Musical Intelligence: A Glass-Box Architecture for Computational Models of Music Listening

Amaç Erdem
Independent researcher
Correspondence: amace@bu.edu · ORCID 0009-0000-1711-0036 · Fulbright Scholar (2022–2024)

Abstract

Musical Intelligence is an executable architecture for developing and testing computational accounts of music listening. It exposes acoustic coordinates, temporal requests and named calculations for inspection and revision. Reconstruction recovered a selected output, and a targeted temporal intervention remained within its declared route. Human-observation comparisons showed consonance correspondence, fitted tension prediction close to an acoustic baseline and cross-corpus arousal associations. Across 1,146 renderings of one blackbird phrase, BCH, PSCL, IACM and TAR retained upper-quartile responses across five transformation paths. TAR predicted increased stress reduction from transcription to flute (+0.049) and decreased reduction to piano (−0.047); exact decomposition exposed opposing roughness and energy terms. In original synthetic speech and song, individual MI mechanisms exceeded a matched 15-descriptor acoustic baseline. RASN reached 100% (27/27 correct) in three-tag speech versus 77.8% (21/27); CSSL reached 96.3% (26/27) in three-tag song versus 70.4% (19/27); IACM reached 82.2% (37/45) in five-tag speech versus 57.8% (26/45). These are gains of 22.2, 25.9 and 24.4 percentage points under the same held-out-text readout. Production tags are generation instructions, not independently measured emotion. The architecture supplies explicit computational predictions for behavioral and neuroscientific tests; complete independent reconstruction currently requires controlled-access implementation records.

Resources: musicalintelligence.app · Live engine Engine code (request access) Results GitHub
Data and code: Zenodo v2.0.0 · 10.5281/zenodo.20457643 Archive · 10.5281/zenodo.19744623 OSF C8NE4

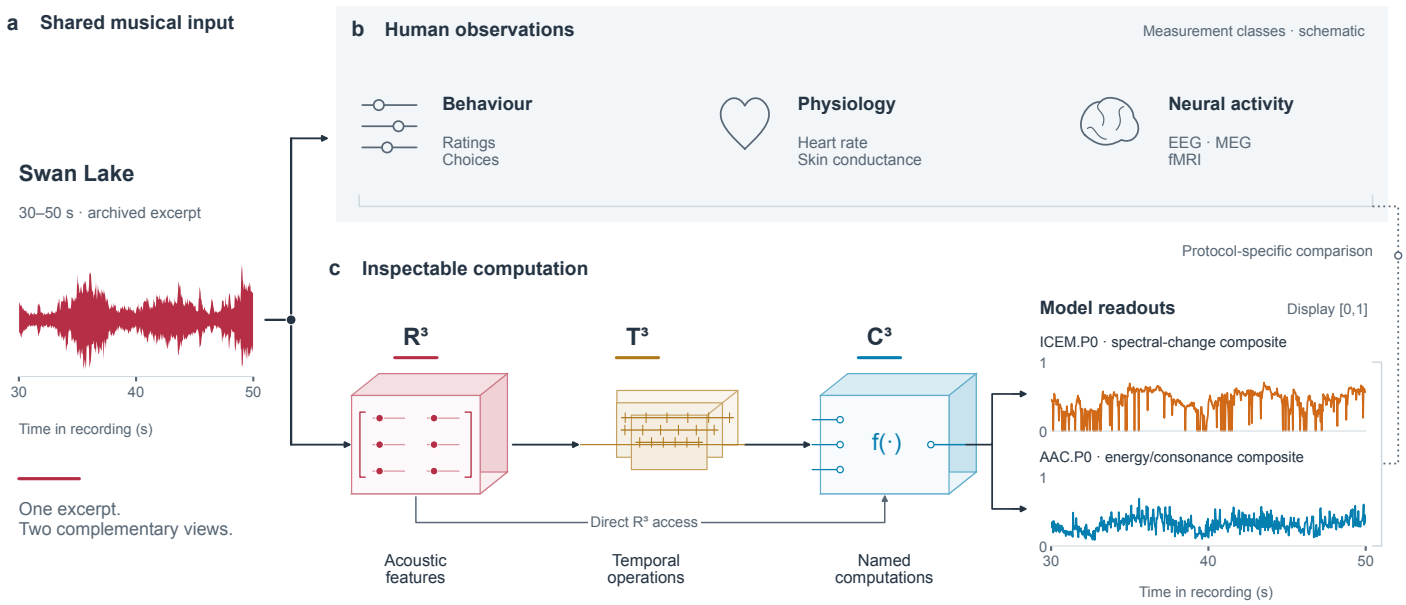


Figure 1. Music as a setting for inspectable computation. The Swan Lake excerpt supplies both paths. Observation symbols denote measurement classes, not collected participant data. R^3 features enter T^3 operations and C^3 calculations, including direct R^3 access. Depth is schematic. Curves retain their archived display scaling. The dotted link denotes a protocol-specific comparison; model outputs and human observations have separate status.

1. Introduction

Developing a model of music listening requires specifying which acoustic features, temporal operations and available states enter a calculation. Musical Intelligence addresses this model-building problem by making those choices accessible to inspection, intervention and comparison. Its scope includes what happens when a music-oriented representation encounters natural sound and speech.

Formal models expose theoretical assumptions and their consequences (Guest and Martin, 2021); cognitive architectures organise components and their interactions (Langley et al., 2009; Laird et al., 2017). Audio libraries provide reusable analysis algorithms (Bogdanov et al., 2013), while probabilistic models specify musical expectation (Pearce and Wiggins, 2012). MI brings named acoustic coordinates, explicit temporal requests and listening-related calculations into a shared execution.

R^3 describes sound; T^3 supplies requested temporal operations; C^3 combines acoustic, temporal and available upstream inputs. A fixed configuration gives deterministic outputs. A response pattern, its native summary and information recovered by a fitted readout are distinct objects of evaluation. Reconstruction and controlled changes identify how the implemented choices shape each one.

The evidence follows this organisation. Worked traces test local revision; human-observation comparisons assess selected readouts; birdsong transformations identify shared mechanisms and traceable directional affective predictions; and the comprehensive speech–song study identifies mechanisms that recover expressive production tags more accurately than matched acoustic descriptors. Together, they show how MI can expose the assumptions behind an output and turn them into controlled comparisons, while retaining the observation requirements of each interpretation.

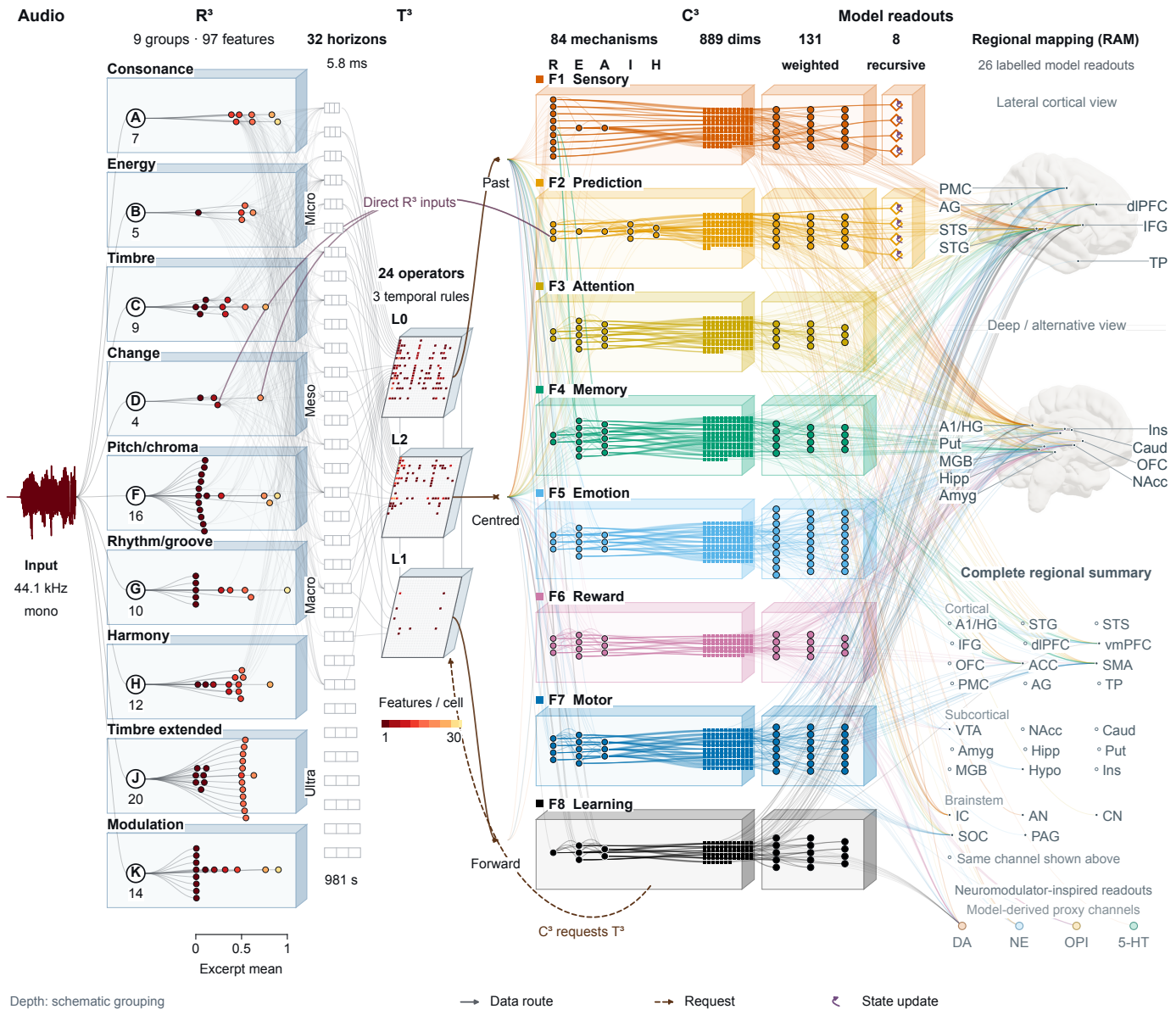


Figure 2. Musical Intelligence architecture. The current configuration contains 84 mechanisms, 889 output dimensions, 131 weighted readouts and eight selected recursive classes. Surfaces group computations; depth is schematic. Solid paths show data routes, dashed arrows requests and loops local state. Two anatomical views use representative registry coordinates; the complete regional summary is schematic. Group labels follow the implementation registry (Ins is anatomically cortical). Open markers repeat channels shown above, keeping 26 regional readouts. Anatomical and neuromodulator labels identify model projections with proposed interpretations; these paths are not measured neural connectivity or transmitter concentrations. R³ means and waveform retain the archived illustrative excerpt.

2. Architecture

The architecture separates acoustic description, temporal context and listening-related calculations, with declared inputs and an explicit execution order. The current configuration has 97 acoustic channels and 1,337 temporal declarations that reduce to 644 unique addresses. Its 84 mechanisms yield 889 output dimensions, alongside 131 weighted readouts and eight selected recursive classes. These counts describe one implemented repertoire within the shared organisation.

Configuration and observation

Evaluation requires both an engine configuration and an observation protocol. Configuration θ specifies implementations, equations, requests, initialization and execution order; protocol π specifies outputs, observations, aggregation and comparison. Their relationship is written as

$$\tilde{y}_\pi = O_\pi(F_\theta(a), c_\pi)$$

where a is audio, F_θ the engine, O_π the observation mapping and c_π a covariate. When O_π is fitted, prediction depends on engine and readout. Regression estimates the readout weights separately; an observation mapping need not forecast events.

A mechanism links a computational definition to a proposed interpretation. Its inputs and rule specify what it calculates; an observation protocol specifies how that calculation is evaluated. Depth roles determine processing order, while functional groups bring related proposals together for inspection. Anatomical and neuromodulator links project the outputs onto separately labelled channels with their own interpretation and measurement requirements.

The executor visits each mechanism once, in processing-depth order, retaining discovery order when depths coincide. A mechanism can use only the upstream outputs already available at that point; unavailable inputs invoke its implementation's fallback. A listed connection therefore affects the calculation only if its source is available and its value is consumed. Both conditions depend on the execution.

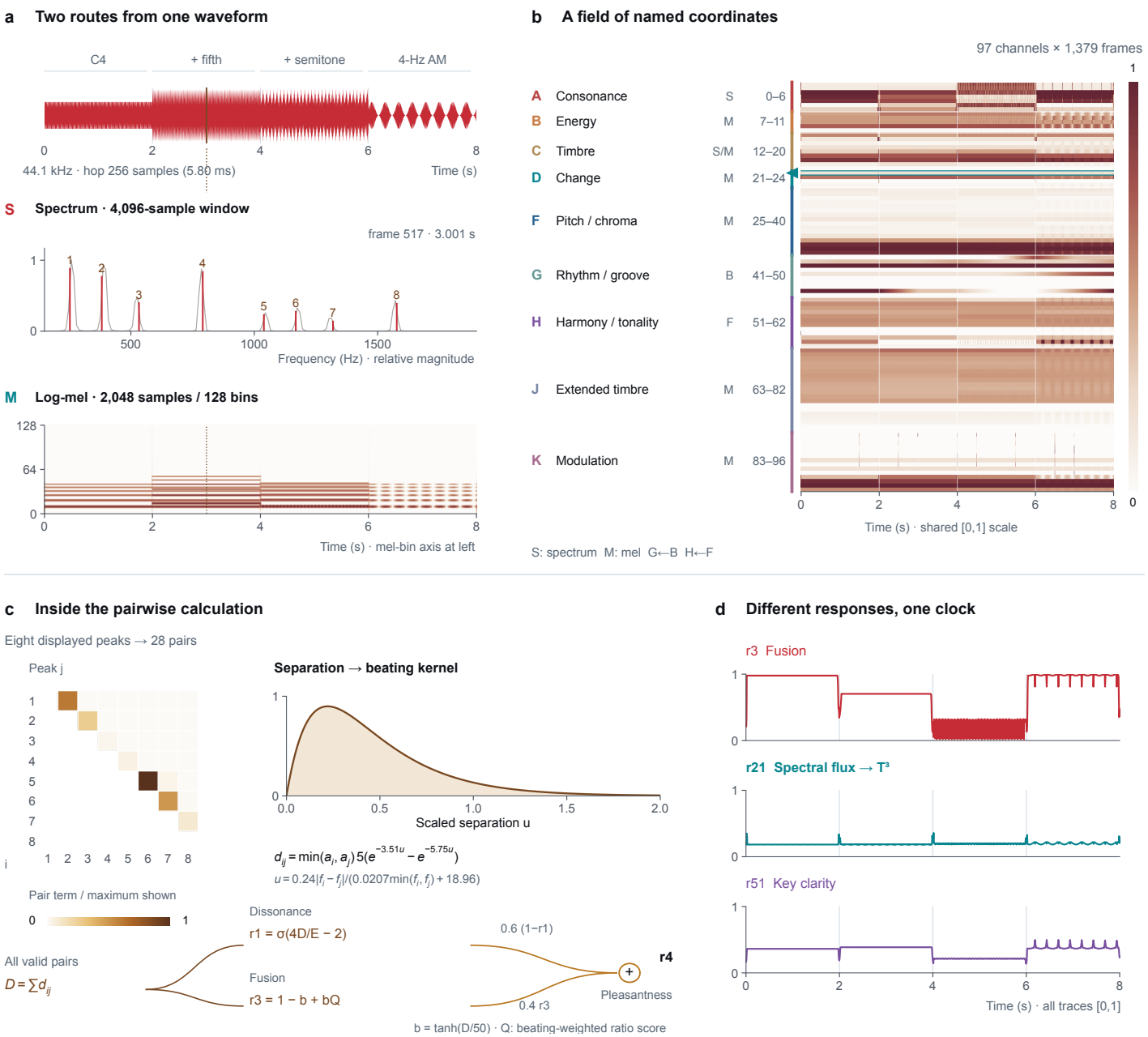


Figure 3. From sound to named acoustic coordinates. (a) The preserved 8-s waveform, saved spectrum at 3.001 s and log-mel field distinguish spectral (S) and mel (M) routes; window lengths differ from the 256-sample hop. (b) All 97 channels and 1,379 frames retain their [0,1] scale. S supplies A and C's warmth/sharpness; G derives from B and H from F. (c) The upper triangle displays 28 pair terms from eight numbered peaks, relative to their displayed maximum. The analytic kernel and computation diagram expose the full-pair pathways without estimating F0: raw beating D, energy-normalized dissonance $r1$, gated fusion $r3$ and their fixed pleasantness combination $r4$. E is the regularized root-sum-square of pairwise minimum amplitudes; Q is the beating-weighted, tiered ratio-eligibility score. The full computation uses all valid peaks, not this displayed subset. (d) Three unchanged feature series share one clock; spectral flux ($r = 21$) continues into Figure 4. Recording-wide normalization can extend context beyond the labelled windows.

