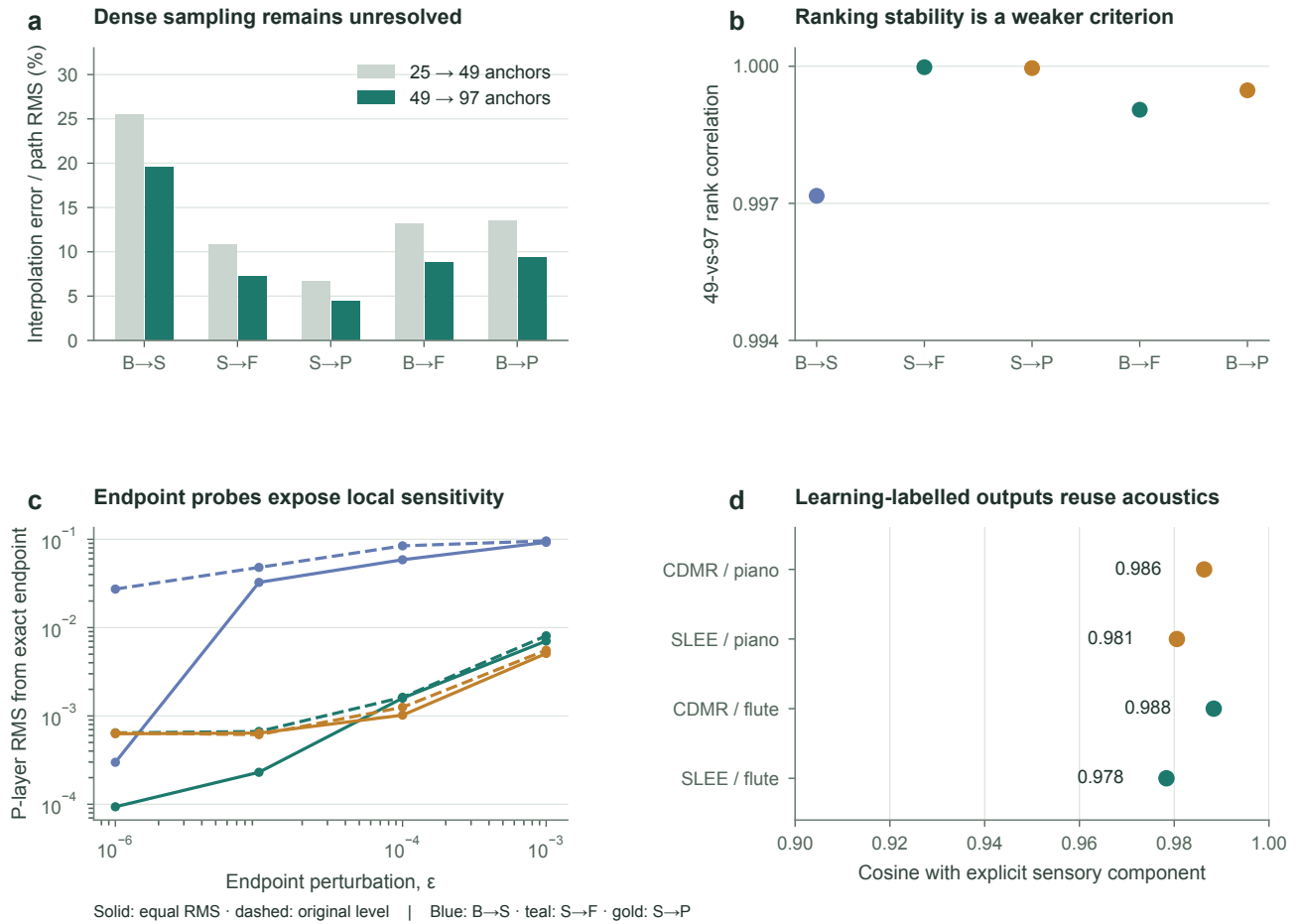


Fig. 5 | Sampling resolution and formula grounding are separate requirements. **a**, RMS error when omitted midpoint tensors are predicted by linear interpolation of neighbouring anchors, divided by full-path RMS. **b**, Rank correlation between 49- and 97-anchor mechanism sensitivities. **c**, Diagnostic perturbations $\epsilon=10^{-6}$ – 10^{-3} from the score endpoint on B→S and the score-origin endpoints on S→F and S→P; log-arithmetic axes. All 24 probes are shown. **d**, Cosine alignment of SLEE/CDMR output changes with their explicitly coded sensory component along score–instrument paths. Alignment is neither a causal neural attribution nor an explained-variance estimate. Sensory information may also enter the residual through upstream terms. All 84 mechanisms enter panel b; the four comparisons in d were selected after the original rankings and source audit.

