

Shared mechanisms and interpretable affective predictions in birdsong-to-music transformations

Amaç Erdem

Independent researcher

Abstract

How does converting natural sound into a musical representation reorganize an auditory model's responses? We analysed a frozen Musical Intelligence implementation across 1,146 dependent renderings of one 2.45-s blackbird phrase, retaining 84 mechanisms and all 889 declared output dimensions. Five transformation paths linked birdsong, a synthetic transcription, flute and piano. Four mechanisms—Brainstem Consonance Hierarchy, Pitch Salience in Cortical Lateralization, Inharmonicity-Attention Capture Model and Therapeutic Affective Resonance (TAR)—remained in the upper response quartile across all five paths under input-level, temporal-window and sampling-grid controls. This shared set depends on the defined output scale. TAR generated a reproducible directional contrast: its mean stress-reduction prediction increased from transcription to flute but decreased from transcription to piano. Exact terminal decomposition showed that similar bird and flute means arose from opposing roughness-trend and energy-change contributions, despite large framewise differences. Isolated timing and spectral transformations produced opposite signs. Combining layers also improved one coordinate-readout contrast, while failing to outperform the best single layer consistently across routes and controls. The architecture therefore supplies an inspectable account of shared sensitivity, conditional complementarity and divergent affective predictions. These are computational findings that specify falsifiable perceptual and stress-regulation hypotheses; they are not measurements of therapeutic efficacy.

Introduction

Natural sounds carry fine frequency trajectories, fluctuations and event relations that musical notation selectively reorganizes. Messiaen's birdsong transcriptions make this tension concrete: sonographic comparisons reveal both correspondences with source recordings and compositional transformation.¹ Asking how natural sound becomes musical structure therefore requires specifying what is retained, regularized or replaced. Here, a controlled transformation space makes those decisions explicit rather than treating “birdsong” and “music” as unrelated stimulus categories.

Speech-modulation, sound-texture and task-optimized modelling studies show that the usefulness of auditory detail depends on the target.^{2,3,4} The relevant question extends beyond acoustic distance. Spectrotemporal features can explain aspects of perceived timbre when calibrated against listeners' judgments,⁵ whereas successful auditory models can still possess invariances unlike those of people.⁶ A mechanism-labelled architecture adds a further opportunity: its computations can be inspected to ask which response

components recur across transformations and which directional hypotheses follow from them. High output magnitude alone does not answer either question.

Affective predictions are especially informative because both natural sounds and music have been studied in relation to stress recovery.^{7,8,9} Recent work contrasts music with an active natural-sound control and links recovery to musical structure.¹⁰ Those studies motivate a within-source question: which spectrotemporal changes preserve or alter a model's proposed stress-regulation signal? They do not establish that a therapeutic label in software measures that human outcome.

We analyse all declared C³ outputs of Musical Intelligence (MI), including each mechanism's actual layer schema. The study identifies a shared four-mechanism response set, traces TAR's signed stress-prediction contrast to exact terminal terms, and tests whether combined layers improve readout of the known transformation coordinate. Reference-dependent geometry and all earlier present-state, resolution and projection audits are retained. The resulting contribution is an executable, falsifiable account of what this architecture computes when one natural sound is made musical.

A controlled transformation space

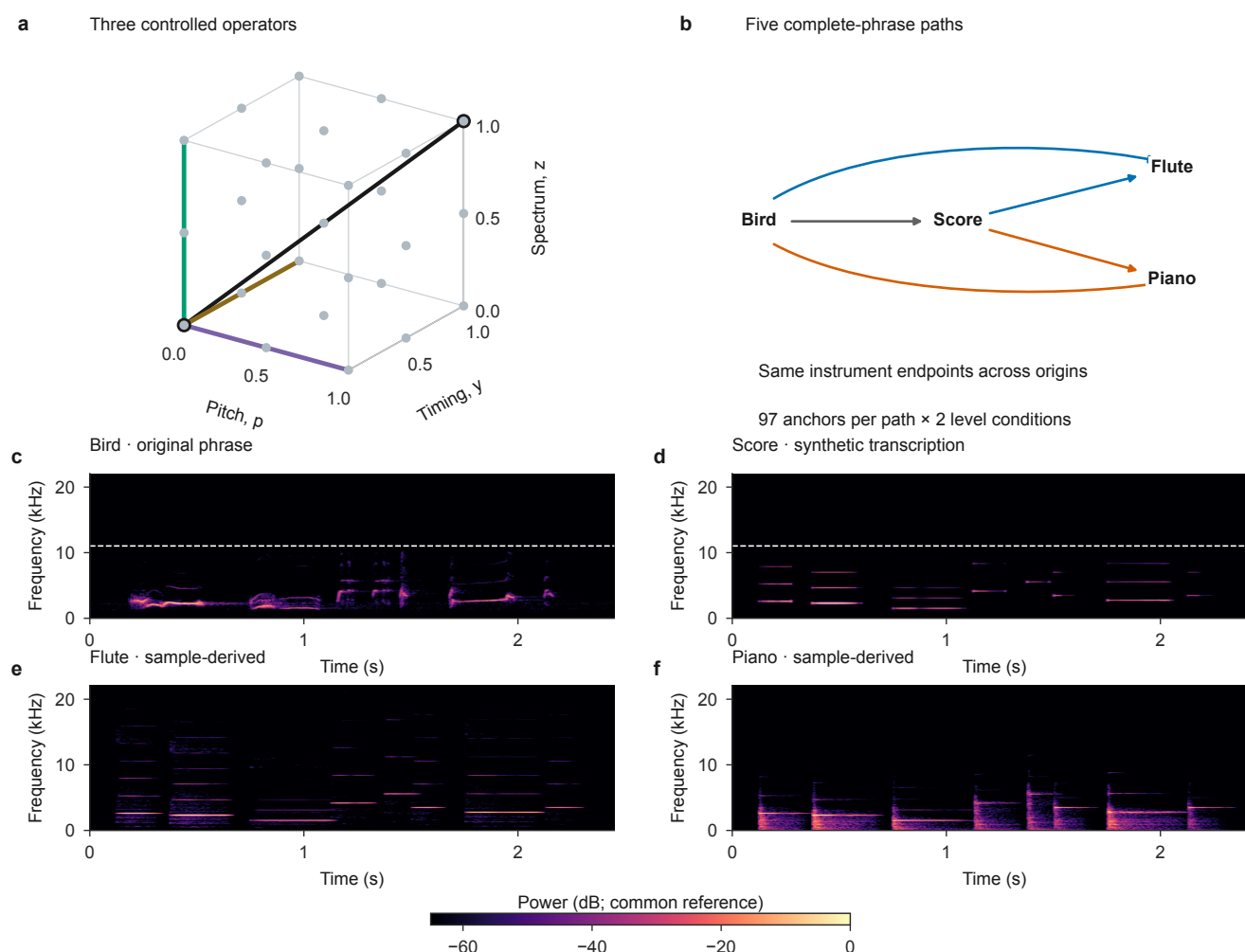


Fig. 1 | A three-operator space connects birdsong to musical endpoints. **a**, Pitch, timing and spectral coordinates each take values 0, 0.5 and 1 in the 27-setting factorial grid. Coloured axes show isolated operators; the black diagonal is the joint transformation from bird (B; 0,0,0) to score (S; 1,1,1). Coordinates are renderer parameters, not perceptual distances. **b**, Five primary paths, each with 97 anchors per level condition, share fixed instrument endpoints. **c–f**, Common-RMS spectrograms with one power reference and colour scale. Dashed lines in c and d mark the source's 11.025-kHz Nyquist limit. High-register sample resynthesis adds energy above that band. 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 full-profile reanalysis retains all 889 declared channels, alongside the original 202-dimensional present-state assay. These are complementary observation choices over the same frozen computations.

The reference reverses the ordering



Fig. 2 | Changing the reference reverses the flute–piano ordering. **a**, Global path RMS pools all 202 present-state dimensions. Identical instrument endpoints are compared from the score and bird references. Values are shown to five decimal places. **b**, Mechanism path RMS for the union of the five highest-ranked mechanisms on each path (13 mechanisms). Squared changes are averaged over anchors, frames and each mechanism’s present-state dimensions before taking the square root. The complete 84-mechanism table accompanies the paper. **c**, Spearman correlations of all 84 mechanism rankings. These describe agreement among model profiles, not behavioural prediction. All panels use common-RMS audio. Anchors are dependent transformations; no error bars or significance tests are assigned. B, bird; S, score; F, flute; P, piano.

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. 2a). 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.

The reference effect also survives extending the observation to all 889 channels. Endpoint RMS is 0.13015 for score→flute and 0.18617 for score→piano, versus 0.14466 for bird→flute and 0.11378 for bird→piano. Relative endpoint ordering reverses in 69 of 84 mechanisms and 274 of 326 recorded layer groups; 74 mechanisms reverse at native audio levels. Present-state rank correlations are 0.950 between score-origin instrument paths and 0.855 between direct paths, but only 0.471 and 0.455 for the same instrument across origins. The effect thus concerns response geometry, with its scale and coordinate set stated explicitly.