Acoustic coordinates

R³ transforms each audio frame into 97 named channels across nine families, including consonance, energy, timbre, change, pitch, rhythm and harmony. The channels have no learned parameters or cross-family state. Published computations include Sethares roughness, Plomp–Levelt critical-band integration, Krumhansl–Kessler key profiles, Zwicker–Fastl sharpness and mel-frequency cepstral coefficients. Some channels share or derive inputs.

Each R³ coordinate also has a normalization rule and temporal support, ranging from an analysis frame to the full recording. These definitions accompany both the direct inputs to C³ and the series processed by T³.

The spectral consonance route operates on detected peaks without estimating a fundamental frequency. Peak separation and amplitude determine beating; ratio quality and a beating-dependent gate determine fusion. Pleasantness combines dissonance and fusion with fixed weights. The construction exposes each term and the algebraic link between inharmonicity and fusion. These coordinates cannot be interpreted as independent quantities. The route specifies how spectral relationships enter MI; it does not establish an advantage over pitch-based alternatives.

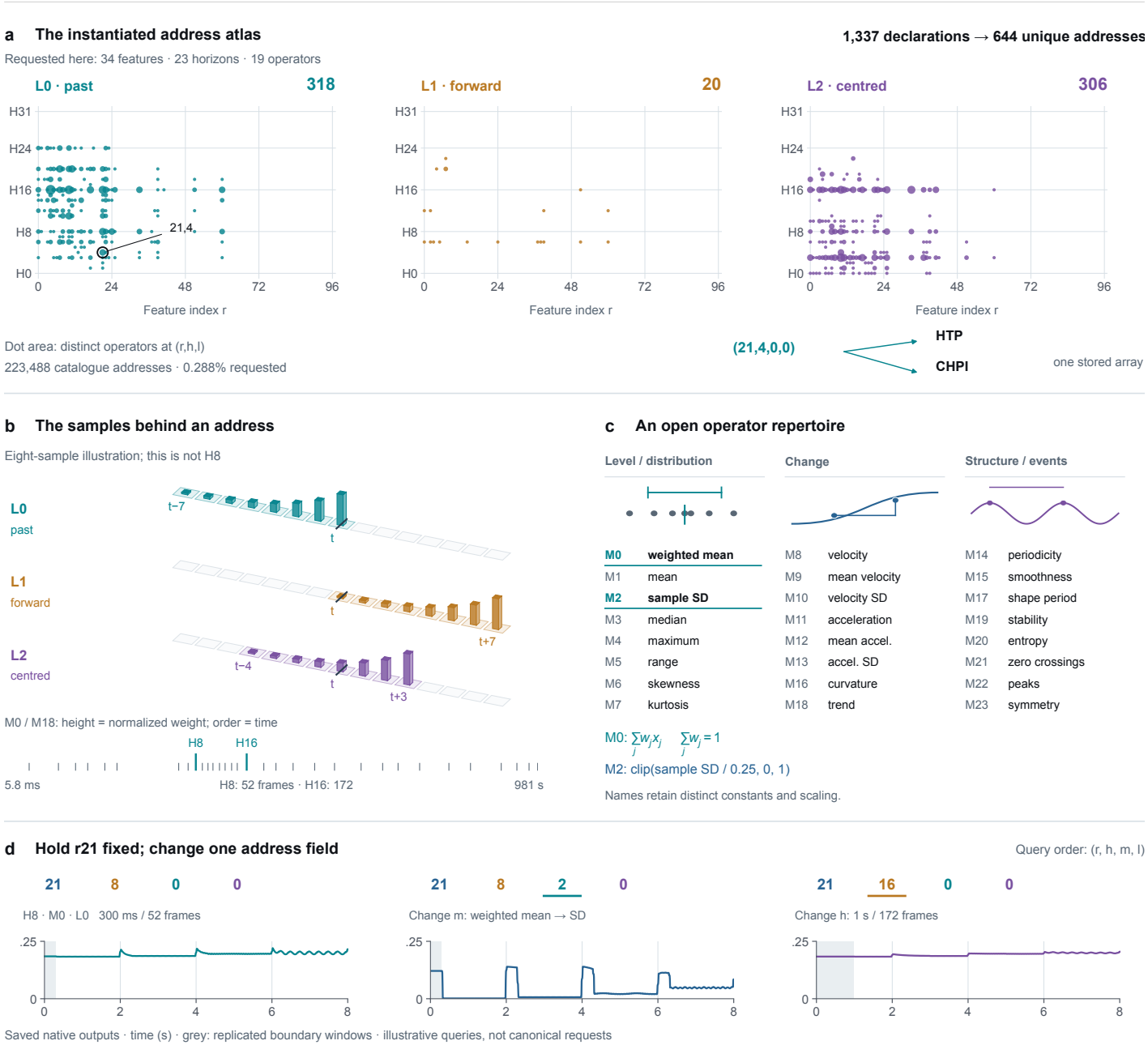


Figure 4. A temporal address opens into a support geometry and a specific operation. (a) The canonical 1,337 declarations reduce to 644 unique addresses. Maps preserve feature--horizon positions; dot area counts distinct operators at each position. Law totals count addresses. HTP and CHPI share $(21, 4, 0, 0)$. (b) Three-dimensional eight-sample teaching windows show interior placement, not H8. Heights encode normalized M0/M18 weights increasing chronologically; depth separates samples. Horizons are nonuniform, effective support is capped by recording length, and boundaries reuse the nearest complete window. R³ processing can already be recording-wide. (c) All 24 operators appear in editorial groups with schematic glyphs. M0 is a weighted mean; M2 divides sample standard deviation by 0.25 and clips to [0,1]. (d) Three saved spectral-flux queries retain native values on a common 0--0.25 display axis, changing one address field at a time. Grey spans mark replicated boundary outputs. These illustrative queries are supported but absent from canonical demands.

Requested temporal operations

T³ makes the temporal context of a calculation explicit. Each request names an R³ feature, horizon, operator and window orientation, with no learned parameters. Each mechanism declares its requests before extraction. Matching requests share one stored array, while both callers remain identifiable. The current configuration has 1,337 declarations and 644 unique addresses, drawn from 32 horizons and 24 operators. Horizons span milliseconds to the scale of a long recording. Requests identify both the computed quantity and its consumers, connecting temporal assumptions to the equations that use them.

Window orientation selects preceding, following or surrounding samples. Recording length and boundary handling determine the available support. Changing an address can alter either the samples included or the operation applied. Temporal direction must be assessed across the whole calculation: even a past-oriented T³ window can depend on later audio when its R³ inputs use normalization over the complete recording.

C³ · ICEM computation and readouts

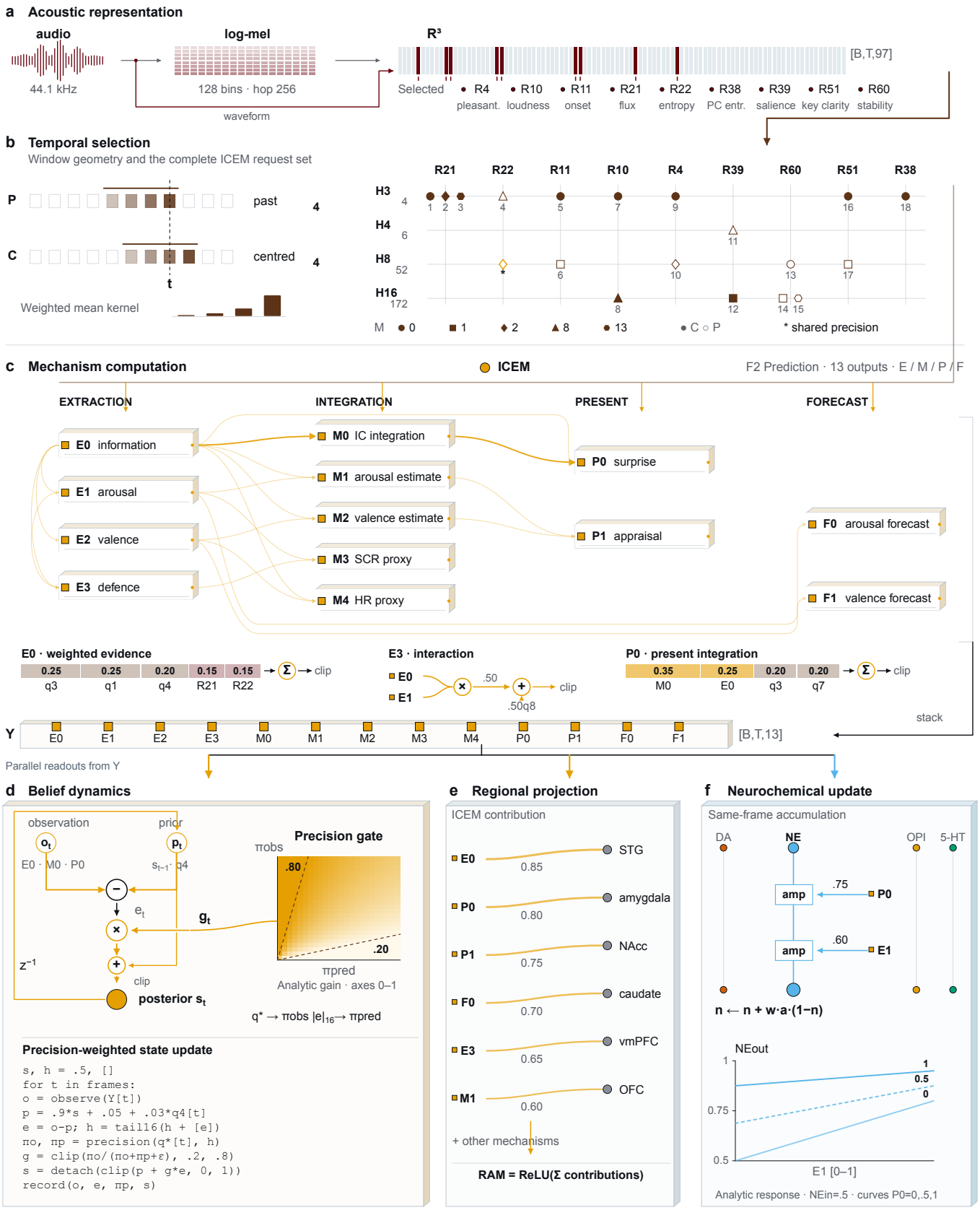


Figure 5. ICEM computation and parallel readouts. (a,b) Nine R³ channels and 18 temporal requests supply (c) 13 named outputs; insets expose weights and interactions. (d--f) Recursive belief, regional accumulation and ordered NE updates form parallel branches. The starred q^* precision input belongs to the source-audited all-mechanism context; availability and fallbacks are configuration-dependent (S1). Block depth and signal glyphs are schematic; gain and NE curves are analytic, with incoming NE $= 0.5$. Physiological labels denote model quantities; E0 is not information in bits and forecast names are not validated horizons. Supplement S1 provides the recorded trace, intervention and protocol.

Named mechanisms

C^3 has 84 mechanisms and 889 output dimensions. Each declares temporal requests, a depth role and a functional group. Depth orders execution from relay to hub. Sensory, prediction, attention, memory, emotion, reward, motor and learning groups organise related proposals. These categories organise the model; biological interpretations require independent evidence.

Mechanisms combine acoustic channels, temporal descriptors and available upstream outputs. Parallel belief-update branches read the resulting dimensions. The 131 weighted readouts and eight selected recursive classes form distinct mappings and updates. Following an output through its executed equation and inputs, including fallbacks, exposes how these choices determine its value.

For each named output, the implemented equation, inputs actually consumed and proposed cognitive interpretation form distinct parts of the specification. An audit can confirm what an output computes while leaving its interpretation unresolved. Listener history, task context or adaptation must enter the calculation or observation model explicitly when required by an interpretation. The acoustic assay below retains the 131 weighted readouts but does not execute the eight separate recursive classes (S8).

Regional and neuromodulator readouts

Regional activation mapping projects mechanism outputs onto 26 labelled regions through fixed links. Four further channels provide neuromodulator-inspired proxies. These mappings express biological hypotheses for comparison with measurements. They can be omitted from behavioural tests or evaluated through separate observation models. The regional links and proxy values are model definitions; independent neural or physiological data are required to assess their proposed biological meanings.

A worked mechanism

ICEM illustrates the acoustic, temporal and mechanism layers together. Nine of the 97 acoustic channels and eighteen temporal requests supply thirteen outputs, labelled from extraction through forecast. Observation, prior and precision terms update a belief, while regional and neuromodulator-proxy branches read the outputs in parallel. Gain and accumulation curves expose the equations' responses; physiological interpretations remain to be tested.

Two checks follow. The first reconstructs an output from saved inputs. The second changes one temporal request while keeping other calculations fixed. Effects in the recorded example stay within the declared route, showing how an individual modelling choice can be revised and its consequences traced.

3. Architectural evaluation

The Swan Lake example evaluates reconstruction and revision within an execution. Section 4 extends this approach to human-observation comparisons, controlled birdsong transformations and mechanism-resolved speech and song. These studies distinguish correspondence, represented information and contributions of stored output layers. Supplement S1 reports the recorded intervention and its protocol. Reconstruction records require controlled access.

An end-to-end trace

ICEM.E0 was reconstructed from three temporal descriptors and two direct acoustic quantities saved for the 182.597-second Swan Lake recording. Within the pre-selected 30–50-second interval, its value at 40.0022 seconds is 0.25011. Parallel surprise and superior-temporal readouts follow the same execution. Recovering E0 from these inputs establishes the weighted computation behind its unexpectedness label. E0 is a proxy, not information in bits; relating it to experienced surprise requires observations of that human response.

A targeted temporal intervention

Only ICEM's centred spectral-flux weighted-mean request is changed, from a 23.2-millisecond horizon of four frames to a 46.4-millisecond horizon of eight frames. Every other input, equation and mapping stays fixed. Over 30–50 seconds, the maximum absolute difference in E0 is 0.003632 and the root-mean-square difference is 0.000643. Seventeen of 889 dimensions exceed the recorded tolerance. All 805 dimensions outside the declared reachable graph remain exactly equal. Observer and identity-adapter controls reproduce the reference outputs exactly. The change is therefore local to the declared route, and the instrumentation is not the source of the difference.

Computational consequences

With the other terms fixed, non-expansive clipping and lower rectification give

$$|\Delta E_0| \leq 0.25|\Delta q|, \qquad |\Delta A_{\mathrm{STG}}| \leq 0.85|\Delta E_0|.$$

Of eight reachable mechanisms, only ICEM, PWUP and CHPI change; the illustrated consumers do not read the altered PWUP output. The bounds follow with other terms fixed. This traces local computational effects, without establishing whole-system stability or biological validity.

Computations before cognitive interpretations

Equation inspection identifies inputs relevant to testing each interpretation. DAP.P1 combines inverse roughness, tonalness, sensory pleasantness, inverse spectral entropy and an upstream composite, without developmental-history inputs. MMP.P2 combines past warmth stability, a rhythmic summary and centred loudness, without listener-history inputs. AAC.F1 combines acoustic intermediates with a 15-second forward amplitude context; its two-second heart-rate forecast label is unvalidated. Source-linked restatements reproduce these outputs in the worked trace. For AAC, the four-term restatement agrees with all 38 saved TenseMusic excerpts to a maximum absolute error of 3.46×10^{-8}, conditional on the cached reconstruction inputs.

These definitions lead to specific empirical comparisons. A composite can be tested against its acoustic constituents on the same human outcome. A proposed forecast can be restricted to past information. Familiarity or developmental accounts can incorporate measured listener history. With waveform, configuration, initial state and consumed inputs fixed, a calculation cannot distinguish listener attributes that never enter it. The architecture makes such dependencies explicit, identifying both what the present acoustic representation can test and where an additional input is needed.