Refining the primary grid from 49 to 97 anchors barely changed mechanism rankings: $\rho=0.99717$ – 0.99998 across the five paths. Yet 49→97 midpoint interpolation errors were 19.56% of path RMS for bird→score, 7.26% for score→flute, 4.39% for score→piano, 8.81% for bird→flute and 9.36% for bird→piano. Rank convergence therefore supports an ordering of responses within this assay, not a claim that continuous trajectories are resolved (Fig. 5).

The largest adjacent-anchor changes occurred in the final interval for bird-origin paths and the first interval for score–instrument paths. At $\epsilon=10^{-6}$ from the score endpoint, the joint path had an original-level relative waveform RMS difference of 0.000656 but a C^3 difference of 0.02730. After level matching, corresponding values were 0.000617 and 0.000298. This sensitivity to level treatment and tiny endpoint perturbations warrants further numerical localization; it does not establish a mathematical discontinuity or a perceptual threshold.

Source inspection explains why prominent outputs require restrained naming. SDNPS explicitly combines inverse roughness, dissonance, inharmonicity and tristimulus balance. Roughness-based consonance models have a substantial psychoacoustic foundation,^{8,9} but their presence inside a module does not independently validate that module’s neural label. SLEE and CDMR include large direct pleasantness and inverse-roughness terms. Along score–instrument paths, their sensory-component alignment was 0.978–0.988; residual-to-total RMS ratios were 0.302–0.355.

The audit also found substantive discrepancies between comments and executable feature indices: SLEE selects distribution flatness where a comment names pitch change, and CDMR selects A-flat chroma where a comment describes pattern coupling. These formulas were frozen, not repaired during analysis. A revised implementation requires a separately versioned replication. High sensitivity of a learning-labelled output in this experiment cannot be read as evidence that learning occurred.