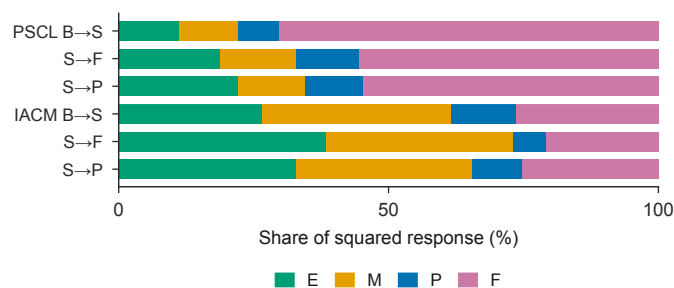
A shared mechanism set across five routes

a Common response profiles | ranks among 84 mechanisms

		B → S	S → F	S → P	B → F	B → P
BCH	Brainstem Consonance Hierarchy	7	11	14	7	10
PSCL	Pitch Salience in Cortical Lateralization	8	15	12	4	2
IACM	Inharmonicity-Attention Capture Model	4	16	9	8	11
TAR	Therapeutic Affective Resonance	1	3	13	3	1
SDNPS	Stimulus-Dependent Neural Pitch Salience	9	4	5	44	29
DAED	Dopamine Anticipation-Experience Dissociation	14	10	15	11	15
MEAMR	Music-Evoked Autobiographical Memory Reward	6	12	18	15	12
CDMR	Context-Dependent Mismatch Response	16	5	4	36	47
SLEE	Synaptic Long-term Encoding Engine	11	2	2	30	32
ESME	Error-Signal Modulated Encoding	20	6	6	46	55

Green marks: four mechanisms retained across all five routes and level, time-window and grid controls.

b Layer contributions reveal present-only omissions



c Strongest additional candidate per route

B → S	PCCR Pitch Chroma Cortical Representation
S → F	CSG Consonance-Salience Gradient
S → P	MSPBA Musical Syntax Processing in Broca's Area
B → F	PCCR Pitch Chroma Cortical Representation
B → P	PCCR Pitch Chroma Cortical Representation

Fig. 3 | Full-profile observation reveals shared and route-dependent responses. **a**, Ranks by framewise, all-layer path RMS in unrescaled model units; smaller numbers indicate larger responses. The ten rows are the common upper quartile (top 21 of 84) on B→S, S→F and S→P. Green marks identify BCH, PSCL, IACM and TAR, retained across all five routes under native/equal-RMS input, full/trimmed time windows and 97/49/25-anchor controls. **b**, Layer shares of squared change for PSCL and IACM. Shares reflect channel count and scale, not independent explained variance. **c**, Highest-ranked additional candidate outside the four-mechanism set on each route. Selection does not establish specificity. B, bird; S, synthetic score; F, flute; P, piano.

A mechanism was designated shared when it ranked in the top 21 of 84 on every selected route. Ten met this descriptive criterion on the three sequential routes; six met it on all five at equal RMS. Four survived all specified level, temporal-window and grid controls: Brainstem Consonance Hierarchy (BCH), Pitch Salience in Cortical Lateralization (PSCL), Inharmonicity-Attention Capture Model (IACM) and Therapeutic Affective Resonance (TAR; Fig. 3). This is a reproducible intersection of response rankings in the frozen implementation.

Retaining the actual output schemas is consequential. Sixty mechanisms use literal E/M/P/F layers, whereas TAR uses E/TI/P/F, with six temporal-integration channels; Dopamine Anticipation-Experience Dissociation uses E/M/P and Music-Evoked Autobiographical Memory Reward uses E/P/F. Restricting the comparison to a four-letter template would omit relevant computations. The full-profile analysis includes these mechanisms without relabelling their schemas.

PSCL and IACM illustrate what present-only observation misses. PSCL moves from present-state ranks 28/28/30 to full-profile ranks 8/15/12 on the sequential routes. Its F layer accounts for 54.7–70.3% of squared response. IACM moves from 10/37/24 to 4/16/9, with 88.0–94.0% outside P. These allocations expose where implemented responses occur; they do not establish successful forecasting or independently measured attention.

Beyond the shared set, Pitch Chroma Cortical Representation leads the additional bird-origin candidates, Consonance-Salience Gradient the score→flute candidate, and Musical Syntax Processing in Broca's Area the score→piano candidate. The latter uses S/M/P/F. The shared set is stable under the listed acquisition and sampling controls but not invariant to output units: endpoint-based rescaling leaves no common top-quartile set on the three sequential routes. We therefore claim shared raw-unit sensitivity, not universal mechanism importance.

TAR produces a directional affective contrast

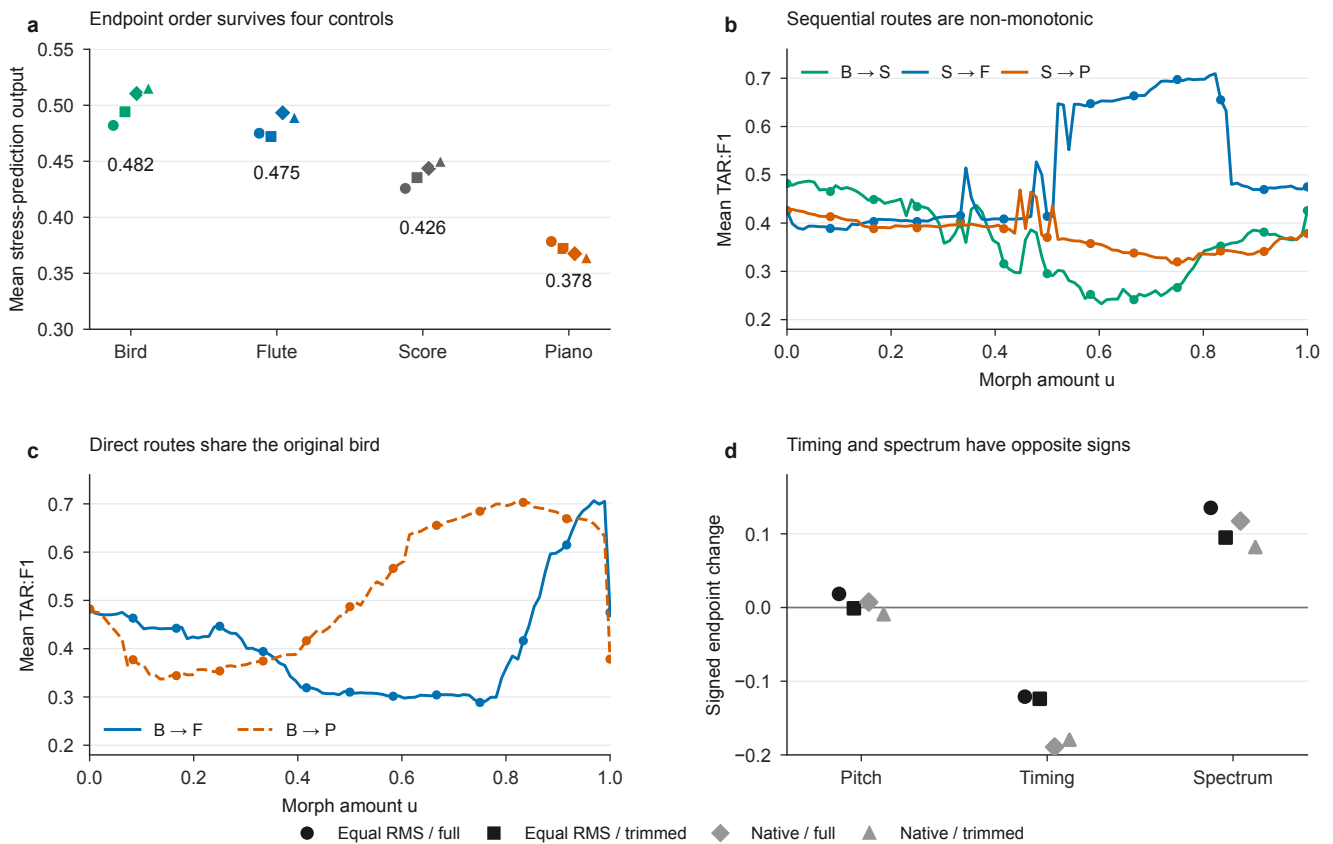


Fig. 4 | A named affective prediction separates the transformation endpoints. **a**, Phrase-mean TAR:F1 stress-reduction predictions at four endpoints under four control conditions; printed values are equal-RMS/full-window means. Markers are controls, not independent samples. **b,c**, All 97 computed anchors for sequential and direct routes; sparse markers aid inspection and connecting lines do not imply evaluated continuity. **d**, Signed endpoint effects for isolated operators under the same controls. Trimming removes 0.3 s at each end. Scores are bounded model outputs, not probabilities or percentages of stress reduction. Non-monotonic interior responses preclude interpreting the endpoint order as a monotonic transition or an optimal intervention.