4. Empirical case studies

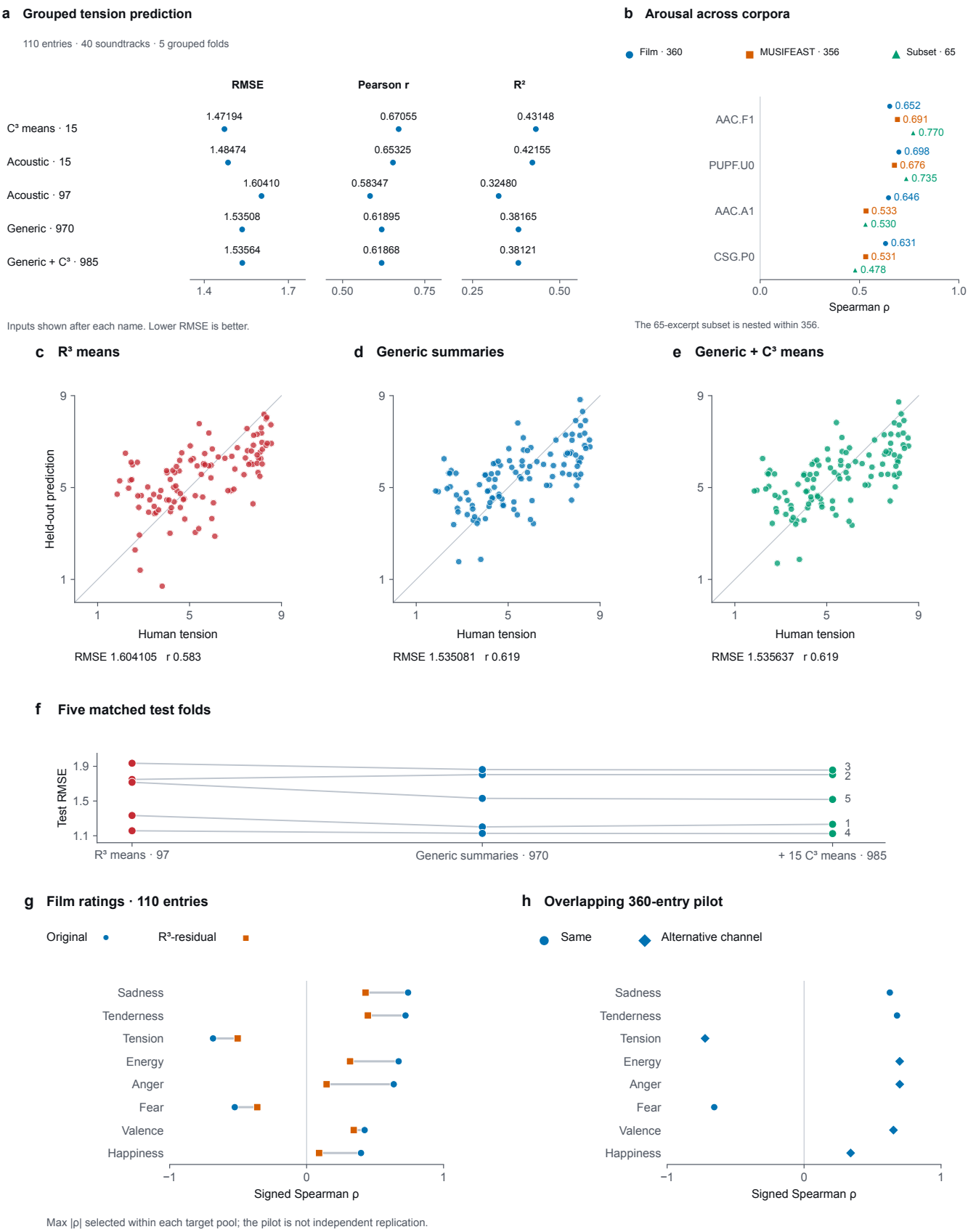


Figure 6. Excerpt-level prediction and cross-corpus arousal. (a) Five representations on 110 entries and five soundtrack-grouped folds; input counts follow each name. Acoustic 15 is the fixed comparator; generic 970 denotes acoustic/temporal descriptors, not canonical T³. RMSE is primary. (b) All four transferred arousal readouts: direct Spearman correlations in the film set, MUSIFEAST and its nested 65-excerpt subset; these are not three independent replications. (c–e) All 110 held-out tension predictions per representation, with identity diagonals and complete prediction ranges. (f) The same five folds; lines pair folds and lower RMSE is better. (g) Eight pool-leading film-emotion channels: original (circles) and R³-residual (squares) correlations. (h) Leading channels in the overlapping pilot: identical (circles) or alternative (diamonds) selections. Selection uses maximum absolute original association; shapes do not encode significance. Prior corpus exposure and exploratory selection limit generalisation. Supplements S2, S3 and S7 retain full tables and protocols.

Readouts, observations and controlled inputs

The case studies examine three connected questions: whether a readout corresponds to an observation, which distinctions are retained across its channels, and which implemented operations determine a response. Human-rating datasets address the first question; natural-sound transformations and speech instructions extend the second and third. A scalar score, a fitted multichannel prediction and a distance between response patterns are different outcomes. Keeping their observation rules explicit allows a positive result at one level to coexist with a limitation at another.

The comparisons retain their original units: interval and excerpt ratings, within-piece trajectories, chord rows and participant-level encoding. Birdsong samples transformations of one phrase; speech varies instructions across three texts and one synthetic voice. These sources are not pooled into a common accuracy or sample size. Figures 6–10 connect their complementary uses of the architecture.

Film tension

Fifteen C^3 means supported prediction of held-out soundtrack tension. On 110 excerpts, grouped ridge cross-validation gave Pearson $r = 0.671$, $R^2 = 0.431$ and RMSE 1.472; the fixed 15-descriptor acoustic subset gave $r = 0.653$ and RMSE 1.485. Generic 970-descriptor models gave $r = 0.619$. C^3 had lower error on three of five folds against either comparator, a descriptive ordering. Appending C^3 to generic descriptors did not reduce pooled error, leaving its incremental predictive value unresolved.

The compact tension result establishes that an explicit set of listening-related computations can support a fitted observation mapping. It does not require treating each component as an independently validated process. Conversely, closeness to the acoustic comparator does not establish equivalence: the paired fold estimates are descriptive, and the comparison was not designed as an equivalence test. The held-out predictions, their full range and all five folds remain visible in Figure 6. This separates the capacity to construct a useful mapping from a claim that the entire mechanism repertoire is necessary.

Excerpt emotion

Across 110 film excerpts, all eight emotion pools had Bonferroni-significant associations. Acoustic residualization changed leading correlations: sadness +0.741 to +0.430, tenderness +0.722 to +0.447, tension −0.683 to −0.504. Seven of eight leading readouts retained $|\rho| \geq 0.10$; every pool had one. This residual analysis does not test incremental prediction. Three channels recurred in the overlapping 360-excerpt pilot, documenting cross-set correspondence rather than independent replication.

The emotion analysis already treats affect as a collection of response coordinates. Pool-leading correlations select a channel for a specified target; they do not define eight mutually exclusive emotion probabilities. Residualization asks how much association remains after the specified acoustic adjustment, whereas prediction asks whether an added representation improves unseen outcomes. These operations need not agree. The speech–song study extends this distinction by measuring each mechanism, layer and channel separately: a complete-vector score can obscure strong distinctions available within a specific computation.

Arousal transfer

After the MUSIFEAST pilot, channels selected from film ratings were evaluated without test-subset reselection. AAC.F1 arousal correlations were +0.652 across 360 film excerpts, +0.691 across 356 MUSIFEAST excerpts and +0.770 in its nested 65-excerpt subset; PUPF.U0 gave +0.698, +0.676 and +0.735. These overlapping rank comparisons show cross-corpus correspondence. Valence associations weakened: SRP.P1 declined from +0.622 to +0.231, and CMAT.S0 from +0.653 to +0.172.

The transferred channels retain their identities, but the observation protocols and engine snapshots differ. Their agreement is therefore evidence about particular ranks under those protocols. The updated speech–song study examines all registered mechanisms and channels using the metadata of its own frozen engine. It does not select candidates from arousal-related names or import a historical affect pool as the definition of the available representation. Human-rating correspondence and production-tag agreement retain their separate targets.

From correspondence to experimental contrasts

Birdsong transformations identify four shared mechanisms across five routes and opposite TAR stress-reduction predictions for flute and piano from one transcription. Speech and song identify individual mechanisms that recover production tags more accurately than conventional acoustic descriptors under the same held-out-text procedure. Both assays make the observation mapping part of the question alongside the acoustic representation and temporal computation.

This approach connects MI to established auditory modelling methods. Sound-texture synthesis tests which statistics preserve recognizable structure (McDermott and Simoncelli, 2011); representational similarity analysis compares patterns of distinctions across models and observations (Kriegeskorte et al., 2008). MI adds explicit access to the selected coordinates, temporal requests and downstream equations. One can therefore compare a response pattern, reconstruct its readout and intervene on an input without changing the interpretation after each result. The current assays exercise these operations; perceptual comparison is a subsequent, separately specified test.

The temporal and perceptual comparisons that follow retain the same distinction. An excerpt average can correspond to a rating while losing within-phrase changes. A physiological label can organize a proposed comparison without itself supplying a physiological measurement. Figures 7 and 8 preserve both the positive associations and the conditions that limit their interpretation.

Compare the representation and the observation rule

The grouped film analysis illustrates why both parts of the comparison matter. Its target is a continuous human tension rating; the speech–song targets comprise three or five production tags. The former uses nested regression and soundtrack groups, while the latter evaluates external readouts with texts held out. Numerical improvements belong to their own tasks and denominators. What transfers is the method of comparison: freeze the representation, specify the target, keep related observations together and estimate the observation mapping only from the remaining groups.

Selected channels do not define the entire representation. The updated speech–song analysis enumerates the complete engine inventory before highlighting informative outputs. Layer exclusions then ask which components add information to the same external readout. The architecture itself remains fixed throughout. Its 889-channel comparator concatenates outputs; it is not an arithmetic average across mechanisms. This preserves the difference between a comprehensive response map and a separate procedure for automatically selecting a predictor.

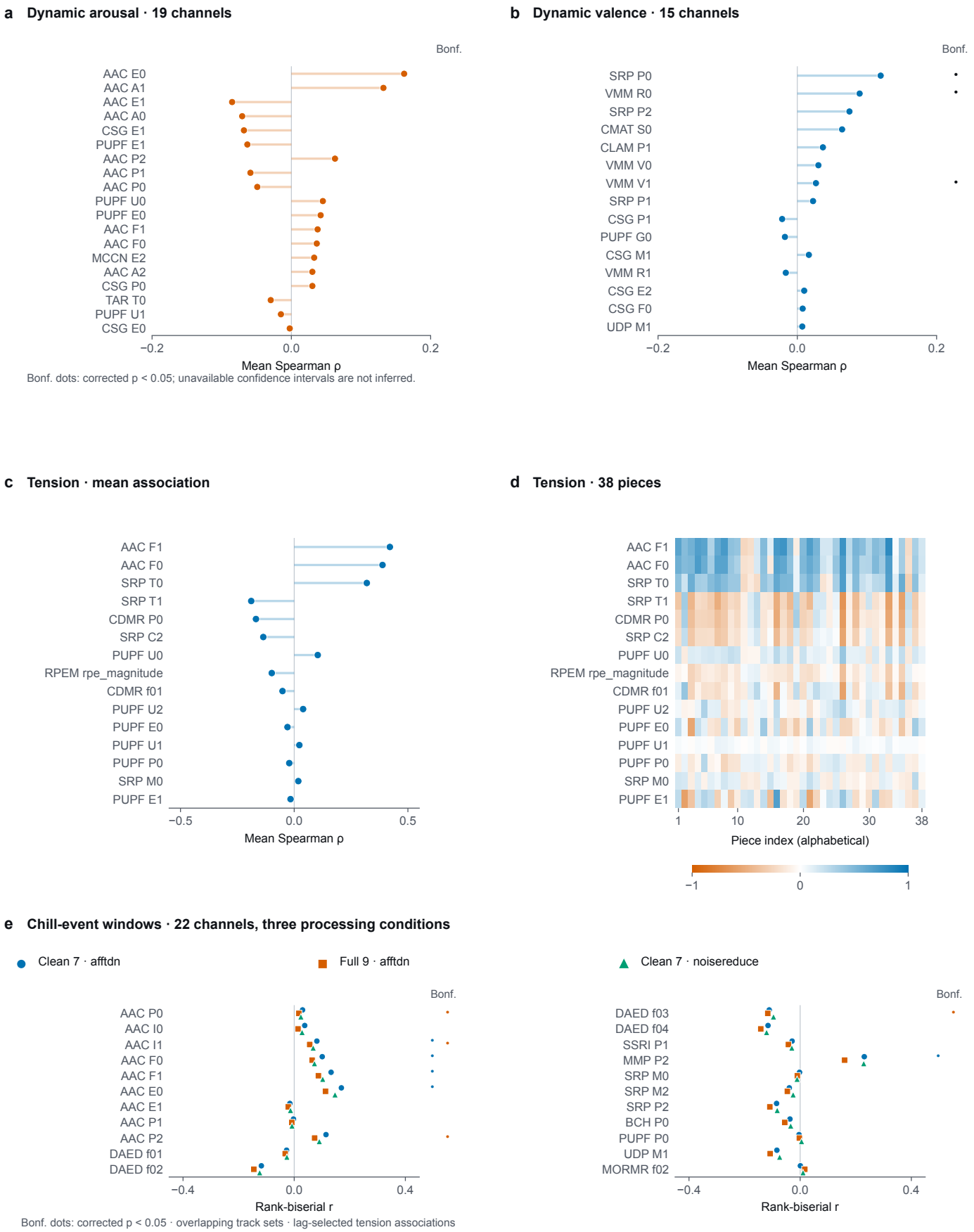
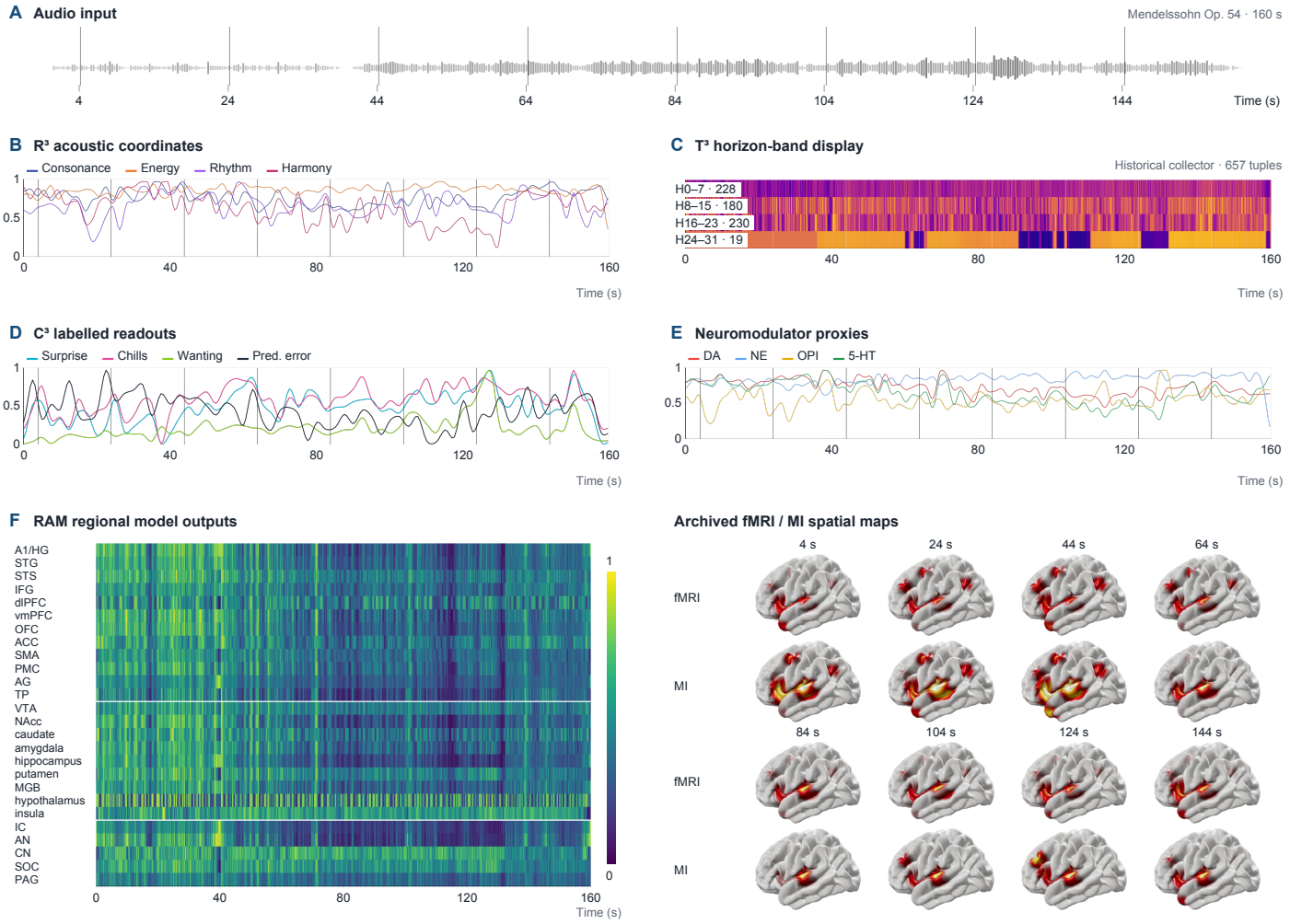


Figure 7. Time-resolved emotion, tension and reported chills. (a,b) All 19 arousal and 15 valence channels in the dynamic-emotion pilot, including negative and non-significant results. Points are saved Fisher-z means of lag-selected within-excerpt correlations. (c,d) All 15 mean tension associations and 570 channel–piece correlations across 38 alphabetically ordered pieces; lags were selected within ±5 s. (e) All 22 chill channels under three overlapping processing conditions. Marker shapes identify conditions; separate Bonf. columns mark each table's corrected $p < 0.05$. Tension and emotion correlations differ from chill rank-biserial effects; no common effect scale or participant-level uncertainty is implied. Archived values retain their timing, execution-identity and selection qualifications. These families address different targets from the pooled prediction in Figure 6. Supplement S7 and Source Data S4.2–S4.3 retain complete records; S6 separately reports all twelve recent-dataset comparisons, including small or negative increments.