Results | Regional and neurochemical outputs are derived projections

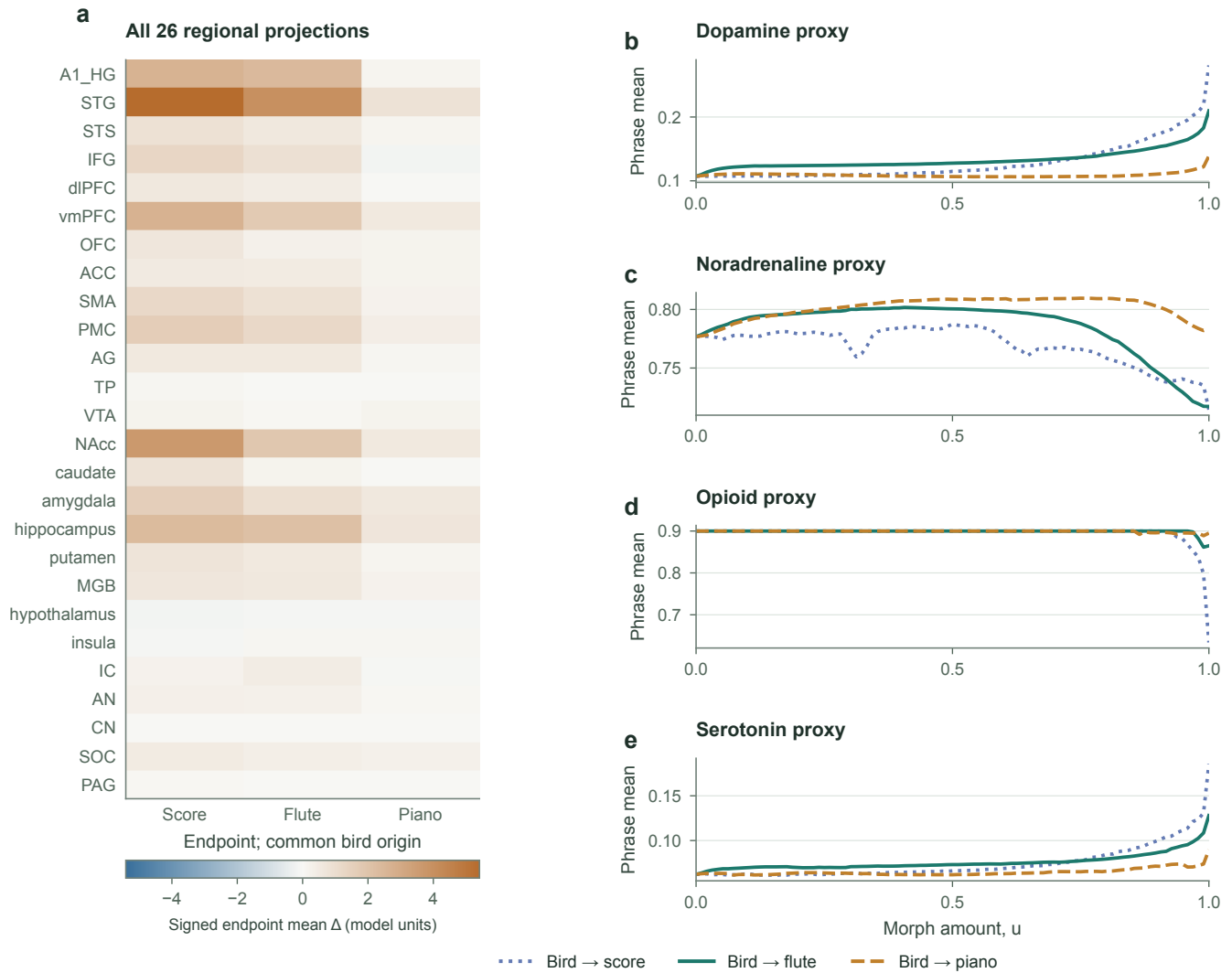


Fig. 6 | The full regional projection and four neuromodulator-labelled readouts. **a**, Signed endpoint-minus-bird differences in phrase-mean values for all 26 regional projections. Each column uses the same original bird reference and common-RMS audio. Region identifiers follow the frozen architecture; the colour scale is shared. Projection magnitudes depend on connection counts and weights and cannot be compared as measured regional activation strengths. **b–e**, Phrase means of four bounded model proxies along bird $\rightarrow$ score, bird $\rightarrow$ flute and bird $\rightarrow$ piano paths. All channels are bounded between zero and one, but panels use different local y scales to show their trajectories. Regional and neuromodulator outputs are deterministic transformations of C^3 outputs. They are not brain recordings, transmitter concentrations, or independent corroborating measurements.

The regional activity mapping (RAM) applies the frozen dimension-to-region weights and rectifies the resulting sums. From the common bird origin, the superior temporal gyrus-labelled projection increased by 5.411 units at the score endpoint, 3.985 at flute and 0.809 at piano (Fig. 6a). The primary auditory/Heschl’s gyrus-labelled projection changed by 2.63979 for the score and 2.37742 for flute. These magnitudes describe the architecture’s aggregation rules; unequal connectivity alone can produce unequal regional scales.

The four neuromodulator-labelled readouts likewise distinguish endpoints. The dopamine proxy’s phrase mean changed from 0.107 at the bird origin to 0.281 for the score, 0.210 for flute and 0.139 for piano. The noradrenaline proxy moved from 0.777 to

0.714, 0.717 and 0.784, respectively. The opioid proxy moved from 0.900 to 0.634, 0.865 and 0.895; the serotonin proxy moved from 0.062 to 0.186, 0.128 and 0.090. Small changes near a bound deserve different interpretation from unconstrained response growth.

The original analysis verified exact agreement of reconstructed RAM and neuromodulator outputs against four fresh full-engine evaluations. This establishes computational fidelity of the readout reconstruction for those checks. It does not validate a correspondence with living tissue. The 131 unstretched weighted aggregates retained with the experiment are additional views of the same model pool, not 131 independent tests of a cognitive hypothesis.

Discussion | From model sensitivity to a testable perceptual account

A computational response is a candidate predictor

This study answers a restricted but necessary question: which parts of one frozen MI implementation change when a particular naturalistic phrase is transformed towards specified musical endpoints? The answer depends on the reference, the rendering operation and the numerical resolution. The clearest result is the reversed flute–piano ordering across bird and score origins, with identical instrument endpoints. It identifies a stimulus contrast on which competing perceptual accounts can disagree. It does not demonstrate that listeners share the same ordering, nor that a larger model displacement means a more salient musical transition.