TAR is relevant because its forecast-labelled output makes a directional statement that an unsigned response norm cannot. At common RMS, mean stress-reduction prediction is 0.48203 for bird, 0.47503 for flute, 0.42577 for synthetic transcription and 0.37838 for piano (Fig. 4a). The order bird > flute > transcription > piano holds under all four combinations of native/equal-RMS level and full/trimmed time window. The transcription→flute contrast is +0.04926; transcription→piano is −0.04739.

The model does not compute a therapeutic relationship between two sounds internally. It evaluates each complete recording separately; the controlled contrasts reveal how its prediction changes. TAR's other outputs also separate: transcription has the highest present therapeutic-reward mean (0.26955), although its stress-prediction mean is below bird and flute. Mood-improvement means instead order transcription (0.18996), piano (0.18149), flute (0.16901) and bird

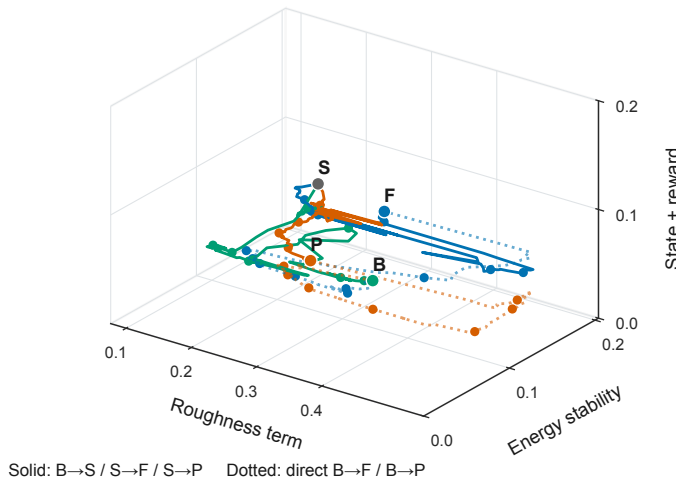
(0.13468). A single favourable-versus-unfavourable interpretation would erase these distinctions.

The full curves are strongly non-monotonic. On the score→flute path, mean F1 reaches 0.70943 at $u=0.82292$, above either endpoint. Direct bird→flute and bird→piano paths likewise attain high interior values. These extrema are properties of mixtures and the frozen formulas, not evidence that intermediate sounds maximize human recovery. Endpoint comparisons and complete trajectories are therefore reported together.

Isolated timing lowers F1 by 0.12100, whereas isolated spectral replacement raises it by 0.13526. Both signs survive level and temporal-window controls. The smaller pitch-only increase (+0.01846) becomes slightly negative after trimming and is not a robust directional result. TAR thus supplies a precise candidate hypothesis: temporal reorganization and spectral replacement can act in opposing directions even when they contribute to the same musical transformation.

Similar means can arise from opposing computations

a Literal terminal-term space



b Audited terminal computation

$$\text{TAR:F1} = \text{clip}(A + B + C + D, 0, 1)$$

$$A = 0.20 \times (E0 \times T2)$$

$$B = 0.20 \times P0$$

$$C = 0.35 \times (1 - r)$$

$$D = 0.25 \times [1 - \text{clip}(|v|, 0, 1)]$$

r : roughness-trend feature
 v : amplitude-change velocity feature
 138 frames $\approx$ 0.801 s per H³ window

466,146 frame checks

Maximum error: 1.19×10^{-7}

c Signed endpoint attribution | opposing terms can cancel

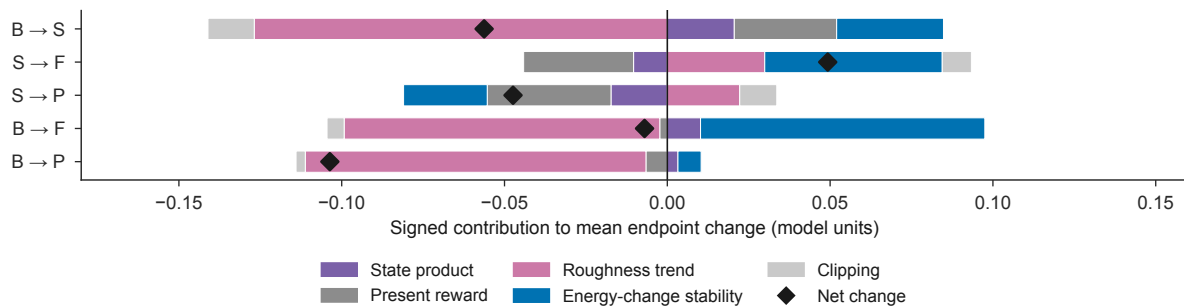


Fig. 5 | Terminal decomposition connects the affective contrast to executable terms. **a**, Three-dimensional trajectories use literal phrase-mean additive coordinates: roughness-trend term, energy-change stability term, and the sum of state-product and present-reward terms. No embedding is fitted; distances are not global C³ distances. Clipping is excluded from these axes and retained explicitly in **c**. **b**, Frozen TAR:F1 formula. E0, T2 and P0 are TAR therapeutic-state, anxiety-reduction and therapeutic-reward outputs; r and v are the specified H³ features. **c**, Signed endpoint contributions from the four terms and final clipping; diamonds show net changes. Contributions are computed framewise before averaging. Formula replay covers 1,102 unique included recordings, not all 1,146 archive entries.

The stress-prediction difference can be traced exactly through the terminal computation (Fig. 5). The coefficients 0.20, 0.20, 0.35 and 0.25 are fixed in the code, not fitted to stress outcomes. The roughness term uses a temporal trend descriptor; the energy-change term uses the velocity transform of an amplitude-derivative feature. Although a code comment calls the latter a tempo proxy, it is not a BPM estimate. Reconstructing these H³ features from saved R³ inputs reproduces F1 across 466,146 frames with maximum error 1.19×10^{-7} .

For transcription→flute, the energy-change contribution is +0.05451 and the roughness-trend contribution +0.02993. They overcome negative state-product (−0.01036) and present-reward (−0.03376) terms. For transcription→piano, the energy-change contribution instead falls by 0.02569, helping produce the opposite total sign. This identifies the arithmetic that separates the instruments without inferring that either term uniquely causes a listener's response.

The near equality of bird and flute means is particularly revealing. Their roughness-trend contrast is −0.09694 and their energy-change contrast +0.08731, largely offsetting one another. Mean F1 differs by only −0.00699, yet framewise RMS difference is 0.35086. Similar aggregate predictions therefore do not imply preserved trajectories or preserved computations. A study that measured only phrase means would miss this distinction.

Across the five routes, recommended consonance (I1) and stress prediction (F1) jointly account for 61.5–87.2% of TAR's squared change. Both reuse roughness-related dynamics, so their agreement is not independent support. The implemented 0.801-s H³ windows and boundary replication are also part of the explanation. TAR's clinical names and forecast comments do not establish calibrated therapeutic efficacy or a validated future horizon; the value is a testable, explicitly decomposed hypothesis.

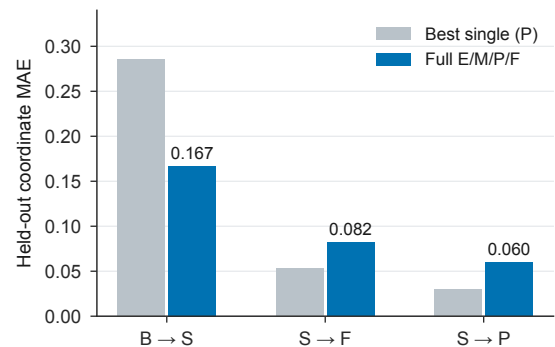
Layer integration yields conditional readout gains

a Best-single MAE minus full-profile MAE

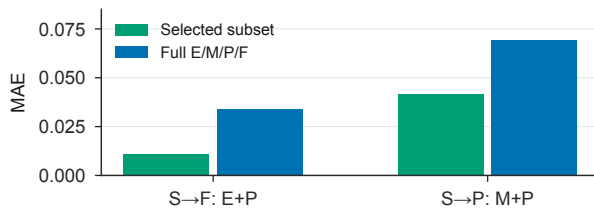
	B → S	S → F	S → P
BCH	-0.013	-0.016	+0.013
PSCL	+0.001	-0.012	-0.016
IACM	-0.006	-0.009	-0.033
TAR	-0.108	-0.002	+0.056
SDNPS	+0.118	-0.029	-0.030
SLEE	-0.076	-0.013	-0.016
CDMR	-0.036	-0.007	-0.008

Positive: full profile improves coordinate readout.

b SDNPS: a route-dependent gain



c PSCL: subsets can outperform the full profile



d Complementarity depends on the task

SDNPS on B→S
41.5% lower MAE
 All three routes, primary
0 of 84 improve
 Fixed full-dimension scaling
4 of 84 improve
 Across every control
0 robust advantages

Five held-out contiguous blocks; four-anchor buffer; training-only scaling; ridge $\alpha = 1$. All 41,760 evaluations retained.