G Acoustic channels across nine references

Reference / n

Anchors		Roughness	Sethares dissonance	Helmholtz–Kang	Stumpf fusion	Sensory pleasantness	Inharmonicity	Harmonic deviation
Harmonic dyads	13	-0.84	-0.81	+0.37	+0.81	+0.89	-0.81	+0.13
Flute rating	13	-0.77	-0.80	+0.26	+0.53	+0.77	-0.53	+0.45
Guitar rating	13	-0.53	-0.55	+0.23	+0.44	+0.60	-0.44	+0.30
Piano rating	13	-0.74	-0.58	+0.15	+0.62	+0.69	-0.62	+0.47
Pure tones	13	-0.73	-0.32	+0.17	+0.45	+0.36	-0.45	-0.52
Indian tension*	12	-0.52	-0.40	+0.36	+0.80	+0.58	-0.80	+0.41
Bidelman ratings	6	-0.89	-0.94	+0.09	+1.00	+1.00	-1.00	+0.03
Speech harmonics	13	-0.81	-0.77	+0.29	+0.92	+0.88	-0.92	+0.23
Sethares reference*	13	-0.81	-0.77	+0.29	+0.92	+0.88	-0.92	+0.23

Spearman ρ: −1 … 0 … 1

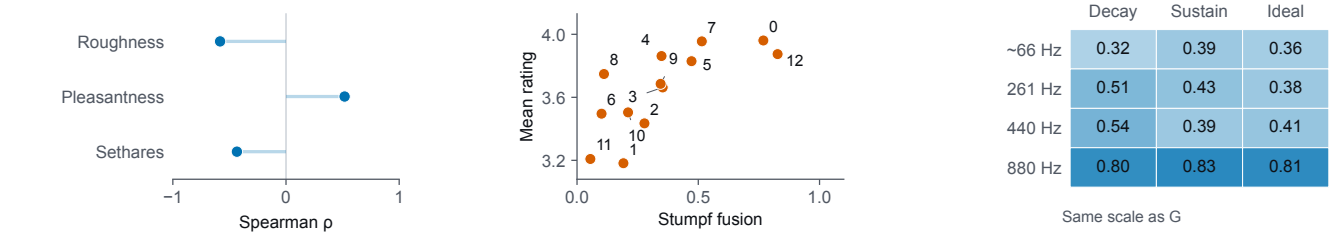


Figure 8. Acoustic representations and perceptual correspondence. (A–E) Archived 160-s Mendelssohn model displays retain source scaling; their 657-tuple collector differs from the current 644-address configuration. (F) RAM (left) and sixteen original brain views (right), pairing archived fMRI and MI projections at eight times; image-level validation remains unresolved. (G) All 63 signed correlations: seven rating references and two analytical anchors; n counts intervals/bins and * denotes target inversion. Instrument names denote rating conditions, not reconstructed recordings. (H) Three acoustic coordinates across 617 chords. (I) All 13 harmonic-dyad means, labelled in semitones. (J) All 12 Carillon register/synthesis effects. Scales and observed units remain distinct. Supplement S4 and Source Data S4.1 retain protocols and values.

Time-resolved and perceptual comparisons

Figures 7 and 8 retain the original experimental displays. Their panels address different observation scales, so their correlations, rank-biserial effects and encoding scores are reported under their own protocols. The aim is to locate correspondence and computational dependence rather than rank heterogeneous datasets on a single scale.

Within-piece tension

Across 38 pieces rated by 30 listeners, historical AAC.F1 analysis gave Fisher-z mean Spearman $\rho = +0.421$ with consensus tension, positive on 34 pieces. All 15 pool channels met the archived Bonferroni criterion. Physical clock alignment remains unresolved after normalized-time interpolation and lag selection. The listener-agreement value, +0.386, follows a different protocol and cannot rank performance. Reconstruction agreed across 38 cached excerpts to 3.46×10^{-8}, verifying arithmetic, with temporal alignment and autonomic forecasting untested.

Chills

Within ±5 s of chills, MMP.P2 (familiarity label) rose in all seven stimuli: rank-biserial $r = +0.231$, Bonferroni $p = 0.007$; its nine-stimulus and alternative seven-stimulus results were nonsignificant (S7).

These timing results motivate direct checks of what an aggregation retains. A lag-selected correlation describes association after alignment choices, while a native-frame contrast describes a change in the model itself. The birdsong TAR audit below provides a concrete example: opposing roughness-trend and energy-change terms nearly cancel in the phrase mean despite substantial framewise differences. It supplies a computational reason to preserve trajectories alongside excerpt averages, without retrospectively resolving the physical clock uncertainty in the older tension analysis.

Consonance

The fixed pleasantness score, $0.6(1 - \text{Sethares dissonance}) + 0.4(\text{Stumpf fusion})$, correlated with human ratings at Spearman $\rho = 0.890$ across 13 harmonic-dyad means and $\rho = 0.518$ across 617 chord rows. On the dyads, fusion reached +0.813 and roughness −0.835. Primary channels carried the expected sign on all nine references. Magnitude criteria were met in full on six and in part on three. Interval-subset ranking above 0.85 passed on four. The same coefficients, applied to synthesized bell spectra, kept fusion positive across twelve register and synthesis conditions, from +0.319 to +0.830, highest for sustained tones at 880 Hz.

The acoustic comparison demonstrates correspondence for specified spectral constructions and rating targets. It also supplies a benchmark for interpreting other outputs built from related cues. In the speech–song study, the SDED:P0 channel separates five song tags at 91.1% (41/45 correct) under a nearest-centroid readout. That channel result identifies a precise spectral computation for controlled follow-up; it does not substitute for a human dissonance or emotion rating. The existing consonance comparisons and the vocal result therefore support distinct, testable uses of the same inspectable architecture.

Pleasure

On matched song-grouped folds, mean held-out r was +0.457 for controls including human valence/arousal ratings, +0.462 with audio-native predictors and +0.543 with symbolic predictors. An additive-score regression reached +0.615 with preprocessing before splitting; the +0.565 ridge reference used different inputs, preventing a matched added-value claim. The proxy interaction was −0.158, with published −0.124 inside its interval; the audio-native interaction was −0.060, with a chord-row bootstrap interval excluding zero. Across rhythmic versions, HTP and ICEM had median $r = +0.993$ and +0.828. The last two values concern model stability; they do not measure pleasure in listeners.

Cross-corpus and template transfer

Across seven chord-rating corpora, consonance correlations averaged $\rho = +0.408$, positive on six; the harmonic-timbre readout reached +0.404 across 25 intervals. Agreement with seven Hindustani raga pitch-class templates averaged $\rho = +0.565$, all positive. The raga comparison assesses template agreement, not listener responses or classification accuracy.

Regional encoding

In 17 participants, 16 of 22 mechanism–region pairs met their BH criterion, versus 11 of 22 random controls; medians were +0.162 and +0.058. Separate BH families do not test this difference. Four-participant voxel encoding gave top-voxel means of +0.165 for MI and +0.138 for CLAP, using evaluation-selected voxels and encoder-specific folds that preclude a matched superiority claim. Direct-correlation shuffles passed in 4/4 participants for MI, 1/4 for routing-off and 0/4 for random controls; they do not validate ridge scores. The model-magnitude profiles correlated across audio collections at $r = 0.998$; assigned coordinates were within 10 mm of 28/31 published peaks; 52/56 tracks placed the caudate-labelled channel first when the preceding-peak rule was used. These model configuration and timing checks cannot establish biological mechanisms.

What the studies jointly establish

The architecture supports distinct forms of evidence: reconstruction identifies a computation; a controlled intervention tests its dependence on selected inputs; a fitted readout assesses recoverable information; and a comparison with human observations evaluates an interpretation. The completed studies instantiate all four operations, with different strengths and limits. The birdsong assay extends these checks to controlled sound trajectories; the vocal inventory identifies named mechanisms that outperform the specified acoustic baseline in original speech and song. They make these choices experimentally accessible rather than leaving them implicit in a single score.

The complete comparisons also retain failures to transfer. The dynamic-emotion pilot has no Bonferroni-significant arousal channel among the 19 shown in Figure 7; an excerpt-level association therefore does not establish a temporally aligned prediction. The additional MUSIFEAST arousal family gives $\Delta R^2 = -0.016$ beyond its specified controls, with an interval spanning zero (S6). These findings locate the limits of the observation mappings actually tested. They do not erase the positive excerpt associations, just as the latter cannot substitute for incremental prediction.

For anatomical outputs, the birdsong source audit adds a specific computational control to this reasoning. A region with many incoming weighted links can accumulate a larger raw change than a sparsely connected region. Dividing by its incoming weight sum changes the cross-region comparison while leaving that region's within-contrast sign intact. The model's regional pattern therefore depends on an explicit mapping in addition to its upstream representation. Human encoding analyses and model-weight audits answer different questions; reporting both makes the bridge between them testable. The following natural-sound and vocal plates extend this principle through shared mechanism profiles, traceable affective predictions and systematic comparisons of stored outputs.

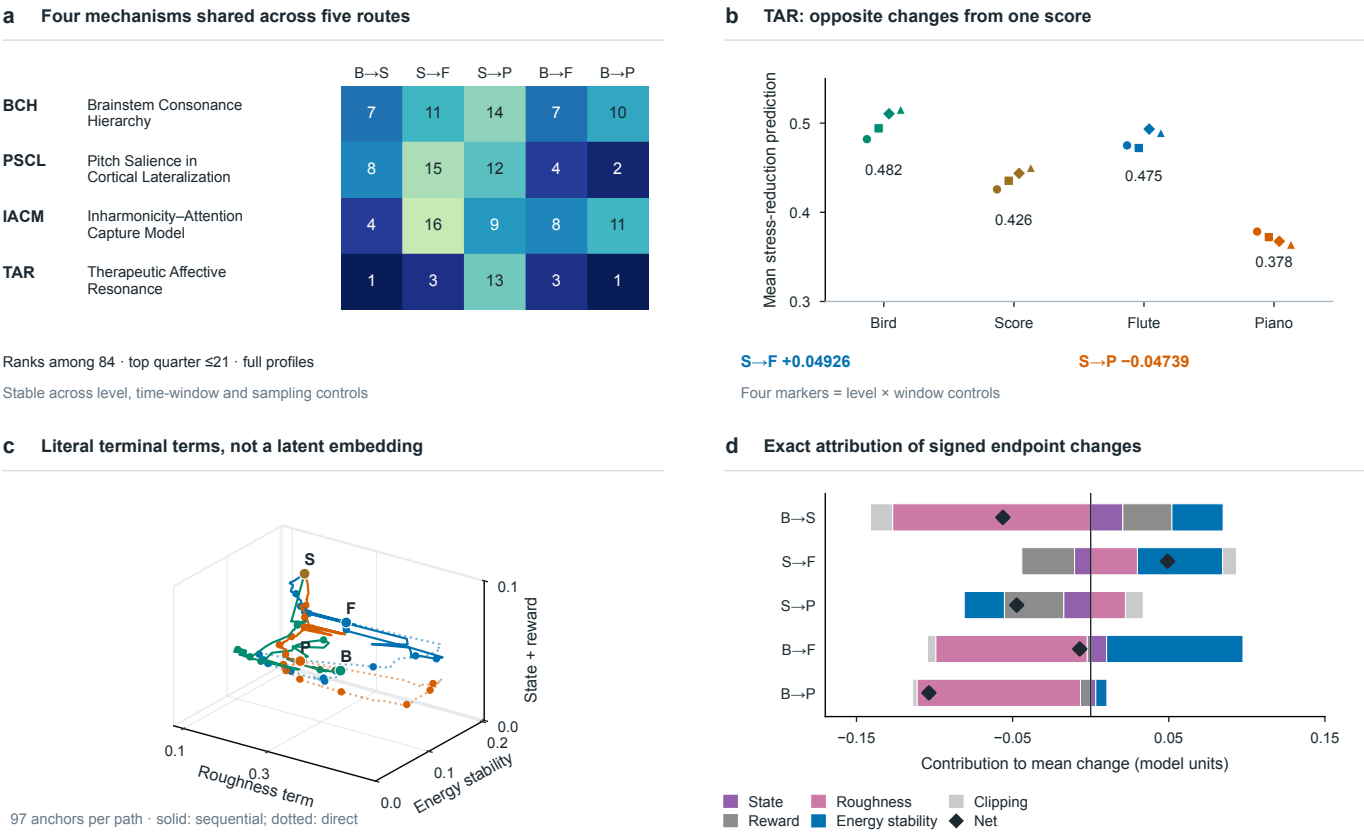


Figure 9. Shared mechanisms and traceable affective predictions in birdsong transformations. (a) Full-profile response ranks among 84 mechanisms, using all declared outputs in original model units. BCH, PSCL, IACM and TAR remain in the top 21 across five paths and level, window and sampling controls. (b) TAR stress-reduction predictions at four endpoints; circles, squares, diamonds and triangles denote equal-RMS/full, equal-RMS/trimmed, native/full and native/trimmed controls, respectively. Printed values use equal-RMS audio and the complete phrase. (c) All 97 anchors per path in three literal terminal-term coordinates; these are computed terms, not latent or perceptual dimensions. (d) Exact contributions to signed mean endpoint changes; diamonds show their sum. Opposing roughness and energy terms almost cancel for bird--flute. B, bird; S, synthetic score; F, flute; P, piano. Rankings depend on output scale; bounded TAR scores are model predictions, not measured stress or therapeutic efficacy (Erdem, 2026a).

Natural sound: shared mechanisms and directional predictions

The birdsong study processed one 2.45-s blackbird phrase through five paths linking the source, a synthetic pitch-and-onset transcription, and fixed sample-derived flute and piano endpoints (Erdem, 2026a). Operator, level and resolution controls yielded 1,146 complete-phrase renderings. Each waveform began from fresh model state. The analysis covers all 84 mechanisms and 889 dimensions alongside the earlier 202-dimensional present-state assay.

Reference and endpoint	Path RMS	End RMS
Transcription → flute	0.09816	0.12373
Transcription → piano	0.13615	0.17155
Bird → flute	0.06488	0.13306
Bird → piano	0.05172	0.09191

Table 1. Common-RMS audio; 202 equally weighted P-layer dimensions. Distances are in model units.

Four mechanisms remain in the upper response quartile (top 21 of 84) across all five paths: Brainstem Consonance Hierarchy (BCH), Pitch Salience in Cortical Lateralization (PSCL), Inharmonicity-Attention Capture Model (IACM) and Therapeutic Affective Resonance (TAR; Figure 9a). This set survives native/equal-RMS audio, full/trimmed windows and 97/49/25-anchor grids. Endpoint-based channel rescaling removes this raw-unit common set. Forecast outputs carry 54.7–70.3% of PSCL's squared response on the sequential routes, illustrating what present-only observation omits.

TAR supplies a directional hypothesis. Its mean stress-reduction prediction is 0.48203 for bird, 0.47503 for flute, 0.42577 for transcription and 0.37838 for piano. The transcription-to-flute change is +0.04926; transcription-to-piano is −0.04739 (Figure 9b). The endpoint order survives all four level/window conditions. Isolated timing lowers the output (−0.12100), whereas spectral replacement raises it (+0.13526). Interior paths are non-monotonic. These bounded scores propose a contrast for independent perceptual and stress-regulation tests; no human stress response was measured.

Exact terminal decomposition explains the prediction rather than assigning importance after fitting (Figure 9c,d). From bird to flute, the roughness-trend term changes by −0.09694 and energy-change stability by +0.08731. State, present-reward and clipping terms complete a net mean change of −0.00699, despite frame-wise RMS 0.35086. An audit of 466,146 frames reconstructs TAR's terminal output with maximum absolute error 1.19×10^{-7}. Similar averages can therefore arise from different, inspectable computations.

Joint outputs also have a selective readout benefit. SDNPS's complete profile reduces blocked bird-to-score coordinate MAE from 0.28546 to 0.16709, but loses to its best single layer on the instrument routes. No mechanism beats its best single layer across all three primary routes and every control. This is conditional complementarity, not universal superiority from adding outputs.

Reference-dependent geometry remains a complementary result (Table 1). Across all outputs, endpoint ordering reverses in 69 of 84 mechanisms and 274 of 326 layer groups. Earlier present-state rank and interpolation controls are retained in the companion.

They describe numerical sensitivity independently of the shared set and TAR's signed hypothesis.

Speech and song: gains over an acoustic baseline

Individual MI mechanisms recover speech and song production tags more accurately than the same readout of 15 acoustic descriptors (Figure 10). The baseline summarizes duration, level, pitch, spectral shape and envelope rhythm. Both representations use identical request units, held-out texts and fixed ridge readouts; MI itself is not fitted to tags. The companion inventories all 84 mechanisms, 326 layers and 889 channels across 342 original ElevenLabs recordings from 172 requests. Its matched core contains 90 requests and 179 alternatives across three texts, two delivery modes and five intended-production tags (Erdem, 2026b).