Reference dependence is expected for distances in a structured representation, but its magnitude and mechanism-level consequences must be measured rather than assumed. Here, rank agreement fell from 0.950 between score-origin instruments to approximately 0.46 when the same instrument was compared across origins. Consequently, mechanism selection should follow the scientific contrast. A module useful for separating a synthetic score from piano need not be useful for explaining how natural birdsong becomes piano-like.

Abstraction and induction require different experiments

The present transcription is an abstraction in an operational sense: selected event boundaries and pitches are retained or regularized while much of the source waveform is replaced. The target sequence is provided by the experimenter. Induction, understood as deriving musical structures from regularities in natural sounds, would require learning or estimating those regularities across independent examples and testing transfer to new sounds. This experiment cannot decide whether music in general is an abstraction of nature or an inductive product of exposure. It supplies a controllable bridge on which those alternatives could be tested.

Three routes to empirical value

1. Predict feature-specific discrimination. Across multiple birds, phrases and registers, measure discrimination or dissimilarity for anchor pairs. Select contrasts with comparable acoustic distance but divergent C³ predictions, including contrasts near the present endpoint sensitivities only after numerical refinement. Compare a preregistered, small mechanism subset with log-mel, envelope and spectrotemporal-modulation baselines. Auditory Feature-based Perceptual Distance (APD), a behaviour-linked birdsong metric described in a preprint, is a relevant additional comparator when its implementation and

evaluation conditions can be reproduced.¹⁰ Added value should mean lower held-out predictive error, not higher within-path correlation.

2. Separate acoustic sensitivity from expectation. Sainburg and colleagues' 2025 work makes cue-dependent acuity a particularly relevant test.² Present the same waveform under experimentally varied predictive cues, with stimulus identities and behavioural trials joined to MI outputs. The present audio-only, reset-per-clip assay produces the same representation for that identical input; it cannot by itself explain a cue-induced change. A contextual extension should therefore be compared with the frozen audio baseline on held-out cue–stimulus combinations, separating discrimination sensitivity from response bias. A successful extension must predict the interaction, not merely rediscover the acoustic gradient.

3. Test representation against population recordings. A 2026 preprint links anticipatory organization of auditory population dynamics to behavioural decisions.¹¹ This motivates comparing stimulus-evoked geometry with MI features while modelling anticipatory state separately. Representational similarity analysis offers a principled comparison without assuming a one-to-one mapping from named model units to neurons.¹² Hold out entire stimulus families and animals or participants, fit mappings only within training folds, and benchmark against simpler acoustic models. Claims about online decision timing additionally require a causal implementation; the current centered and forward-looking context cannot establish them.

Limits that determine the next experiment

There is one phrase, one transcription and one rendering recipe per instrument. Spectral bandwidth, envelope, background removal and extreme register remain entangled. Raw output scales favour some dimensions; many declared temporal horizons exceed the available phrase. The 97-anchor grid resolves rankings better than local response curves. These limitations call for independent phrases, physiologically plausible register controls, matched envelopes and bandwidth, and adaptive sampling based on waveform as well as output changes. Sampling should stop according to a predefined numerical tolerance before human thresholds are interpreted.

The contribution is therefore an audited assay, a set of quantified model contrasts and a falsifiable validation programme. The decisive next result would be an improvement in predicting unseen perceptual or neural observations beyond acoustic baselines. Until that comparison exists, MI's cognitive, regional and neurochemical labels remain model hypotheses whose empirical meaning must be earned.

Methods

Design and provenance

The local analysis specification was written after a runtime-only pilot and before inspecting pathway rankings; it was not publicly preregistered. Initial 49-anchor paths were refined to 97 following midpoint-error checks. Direct bird-instrument branches and 24 endpoint probes were documented follow-ups. The source is Oona Räisänen's 2007 public-domain blackbird recording, "Turdus merula 2.ogg"; its description reports reduced car noise. The excerpt spans 1.20–3.65 s. Original audio was 22,050-Hz, 16-bit PCM. Renderings are 44,100-Hz stereo, 24-bit PCM, 108,044 samples. Source, renderer and engine hashes are retained. No human participants or experimental animals were tested.