Fig. 6 | Shared sensitivity and improved readout are different properties. **a**, Best-single-layer MAE minus full-profile MAE for selected mechanisms, with ridge $\alpha=1$ and subset-specific dimension scaling. Positive values favour the full profile. **b**, SDNPS full-profile and best-single-layer errors by route. **c**, Illustrative PSCL subsets versus all four layers. The target is the known renderer coordinate u , with five held-out contiguous blocks and a four-anchor training exclusion buffer. “Best” single layers and displayed subsets are selected descriptively from held-out scores, not by nested model selection. All outcomes, including controls and failures, are retained; no human response is predicted by this test.

A common high-response mechanism need not encode morph amount in a form recoverable by a linear reader. We therefore treated every channel’s phrase mean as a separate feature and predicted the known coordinate u in contiguous held-out blocks. Training-only scaling, near-constant-channel removal and a four-anchor exclusion buffer reduce obvious adjacent-anchor leakage. The 41,760 evaluations compare actual layer subsets, two level conditions, two dimension denominators and ridge penalties 0.1, 1 and 10; they are dependent computational evaluations, not independent experiments.

For SDNPS on bird→transcription, full-profile MAE is 0.16709 versus 0.28546 for the best single layer P, a 41.5% error reduction (Fig. 6). The same full profile is worse on transcription→flute (0.08212 versus 0.05314) and transcription→piano (0.05986 versus 0.02959). PSCL’s E+P subset outperforms its full profile on flute (0.01085 versus 0.03402), and M+P does so on piano (0.04167 versus 0.06906). More layers are therefore useful in some contrasts but not uniformly necessary.

No mechanism beats its own best single layer on all three sequential routes under the primary reader. Four do with a fixed full-dimension denominator, but none gains more than 0.01 MAE on every route, and none preserves a positive advantage across all tested controls. Moreover, the bird→transcription full readers for CDMR, SLEE and ESME are worse than the training-mean baseline (MAE 0.33380), despite their high response ranks. Sensitivity and coordinate recoverability are empirically distinct here.

An earlier cross-instrument task gives SDNPS MAE 0.04172 when training on the flute path and testing on piano; its best single-layer error changes from 0.36858 to 0.12370 under denominator control. This is a different generalization task, with marked direction and scaling sensitivity. Channel exclusion tests access to the computed representation; they do not sever internal connections or demonstrate causal synergy. Integrated architecture is supported here by inspectable, complementary responses and conditional readout gains, not by universal predictive superiority.

Three-dimensional response profiles

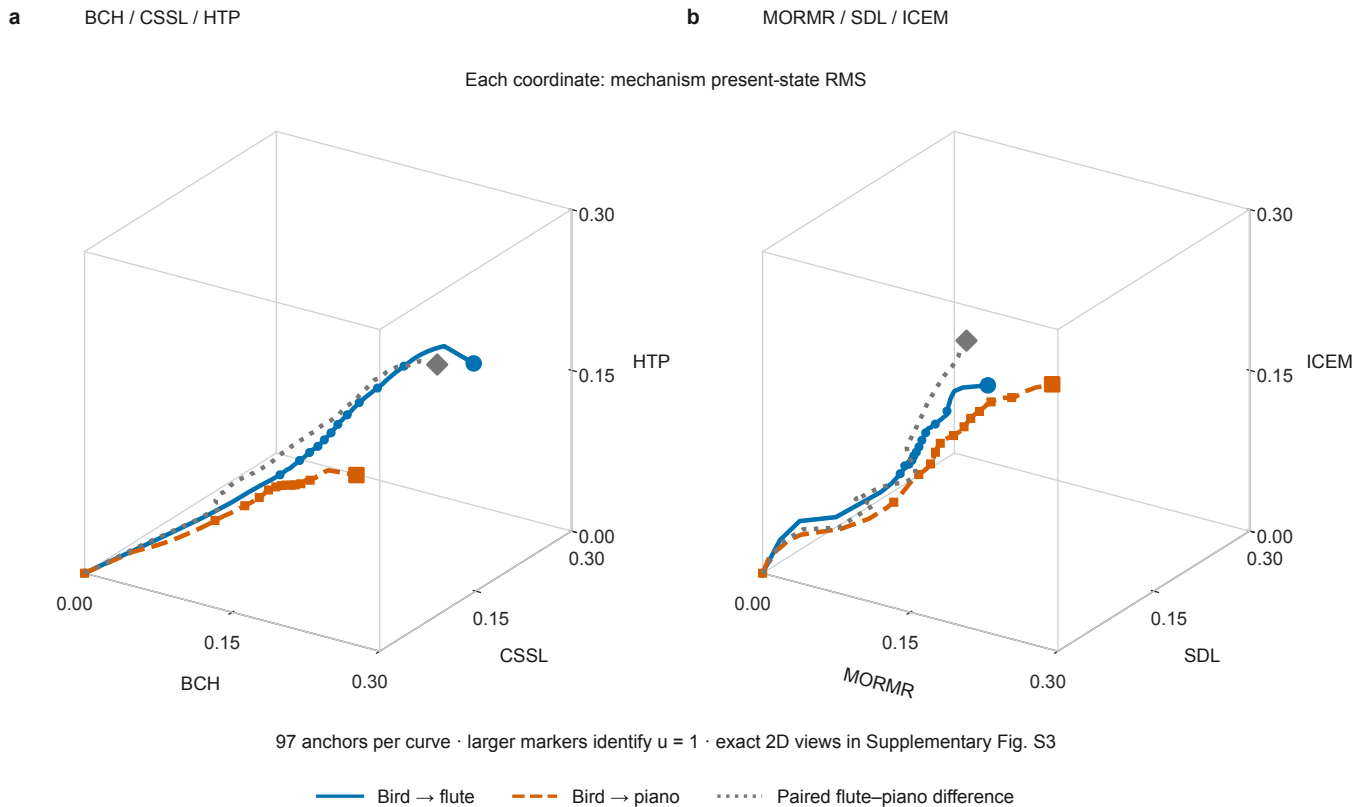


Fig. 7 | Direct instrument paths traced through six named response coordinates. **a**, BCH, CSSL and HTP present-state RMS. **b**, MORMR, SDL and ICEM present-state RMS. The six mechanisms are the union of the top three on each direct path, plus ICEM, the second-ranked paired contrast; Supplementary Table S5 expands their names. Each blue or vermilion coordinate is an RMS displacement from the shared bird origin. Grey coordinates are RMS differences between flute and piano tensors at matched morph amounts, not differences between distances. Each curve includes all 97 anchors; every eighth is marked and larger markers denote $u=1$. The three-dimensional views use named coordinates without fitting or dimensionality reduction; distances within these displays are not global C^3 distances. Supplementary Fig. S3 preserves exact per-coordinate plots against morph amount. An interactive companion permits rotation and anchor inspection.

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. 7; Supplementary Fig. S3). 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.