For three tags—excited, sad and untagged—RASN reaches 100% in speech (27/27 correct), compared with the acoustic baseline's 77.8% (21/27): six additional correct requests, or +22.2 percentage points. In song, CSSL reaches 96.3% (26/27) against 70.4% (19/27): seven additional correct requests, or +25.9 points. These comparisons directly identify the mechanism, its achieved performance and its gain over acoustics (Figure 10b).

With mad and peaceful added, IACM leads five-tag speech at 82.2% (37/45), compared with 57.8% (26/45) for acoustics: +24.4 points and 11 additional correct requests. MMP follows at 80.0% (36/45). In five-tag song, PWUP, CSSL and OII each reach 66.7% (30/45), versus 51.1% (23/45): +15.6 points and seven additional correct requests (Figure 10b). PWUP identifies excited, mad and sad at 100% each (9/9), but peaceful at 22.2% (2/9) and untagged at 11.1% (1/9). The four task-specific gains are observed differences, not a pooled emotion-recognition rate.

Which outputs support those gains?

Output exclusions then localize the information supporting these mechanism-level results. In CSSL and IACM, the four output groups are called Extraction, Memory, Present and Forecast (E, M, P and F). CSSL's complete output achieves 96.3% (26/27) in three-tag song. Removing those groups in turn reduces agreement to 70.4% (19/27), 81.5% (22/27), 88.9% (24/27) and 74.1% (20/27); every proper subset falls below the full output. In five-tag speech, adding IACM's Present outputs raises 64.4% (29/45) to 82.2% (37/45), although that group alone reaches only 40.0% (18/45). These tests explain a mechanism's success after its acoustic-baseline advantage has been established.

The contribution is selective. RASN retains 100% (27/27) without Extraction outputs; PWUP's Memory+Present subset reaches 75.6% (34/45) in five-tag song, above its complete 66.7% (30/45). Across the inventory, 1,304 output-group exclusions and 1,152 detailed subset/readout combinations document these effects. They remove stored outputs and refit the external readout, without disabling upstream computation. Full-mechanism scaling controls preserve the principal IACM and CSSL losses; the companion supplies complete tables.

A separate nearest-centroid control recovers five song tags from SDED:P0 alone at 91.1% (41/45). It uses a different reader and is not included in Figure 10's matched ridge comparisons. Candidate-wide permutations support tag-related structure under their stated assumptions; they do not establish significance for every mechanism–baseline difference. The highlighted mechanisms are discoveries from the exhaustive exploratory map. One synthetic voice and three texts limit generalization. Within this corpus and protocol, however, the result is concrete: specific MI computations improve discrimination over the specified acoustic descriptors.

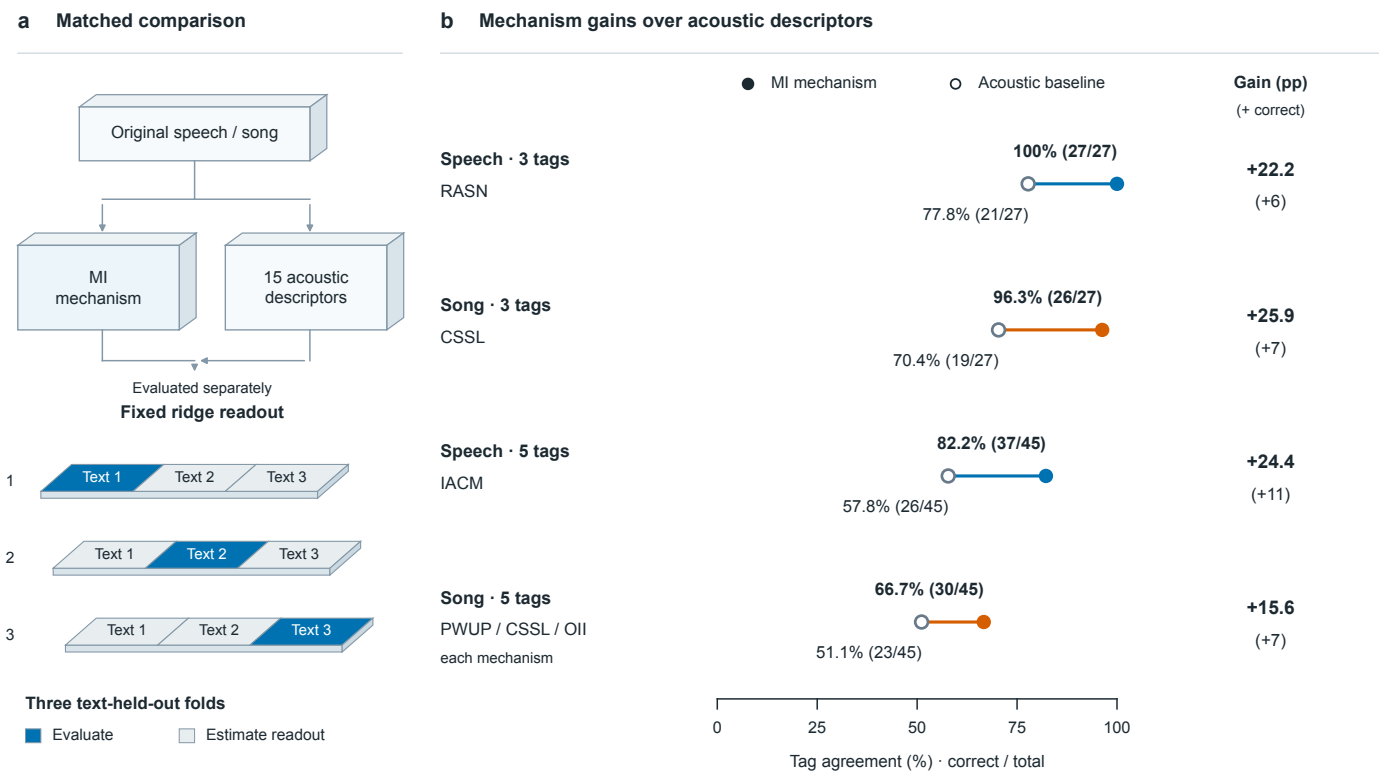


Figure 10. Mechanism gains over a matched acoustic baseline. (a) Both representations use identical waveforms and the same fixed ridge procedure. Three stacked planes show the text-held-out folds; the marked cell is evaluated and the other two estimate the readout. (b) Each row reports production-tag agreement as percentage and correct/total; right-hand labels give percentage-point gains and additional correct requests. PWUP, CSSL and OII each achieve the displayed five-tag song result. Three tags are excited, sad and untagged; five tags add mad and peaceful. These are observed results from the full 84-mechanism inventory, not listener-measured emotion (Erdem, 2026b).

Figure 11. Literature origin and mechanism evidence

The repertoire grew from the Cognitive Consonance Circuit literature-synthesis and exploratory meta-analysis programme. Published findings, extracted claims and theoretical proposals provided starting points for the design. Turning those ideas into computations required further choices about inputs, coefficients and temporal support. Figure 11 and Supplement S5 make that translation inspectable by connecting mechanisms to literature anchors, implemented operations and evaluation roles. Worked examples follow a proposal through its source observation to a specified test. The original quantitative synthesis motivates hypotheses; it supplies neither validated equation coefficients nor performance estimates. Its extraction and dependence limitations remain documented in S5. Unresolved translations identify hypotheses for further adjudication.

Stable mechanism names link the implementation to its design history. Supplement S5b distinguishes topic motivation, component-level background and unsupported extensions of a label. The source-role review examined primary abstracts and accessible full texts, recording access limits and preserving historical anchors. This separates what a source establishes from what a computation proposes. Brainstem, synaptic, memory and transfer interpretations require corresponding measurements; the stored falsification criteria describe tests still to be performed.

F1 Sensory	BCH	F1 H	Brainstem Consonance Hierarchy; Bidelman 2009 · Bidelman 2013
	CSG	F1 A	Consonance-Salience Gradient; Bravo 2017 · Sarasso 2019
	MIAA	F1 U	Musical Imagery Auditory Activation; Kraemer 2005 · Halpern 2004
	MPG	F1 U	Melodic Processing Gradient; Rupp 2022 · Patterson 2002
	PNH	F1 U	Pythagorean Neural Hierarchy; Bidelman 2009 · Tabas 2019
	SDNPS	F1 U	Stimulus-Dependent Neural Pitch Salience; Cousineau 2015 · Penagos 2004
	SDED	F1 U	Sensory Dissonance Early Detection; Crespo-Bojorque 2018 · Fishman 2001
	TPIO	F1 U	Timbre Perception-Imagery Overlap; Halpern 2004 · Zatorre 2005
	TPRD	F1 U	Tonotopy-Pitch Representation Dissociation; Allen 2022 · Norman-Haignere 2013
	PSCL	F1 U	Pitch Salience in Cortical Lateralization; Penagos 2004 · Allen 2022
	PCCR	F1 U	Pitch Chroma Cortical Representation; Briley 2013 · Allen 2022
F2 Prediction	HTP	F2 H	Hierarchical Temporal Prediction; Norman-Haignere 2022 · Bonetti 2024
	SPH	F2 U	Spatiotemporal Prediction Hierarchy; Bonetti 2024 · Fernandez-Rubio 2022
	ICEM	F2 CH	Information Content Emotion Model; Egermann 2013 · Cheung 2019
	PWUP	F2 C	Precision-Weighted Uncertainty Processing; Mencke 2019 · Cheung 2019
	WMED	F2 U	Working Memory-Entrainment Dissociation; Noboa 2025 · Yuan 2025
	UDP	F2 U	Uncertainty-Driven Pleasure; Mencke 2019 · Gold 2019b
	CHPI	F2 C	Cross-Modal Harmonic Predictive Integration; Moller 2021 · Cheung 2019
	IGFE	F2 U	Individual Gamma Frequency Enhancement; Yokota 2025 · Bolland 2025
	MAA	F2 U	Multifactorial Atonal Appreciation; Mencke 2019 · Gold 2019b
	PSH	F2 U	Prediction Silencing Hypothesis; Millidge 2022 · Auksztulewicz 2016
	SNEM	F3 U	Sensory Novelty and Expectation Model; Nozaradan 2011 · Nozaradan 2012
F3 Attention	IACM	F3 U	Inharmonicity-Attention Capture Model; Basinski 2025 · Koelsch 1999
	BARM	F3 U	Brainstem Auditory Response Modulation; Rathcke 2024 · Grahn 2007
	STANM	F3 U	Spectrotemporal Attention Network Model; Haiduk 2024 · Zatorre 2022
	AACM	F3 U	Aesthetic-Attention Coupling Model; Sarasso 2019 · Sarasso 2021
	AMSS	F3 U	Attention-Modulated Stream Segregation; Hausfeld 2021 · Wikman 2025
	ETAM	F3 U	Entrainment, Tempo and Attention Modulation; Hausfeld 2021 · Doelling 2015
F5 Emotion	DGTP	F3 U	Domain-General Temporal Processing; Rathcke 2024 · Grahn 2007
	SDL	F3 U	Salience-Dependent Lateralization; Haiduk 2024 · Zatorre 2022
	NEWMD	F3 U	Neural Entrainment-Working Memory Dissociation; Noboa 2025
	PWSM	F3 U	Precision-Weighted Salience Model; Basinski 2025 · Millidge 2022
F4 Memory	MEAMN	F4 U	Music-Evoked Autobiographical Memory Network; Janata 2009 · Sakakibara 2025
	MMP	F4 CH	Musical Mnemonic Preservation; Jacobsen 2015 · Fang 2017
	RASN	F4 U	Rhythmic Auditory Stimulation Neuroplasticity; Thaut 2015 · Grahn 2007
	PMIM	F4 U	Predictive Memory Integration Model; Koelsch 2000 · Koelsch 2009
	OII	F4 U	Oscillatory Intelligence Integration; Bruzzone 2022 · Samiee 2022
	HCMC	F4 U	Hippocampal-Cortical Memory Circuit; Squire 1995 · McClelland 1995
	RIRI	F4 U	RAS-Intelligent Rehabilitation Integration; Thaut 2015 · Harrison 2025
	MSPBA	F4 U	Musical Syntax Processing in Broca's Area; Maess 2001 · Kim 2021
	VRIAP	F4 U	VR-Integrated Analgesia Paradigm; Deng 2025
	CMAPCC	F4 U	Cross-Modal Action-Perception Common Code; Lahav 2007 · Bangert 2006
	DMMS	F4 U	Developmental Music Memory Scaffold; Partanen 2022 · Qiu 2025
	CSSL	F4 U	Cross-Species Song Learning; Burchardt 2025
	CDEM	F4 U	Context-Dependent Emotional Memory; Sachs 2025
F5 Emotion	SRP	F5 PA	Striatal Reward Prediction; Salimpoor 2011 · Salimpoor 2013
	AAC	F5 CPA	Autonomic-Arousal Circuit; Egermann 2013 · Salimpoor 2011
	VMM	F5 A	Valence-Mode Mapping; Mitterschiffthaler 2007 · Koelsch 2006
	PUPF	F5 PA	Pleasure-Uncertainty Prediction Function; Cheung 2019 · Egermann 2013
	CLAM	F5 U	Closed-Loop Affective Modulation; Ehrlich 2019 · Daly 2016
	MAD	F5 U	Musical Anhedonia Disconnection; Loui 2017 · Martinez-Molina 2016
	NEMAC	F5 U	Nostalgia-Evoked Memory-Affect Circuit; Sakakibara 2025 · Barrett 2010
	STAI	F5 U	Spectral-Temporal Aesthetic Integration; Kim 2019 · Alluri 2012
	DAP	F5 C	Developmental Affective Plasticity; Scholkmann 2024 · Trainor 2012
	CMAT	F5 A	Cross-Modal Affective Transfer; Spence 2011
	TAR	F5 U	Therapeutic Affective Resonance; Kheirkhah 2025 · Ehrlich 2019
F6 Reward	DAED	F6 H	Dopamine Anticipation-Experience Dissociation; Salimpoor 2011 · Gold 2023
	MORMR	F6 H	mu-Opioid Receptor Music Reward; Putkinen 2025 · Salimpoor 2011
	RPEM	F6 P	Reward Prediction Error in Music; Gold 2019a · Gold 2023
	IUCP	F6 U	Inverted-U Complexity Preference; Gold 2019b · Gold 2023
	MCCN	F6 U	Musical Chills Cortical Network; Chabin 2020 · Putkinen 2025
	MEAMR	F6 U	Music-Evoked Autobiographical Memory Reward; Janata 2009 · Salimpoor 2011
	SSRI	F6 U	Social Synchrony Reward Integration; Kokal 2011 · Bigand 2025
	LDAC	F6 U	Liking-Dependent Auditory Cortex; Gold 2023 · Cheung 2019
	IOTMS	F6 U	Individual Opioid Tone Music Sensitivity; Putkinen 2025 · Salimpoor 2011
	SSPS	F6 U	Saddle-Shaped Preference Surface; Cheung 2019 · Gold 2019b
F7 Motor	PEOM	F7 U	Period Entrainment Optimization Model; Thaut 2015 · Thaut 1998
	MSR	F7 U	Musician Sensorimotor Reorganization; Bangert 2006 · Lahav 2007
	GSSM	F7 U	Gait-Synchronized Stimulation Model; Yamashita 2025
	HMCE	F7 U	Hierarchical Musical Context Encoding; Mischler 2025 · Norman-Haignere 2022
	HGSIC	F7 U	Hierarchical Groove State Integration Circuit; Potes 2012 · Grahn 2007
	ASAP	F7 U	Action Simulation for Auditory Prediction; Ross 2022 · Patel 2014
	DDSMI	F7 U	Dyadic Dance Social Motor Integration; Bigand 2025 · Kohler 2025
	VRMSME	F7 U	VR Music Stimulation Motor Enhancement; Liang 2025
	SPMC	F7 U	SMA-Premotor-M1 Motor Circuit; Grahn 2007 · Hoddinott 2024
	NSCP	F7 U	Neural Synchrony Commercial Prediction; Leeuwis 2021
	CTBB	F7 U	Cerebellar Theta-Burst Balance; Sansare 2025 · Okada 2022
	STC	F7 U	Singing Training Connectivity; Zamorano 2023 · Kleber 2013
F8 Learning	EDNR	F8 U	Expertise-Dependent Network Reorganization; Paraskevopoulos 2022 · Leipold 2021
	TSCP	F8 U	Temporal Synaptic Consolidation Processor; Pantev 2001 · Bellmann 2024
	CDMR	F8 P	Context-Dependent Mismatch Response; Rupp 2022 · Crespo-Bojorque 2018
	SLEE	F8 U	Synaptic Long-term Encoding Engine; Paraskevopoulos 2022 · Porfyri 2025
	ESME	F8 U	Error-Signal Modulated Encoding; Koelsch 1999 · Vuust 2012
	ECT	F8 U	Expertise Compartmentalization Trade-off; Paraskevopoulos 2022 · Porfyri 2025

	Mechanism / full name		Literature motivation	Implemented proposal	Current evidence	Discriminating test
F2 Prediction	**ICEM** Information Content Emotion Model	F2	Egermann et al. (2013) Cheung et al. (2019) Expectation violations and uncertainty relate to musical responses.	Acoustic context supplies an unexpectedness-related score; this translation is a model choice.	A recorded calculation audit and temporal intervention (E1; S1).	Expectancy measures and matched acoustic-change alternatives.
F5 Emotion	**AAC** Autonomic-Arousal Circuit	F5	Egermann et al. (2013) Salimpoor et al. (2011) Expectation and pleasurable music evoke psychophysiological arousal responses.	A dimensionless audio-derived score is proposed as an arousal-related readout.	Arousal: F1 ρ = 0.691; A1 ρ = 0.533 (356 excerpts).	Synchronized physiology and acoustic controls for an autonomic claim.
F5 Emotion	**PUPF** Pleasure-Uncertainty Prediction Function	F5	Cheung et al. (2019) Egermann et al. (2013) Uncertainty, surprise and expectation motivate an affective proposal.	An audio-derived uncertainty-labelled quantity is compared with human arousal.	Arousal: U0 ρ = 0.676 (356 excerpts). This is not a pleasure test.	Pleasure-specific targets and alternatives for the uncertainty interpretation.
F1 Sensory	**CSG** Consonance-Salience Gradient	F1	Bravo et al. (2017) Sarasso et al. (2019) Interval acoustics relate to affective appraisal and attention.	Consonance cues supply salience- and affect-labelled model outputs.	Arousal: P0 ρ = 0.531 (356 excerpts). Salience itself is not measured.	Attention/salience targets; separation from shared acoustic cues.
F4 Memory	**MMP** Musical Mnemonic Preservation	F4	Jacobsen et al. (2015) Fang et al. (2017) Preserved musical memory motivates the proposal; Fang is a review.	The inspected familiarity-labelled path is audio-derived, without individual history.	Inspected pathway and historical comparison (S5/S7); recognition not measured.	Listener history or exposure inputs, and recognition or recall observations.
F6 Reward	**IUCP** Inverted-U Complexity Preference	F6	Gold et al. (2019b) Predictability and uncertainty relate non-monotonically to liking.	A non-monotonic response is proposed; its shape is an additional assumption.	No additional outcome judgement assigned in the S5 crosswalk (U).	Compare alternative response shapes and acoustic controls on held-out liking.