Event and waveform transformations

Eight source-event starts were 0.180, 0.335, 0.755, 1.165, 1.335, 1.455, 1.695 and 2.115 s; ends were 0.335, 0.570, 1.100, 1.320, 1.445, 1.555, 2.005 and 2.235 s. Target starts were 0.125, 0.375, 0.750, 1.125, 1.375, 1.500, 1.750 and 2.125 s; MIDI targets were 100, 98, 91, 108, 113, 105, 101 and 105. Pitch amount p maps tracked frequency f towards target g by $f^{1-p}g^p$, using analytic quadrature phase rotation. Timing amount y interpolates starts with fractional-sample deposition. Spectral amount z mixes each translated event with a phase-continuous tone containing relative harmonic amplitudes 1, 0.17 and 0.05, with Nyquist taper; outside-event background is multiplied by $1-z$. Event energy matching is capped at gain 3, with 5-ms edge windows. The joint path sets $p=y=z=u$; isolated paths vary one coordinate.

Instrument paths and level controls

Flute uses two high-register VSCO 2 Community Edition samples (roots MIDI 93 and 96; CC0). Piano uses Salamander Grand Piano v3, velocity layer 8 (Alexander Holm; CC BY 3.0), with measured-root tuning correction. Both use nearest-root pitch resampling with a 32-tap, 256-phase windowed-sinc kernel. Score-origin blends weight transcription and instrument by $\cos(\pi u/2)$ and $\sin(\pi u/2)$. Instrument release depends on u (65 ms flute; 120 ms piano); attacks are 3 and 0.5 ms. Sample gain follows the frozen event-energy rule, capped at 8; peak limiting caps mixtures at 0.94 when required. Direct paths instead blend the fixed original bird and fully rendered instrument endpoint with those same equal-power coefficients, with exact endpoint branches. Primary audio is also rendered at a shared mono RMS of 0.0188348729. Equal-power mixing does not guarantee equal waveform RMS, and RMS matching does not match perceived loudness.

Frozen computation and output registry

Strict soundfile float32 decoding averages stereo channels before inference. The front end uses a 2,048-sample FFT, 256-sample hop, 128 mel bands, power 2, log1p and per-clip maximum normalization (floor 10^{-8}). This yields 423 frames at 5.805-ms spacing; the FFT window is 46.44 ms. Fresh extractors and mechanism instances process each complete phrase, with no loop padding or inter-anchor state. The first-function-wins registry retains 84 of 88 exported named variants: F2 IGFE, F1 PNH, F1 TPRD and F2 MAA override duplicate names. All 889 declared outputs are

retained; 202 P-labelled dimensions define the primary assay. Eight separate recursive classes are not executed. Default within-clip percentile stretching is disabled for the 131 weighted aggregates. R^3 and temporal features remain computational inputs, not primary outcomes. C^3 combines acoustic descriptors, multihorizon summaries and upstream module outputs through fixed formulas. Its function labels are F1 sensory processing; F2 pattern recognition and prediction; F3 attention and salience; F4 memory; F5 emotion and valence; F6 reward and motivation; F7 motor and timing; and F8 learning and plasticity. E/M/P/F output labels denote extraction, memory, present and forecast, without independently establishing these capacities.

Response estimands

Let $Y_m(u, t, d)$ denote mechanism m at morph amount u , time t and dimension d . Define non-negative RMS values by:

$$r_m(u)^2 = \text{mean}\{[Y_m(u, t, d) - Y_m(0, t, d)]^2\},$$

$$S_m^2 = \text{mean}\{r_m(u)^2\}.$$

The first mean spans all frames and the mechanism's P-labelled dimensions; the second spans all sampled anchors. Global RMS pools all 202 P dimensions, so mechanisms with more dimensions receive more weight. Paired instrument RMS replaces the reference difference with $Y_F(u, t, d) - Y_P(u, t, d)$. Signed endpoint means are computed separately. All frames are compared at their original indices; there is no time warping or learned rescaling. Ranks and correlations are descriptive; no p-values, confidence intervals or resampling of dependent anchors are used.

Resolution, attribution and derived readouts

Nested-grid error compares actual odd-anchor tensors with the mean of adjacent even-anchor tensors, using all frames and P dimensions, normalized by full-path RMS. A boundary check removes 0.3 s from each end. Factorial additivity predicts $Y(p, 0, 0) + Y(0, y, 0) + Y(0, 0, z) - 2Y(0, 0, 0)$; its residual is evaluated over all 27 settings. Sensory attribution compares output-change tensors with explicit SLEE/CDMR sensory terms, reporting cosine alignment and residual RMS without causal interpretation. RAM uses the frozen weighted dimension sums followed by ReLU. The ordered neuromodulator manager executes production, amplification and inhibition operations with outputs clamped to $[0, 1]$. All 26 regional and four chemical-labelled channels are retained.