Operator and level controls

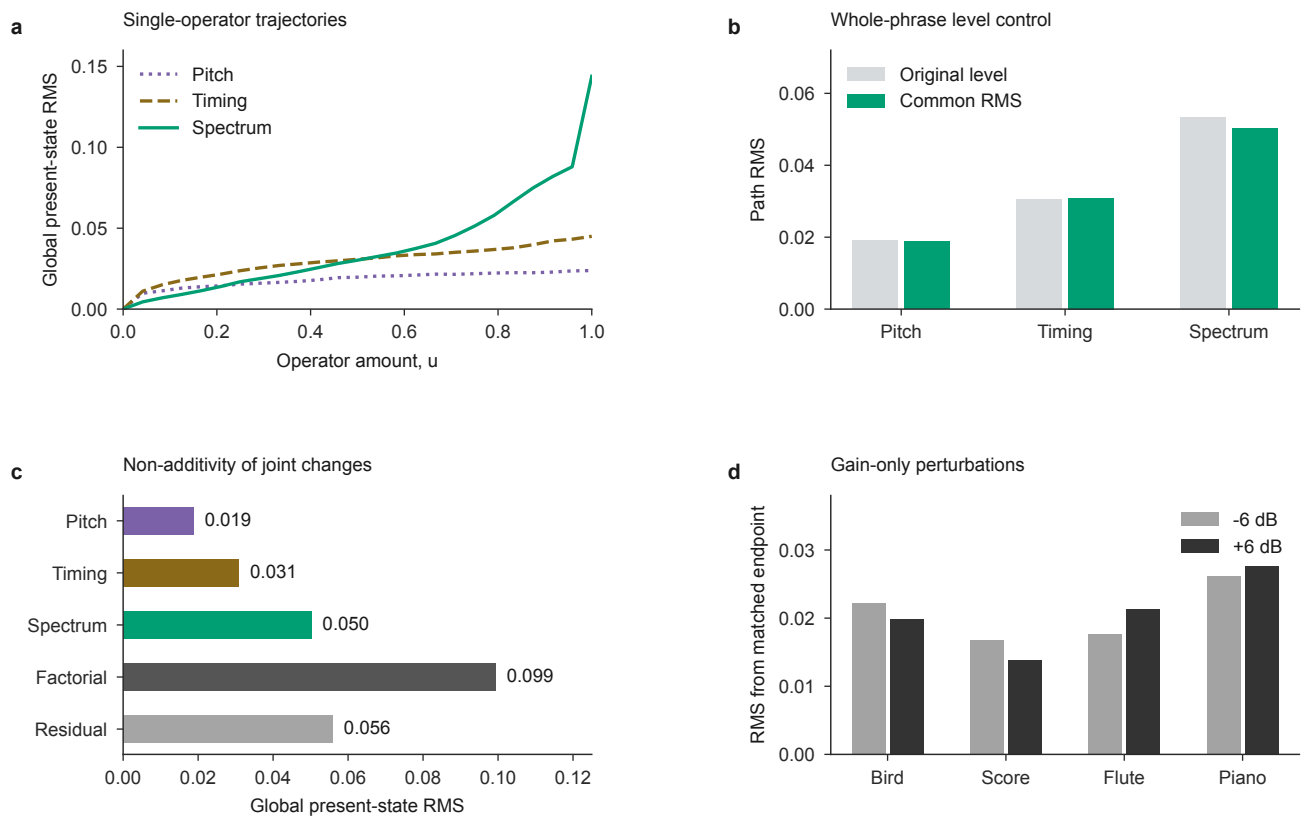


Fig. 8 | Operator and level controls constrain cognitive interpretations. **a**, Twenty-five-anchor, common-RMS paths varying only pitch mapping, onset displacement or spectral replacement. **b**, Path RMS before and after whole-phrase RMS matching. **c**, Single-operator responses compared with the 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. 8a,b). Spectral replacement was therefore 2.68 times the pitch response and 1.63 times the timing response under these particular operator definitions. The operations differ physically; spectral replacement also removes background and changes within-event structure. With all 889 dimensions, the corresponding values are 0.02423, 0.03804 and 0.05889; 75 of 84 mechanisms respond more to spectral replacement than to either isolated alternative.

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 are output labels; no listener memory or reward report was supplied.

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. 8c). 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. The all-channel residual-to-total RMS ratio is 0.5979; neither ratio is an explained-variance percentage.

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.

Rank stability and local sensitivity

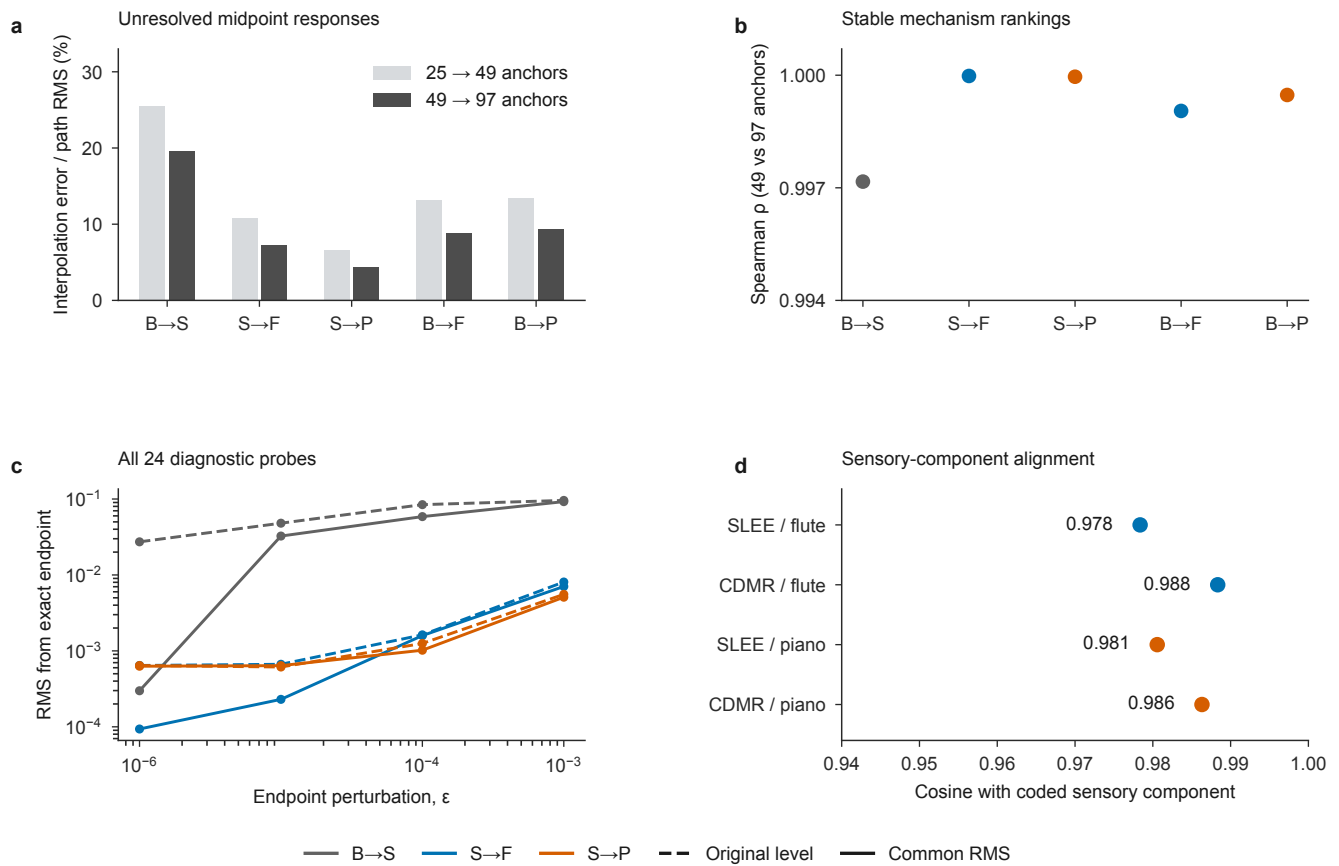


Fig. 9 | Sampling resolution and formula grounding are separate requirements. **a**, RMS error when omitted midpoint tensors are predicted by linear interpolation of neighbouring anchors, divided by full-path RMS. **b**, Rank correlation between 49- and 97-anchor mechanism sensitivities. **c**, Diagnostic perturbations $\epsilon=10^{-6}$ – 10^{-3} from the score endpoint on B→S and the score-origin endpoints on S→F and S→P; logarithmic 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. 9).

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. These level-sensitive endpoint changes require numerical localization, not a perceptual-threshold interpretation. Full-profile local-change peaks are likewise often boundary-dominated. The MSPBA piano peak shifts under coarse sampling, precluding a stable threshold claim.

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,^{11,12} 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. The formulas remain frozen; correction requires a versioned replication. High sensitivity of a learning-labelled output in this experiment cannot be read as evidence that learning occurred.