5. Discussion

MI provides an executable architecture in which listening-related calculations can be reconstructed, revised and compared across observation scales. The combined evidence establishes more than a repertoire of labelled outputs. A temporal intervention can be localized; a controlled change of acoustic form can be followed through a multichannel representation; and a readout's behaviour can be attributed to explicit feature and aggregation rules. Human-rating comparisons then ask whether the resulting variables correspond to listening observations. The bird and speech studies extend this programme beyond conventional music and make the relation between representation and readout an empirical property of the model.

The architecture as a research contribution

Architectural research asks how a common organisation supports the construction and evaluation of models (Langley et al., 2009; Laird et al., 2017). For MI, the central commitment is to make the acoustic and temporal assumptions of a listening calculation accessible within its executed dependencies. A temporal request can be altered while retaining the surrounding computation; inspection can also reveal that an input required by an interpretation is absent. This is one way formalisation can sharpen theoretical questions (Guest and Martin, 2021). The worked trace, birdsong assay and speech–song inventory demonstrate this use through temporal intervention, controlled sound transformations and comparisons of identifiable output layers. Their shared organisation makes different hypotheses comparable while preserving the identity of each frozen execution.

The 84 mechanisms and 889 dimensions form the current repertoire of computational proposals. Their number records scope; it does not establish a minimal set of human processes or a need for every mechanism. Fifteen output means support fitted tension prediction, but their RMSE differs from the fixed acoustic subset by 0.013, and appending them to generic descriptors adds no pooled improvement. These findings specify the evidence for that task. Speech and song provide a distinct positive result against a matched acoustic baseline: RASN reaches 100% (27/27) versus 77.8% (21/27) in three-tag speech, CSSL 96.3% (26/27) versus 70.4% (19/27) in three-tag song, and IACM 82.2% (37/45) versus 57.8% (26/45) in five-tag speech. These gains of 22.2–25.9 percentage points show that specific computations make production-tag distinctions more accessible to the same readout. Efficiency, upstream cost and generalization beyond the observed corpus remain separate questions.

The companion studies show why mechanism identity and the readout operator belong inside the scientific account. Birdsong reveals a four-mechanism response set and opposite TAR stress-reduction predictions for flute and piano. Exact decomposition traces that contrast to executed roughness, energy and state terms; a small mean can conceal substantial temporal change. In song, CSSL requires its complete output to attain its strongest observed three-tag result; in speech, IACM P is useful in combination despite weak standalone agreement. A different reader reveals five-tag song structure in a single SDED channel. These demonstrated properties locate compact and complementary information in specific computations. They supply predictions about representation and measurement without equating mechanism names with human processes.

Reusable components and explicit graphs have precedents. Essentia supports audio-analysis networks (Bogdanov et al., 2013); the IPEM Toolbox provides perception-based analysis; AMT organises auditory models and reproduction experiments (Majdak et al., 2022); PsyNeuLink connects psychological and neural components; and ModECI specifies model structure and execution (Gleeson et al., 2023). MI's contribution lies in its particular organisation of named acoustic coordinates, explicit temporal requests and listening-related computations with identifiable consumers. The studies demonstrate uses of that organisation; comparative development effort, efficiency and the necessity of particular components remain open.

Correspondence across observations

The observation scale helps explain the empirical pattern. Acoustic cues show consonance correspondence, and arousal ranks transfer across excerpts more strongly than valence. Yet the PMEmo dynamic pilot had no Bonferroni-significant arousal channel among 19 (S7). Strong excerpt associations also need not add prediction: the specified MUSIFEAST arousal family gave $\Delta R^2 = -0.016$ beyond its controls, with an interval spanning zero (S6). Matched pleasure prediction gained about 0.005 in mean fold r. Excerpt means discard the timing of changes, motivating tests of temporal support and the observation mapping. These outcomes identify where the representation is useful and where matched alternatives remain necessary. Regional encoding tests a further observation level; interpreting it as brain computation requires evidence beyond encoding success (Kriegeskorte and Douglas, 2019).

Transfer between music, natural sound and speech

The two companion studies strengthen the architecture's experimental role through different controls. Birdsong identifies shared full-profile sensitivity and traceable directional affective predictions; speech and song locate production-tag information and output combinations that add discrimination. IACM participates in the common birdsong set and reaches 82.2% in five-tag speech against 57.8% for acoustics. This convergence identifies a reusable calculation for cross-domain tests, without establishing a common human mechanism. The studies retain their units: birdsong ranks pool framewise changes over all outputs, while vocal readers use full-recording summaries combined within request. Historical P-only distances remain separately labelled.

Vocal and musical expression share acoustic cues (Juslin and Laukka, 2003; Coutinho and Schuller, 2017), but spectral–temporal information can support speech and melody differently (Albouy et al., 2020). The vocal results identify concrete comparisons for investigating this organization: CSSL improves three-tag song agreement by 25.9 percentage points over acoustics, and IACM improves five-tag speech by 24.4 points. Output-group exclusions then localize the calculations supporting these gains. Controlled acoustic manipulations can test which cues sustain each pattern. New voices, independently generated texts and listener ratings can test whether production-tag distinctions extend to perceived expression.

The positive result is precise: individual MI mechanisms exceed the matched acoustic-descriptor baseline in all four core speech and song tasks, by 15.6–25.9 percentage points. This is the direct evidence of added representational value under the specified readout. Output-group comparisons supply the next explanation: CSSL needs its complete output for its best three-tag song result, whereas PWUP improves with a smaller subset. Diagnostic readouts expose these properties without changing MI (Hewitt and Liang, 2019). The advantage is relative to the tested 15 descriptors; it does not establish superiority over every acoustic

model. Biological interpretation requires separate observations, as model-metamer work also emphasizes (Feather et al., 2023).

Research use and the next tests

Further comparisons should test whether a specified composite is sufficient for an observation under the same acoustic and temporal information as its alternatives. For birdsong, the immediate prediction is whether flute and piano derived from a shared transcription differ in perceived calming effect when pitch, timing and level are held constant. Multiple source phrases and independent listeners can test the direction, while stress measures can separately evaluate TAR's named interpretation. Matched R^3, log-mel and envelope alternatives, with interventions on the audited terminal terms, can test whether the predicted contrast adds explanatory value. A compact acoustic computation may suffice for a particular observation. Interpretations predicting learning or task-context effects for an identical probe require explicit state inputs and controlled manipulations. Forecast tests must restrict upstream normalization as well as T^3 windows to available past information. These proposed tests can support, restrict or replace an interpretation while retaining inspectable computational alternatives.

A next step is to test whether listener-specific experience improves prediction while the shared MI computations remain fixed. A proposed study would compare a common MI readout, a history-informed personal readout held fixed during the session, and an identically initialized readout updated only from completed listening blocks. All three would predict the same continuous pleasure reports, with predictions committed before each block and transfer assessed on new source recordings. Matched acoustic predictors and simple personal-calibration baselines would test whether any gain depends on the selected MI computations. This design separates the contribution of listening history from that of subsequent feedback. Model personalization would remain distinct from listener learning: a subsequent experiment could manipulate transition histories while matching sound exposure, then measure expectation and pleasure separately for identical probes. Such tests would motivate comparisons between adaptive readouts, adaptive temporal-context weighting and learned representations, using MI's explicit component boundaries to formulate alternatives.

Independent reconstruction is also central to an architectural contribution. Implementation and reconstruction records are currently available through controlled access, and the public description does not specify all 84 mechanisms. A sufficiently detailed specification and independently executable worked example would strengthen external evaluation, as reproducible auditory modelling and model exchange illustrate (Majdak et al., 2022; Gleeson et al., 2023). MI provides a working basis for this research programme: an organisation in which listening hypotheses can be expressed as calculations, revised at identified points and assessed against appropriate observations.

Data and code availability

Public results and analysis records are linked through the Musical Intelligence Results repository and the v2.0.0 archive. The Scientific Supplement supplies the mechanism crosswalk, dataset decisions and evaluation tables. The complete engine is available on request to amace@bu.edu under the stated research-access terms. Detailed source, configuration and reconstruction records require controlled access and are separate from the earlier deposit. The trace demonstrates reconstruction from its retained inputs; it does not establish unrestricted end-to-end reproducibility of the full engine from public files. Raw stimuli and ratings retain their publishers' licences. OSF C8NE4 links protocol and analysis-plan records; musicalintelligence.app provides the project interface. Independent reruns require the matching implementation, inputs, configuration and access permissions.

The companion studies are separately versioned research manuscripts: birdsong, 16 October 2026 and speech and song, 16 October 2026. The updated birdsong manuscript and Supplementary Information retain the original assay and add shared-mechanism controls, TAR terminal decomposition and coordinate-readout tests. The vocal companion reports its complete mechanism, layer and channel inventory. This edition integrates the new birdsong findings into the Abstract, evidence plate, interpretation and Methods while retaining the other ten figures and existing human-data and vocal results. The architectural Scientific Supplement remains the historical R3 record; each companion supplies its current detailed methods. Dated prior editions remain accessible through the publication history.

Scientific Supplement

Implementation records: controlled access

References

Albouy, P., Benjamin, L., Morillon, B. & Zatorre, R. J. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. Science 367, 1043–1047 (2020). Source

Allen, E. J., Mesik, J., Kay, K. N. & Oxenham, A. J. Distinct Representations of Tonotopy and Pitch in Human Auditory Cortex. The Journal of Neuroscience 42, 416-434 (2022). Source

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M. & Brattico, E. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. NeuroImage 59, 3677-3689 (2012). Source

Auksztulewicz, R. & Friston, K. Repetition suppression and its contextual determinants in predictive coding. Cortex 80, 125-140 (2016). Source

Bangert, M. et al. Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. NeuroImage 30, 917-926 (2006). Source

Barrett, F. S., Grimm, K. J., Robins, R. W., Wildschut, T., Sedikides, C. & Janata, P. Music-evoked nostalgia: Affect, memory, and personality. Emotion 10, 390-403 (2010). Source

Basiński, K., Celma-Miralles, A., Quiroga-Martinez, D. R. & Vuust, P. Inharmonicity enhances brain signals of attentional capture and auditory stream segregation. Communications Biology 8, 1584 (2025). Source

Bellmann, O. T. & Asano, R. Neural correlates of musical timbre: an ALE meta-analysis of neuroimaging data. Frontiers in Neuroscience 18, 1373232 (2024). Source

Bidelman, G. M. & Krishnan, A. Neural Correlates of Consonance, Dissonance, and the Hierarchy of Musical Pitch in the Human Brainstem. The Journal of Neuroscience 29, 13165-13171 (2009). Source

Bidelman, G. M. The Role of the Auditory Brainstem in Processing Musically Relevant Pitch. Frontiers in Psychology 4 (2013). Source

Bigand, F., Bianco, R., Abalde, S. F., Nguyen, T. & Novembre, G. EEG of the Dancing Brain: Decoding Sensory, Motor, and Social Processes during Dyadic Dance. The Journal of Neuroscience 45, e2372242025 (2025). Source

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J. R., and Serra, X. (2013). Essentia: An audio analysis library for music information retrieval. Proceedings of ISMIR, 493–498. Source

Bolland, E., De Burca, A., Wang, S. H., Khalil, A. & McLoughlin, G. Efficacy of auditory gamma stimulation for cognitive decline: a systematic review of individual and group differences across cognitively impaired and healthy populations. npj Aging 12, 8 (2025). Source

Bonetti, L. et al. Spatiotemporal brain hierarchies of auditory memory recognition and predictive coding. Nature Communications 15, 4313 (2024). Source

Bravo, F., Cross, I., Stamatakis, E. A. & Rohrmeier, M. Sensory cortical response to uncertainty and low salience during recognition of affective cues in musical intervals. PLOS ONE 12, e0175991 (2017). Source

Briley, P. M., Breakey, C. & Krumbholz, K. Evidence for Pitch Chroma Mapping in Human Auditory Cortex. Cerebral Cortex 23, 2601-2610 (2013). Source

Bruzzone, S. E. P., Lumaca, M., Brattico, E., Vuust, P., Kringelbach, M. L. & Bonetti, L. Dissociated brain functional connectivity of fast versus slow frequencies underlying individual differences in fluid intelligence: a DTI and MEG study. Scientific Reports 12, 4746 (2022). Source

Burchardt, L. S., Varkevisser, J. M. & Spierings, M. J. Zebra finch tutees not only share the melody but also the rhythm of their tutor's song. Scientific Reports 15, 35573 (2025). Source

Cawley, G. C. & Talbot, N. L. C. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research 11, 2079–2107 (2010). Source

Chabin, T. et al. Cortical Patterns of Pleasurable Musical Chills Revealed by High-Density EEG. Frontiers in Neuroscience 14, 565815 (2020). Source

Cheung, V. K., Harrison, P. M., Meyer, L., Pearce, M. T., Haynes, J. D. & Koelsch, S. Uncertainty and Surprise Jointly Predict Musical Pleasure and Amygdala, Hippocampus, and Auditory Cortex Activity. Current Biology 29, 4084-4092.e4 (2019). Source

Cousineau, M., Bidelman, G. M., Peretz, I. & Lehmann, A. On the Relevance of Natural Stimuli for the Study of Brainstem Correlates: The Example of Consonance Perception. PLOS ONE 10, e0145439 (2015). Source

Coutinho, E. & Schuller, B. Shared acoustic codes underlie emotional communication in music and speech—Evidence from deep transfer learning. PLoS ONE 12, e0179289 (2017). Source

Crespo-Bojorque, P., Monte-Ordoño, J. & Toro, J. M. Early neural responses underlie advantages for consonance over dissonance. Neuropsychologia 117, 188-198 (2018). Source

Daly, I. et al. Affective brain–computer music interfacing. Journal of Neural Engineering 13, 046022 (2016). Source

Deng, X., Jiang, N., Huang, Z. & Wang, Q. Cortical Activation and Functional Connectivity Response to Different Interactive Modes of Virtual Reality (VR)-Induced Analgesia: A Prospective Functional Near-Infrared Spectroscopy (fNIRS) Study. Journal of Pain Research 18, 1095-1108 (2025). Source

Doelling, K. B. & Poeppel, D. Cortical entrainment to music and its modulation by expertise. Proceedings of the National Academy of Sciences 112 (2015). Source

Eerola, T., and Lahdelma, I. (2021). The anatomy of consonance/dissonance: Evaluating acoustic and cultural predictors across multiple datasets with chords. Music & Science, 4, 1–19. Source

Eerola, T., and Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. Psychology of Music, 39(1), 18–49. Source

Egermann, H., Pearce, M. T., Wiggins, G. A. & McAdams, S. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. Cognitive, Affective, & Behavioral Neuroscience 13, 533-553 (2013). Source

Ehrlich, S. K., Agres, K. R., Guan, C. & Cheng, G. A closed-loop, music-based brain-computer interface for emotion mediation. PLOS ONE 14, e0213516 (2019). Source