Integrity, reproducibility and temporal scope

This manuscript reanalysis verified all 1,146 waveform hashes and 855 frozen engine-file hashes with zero mismatches. The original receipts record nine exact repeat checks and four exact regional/neuromodulator reconstruction checks; these are computational checks, not independent experiments. Twenty of 84 mechanisms directly request horizons longer than the phrase, and 78 use forward or centered context; requested horizons reach 36 s. Boundary trimming preserves main-path ranks ($p=0.992-0.997$) but cannot recover missing long context. The companion analysis exposes every mechanism, dimension, region and proxy rather than only figure-selected leaders.

References and research record

1. Thoret, E., Caramiaux, B., Depalle, P. & McAdams, S. Learning metrics on spectrotemporal modulations reveals the perception of musical instrument timbre. *Nat. Hum. Behav.* **5**, 369–377 (2021). doi:10.1038/s41562-020-00987-5
2. Sainburg, T. *et al.* Expectation-driven sensory adaptations support enhanced acuity during categorical perception. *Nat. Neurosci.* **28**, 861–872 (2025). doi:10.1038/s41593-025-01899-1
3. Feather, J., Leclerc, G., Mądry, A. & McDermott, J. H. Model metamers reveal divergent invariances between biological and artificial neural networks. *Nat. Neurosci.* **26**, 2017–2034 (2023). doi:10.1038/s41593-023-01442-0
4. Taylor, H. Whose Bird Is It? Messiaen's Transcriptions of Australian Songbirds. *Twentieth-Century Music* **11**, 63–100 (2014). doi:10.1017/S1478572213000194
5. Elliott, T. M. & Theunissen, F. E. The Modulation Transfer Function for Speech Intelligibility. *PLoS Comput. Biol.* **5**, e1000302 (2009). doi:10.1371/journal.pcbi.1000302
6. McDermott, J. H. & Simoncelli, E. P. Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* **71**, 926–940 (2011). doi:10.1016/j.neuron.2011.06.032
7. Saddler, M. R. & McDermott, J. H. Models optimized for real-world tasks reveal the task-dependent necessity of precise temporal coding in hearing. *Nat. Commun.* **15**, 10590 (2024). doi:10.1038/s41467-024-54700-5
8. Plomp, R. & Levelt, W. J. M. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**, 548–560 (1965). doi:10.1121/1.1909741
9. Sethares, W. A. Local consonance and the relationship between timbre and scale. *J. Acoust. Soc. Am.* **94**, 1218–1228 (1993). doi:10.1121/1.408175
10. Chen, S., Thielk, M. & Gentner, T. Q. Auditory Feature-based Perceptual Distance. *bioRxiv*, preprint (2024). doi:10.1101/2024.02.28.582631
11. Gorman, J. C., Sainburg, T., McPherson, T. S. & Gentner, T. Q. Anticipatory organization of neural population dynamics speeds behavioral decisions. *bioRxiv*, preprint (2026). doi:10.64898/2026.06.30.735699
12. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008). doi:10.3389/neuro.06.004.2008

Data and code availability

The accompanying local research package contains the editable manuscript, six vector figures, figure-source arrays, complete mechanism/dimension and projection tables, analysis scripts, source metadata, reference records and a SHA-256 manifest. Full waveforms, frozen engine sources and per-recording tensors remain in the adjacent experiment archive; its local paths and exact rebuild commands are documented in README.txt. No public repository accession or permanent dataset DOI has been assigned. The compact package reproduces figures and typesetting; raw-output reanalysis additionally requires that archive.

Attribution and research transparency

Bird audio: Oona Räisänen, public domain (2007). Flute samples: VSCO 2 Community Edition, Versilian Studios/Sam Gossner, CC0. Piano samples: Salamander Grand Piano v3, Alexander Holm, CC BY 3.0. Third-party asset links and transformation details are included with the source record. MI is developed by the author; this is an internal evaluation of that research prototype. OpenAI Codex assisted source inspection, reanalysis, figure generation and drafting. This document is a research manuscript, not a peer-reviewed publication; no collaborator authorship or endorsement is implied.

Companion analysis inventory

File in data/	Content
mechanism-responses.csv	All 84 mechanisms; path, endpoint and layer summaries
dimension-responses.csv	All 889 declared dimensions, including 202 present-state dimensions
regional-responses.csv	All 26 regional projections; both level conditions
neurochemical-responses.csv	All four model proxies; both level conditions
curves.npz / analysis.json	Anchor trajectories, rankings, grid resolution and integrity receipts