What derived readouts conceal

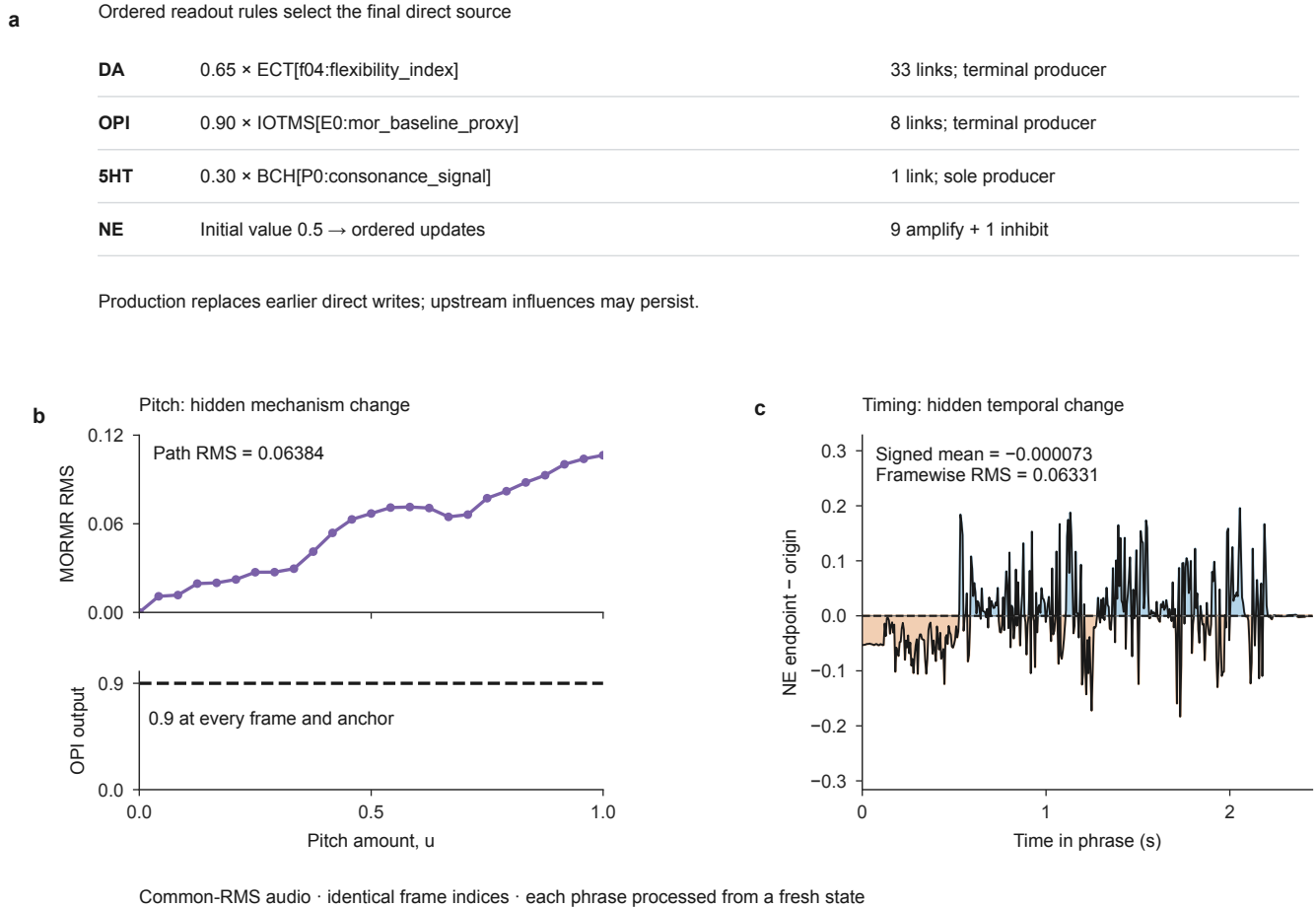


Fig. 10 | A derived readout can conceal a changing representation. **a**, Terminal mappings in the frozen neuromodulator manager. Later *produce* operations overwrite earlier direct writes; NE retains ordered amplification and inhibition. Upstream dependencies can still transmit indirect influences. **b**, On the 25-anchor common-RMS pitch path, MORMR present-state RMS changes (upper axis), while every OPI frame remains at 0.9 (lower axis). Aligned panels separate the two quantities and scales. **c**, NE endpoint minus origin on the common-RMS timing path, at the same 423 frame indices. The signed mean is -0.000073 , but frame-wise RMS is 0.06331; shaded signs expose temporal cancellation. This is a within-phrase contrast, not sequential adaptation. Panels b and c were selected after source inspection. Supplementary Fig. S1 and Table S4 retain the complete path summaries.

The final DA, OPI and 5HT outputs equal weighted source dimensions from ECT, IOTMS and BCH, with coefficients 0.65, 0.90 and 0.30, respectively (Fig. 10a). These identities matched a float32 replay for all 484,758 frames from 1,146 archived recordings; reconstructed channel means agreed with stored means within 2.98×10^{-8} . Earlier direct production writes are overwritten, although their mechanisms may still affect later source dimensions through C^3 dependencies. The inspected executor does not feed these channel outputs back into mechanism computation.

On the common-RMS pitch path, MORMR present-state path RMS was 0.06384, yet OPI was exactly 0.9 at every frame and anchor (Fig. 10b). The terminal IOTMS source was already clipped at 1.0; multiplying by 0.9 produces this effective ceiling. Constancy of the opioid-labelled readout

therefore does not imply invariance of a reward-labelled mechanism. Timing also retained OPI=0.9 throughout. For NE, the timing endpoint mean difference was -0.000073 while the same endpoint's frame-wise RMS difference was 0.06331 (Fig. 10c). Temporal averaging can hide opposing changes.

The observation rule is therefore part of the scientific result. A mechanism-level displacement, a bounded channel value and a phrase-averaged channel are different estimands. Their agreement or disagreement must be traced through the implemented source dimensions and update order before assigning an interpretation. The full ordered link registry and matched temporal contrasts are retained in Supplementary Tables S3–S4.

Regional scale depends on connectivity

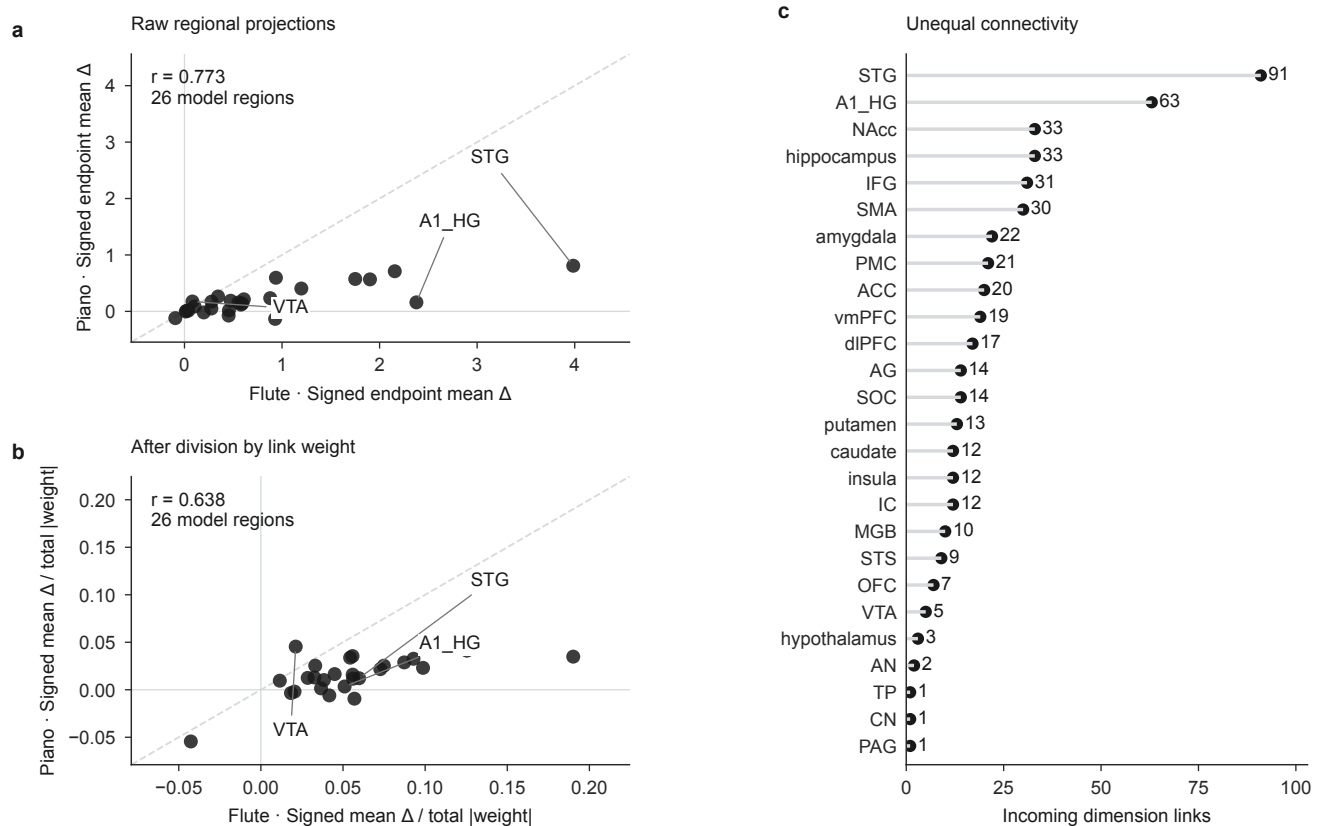


Fig. 11 | Connectivity changes the scale of regional concordance. **a**, Signed bird→flute and bird→piano endpoint mean changes in all 26 RAM regions (Pearson $r=0.773$). **b**, The same values divided by each region's sum of absolute incoming link weights ($r=0.638$). Each dot represents one dependent model region; dashed diagonals indicate equality, not a fitted regression. STG, A1_HG and VTA are labelled to orient the two views. **c**, Every region's incoming dimension-link count, ordered from largest to smallest. These counts describe the frozen registry, not biological connectivity measurements. Audio uses common whole-phrase RMS and the same bird origin. Supplementary Fig. S2 and Tables S1–S2 retain every signed regional value, source count, weight and leading contribution.

Regional response summaries were concordant but sensitive to the observation scale. Direct flute and piano endpoint-change vectors correlated at $r=0.772922$ across all 26 regions. Dividing each projection by its total absolute incoming link weight reduced this value to 0.638289 (Fig. 11a,b). No weights were fitted. The normalization asks how much the cross-region pattern depends on heterogeneous weighted accumulation; it does not calibrate the outputs to neural measurements.

The source inventory makes that heterogeneity explicit. STG receives 91 dimension links from 57 mechanisms, and A1_HG receives 63 from 44. CN, TP and PAG each receive one link (Fig. 11c; Supplementary Table S1). A larger unnormalized regional change can therefore reflect the number, magnitude and composition of summed sources. It cannot establish that one anatomical region would be more strongly recruited in a listener.