Erdem, A. Expressive information across complementary processing layers in speech and song. Research manuscript, 16 October 2026 mechanism-and-layer edition (2026b). Source

Erdem, A. Shared mechanisms and interpretable affective predictions in birdsong-to-music transformations. Research manuscript, 16 October 2026 (2026a). Source

Eyben, F., Wöllmer, M., and Schuller, B. (2010). openSMILE: The Munich versatile and fast open-source audio feature extractor. Proceedings of ACM Multimedia, 1459–1462. Source

Fang, R., Ye, S., Huangfu, J. & Calimag, D. P. Music therapy is a potential intervention for cognition of Alzheimer's Disease: a mini-review. Translational Neurodegeneration 6, 2 (2017). Source

Feather, J., Leclerc, G., Mądry, A. & McDermott, J. H. Model metamers reveal divergent invariances between biological and artificial neural networks. Nature Neuroscience 26, 2017–2034 (2023). Source

Fernández-Rubio, G., Brattico, E., Kotz, S. A., Kringelbach, M. L., Vuust, P. & Bonetti, L. Magnetoencephalography recordings reveal the spatiotemporal dynamics of recognition memory for complex versus simple auditory sequences. Communications Biology 5, 1272 (2022). Source

Fishman, Y. I. et al. Consonance and Dissonance of Musical Chords: Neural Correlates in Auditory Cortex of Monkeys and Humans. Journal of Neurophysiology 86, 2761-2788 (2001). Source

Gleeson, P., Crook, S., Turner, D., Mantel, K., Raunak, M., Willke, T. & Cohen, J. D. Integrating model development across computational neuroscience, cognitive science, and machine learning. Neuron 111, 1526–1530 (2023). Source

Gold, B. P., Mas-Herrero, E., Zeighami, Y., Benovoy, M., Dagher, A. & Zatorre, R. J. Musical reward prediction errors engage the nucleus accumbens and motivate learning. Proceedings of the National Academy of Sciences 116, 3310-3315 (2019a). Source

Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A. & Zatorre, R. J. Predictability and Uncertainty in the Pleasure of Music: A Reward for Learning?. The Journal of Neuroscience 39, 9397-9409 (2019b). Source

Gold, B. P., Pearce, M. T., McIntosh, A. R., Chang, C., Dagher, A. & Zatorre, R. J. Auditory and reward structures reflect the pleasure of musical expectancies during naturalistic listening. Frontiers in Neuroscience 17, 1209398 (2023). Source

Grahn, J. A. & Brett, M. Rhythm and Beat Perception in Motor Areas of the Brain. Journal of Cognitive Neuroscience 19, 893-906 (2007). Source

Guest, O., and Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. Perspectives on Psychological Science, 16(4), 789–802. Source

Haiduk, F., Zatorre, R. J., Benjamin, L., Morillon, B. & Albouy, P. Spectrotemporal cues and attention jointly modulate fMRI network topology for sentence and melody perception. Scientific Reports 14, 5501 (2024). Source

Halpern, A. R., Zatorre, R. J., Bouffard, M. & Johnson, J. A. Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia 42, 1281-1292 (2004). Source

Harrison, E. C. et al. Neural mechanisms underlying synchronization of movement to musical cues in Parkinson disease and aging. Frontiers in Neuroscience 19, 1550802 (2025). Source

Harrison, P. M. C., and MacConnachie, J. M. C. (2024). Consonance in the carillon. Journal of the Acoustical Society of America, 156(2), 1111–1122. Source

Hausfeld, L., Disbergen, N. R., Valente, G., Zatorre, R. J. & Formisano, E. Modulating Cortical Instrument Representations During Auditory Stream Segregation and Integration With Polyphonic Music. Frontiers in Neuroscience 15, 635937 (2021). Source

Hewitt, J. & Liang, P. Designing and interpreting probes with control tasks. Proceedings of EMNLP-IJCNLP, 2733–2743 (2019). Source

Hoddinott, J. D. & Grahn, J. A. Neural representations of beat and rhythm in motor and association regions. Cerebral Cortex 34, bhae406 (2024). Source

IPEM Toolbox contributors. IPEM Toolbox: An open source auditory toolbox for perception-based music analysis. Project documentation (accessed 8 October 2026). Source

Jacobsen, J. H., Stelzer, J., Fritz, T. H., Chételat, G., La Joie, R. & Turner, R. Why musical memory can be preserved in advanced Alzheimer's disease. Brain 138, 2438-2450 (2015). Source

Janata, P. The Neural Architecture of Music-Evoked Autobiographical Memories. Cerebral Cortex 19, 2579-2594 (2009). Source

Juslin, P. N. & Laukka, P. Communication of emotions in vocal expression and music performance: different channels, same code? Psychological Bulletin 129, 770–814 (2003). Source

Kheirkhah, M. et al. Mindfulness, music, visual occlusion in ketamine therapy for depression: do they change outcomes? A qualitative and quantitative analysis of a randomized controlled trial. Frontiers in Psychiatry 16, 1642025 (2025). Source

Kim, C. H., Jin, S. H., Kim, J. S., Kim, Y., Yi, S. W. & Chung, C. K. Dissociation of Connectivity for Syntactic Irregularity and Perceptual Ambiguity in Musical Chord Stimuli. Frontiers in Neuroscience 15, 693629 (2021). Source

Kim, S. G., Mueller, K., Lepsien, J., Mildner, T. & Fritz, T. H. Brain networks underlying aesthetic appreciation as modulated by interaction of the spectral and temporal organisations of music. Scientific Reports 9, 19446 (2019). Source

Kleber, B., Zeitouni, A. G., Friberg, A. & Zatorre, R. J. Experience-Dependent Modulation of Feedback Integration during Singing: Role of the Right Anterior Insula. The Journal of Neuroscience 33, 6070-6080 (2013). Source

Koelsch, S. Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. Psychophysiology 46, 179-190 (2009). Source

Koelsch, S., Fritz, T., v. Cramon, D. Y., Müller, K. & Friederici, A. D. Investigating emotion with music: An fMRI study. Human Brain Mapping 27, 239-250 (2006). Source

Koelsch, S., Gunter, T., Friederici, A. D. & Schröger, E. Brain Indices of Music Processing: "Nonmusicians" are Musical. Journal of Cognitive Neuroscience 12, 520-541 (2000). Source

Koelsch, S., Schröger, E. & Tervaniemi, M. Superior pre-attentive auditory processing in musicians. NeuroReport 10, 1309-1313 (1999). Source

Kohler, N. et al. Distinct and content-specific neural representations of self- and other-produced actions in joint piano performance. Frontiers in Human Neuroscience 19, 1543131 (2025). Source

Kokal, I., Engel, A., Kirschner, S. & Keysers, C. Synchronized Drumming Enhances Activity in the Caudate and Facilitates Prosocial Commitment - If the Rhythm Comes Easily. PLOS ONE 6, e27272 (2011). Source

Kraemer, D. J. M., Macrae, C. N., Green, A. E. & Kelley, W. M. Sound of silence activates auditory cortex. Nature 434, 158-158 (2005). Source

Kriegeskorte, N. & Douglas, P. K. Interpreting encoding and decoding models. Current Opinion in Neurobiology 55, 167–179 (2019). Source

Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis—connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience 2, 4 (2008). Source

Lahav, A., Saltzman, E. & Schlaug, G. Action Representation of Sound: Audiomotor Recognition Network While Listening to Newly Acquired Actions. The Journal of Neuroscience 27, 308-314 (2007). Source

Laird, J. E., Lebiere, C. & Rosenbloom, P. S. A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. AI Magazine 38(4), 13–26 (2017). Source

Langley, P., Laird, J. E. & Rogers, S. Cognitive architectures: Research issues and challenges. Cognitive Systems Research 10, 141–160 (2009). Source

Leeuwis, N., Pistone, D., Flick, N. & van Bommel, T. A Sound Prediction: EEG-Based Neural Synchrony Predicts Online Music Streams. Frontiers in Psychology 12, 672980 (2021). Source

Leipold, S., Klein, C. & Jäncke, L. Musical Expertise Shapes Functional and Structural Brain Networks Independent of Absolute Pitch Ability. The Journal of Neuroscience 41, 2496-2511 (2021). Source

Liang, J. et al. The Brain Mechanisms of Music Stimulation, Motor Observation, and Motor Imagination in Virtual Reality Techniques: A Functional Near-Infrared Spectroscopy Study. eneuro 12, ENEURO.0557-24.2025 (2025). Source

Livingstone, S. R. & Russo, F. A. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in North American English. PLoS ONE 13, e0196391 (2018). Source

Loui, P., Patterson, S., Sachs, M. E., Leung, Y., Zeng, T. & Przysinda, E. White Matter Correlates of Musical Anhedonia: Implications for Evolution of Music. Frontiers in Psychology 8, 1664 (2017). Source

Maess, B., Koelsch, S., Gunter, T. C. & Friederici, A. D. Musical syntax is processed in Broca's area: an MEG study. Nature Neuroscience 4, 540-545 (2001). Source

Majdak, P., Hollomey, C. & Baumgartner, R. AMT 1.x: A toolbox for reproducible research in auditory modeling. Acta Acustica 6, 19 (2022). Source

Marjieh, R., Harrison, P. M. C., Lee, H., Deligiannaki, F., and Jacoby, N. (2024). Timbral effects on consonance disentangle psychoacoustic mechanisms and suggest perceptual origins for musical scales. Nature Communications, 15, 1482. Source

Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R. J. & Marco-Pallarés, J. Neural correlates of specific musical anhedonia. Proceedings of the National Academy of Sciences 113 (2016). Source

McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychological Review 102, 419-457 (1995). Source

McDermott, J. H. & Simoncelli, E. P. Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron 71, 926–940 (2011). Source

Mencke, I., Omigie, D., Wald-Fuhrmann, M. & Brattico, E. Atonal Music: Can Uncertainty Lead to Pleasure?. Frontiers in Neuroscience 12, 979 (2019). Source

Millidge, B., Seth, A. K. & Buckley, C. L. Predictive coding: a theoretical and experimental review. arXiv:2107.12979v4 (2022). Source

Mischler, G., Li, Y. A., Bickel, S., Mehta, A. D. & Mesgarani, N. The impact of musical expertise on disentangled and contextual neural encoding of music revealed by generative music models. Nature Communications 16, 8874 (2025). Source

Mitterschiffthaler, M. T., Fu, C. H., Dalton, J. A., Andrew, C. M. & Williams, S. C. A functional MRI study of happy and sad affective states induced by classical music. Human Brain Mapping 28, 1150-1162 (2007). Source

Møller, C. et al. Audiovisual structural connectivity in musicians and non-musicians: a cortical thickness and diffusion tensor imaging study. Scientific Reports 11, 4324 (2021). Source

Noboa, M. d. L., Kertész, C. & Honbolygó, F. Neural entrainment to the beat and working memory predict sensorimotor synchronization skills. Scientific Reports 15, 10466 (2025). Source

Norman-Haignere, S. V. et al. Multiscale temporal integration organizes hierarchical computation in human auditory cortex. Nature Human Behaviour 6, 455-469 (2022). Source

Norman-Haignere, S., Kanwisher, N. & McDermott, J. H. Cortical Pitch Regions in Humans Respond Primarily to Resolved Harmonics and Are Located in Specific Tonotopic Regions of Anterior Auditory Cortex. The Journal of Neuroscience 33, 19451-19469 (2013). Source

Nozaradan, S., Peretz, I. & Mouraux, A. Selective Neuronal Entrainment to the Beat and Meter Embedded in a Musical Rhythm. The Journal of Neuroscience 32, 17572-17581 (2012). Source

Nozaradan, S., Peretz, I., Missal, M. & Mouraux, A. Tagging the Neuronal Entrainment to Beat and Meter. Journal of Neuroscience 31, 10234-10240 (2011). Source

Okada, K. i., Takeya, R. & Tanaka, M. Neural signals regulating motor synchronization in the primate deep cerebellar nuclei. Nature Communications 13, 2504 (2022). Source

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A. & Ross, B. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12, 169-174 (2001). Source

Paraskevopoulos, E., Chalas, N., Anagnostopoulou, A. & Bamidis, P. D. Interaction within and between cortical networks subserving multisensory learning and its reorganization due to musical expertise. Scientific Reports 12, 7891 (2022). Source

Partanen, E., Mårtensson, G., Hugoson, P., Huotilainen, M., Fellman, V. & Ådén, U. Auditory Processing of the Brain Is Enhanced by Parental Singing for Preterm Infants. Frontiers in Neuroscience 16, 772008 (2022). Source

Patel, A. D. & Iversen, J. R. The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. Frontiers in Systems Neuroscience 8 (2014). Source

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S. & Griffiths, T. D. The Processing of Temporal Pitch and Melody Information in Auditory Cortex. Neuron 36, 767-776 (2002). Source

Pearce, M. T. & Wiggins, G. A. Auditory expectation: The information dynamics of music perception and cognition. Topics in Cognitive Science 4, 625–652 (2012). Source

Penagos, H., Melcher, J. R. & Oxenham, A. J. A Neural Representation of Pitch Salience in Nonprimary Human Auditory Cortex Revealed with Functional Magnetic Resonance Imaging. The Journal of Neuroscience 24, 6810-6815 (2004). Source

Porfyri, I., Paraskevopoulos, E., Anagnostopoulou, A., Styliadis, C. & Bamidis, P. D. Multisensory vs. unisensory learning: how they shape effective connectivity networks subserving unimodal and multimodal integration. Frontiers in Neuroscience 19, 1641862 (2025). Source

Potes, C., Gunduz, A., Brunner, P. & Schalk, G. Dynamics of electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music listening. NeuroImage 61, 841-848 (2012). Source

PsyNeuLink contributors. Welcome to PsyNeuLink: Purpose and environment overview. Project documentation (accessed 8 October 2026). Source

Putkinen, V., Seppälä, K., Harju, H., Hirvonen, J., Karlsson, H. K. & Nummenmaa, L. Pleasurable music activates cerebral μ-opioid receptors: a combined PET-fMRI study. European Journal of Nuclear Medicine and Molecular Imaging 52, 3540-3549 (2025). Source

Qiu, R. et al. The impact of musical intervention during fetal and infant stages on social behavior and neurodevelopment in mice. Translational Psychiatry 15, 408 (2025). Source

Rathcke, T., Smit, E., Zheng, Y. & Canzi, M. Perception of temporal structure in speech is influenced by body movement and individual beat perception ability. Attention, Perception, & Psychophysics 86, 1746-1762 (2024). Source

Ross, J. M. & Balasubramaniam, R. Time Perception for Musical Rhythms: Sensorimotor Perspectives on Entrainment, Simulation, and Prediction. Frontiers in Integrative Neuroscience 16, 916220 (2022). Source

Rupp, A., Englitz, B., Balaguer-Ballester, E. & Andermann, M. Editorial: Early neural processing of musical melodies. Frontiers in Human Neuroscience 16, 1109500 (2022). Source

Sachs, M. E., Kozak, M. S., Ochsner, K. N. & Baldassano, C. Emotions in the Brain Are Dynamic and Contextually Dependent: Using Music to Measure Affective Transitions. eneuro 12, ENEURO.0184-24.2025 (2025). Source

Sakakibara, Y. et al. A Nostalgia Brain-Music Interface for enhancing nostalgia, well-being, and memory vividness in younger and older individuals. Scientific Reports 15, 32337 (2025). Source

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A. & Zatorre, R. J. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. Nature Neuroscience 14, 257-262 (2011). Source

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A. & Zatorre, R. J. Interactions Between the Nucleus Accumbens and Auditory Cortices Predict Music Reward Value. Science 340, 216-219 (2013). Source

Samiee, S., Vuvan, D., Florin, E., Albouy, P., Peretz, I. & Baillet, S. Cross-Frequency Brain Network Dynamics Support Pitch Change Detection. The Journal of Neuroscience 42, 3823-3835 (2022). Source

Sansare, A., Weinrich, M., Bernard, J. A. & Lei, Y. Enhancing Balance Control in Aging Through Cerebellar Theta-Burst Stimulation. The Cerebellum 24, 161 (2025). Source

Sarasso, P. et al. Aesthetic appreciation of musical intervals enhances behavioural and neurophysiological indexes of attentional engagement and motor inhibition. Scientific Reports 9, 18550 (2019). Source

Sarasso, P., Perna, P., Barbieri, P., Neppi-Modona, M., Sacco, K. & Ronga, I. Memorisation and implicit perceptual learning are enhanced for preferred musical intervals and chords. Psychonomic Bulletin & Review 28, 1623-1637 (2021). Source

Scholkmann, F. et al. Creative music therapy in preterm infants effects cerebrovascular oxygenation and perfusion. Scientific Reports 14, 28249 (2024). Source

Spence, C. Crossmodal correspondences: A tutorial review. Attention, Perception, & Psychophysics 73, 971-995 (2011). Source

Squire, L. R. & Alvarez, P. Retrograde amnesia and memory consolidation: a neurobiological perspective. Current Opinion in Neurobiology 5, 169-177 (1995). Source