Signs agreed between the two instrument contrasts in 22 of 26 regions; flute changes were positive in 25 and piano changes in 21. Flute exceeded piano in signed change in 25 regions, with VTA the exception. Division by a positive region-specific weight sum preserves each region's sign and within-region instrument order, while changing the relative scale across regions. Thus the broad shared direction and the reduced correlation after normalization are compatible descriptions of the same outputs.

The four audited endpoints had strictly positive sums before the regional rectification step (minimum 0.062835). Their signed changes can consequently be decomposed into additive weighted source contributions without crossing that non-linearity. Supplementary Table S2 reports the three largest absolute contributors per region with their signs; the complete contribution table retains cancellation among all sources. These decompositions trace the model's arithmetic and do not establish causal neural attribution.

Discussion

What the architecture contributes

Three findings now organize the essay. First, four mechanisms remain shared high-response candidates across five natural-to-musical routes under the specified acquisition and sampling controls. Second, signed TAR predictions separate timing, spectral replacement and instrument rendering in ways that can be reconstructed from the code. Third, combining layers improves some coordinate readouts but is not universally beneficial. Together, these results establish an interpretable computational phenotype: what changes, which operations carry the change, and where the observation rule matters. The contribution is stronger than an unqualified report that MI “reacts” to the sounds.

From a therapeutic name to a falsifiable hypothesis

TAR deserves attention because its outputs state a specific functional hypothesis, not because its RMS is maximal. The observed transcription→flute increase and transcription→piano decrease can be specified before a new empirical study. The bird–flute near-equality is equally informative: opposing terminal terms conceal large time-resolved differences. A useful validation must therefore distinguish mean recovery, moment-to-moment affective response and acoustic resemblance, rather than treating them as interchangeable endpoints.

Human studies make this question plausible but do not answer it for these renderings. Alvarsson and colleagues reported faster skin-conductance recovery during a fountain-and-bird soundscape than noise, with no corresponding high-frequency heart-rate-variability effect; exposure lasted four minutes.⁷ Stobbe and colleagues found reduced anxiety and paranoia after six minutes of birdsong in healthy adults, without improved cognitive-task performance.⁸ A broader synthesis supports aggregate benefits of natural sounds while retaining substantial differences among outcomes and comparisons.⁹ In 2026, You and colleagues reported advantages of music over an active natural-sound control and associations between musical tension and stress-network dynamics.¹⁰ These experiments use different stimuli and endpoints; none validates the present TAR coefficients or the four-endpoint ordering.

The decisive next question is whether a mechanistic score adds information beyond its ingredients. Roughness can relate to aversion and salience,¹³ but a roughness-trend feature in MI is neither a direct aversion rating nor a stress biomarker. Test TAR against roughness, envelope dynamics, loudness, pleasantness and arousal baselines, then compare the complete formula with its terminal terms. If the full formula adds no held-out value, the interpretable result remains useful but its therapeutic interpretation should be narrowed.

If it adds reproducible value across independent sources and listeners, the architecture gains evidence for the proposed functional composition.

A targeted empirical programme

The first experiment should use independent birds, phrases, registers and renderers, with source-matched bird, transcription and instrument conditions. Register, bandwidth, envelope and source recognition need explicit controls. Preregister the signed endpoint and isolated-operator contrasts, while retaining interior anchors that challenge the predictions rather than selecting only supportive examples. Measure perceptual dissimilarity and affective judgments before undertaking a recovery study with adequate exposure, a baseline stress measure, randomized conditions and prespecified physiological outcomes. The 2.45-s source is a controlled computational probe, not an established therapeutic dose.

For perceptual discrimination, compare these response features with spectrotemporal metrics and Auditory Feature-based Perceptual Distance where reproducible.^{5,14} Cue-dependent changes in acuity provide a separate task: identical waveforms under different expectations require a contextual model extension, since the present reset-per-clip computation has no manipulated cue state.¹⁵ Anticipatory population dynamics described in a 2026 preprint similarly motivate separating pre-stimulus state from stimulus-evoked geometry.¹⁶ Representational similarity analysis can compare measured and computed patterns without equating mechanism names to neurons.¹⁷ Hold out source families and participants or animals, not nearby morph anchors.

Scope of the integrated-mechanism claim

The shared set depends on the raw-unit response definition. Changing per-channel scale changes the selected mechanisms; shared acoustic inputs also make some outputs highly redundant. Full-profile readout gains may reflect useful non-linear feature construction or altered regularization geometry rather than independent information supplied by each layer. Demonstrating causal benefit of architecture connectivity would require a separately versioned internal-ablation study, matched-capacity alternatives and external targets. The present result supports integrated inspection and conditional complementary readout, while making that stronger test explicit.

One source phrase, one transcription and extreme-register sample resynthesis limit generalization. Forward or centered context, horizons exceeding the phrase, endpoint sensitivity, feature-index mismatches and readout ceilings remain part of the frozen model. All were retained rather than repaired selectively after observing results. Within this scope, the conclusions are concrete: reference geometry changes instrument ordering; common mechanisms coexist with route-specific responses; and a clinically named prediction

can be reduced to opposing, auditable spectrotemporal computations. The next advance is to test these precise contrasts in independent sounds and listeners.

References

1. Taylor, H. Whose Bird Is It? Messiaen's Transcriptions of Australian Songbirds. *Twentieth-Century Music* **11**, 63–100 (2014). doi:10.1017/S1478572213000194
2. 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
3. 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
4. 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
5. 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
6. Feather, J., Leclerc, G., Mađry, 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
7. Alvarsson, J. J., Wiens, S. & Nilsson, M. E. Stress recovery during exposure to nature sound and environmental noise. *Int. J. Environ. Res. Public Health* **7**, 1036–1046 (2010). doi:10.3390/ijerph7031036
8. Stobbe, E., Sundermann, J., Ascone, L. & Kühn, S. Birdsongs alleviate anxiety and paranoia in healthy participants. *Sci. Rep.* **12**, 16414 (2022). doi:10.1038/s41598-022-20841-0
9. Buxton, R. T., Pearson, A. L., Allou, C., Fristrup, K. & Wittemyer, G. A synthesis of health benefits of natural sounds and their distribution in national parks. *Proc. Natl Acad. Sci. USA* **118**, e2013097118 (2021). doi:10.1073/pnas.2013097118
10. You, S., Zhang, L., Wu, G. & Du, Y. The structural dynamics of music drive acute stress recovery through functional reorganization of stress-regulation networks. *Proc. Natl Acad. Sci. USA* **123**, e2611584123 (2026). doi:10.1073/pnas.2611584123
11. Plomp, R. & Levelt, W. J. M. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**, 548–560 (1965). doi:10.1121/1.1909741
12. 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
13. Arnal, L. H., Kleinschmidt, A., Spinelli, L., Giraud, A.-L. & Mégevand, P. The rough sound of salience enhances aversion through neural synchronisation. *Nat. Commun.* **10**, 3671 (2019). doi:10.1038/s41467-019-11626-7
14. Chen, S., Thielk, M. & Gentner, T. Q. Auditory Feature-based Perceptual Distance. *bioRxiv*, preprint (2024). doi:10.1101/2024.02.28.582631
15. 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
16. 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
17. 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
18. Roberts, D. R. *et al.* Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Eco-graphy* **40**, 913–929 (2017). doi:10.1111/ecog.02881

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. The full-layer, integration and TAR analyses were documented follow-ups on 16 October 2026 after the earlier response audit. Shared-set selection and TAR interpretation are exploratory; no public preregistration is claimed.

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; The revised primary response analysis uses all 889 declared channels; 202 P-labelled dimensions reproduce the earlier observation. Three reserved dimensions are constant. 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. Layer schemas are taken from the actual

registry. M denotes memory or temporal integration depending on the module; TAR uses E/TI/P/F, MSPBA S/M/P/F, DAED E/M/P and MEAMR E/P/F. Labels do not independently establish the named 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 declared dimensions (or only its P-labelled dimensions for the retained present-state figures); the second spans all sampled anchors. Full-profile global RMS pools all 889 declared dimensions; the retained present-state global RMS pools 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. Three-dimensional response views group the same six displayed mechanism RMS curves into two fixed triplets. Axes retain their original units, with identical limits across panels; no latent embedding or fitted surface is introduced. Larger markers denote endpoints, while the companion two-dimensional panels preserve exact coordinate comparisons.

Shared responses and route additions

All-layer sensitivity uses the same squared-change estimand with no learned per-channel scaling. A shared mechanism ranks in the top 21 of 84 on every specified path. The three sequential routes are bird→score, score→flute and score→piano; the five-route set also includes direct bird→flute and bird→piano. Robust intersection requires native/equal-RMS audio, full/trimmed frames and nested 97/49/25-anchor grids. This is a descriptive intersection, not a probability of replication. Layer energy equals dimension count times layer RMS squared; shares divide by the mechanism total. Endpoint-scaled sensitivity is a separate control. The highest-ranked mechanism outside the robust four is selected as one additional candidate per route; repeated selection of the same candidate is allowed. Full names, schemas, all ranks and non-selected mechanisms are retained.

Local intervals and endpoint scaling

For adjacent anchors, local change is the RMS of their tensor difference divided by Δu , distinct from RMS distance to the

path origin. The archive retains 92,736 local records, 4,080 peak-control records and 473 contiguous winner intervals. Coarse-grid peak overlap is descriptive and does not resolve an exact transition. Endpoint-based channel scaling uses the root-mean-square spread across the four endpoint tensors around their per-frame endpoint mean, excluding channels with scale $\leq 10^{-6}$; its purpose is sensitivity to output units, not clinical calibration. It removes the three-route common top-quartile set. No local maximum is called a human threshold.

Resolution and sensory attribution

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.

Regional projections and source weights

For region r , $RAM_r(t)=\max(0, \sum_l w_l Y_l(t))$, where l indexes incoming dimension links. These links can select from the full 889-dimensional pool; primary C^3 metrics use only the 202 P dimensions. The weight-sum control divides RAM_r by $W_r=\sum_l |w_l|$, using the same fixed denominator at both endpoints. Pearson correlations compare the 26 signed endpoint-change means. Regions share sources and are not independent observations. For the audited endpoints, all pre-ReLU sums are positive; signed differences therefore equal the sum of per-link weighted differences, up to numerical precision. Contributions are grouped by mechanism and ranked by absolute magnitude, with their signs retained. No physiological normalization, fitting or inferential test is performed.

Ordered channel updates

Nuclei execute in ascending processing depth, with stable registry order within a depth. The initial [DA, NE, OPI, 5HT] vector is [0, 0.5, 0, 0]. For source dimension x and weight w , produce sets $c \leftarrow wx$; amplify sets $c \leftarrow c+wx(1-c)$; inhibit sets $c \leftarrow c-wxc$. Channels are clipped to [0,1] after each nucleus. Terminal identities are $DA=0.65 \times ECT[f04:flexibility_index]$, $OPI=0.90 \times IOTMS[E0:mor_baseline_proxy]$ and $5HT=0.30 \times BCH[P0:consonance_signal]$. NE retains nine amplify links and one inhibit link; the last is $TAR[T2:anxiety_reduction]$, weight 0.45. The executor does not feed these channel outputs back into mechanism computation. All 52 valid links are retained in execution order in Supplementary Table S3.

Ceilings, temporal contrasts and replay

The IOTMS terminal source clips $0.55 \times pleasantness_mean_1s + 0.45 \times (1 - roughness_skew_1s)$ to [0,1]. A source ceiling of 1 produces $OPI=0.9$. Ceiling fractions count frames exactly equal to $float32(0.9)$, with equal weight across anchors. For NE, $\delta(t)=NE(1,t)-NE(0,t)$ compares the same frame indices; its signed mean and $\sqrt{\text{mean}[\delta^2]}$ quantify different properties of the same endpoint contrast. NumPy float32 replay applies the pinned update rules to all archived C^3 tensors without new audio inference. Supplementary Methods and Table S4 give exact replay tolerances and all-path readout summaries.

Integrity and temporal scope

The original manuscript audit 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 ($\rho=0.992-0.997$) but cannot recover missing long context. The companion analysis exposes every mechanism, dimension, region and proxy rather than only figure-selected leaders.

TAR layers and exact terminal audit

TAR has ten channels: E0 therapeutic; T0 arousal modulation target; T1 valence modulation target; T2 anxiety reduction; T3 depression improvement; I0 recommended normalized tempo; I1 recommended consonance; P0 therapeutic reward; F0 mood-improvement prediction; and F1 stress-reduction prediction. Names are literal schema descriptions. $F1=\text{clip}[0.20(E0 \times T2)+0.20P0+0.35(1-r)+0.25\{1-\text{clip}(|v|,0,1)\},0,1]$. The requested H^3 tuples are (feature 0, horizon 15, morphology 18, law 0) for r and (feature 8, horizon 15, morphology 8, law 0) for v . These select roughness/trend/memory and velocity_A/velocity/memory, respectively; each uses 138 frames, approximately 0.801088 s. velocity_A derives from amplitude changes, not beat rate. The first incomplete window is replaced by the first full-window value. Other TAR inputs may use centered context, so trimming does not establish a causal online forecast.

Signed attribution and reproducibility

The audit uses the unique union of eight path families at both input levels: 1,102 recordings and 466,146 frames. It excludes archive entries not on those paths, including factorial-only and gain-control cases. Two H^3 features are recomputed from archived R^3 tensors using the frozen engine; no audio or complete C^3 inference is repeated. Products and clipping are evaluated framewise, then averaged. Four pre-clipping terms and the residual

clip(sum)–sum form an exact additive identity. Signed end-point attribution agrees within 2.11×10^{-9} ; the per-frame F1 reconstruction maximum error is 1.1921×10^{-7} . Every TAR channel, term and curve is retained. Changes in $E0 \times T2$ must not be replaced by changes in the product of separate phrase means. TAR warmth comments refer to a feature index resolving to spectral smoothness; the original code and mismatch are preserved.

Blocked coordinate readers

Per-recording channel means, rather than an average of four layer scores, form the reader features. The target u ranges from 0 to 1. Five contiguous test blocks partition each 97-anchor path; four neighbouring anchors on each side are purged from training. Training data determine means and standard deviations. A channel is retained only when its training standard deviation exceeds 16 times float32 epsilon times $\max(1, \text{its training maximum absolute value})$. This respects the ordered design more than random anchor splitting, but does not create independent source samples or fully remove smooth-path dependence.¹⁸ Ridge minimizes squared error plus α times coefficient norm squared, with an unpenalized intercept. The primary α is 1; 0.1 and 10 are controls. Standardized features are divided either by the square root of active subset dimensionality or by that of the full active mechanism. All actual layer subsets are tested at $\alpha=1$; complete profiles and single layers are retained at all three penalties. Native and equal-RMS conditions are separate.

Reader comparisons and integration limits

Held-out predictions are assembled for all 97 anchors and scored by mean absolute error and R^2 . The empty reader predicts the training-fold mean. Gain equals the minimum single-layer MAE minus full-profile MAE. Selecting the minimum on these same held-out scores is descriptive and optimistic for the single-layer comparator; it is not nested model selection. The 0.01 gain criterion denotes one percent of the renderer coordinate range, not statistical significance. All 41,760 evaluations and 4,050,720 predictions are preserved. Saved predictions reproduce MAE within 7.21×10^{-8} ; 100 fold computations agree with independent sklearn fits within 1.9×10^{-14} . Channel exclusion affects the external reader only; internal MI dependencies remain intact. Some forecast channels are deterministic functions of earlier outputs, so additional dimensions can alter regularized readout geometry without adding independent information.

Separate cross-instrument transfer task

The earlier analysis trains on all anchors of one score–instrument path and tests on the other. Direction, z-scaling, denominator and regularization controls are retained separately from blocked within-path prediction. SDNPS's primary flute-trained/piano-tested MAE is

0.04172; reversing direction gives 0.2321. A standardization failure can make an acoustic baseline's error extremely large under variance shift; center-only acoustic scaling yields approximately 0.23 instead. Single-layer readers can also outperform the complete C^3 pool. These outcomes prevent an architecture-wide superiority claim. Predictions and all 70,690 earlier reader outcomes remain in the evidence package, alongside the newer within-path subset analysis.

Scientific graphics and availability

New figures derive from saved CSV/NPZ outputs as editable vectors. TAR's orthographic 3D view uses literal terminal contributions; the accompanying signed attribution includes clipping. All seven previous main figures and three supplementary figures are retained. The source package includes figure data, failed reader comparisons, protocols, manifests and a claim-to-data ledger.

Research record

Data and code availability

The versioned local research package includes editable manuscript sources, vector figures, original analysis tables, complete new stage/TAR tables, layer-reader scores and saved predictions, protocols and SHA-256 manifests. The frozen waveform and engine archive is linked by absolute path and source hashes; no public DOI for this revision is asserted. The published 15 October edition remains identifiable separately. Independent source replication and public archival release are future steps.

Funding

This research received no external funding.

Competing interests

The author develops Musical Intelligence. This manuscript reports an internal evaluation of that research prototype.

Materials and attribution

Bird audio: Oona Räisänen, public domain (2007). Flute: VSCO 2 Community Edition, Versilian Studios/Sam Gossner, CC0. Piano: Salamander Grand Piano v3, Alexander Holm, CC BY 3.0. Source links, licences and transformations are retained with the research package.

Research transparency

No human participants or experimental animals were tested. The researcher set the questions, interpretation and evidence boundaries; MI supplied frozen numerical outputs; OpenAI Codex assisted source tracing, reanalysis, drafting and visual production. This 16 October major revision integrates the documented layer, stage and TAR audits while preserving previous editions and all seven main figures. The manuscript has not been peer reviewed; no collaborator endorsement is implied.

Correspondence

Amaç Erdem · amace@bu.edu · ORCID 0009-0000-1711-0036.