Tabas, A., Andermann, M., Schuberth, V., Riedel, H., Balaguer-Ballester, E. & Rupp, A. Modeling and MEG evidence of early consonance processing in auditory cortex. PLOS Computational Biology 15, e1006820 (2019). Source

Thaut, M. H., McIntosh, G. C. & Hoemberg, V. Neurobiological foundations of neurologic music therapy: rhythmic entrainment and the motor system. Frontiers in Psychology 5 (2015). Source

Thaut, M. H., Miller, R. A. & Schauer, L. M. Multiple synchronization strategies in rhythmic sensorimotor tasks: phase vs period correction. Biological Cybernetics 79, 241-250 (1998). Source

Trainor, L. J. & Unrau, A. Development of Pitch and Music Perception. Springer Handbook of Auditory Research, 223-254 (2012). Source

Van Der Donckt, J., Van Der Donckt, J., Deprost, E., and Van Hoecke, S. (2022). tsflex: Flexible time series processing & feature extraction. SoftwareX, 17, 100971. Source

van der Walle, H. A., Wu, W., Margulis, E. H., and Jakubowski, K. (2025). MUSIFEAST-17: MUsic Stimuli for imagination, familiarity, emotion, and Aesthetic STudies across 17 genres. Behavior Research Methods, 57, 204. Source

Varoquaux, G. Cross-validation failure: small sample sizes lead to large error bars. NeuroImage 180, 68–77 (2018). Source

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R. & Tervaniemi, M. The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. Neuropsychologia 50, 1432-1443 (2012). Source

Wikman, P. et al. Selective Attention Shapes Neural Representations of Complex Auditory Scenes: The Roles of Object Identity and Scene Composition. The Journal of Neuroscience 45, e0506252025 (2025). Source

Yamashita, K. et al. A pilot study on simultaneous stimulation of the primary motor cortex and supplementary motor area using gait-synchronized rhythmic brain stimulation to improve gait variability in post-stroke hemiparetic patients. Frontiers in Human Neuroscience 19, 1618758 (2025). Source

Yokota, Y. et al. Auditory stimulation at individual gamma frequency enhances cognitive performance. Scientific Reports 15, 38697 (2025). Source

Yuan, Y., Gayet, S., Wisman, D. C., van der Stigchel, S. & van der Stoep, N. Decoding Auditory Working Memory Load From EEG Alpha Oscillations. Psychophysiology 62, e70210 (2025). Source

Zamorano, A., Zatorre, R., Vuust, P., Friberg, A., Birbaumer, N. & Kleber, B. Singing training predicts increased insula connectivity with speech and respiratory sensorimotor areas at rest. Brain Research 1813, 148418 (2023). Source

Zatorre, R. J. & Halpern, A. R. Mental Concerts: Musical Imagery and Auditory Cortex. Neuron 47, 9-12 (2005). Source

Zatorre, R. J. Hemispheric asymmetries for music and speech: Spectrotemporal modulations and top-down influences. Frontiers in Neuroscience 16, 1075511 (2022). Source

6. Methods

Computational configuration and outputs

The study uses saved executions, acoustic and mechanism arrays, and existing human-rating datasets; no new participants were recruited. Configuration records identify implementations, requests, execution order and output mappings. Mechanisms run in processing-depth order, with discovery order retained at equal depths, and upstream values contribute only when consumed. The Swan Lake trace records ordered sources and inputs. Earlier caches have separate execution records; equal output counts do not identify the code that produced them. Supplement S5 links mechanism names to literature and evaluation roles, distinguishing motivation for a label from derivation of its equation. Independent reproduction requires implementation and reconstruction records, available through controlled access.

At 44,100 Hz and a 256-sample hop, the inspected adapter returns 172.265625 frames per second. Individual R^3 analysis windows, longer temporal operations and recording-wide normalization determine the support of each coordinate. The behavioural analyses below use full-excerpt context and output means. Forward and centred operations are part of those offline representations. Numerical coefficients and channel names are documented with their source definitions; fitted observation models are specified separately.

Worked execution and intervention

The source Swan Lake recording lasts 182.597 s. Its 30–50-s display interval was selected before inspecting intervention effects. The saved execution identifies the input, selected configuration, outputs and observed consumption. The intervention changes one temporal-context requirement with other computations held fixed. A dimension is counted as changed if any common valid frame exceeds the recorded absolute tolerance of one ten-millionth plus a relative tolerance of one millionth. Observer and identity-adapter controls assess instrumentation. Native-frame results determine the counts; display averaging is used only for visualization. S1 records the scope and results of these checks; the implementation recipe remains under controlled access.

Controlled acoustic assay

The birdsong assay retained its own frozen engine and adapter identity. Five primary paths used 97 anchors; three 25-anchor operator paths, a 27-setting factorial grid and level controls contributed to 1,146 recordings. Twenty-four diagnostic endpoint probes were additional follow-ups. Full-phrase waveforms were decoded to float32 arrays, averaged across stereo channels and processed without inter-anchor state. The primary outcome used the 202 dimensions assigned to the P layer by registry ranges; dimension-name prefixes were not used. Path RMS pooled squared differences from the first anchor across frames, dimensions and anchors. Global pooling weighted dimensions equally; signed endpoint means were retained separately. Original-level and common whole-phrase RMS conditions were analysed separately. The adapter retained first-function-wins registry selection, depth-ordered execution and recording-wide mel normalization. It saved 131 weighted readouts without clip-wise percentile stretching and did not execute the eight separate recursive classes. Rank correlations are descriptive, and interpolation error is distinct from rank stability. S8 records renderer settings, temporal support, amendments, execution differences and the archived verification checks.

Birdsong full-profile and TAR analyses

The revised assay retains all 889 declared dimensions and the actual layer schemas, including TAR's E/TI/P/F groups. Full-profile path RMS pools squared changes from the first anchor across frames, dimensions and anchors. A shared mechanism must rank in the top 21 of 84 on all five paths; native/equal-RMS level, full/trimmed windows and 97/49/25-anchor sampling are checked separately. Rankings use unrescaled model units. Endpoint-based channel rescaling is an explicit sensitivity analysis and yields no common upper-quartile set on the sequential routes. Layer shares divide their squared change by the mechanism's full squared change; they reflect channel count and scale rather than independent explained variance.

For TAR's stress-reduction output, the audited terminal rule is $\mathrm{clip}(0.20E_0T_2 + 0.20P_0 + 0.35(1 - r) + 0.25[1 - \mathrm{clip}(|v|, 0, 1)], 0, 1)$, where r is the executed roughness-trend feature and v the amplitude-change velocity feature. Signed contrasts are evaluated at identical frame indices. Four additive terms and a separate clipping residual exactly reconstruct the mean endpoint change; 466,146 frame checks give maximum absolute error 1.19×10^{-7}. Scores are computed independently for each complete recording, without predicting an inter-sound therapeutic relationship or observing clinical outcomes.

Coordinate readers use five held-out contiguous blocks with a four-anchor buffer, training-only scaling and fixed ridge regularization ($\alpha = 1$). Full profiles, every permitted layer subset and descriptive best-single comparisons are retained with alternative scaling and level controls: 41,760 evaluations and 4,050,720 predictions. Selecting the best single layer from the observed scores is descriptive. Output exclusion removes stored reader features; it does not disable upstream computations or demonstrate causal necessity. Full methods and archived numerical results are supplied by the 16 October companion (Erdem, 2026a).

Birdsong readout and regional audit

The operational transcription maps eight events to supplied pitch and onset targets; it does not learn musical regularities from a corpus. Flute and piano resynthesis retains MIDI 91–113 (G6–F8) without octave folding, requiring pitch-shifted samples above ordinary instrumental ranges. Instrument-source labels therefore do not isolate identity from register, spectral bandwidth or envelope. Common waveform RMS controls overall level, not instantaneous envelope or perceived loudness.

The latest companion audit reconstructs terminal channels for all 484,758 frames in the 1,146 recordings. Ordered production writes replace earlier direct values: the final identities are DA= $0.65\,\mathrm{ECT}[f04]$, OPI= $0.90\,\mathrm{IOTMS}[E0]$ and 5HT= $0.30\,\mathrm{BCH}[P0]$. NE starts at 0.5 and undergoes nine amplification and one inhibition updates, with clipping after each nucleus. Earlier sources can still have indirect downstream effects. The audited executor does not feed these channel values back into mechanisms. Phrase means, framewise RMS and signed trajectories are retained separately. Regional projections sum fixed weighted dimension links and rectify at zero; the weight control divides each region by its sum of absolute incoming weights. All four audited endpoints have positive pre-rectification sums. The complete source inventory and readout checks appear in the 16 October birdsong companion and its supplement (Erdem, 2026a); the original architectural S8 remains a record of the earlier assay.

Original speech and song: corpus and fixed engine

The updated vocal study retains all 342 original ElevenLabs recordings from 172 generation requests. The matched core crosses three intended English texts, two requested modes, five production tags and three requests per cell, giving 90 requests and 179 returned alternatives. The displayed interface identified Eleven v4 and Anna—Young Female Conversational Voice, with stability 0.5 and similarity 0.75. Files, receipts, prompts and hashes define the corpus; no recoverable synthesis seed or exact checkpoint was supplied. Six transcript disagreements remain included. Tags are archived generation directions, not independent emotion classifications; untagged means no explicit emotion instruction. RAVDESS supplies a possible future human comparison (Livingstone and Russo, 2018), not additional observations here.

The pinned adapter processes 44.1-kHz mono float32 audio using a 2,048-sample Fourier window, a 256-sample hop, 128 mel bands, log(1+power) and per-recording normalization. Each waveform starts with freshly instantiated mechanisms. MI is fixed and receives no tags or tag-dependent optimization objective. Some context demands include centered or future samples, so the analysis is offline. First-function-wins enumeration of the registry yields 84 mechanisms, 326 declared layers and 889 channels. Layer meanings follow each mechanism's metadata; M may denote memory or temporal integration, and MMP uses R/P/F/C. No semantic-name filter restricts the candidate inventory.

External readouts and comprehensive response mapping

Each channel is averaged within recording, then available alternatives are averaged with equal weight within generation request. The complete-C^3 comparator concatenates the 889 channel values, rather than averaging unlike mechanisms. Three-tag tasks use excited, sad and untagged, with 27 requests per mode; five-tag tasks add mad and peaceful, with 45 per mode. For each held-out text, readout parameters are estimated from the other two texts. All mechanisms enter the original fixed ridge procedure ($\alpha = 1$), with imputation, constant-channel filtering, standardization and feature-count scaling confined to the estimation texts. The 15-descriptor acoustic baseline covers duration, level, pitch, spectral shape and envelope rhythm, using the same request units, held-out texts and fixed ridge procedure. Figure 10 reports each mechanism and baseline as correct/total and percentage; the gain in percentage points is $100(c_{\mathrm{MI}} - c_{\mathrm{acoustic}})/n$ within each task. Gains are computed from the integer counts before rounding.

Nearest-centroid measurement is additionally applied to the complete mechanism, layer and channel inventory. A centered scalar linear-ridge classifier can favor extreme fitted slopes and miss intermediate classes; centroid distances provide a complementary test of that geometry. Its results are labelled separately throughout. Highlighted outputs are discoveries from the complete exploratory map, not predictors chosen before examining this corpus. The companion reports candidate-wide permutation tests, restricted exchangeability assumptions, full confusion matrices and secondary automatic-selection controls. These tests address tag-related structure; they do not establish superiority of every named mechanism over an acoustic alternative or generalization to a population of voices and texts.

Layer contributions and numerical provenance

Single-layer exclusion removes a mechanism's stored outputs for one declared layer and refits the same external readout. The analysis includes 1,304 exclusions across four core tasks and 1,152 detailed subset/readout combinations for six mechanisms. All subsets, single-layer measurements and a control preserving the complete mechanism's scaling denominator remain available. These comparisons measure the contribution of accessible output information; upstream computations are not disabled, and the results are not causal ablations of the running engine.

Verification reproduced hashes for all 342 originals, reconstructed 1,590 archived prediction matrices and checked 43 task splits. Recomputing all 84 mechanisms across the four core tasks reproduced stored scores to a maximum absolute difference of 1.14×10^{-7}. Figure 10 is regenerated directly from the updated companion's archived primary request-level results, including its acoustic baseline; no layer subset or alternative reader is mixed into these four comparisons. The present architectural revision performs no new audio synthesis, engine fitting or inference. Full numerical provenance and measurement controls are in the 16

October manuscript and its Supplementary Information (Erdem, 2026b).

Grouped tension prediction

The film-tension comparison retains 110 cache entries grouped into 40 source soundtracks; repeated excerpts and segments remain in the same group. Five outer GroupKFold splits provide held-out predictions. Four grouped inner folds select ridge alpha from 0.1, 1, 10, 100, 1,000 and 10,000 by mean squared error. Mean imputation and standardization are fitted inside each training split. Features use the supplied full clip after omission of the first 172 cached frames. The representations comprise 97 R^3 means; 970 acoustic/temporal descriptors; those 970 descriptors plus fifteen C^3 means; the fifteen C^3 means alone; two training-fold PCA15 representations; a fixed fifteen-descriptor acoustic subset; and a training-target-mean reference. Each R^3 channel supplies its mean, standard deviation, 10th/50th/90th percentiles, mean absolute successive difference, normalized-time slope and autocorrelations at 0.35, 1 and 4 s for the 970-descriptor representation. The acoustic subset uses mean, variability and successive change in roughness, pleasantness, loudness, tonalness and spectral flux. The C^3 set includes all channels of the existing tension pool. PCA is fitted within each inner and outer training split to standardized predictors, without whitening or subsequent component rescaling.

Pooled out-of-fold RMSE is the primary score; Pearson r and R^2 summarize the same predictions. R^2 uses squared prediction error relative to variation around the pooled observed target mean. The five paired folds are reported descriptively. The compact extension retained the earlier outer folds and tuning grid; its alternatives were specified before those fits. Full columns, fold identities, predictions and tuning records accompany Supplement S2.

Arousal correspondence

MUSIFEAST-17 supplies perceived arousal and valence ratings for thirty-second excerpts. Its recorded partition comprises seventeen previously examined pilot excerpts, 274 development excerpts and 65 fixed test excerpts, grouping connected artist/composer, source and identical-audio identities. The total of 356 includes both pilot and 65-excerpt test subset. Channel means use the released excerpt after omitting its first 172 cached frames. Each direct Spearman correlation relates a channel mean to a human excerpt rating, without fitting a regression or removing acoustic contributions from the target ratings.

Four arousal and four valence channel identities were transferred from the prior film tables after the initial pilot, without channel reselection on the test subset. Historical film energy/arousal and valence targets, populations and engine snapshots differ from MUSIFEAST. Normative labels had been accessed for curation and quality control. Supplement S3 retains all eight channels and the partition chronology. The main arousal table reports the complete four-channel arousal family.

Consonance comparisons

Consonance analyses correlate model descriptors with published interval or chord means, retaining each target convention. The fixed pleasantness combination is stated in Results; this evaluates established acoustic cues within MI, not the necessity of its cognitive core. The harmonic-dyad comparison uses thirteen means and the pooled-chord analysis 617 rows. The extended panel uses recorded six-partial harmonic synthesis; flute, guitar and piano identify human rating conditions. Two analytical references are distinguished from seven behavioural targets.

The Carillon evaluation applies one extractor with unchanged coefficients to twelve combinations of four registers and three spectral/decay constructions. These inputs are synthesized dyads. All twelve correlations are retained. The spectral route, normalization, synthesis definitions and rating references appear in Supplement S4. Algebraically related coordinates are presented together as properties of the same representation.

Public-data eligibility

A dataset was eligible for a proposed comparison when it supplied the relevant human target, accessible audio or explicitly documented stimulus construction, recoverable stimulus–response mapping and timing adequate to the observation scale. The bounded dataset review inventoried prior use and considered additional releases, prioritizing 2025–2026. Supplement S6 distinguishes processed, partially available, unresolved and deferred cases with dated source records. It also reports completed comparisons whose increments were small or negative. Availability alone does not establish suitability, and an available but deferred resource is not classified as inaccessible.

Analysis scope

These evaluations are exploratory because corpus-guided development shaped the repertoire before comparison. The film corpus and tension pool were already known, and MUSIFEAST transfer followed an inspected pilot. Training-fold preprocessing and nested tuning constrain fitting within a specified model; they cannot undo earlier representation selection (Cawley and Talbot, 2010). Bonferroni and BH corrections concern specified test families, leaving development exposure and effect-size superiority unresolved. Correlated channels and overlapping samples supply dependent evidence. Supplement protocols document negative results, historical execution identities and planning after data inspection. The narrow acoustic-comparator difference has no significance or equivalence test, and selected-output comparisons establish no necessity for all 84 mechanisms. Source definitions, analysis scripts and predictions are indexed in the accompanying record.

Author statements

AI assistance. Claude Code (Anthropic) and Codex (OpenAI) were used to assist software development and analysis, literature searching, manuscript drafting and revision, LaTeX editing, and document checks. The author is responsible for the scientific content and the final manuscript.

Funding and interests. No external funding was received for this work. The author developed Musical Intelligence.